

Editors

Andras Kemeny
Jean-Rémy Chardonnet
Florent Colombet
Stéphane Espié

Proceedings of the Driving Simulation Conference

*“Recent trends in driving simulation technology,
virtual experience and testing”*

DSC 2026 EUROPE^{XR}

Palais des Congrès

Antibes, France

September 16 – 18, 2026

ISBN 978-2-9573777-9-4

Direction

Andras Kemeny
Driving Simulation Association

Editors

Andras Kemeny	<i>Driving Simulation Association</i>
Jean-Rémy Chardonnet	<i>Arts et Métiers Institute of Technology</i>
Florent Colombet	<i>Renault</i>
Stéphane Espié	<i>Gustave Eiffel University</i>

Scientific committee

Chairman:

Andras Kemeny

Scientific committee members:

Mohammad Bahram	BMW Group R&T (Germany)
Gerd Baumann	FKFS (Germany)
Klaus Bengler	Technical University Munich (Germany)
Jost Bernasch	The Virtual Vehicle (Austria)
Jelte Bos	TNO (The Netherlands)
Heinrich H. Bühlhoff	Max Planck Institute (Germany)
Frank Cardullo	State University of NY (United States)
Viola Cavallo	Gustave Eiffel University (France)
Jean-Rémy Chardonnet	Arts et Métiers Institute of Technology (France)
Florent Colombet	Renault (France)
George Drettakis	INRIA (France)
Magnus Eek	VTI (Sweden)
Stéphane Espié	Gustave Eiffel University (France)
Zhou Fang	Renault (France)
Martin Fischer	DLR (Germany)
Massimiliano Gobbi	Polytechnic University of Milan (Italy)
Jens Häcker	DHBW Stuttgart (Germany)
Siddhartha Khastgir	University of Warwick (United Kingdom)
Joseph K. Kearney	University of Iowa (United States)
Franck Mars	CNRS (France)
Philippe Mathieu	University of Lille (France)
Frédéric Merienne	Arts et Métiers Institute of Technology (France)
James Oliver	Iowa State University (United States)
Jean-Christophe Popieul	Hauts-de-France Polytechnic University (France)
Paolo Pretto	The Virtual Vehicle (Austria)
Günther Prokop	Technische Universität Dresden (Germany)
Richard Romano	University of Leeds (United Kingdom)
Joost Venrooij	BMW Group (Germany)

© Driving Simulation Association

ISBN 978-2-9573777-9-4

Preface

It is with great pleasure that the Driving Simulation Association is editing the Driving Simulation Europe (DSC) 2026 XR Conference proceedings, a collection of scientific papers and short notes presented this year in Antibes at the 2025 edition of the DSC Conference.

They are a well-known and significant sample of the most recent works carried out in driving simulation research, using AI and XR, technology, driving simulator design and validation, motion cueing algorithms, visual rendering and DIL experiments but also autonomous driving (AD) and advanced driver-assistance system (ADAS) verification and validation (V&V), including massive simulation technology. This now recurrent and central driving simulation domain of connected and autonomous vehicle (CAV) simulation takes a crucial part of the papers.

This year, AI and ML based technologies, already introduced in the last years, are an essential part of the scientific papers, both in Motion Cueing & Control and Scenario Modelling and Generation. Trustworthiness is a crucial part of AI generated features as well as all the necessary control of the used technics, as we have seen recently with OpenAI AI tools going “rogue” and escaping the service provider’s control.

Motion sickness is a continuously developing research theme brought by driving simulators, as well as the advent of autonomous vehicles which may place users in unexpected motion conditions, without getting useful feedback on driver’s actions. They are completed with digital and immersive twin studies and experiments with a growing role of AI and XR based systems.

These proceedings are completing again that of the Driving Simulation Proceedings, an online, open access scientific journal: <https://proceedings.driving-simulation.org/>, supported by the Driving Simulation Association and its editorial board. The Driving Simulation Proceedings editorial board is composed of the Scientific Committee members, reviewing papers which are submitted both for the DSC conferences. These proceedings give a complete insight of the scientific community of driving simulation.

The DSC Europe 2026 XR Proceedings, integrate both full and short scientific papers, as well as scientific short notes, presented at the DSC Europe 2026 XR conference. In parallel, technical papers presented at this year’s conference are collected in a DSC Europe 2026 XR Product Solution book, to provide you with an overview of the corresponding technical developments.

Scientific and industrial keynotes, showing the weight of the industrial simulation chains, standards, simulator centres and products, as well as connected and autonomous vehicles are not integrated in the proceedings, but the reader may access the conference website, see: <https://dsc2026.org/keynote-speakers/>. Industrial pitches are also directly accessible through the DSC Europe 2022 VR website, <https://dsc2026.org/industrial-pitches/>.

We would like to thank all the reviewers and members of the DSC Scientific Committee; without their work it would not have been possible to guarantee the scientific quality of the published papers. They are too many to cite all here, but you can find them on our conference site - <http://dsc2026.org/committees/>.

We wish you a great pleasure in reading these papers bringing for you an insight in the latest trends in immersive simulation science and technology.

The Driving Simulation Association Executive Board

Andras Kemeny, Jean-Rémy Chardonnet, Florent Colombet, Stéphane Espié

Table of Contents

PERCEPTION & HUMAN FACTORS

9

Advanced Road Feedback in Driving Simulators: The Role of Physical Tire Models <i>Maximilian Böhle, Benjamin Rieff</i>	11
Occupant behaviour, cognition, and Human-Machine Interaction in Level 3 Autonomous Vehicle driving simulator <i>Ali Olayan, Jean-Rémy Chardonnet, Baptiste Sandoz</i>	19
Effects of Cognitive Load, Age and Sex on Driver Control Parameters <i>Victor Nalda-Tárrega, Juan-Manuel Belda-Lois, Begoña Mateo-Martínez</i>	29
In-car occupant state monitoring using RGB- and IR-images – carsickness as an example <i>Jelte E. Bos, Susanne Louvenberg, Jan L. Souman</i>	33
How We Roll: Validating Cycling Simulation Against Real-Life Riding <i>My Weidel, Lennart Ochel, Jakub Sliacan, Katja Kircher</i>	39
Simulation-Based Human-in-the-Loop Assessment of Shifting Comfort in Two Speed Electric Vehicles <i>Omar Fahmy, Safa Kaskaya, Stephan Rinderknecht</i>	47
Usability of a mixed reality head mounted display in a static driving simulator <i>Eva Morgand, Florent Colombet, Nicolas Mignotte, Javier Posselt</i>	55
Human Vision Driven Standard for Automotive Lit Surfaces <i>Victor Loya Garnica</i>	61

MOTION CUEING & CONTROL & AI

65

A Model-Predictive Control Algorithm for Matched Lateral/Yaw Motion <i>Maurice Kolff, Daan M. Pool, Vincenzo Ricciardi, Martin Heiderich, Frank Zantner, Barys Shyrokau</i>	67
Learning-Based Acceleration Prediction for Real-Time Model Predictive Motion Cueing <i>Keno Pape, Kevin Malena, Julia Timmermann, Ansgar Trächtler</i>	75
Comparison of Cost Functions for Automatic Motion Cueing Algorithm Parametrisation <i>Craig Brown, Marco Schlueter, Christian Holzapfel</i>	83
Effects of motion cueing and adaptive procedure on acceleration JND estimation <i>Erik Gustaf Lilljebjörn, Sogol Kharrazi, Jan Åslund</i>	87
Acceptance of Adaptive Cruise Control on a Dynamic Driving Simulator: the Role of Tilt Coordination <i>Xabier Carrera Akutain, Paolo Cattaneo, Francesco Comolli, Massimiliano Gobbi, Giampiero Mastinu, Roger Torm Obradors</i>	95
Physics-Informed Neural Network-based Motion Cueing Algorithm <i>Robert Jacumet, Simon Duell, Sebastian Wagner, Markus Schwienbacher, Marion Leibold</i>	103
Enhancing Longitudinal Motion Cueing via AI-Based Driver Behavior Prediction <i>Zhou Fang, Fabien De-Oliveira, Dongcheng Qiu, Matthieu Sommervogel, Florent Colombet, Stéphane Regnier</i>	113

Evaluating Data-Driven Architectures for High-Fidelity Modeling of Driving Simulator Motion <i>Philipp Legran, Fabian Russ, Lutz Eckstein</i>	119
---	-----

VIRTUAL REALITY FOR DRIVING SIMULATION	123
---	------------

Assessing the Sensitivity of a VR Driving Simulator to Single-Day Highway Work Zones <i>Paolo Pretto, Manuela Prior, Mateo Jukić, Inge Gsellmann, Thorsten Muik, Manfred Rosenberger, Christina Hager, Alexander Maletzky</i>	125
Expectation-Driven Differences in Pedestrian Gap Acceptance toward Automated Vehicles <i>Melina Bergen, Gerald Temme, Wolfgang Einhäuser, Martin Fischer</i>	133
Using Mixed Reality for Safety-Critical Testing in Railway Systems <i>Raphael Kling, Manuel Wild, Marvin Wieschollek, Martin Cichon</i>	137

SCENARIO MODELING, CONTROL AND AUTHORING TOOLS	141
---	------------

A Unified Framework for Scenario Coverage Estimation and Representative Selection in Simulation-Based Testing of Automated Vehicles <i>Anasse Essalih, Abdel-Hakim Arab, Laurent Durville, Mohamed-Cherif Rahal, Mustapha Lebbah, Walid Kherji, Fouad Hadj Selem</i>	143
A Framework for Describing and Controlling Surrounding Traffic in Driving Simulators <i>Sven Tarlowski, Philipp Legran, Lutz Eckstein</i>	151
Applying a Representativity-Based Model Selection Framework to an Industrial ADAS Use Case <i>Yara Hallak, Layale Saab, Dominique Blouin, Baptiste Laborie, Laurent Pautet</i>	159

SIMULATION DESIGN	163
--------------------------	------------

CycloCARLA: Design, Implementation, and Evaluation of a High-Fidelity Bicycle Simulator <i>Patrick Rebling, Yannick Rauch, Nico Sander, Philipp Nenninger, Reiner Kriesten</i>	165
--	-----

AI & MACHINE LEARNING FOR SCENARIO GENERATION	169
--	------------

Open-Source Scenes and Scenarios for various Intelligent Transportation Systems Applications <i>Kevin Malena, Abdelmawla Rizk, Niklas Fittkau, Sandra Gausemeier, Ansgar Trächtler</i>	171
AI-based simulation scenario generation from the ADScene complex scenarios <i>Sofiane Sadat, Hakim Mohellebi, Chouki Sentouh, Eric Vaillant, Emmanuel Arnoux, Stephane Regnier, Jean-Christophe Popieul</i>	175

SHORT NOTES	183
--------------------	------------

Absolute Accuracy Validation of Unity3D Perpendicular Distance Measurements: Impact for Road Safety Studies <i>Machiel van der Stelt, Thomas Bräunl, Paul Roberts</i>	185
Acoustic measures to increase immersion in static driving simulators <i>Arne Düselder, Marius Dute, Mirko Djukic</i>	187
Designing a Balanced Order Driving Simulator Study with OpenSCENARIO <i>Anders Andersson</i>	189

Improving Motion Cueing with High Shelf Filters <i>Craig Brown, Edward Smith, Nikhil Garrett, Chris Williams</i>	191
Fatigue-Inducing Scenario Design for Truck-Driving Simulation: From Evidence to Implementation <i>Tiago Fonseca, Telmo Costa Botelho, Sérgio Pedro Duarte, António Lobo, Sara Ferreira</i>	193
Assessing Communication Behaviors between Excavator Operators, Human Workers and Assistance Systems on a Virtual Construction Site <i>René Reinhard, Franziska Hekele, Jonathan Schneider, Tim Nicolai, Sebastian Emmerich, Michael Burger</i>	197

AUTHORS' INDEX**203**

PERCEPTION & HUMAN FACTORS

Advanced Road Feedback in Driving Simulators: The Role of Physical Tire Models

Maximilian Böhle^{1,2}, Benjamin Rieff¹

(1) Technical University of Berlin, Gustav-Meyer-Allee 25, 13355 Berlin, e-mail: boehle@campus.tu-berlin.de
(2) cosin scientific software, Hans-Stützle-Straße 20, 81249 München, e-mail: mail@cosin.eu

Abstract - With the increasing adoption of physics-based tire models and high-resolution road surface representations in real-time vehicle simulation, combined with the ongoing advancement of driving simulator technologies, the consideration of high-frequency vehicle dynamics in the context of human perception has become essential. This work investigates the effects of using a state-of-the-art physical tire model in direct comparison to a conventional empirical tire model on the physical and behavioral validity of high-frequency road feedback achieved in a dynamic driving simulator using high-fidelity road data. A meta-analysis of three randomized controlled subject studies comparing the results using a physical model of the type FTire and an empirical model of the type MF-Tyre 5.2 is presented. The analysis shows a substantial beneficial effect of the physical tire model on the perception of overall realism and road feedback. The results suggest a substantial effect of drivers' levels of relevant evaluation experience and the available bandwidth of the driving simulator.

Keywords: Road Feedback, Driving Simulator, FTire, MF-Tyre

Copyright: © 2026 by the authors.

Licensee Driving Simulation Association.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.82157/dsa/2026/1>

Introduction

Tire models for vehicle simulation can be separated into three general categories of rising fidelity: empirical models, physical models, and finite-elements-method-based models. Traditionally, real-time simulation environments for driving simulators have been limited to the use of simplified empirical tire models due to computational restrictions. However, in recent years, a rising number of works have demonstrated the feasibility of closed-loop real-time simulations using physics-based tire models (Gallrein, et al., 2014; Rieff, et al., 2022; 2024; Schurmann, et al., 2024). This substantially improves the representation of the tire-road interaction, particularly in combination with high-fidelity road models generated from surface scans of real roads. At the same time, the usable frequency range of modern driving simulators has increased considerably in recent years (Xue, et al., 2023).

As a result, recent works have investigated the effects of using physical tire models in driving simulators on vehicle dynamics (Rupp and Schick, 2025), steering feel (Böhle, Schick, and Müller, 2025b), and driving performance (Böhle, Schick, and Müller, 2025a) at frequencies up to over 100 Hz. Due to the human perception of vibrations above 30 Hz, the phenomena introduced by the increased physical validity of road-induced high-frequency excitations have implications for both the tactile and the acoustic road feedback in driving simulators, see Figure 1.

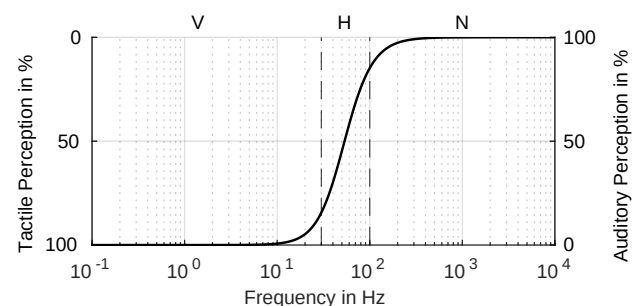


Figure 1: Frequency ranges of human perception of vehicle excitations, as defined in Hieronimus, 1992

Related Work

A limited number of previous works have investigated the effects of tactile feedback at frequencies up into the acoustically relevant range. However, existing works performed on high-fidelity driving simulators have been limited to the use of simplified empirical tire models (Parduzzi, 2021; Shyrokau, et al., 2018). Furthermore, implications for the acoustic perception of road feedback via structure-borne sound in driving simulators have not been addressed, yet.

To analyze the effects of using a state-of-the-art physical tire model on the validity of subjective evaluations obtained in driving simulators, this work presents the findings from a meta-analysis performed

on the results of three randomized controlled single-blind subject studies. An analysis of the effects of isolated frequency components on the validity of steering feedback from the first two subject studies has already been published in Böhle, Schick, and Müller, 2024 and Böhle, Schick, and Müller, 2025b. A meta-analysis of the effects of the simulator modifications on driving performance based on the objective data obtained from all three subject studies has been published in Böhle, Schick, and Müller, 2025a. This work presents a meta-analysis of the effects of the implemented tire models on the behavioral validity, as reflected in the subjective evaluations of simulator realism and road feedback, performed across the subjective data from all three studies.

Method

The subject studies analyzed in this work consisted of randomized single-blind within-subjects comparisons between a reference vehicle and different variants of its virtual representation in a driving simulator. A total of 121 participants completed a subjective evaluation of road feedback in the reference vehicle on a reference country road, followed by evaluation drives of different variants in the driving simulator using digital twins of the reference vehicle, tires, and road. The studies were performed on a dynamic driving simulator equipped with a motion platform covering the representation of excitations up to 50 Hz and structure-borne sound transducers enabling the representation of excitations up to 500 Hz, see Figure 2.



Figure 2: The Advanced Vehicle Driving Simulator (aVDS) at the IFM of Kempten University of Applied Sciences

In the driving simulator, the reference vehicle was represented via a two-track rigid-body model in IPG CarMaker, validated in accordance with ISO 22140 (ISO, 2021) and ISO 19634 (ISO, 2014). Steering feedback was provided by a steering system hardware-in-the-loop (HiL) setup, running the electromechanical steering system in a full vehicle bus simulation, providing external rack force feedback via a linear actuator.

Despite considerable conceptual limitations, particularly in combination with high-fidelity road data, the use of empirical tire models in driving simulators is still very common. To mitigate the fundamental issues arising from the lack of a contact patch representation, a plethora of empirical and semi-physical approaches exist to extend the capabilities of Magic-Formula-based tire models. Hence, the baseline tire model used in this work was an MF-Tyre 5.2 model (Pacejka, 2005) extended by an unweighted contact

patch approximation based on the volumetric intersection of the theoretical tire geometry with the road surface. An FTire model (Gipser, 2007) was implemented as the physical counterpart, representing the only commercially available tire model capable of representing full three-dimensional road surface enveloping via a physical model of the tire structure under real-time conditions. Across the analyzed subject studies, five simulator variants employed the empirical MF-Tyre model, while six variants utilized the physical FTire model. Both tire datasets had been parametrized from the same data source, using dyno measurements of the reference tire. The road models represented the reference road surface with a horizontal resolution of 10 mm. Both the regular grid road model of the type rgr used for the FTire model and the rFpro terrain server model of the reference track were generated from the same set of LiDAR data. All simulator runs were performed using a visual environment representing the reference road in rFpro.

Due to its prevalence across the three subject studies and its interaction with the effects of different tire models, the main effect of the algorithm for rack force frequency augmentation (RFFA) is of particular relevance for the analyses performed in this work. To match the road feedback recorded during participants' reference drives, the RFFA algorithm implemented additional high-frequency steering rack forces between 10 and 30 Hz via the steering HiL system. Table 1 provides an overview of the distribution of RFFA and the two tire models across the simulator variants from all three subject studies.

A summary of all six simulator modifications implemented across the three subject studies is shown in Table 2. These interventions comprised variations in the bandwidth of vehicle and road feedback available in the driving simulator through controller, model, and hardware adaptations. As indicated by the specified frequency ranges, four of the six modifications had implications for both tactile and auditory perception in the driving simulator. To enable an isolation of these mechanisms, participants received auditory feedback generated in an adaptive audio environment via active noise cancelling headphones. Further details regarding the implementation and performance of the individual configurations are provided in Böhle, Schick, and Müller (2024; 2025a; 2025b).

To facilitate the analysis of specific frequency contributions, participants completed all evaluation drives using the self-selected velocity profiles recorded during their respective reference drives. The three individual samples of participants consisted predominantly of experienced simulator drivers which were clustered into three groups based on their relevant prior professional evaluation experience. The resulting sample used in this meta-analysis consisted of 28 steering system experts, 46 vehicle experts, and 38 non-experts. Additional details regarding the participant clusters can be found in Böhle, Schick, and Müller, 2025a.

The subjective evaluation was conducted using two questionnaires covering the perception of road feedback transmitted via the steering wheel and through vehicle body excitations. The present work focuses on the analysis of the results obtained from the first questionnaire across all three subject studies. Established steering feedback characteristics were assessed, supplemented by additional items specif-

Table 1: Overview of the main effects RFFA and Tire Model across all simulator variants

	V1.1	V1.2	V1.3	V2.1	V2.2	V2.3	V2.4	V3.1	V3.2	V3.3	V3.4
RFFA	✓	✗	✓	✓	✓	✓	✓	✗	✗	✗	✗
MF-Tyre	✓	✓	✓	✓	✓	✗	✗	✗	✗	✗	✗
FTire	✗	✗	✗	✗	✗	✓	✓	✓	✓	✓	✓

Table 2: Overview of all simulator modifications across simulator variants. Asterisks mark modifications utilizing FTire outputs.

Modification	Description	Frequency Range
RFFA	Augmentation of high-frequency steering rack forces with measured reference forces	10 - 30 Hz
MCA-LP	Attenuation of high-frequency vehicle body accelerations via motion cueing parameters	10 - 50 Hz
SST-NRI	Representation of non-road-induced vehicle excitations via SSTs	30 - 100 Hz
Tire Model*	Implementation of road-induced high-frequency vehicle excitations via motion platform	30 - 100 Hz
SST-RI*	Representation of road-induced high-frequency vehicle excitations via SSTs	30 - 500 Hz
AA-RI	Implementation of auditory road surface feedback via an adaptive audio environment	100 - 500 Hz

ically designed to capture the separation of low-frequency (LF) and high-frequency (HF) components of both steering and chassis feedback.

All attributes were rated using sign-weighted values on a scale from 1 to 10, using the automotive assessment index (BI) as described in HeiBing and Brandl, 2002. The majority of participants answered the German version of the questionnaire shown in Table 3. Further details regarding the rating scale, questionnaire structure, and individual evaluation criteria are provided in Böhle, Schick, and Müller (2024; 2025b).

Table 3: Questionnaire items related to road feedback

Description	lower	upper
Road Contact <i>Fahrbahnkontakt</i>	too low <i>zu niedrig</i>	- -
Low-Frequency Steering Feedback (Bumps and Isolated Events) <i>Lenkungsrückmeldung niederfrequent</i> (<i>Stöße und Einzelereignisse</i>)	too low <i>zu niedrig</i>	too high <i>zu hoch</i>
High-Frequency Steering Feedback (Vibrations) <i>Lenkungsrückmeldung hochfrequent</i> (<i>Vibrationen</i>)	too low <i>zu niedrig</i>	too high <i>zu hoch</i>
Low-Frequency Chassis Feedback (Bumps and Isolated Events) <i>Chassisrückmeldung niederfrequent</i> (<i>Stöße und Einzelereignisse</i>)	too low <i>zu niedrig</i>	too high <i>zu hoch</i>
High-Frequency Chassis Feedback (Vibrations) <i>Chassisrückmeldung hochfrequent</i> (<i>Vibrationen</i>)	too low <i>zu niedrig</i>	too high <i>zu hoch</i>

Due the consistent methodology applied across all three subject studies, a combined analysis of the data from all 121 participants was conducted. For this purpose, all datasets without missing observations were analyzed using Linear Mixed Models (LMMs) implemented in jamovi version 2.3.28 (The jamovi project, 2024), running R version 4.1.2 (R Core Team, 2021).

Building on the results of the statistical analyses conducted within the individual subject studies, the meta-analysis focused on dependent variables (DVs) associated with road feedback. Only questionnaire items included in all three studies were considered in the

analysis. This yielded a total of 2286 observations obtained from 102 participants over 381 simulator drives.

The Linear Mixed Models (LMMs) incorporated the six simulator modifications as fixed main effects, while also accounting for relevant interaction terms both among main effects and in combination with the variables User Level and Repetition. Inter-subject variability was represented via a random intercept. Furthermore, observations were clustered at the participant level to reflect the repeated-measures design, and at the study level to account for potential confounding differences between the individual datasets. The general structure of the resulting model equation, expressed in jamovi notation, is provided in Equation 1.

$$\begin{aligned}
 DV \sim & 1 + Main_{1...6} + User\ Level + Repetition \\
 & + Main_{1...6} : Main_{6...1} \\
 & + Main_{1...6} : User\ Level + Main_{1...6} : Repetition \\
 & + (1 | Participant)
 \end{aligned}
 \quad (1)$$

Assumption checks covered the normality of model residuals via Kolmogorov-Smirnov tests and a visual inspection of residual densities and QQ-plots. The general model structure had been selected based on overall model fit in terms of the characteristics Akaike information criterion, Bayesian information criterion, Loglikelihood, and both marginal and conditional R^2 .

Results

The following section is structured in three parts. After taking an initial look at the distribution of subjective data, a summary of the findings from the individual subject studies is provided, followed by a discussion of the results obtained from the statistical meta-analysis conducted in this work.

The distribution of BI ratings across all analyzed road-feedback-related characteristics from the three subject studies is presented in Figure 3. In the first two studies, both the reference vehicle and all driving simulator variants were generally perceived as providing excessively high levels of road feedback. In contrast, the third study yielded the opposite trend,

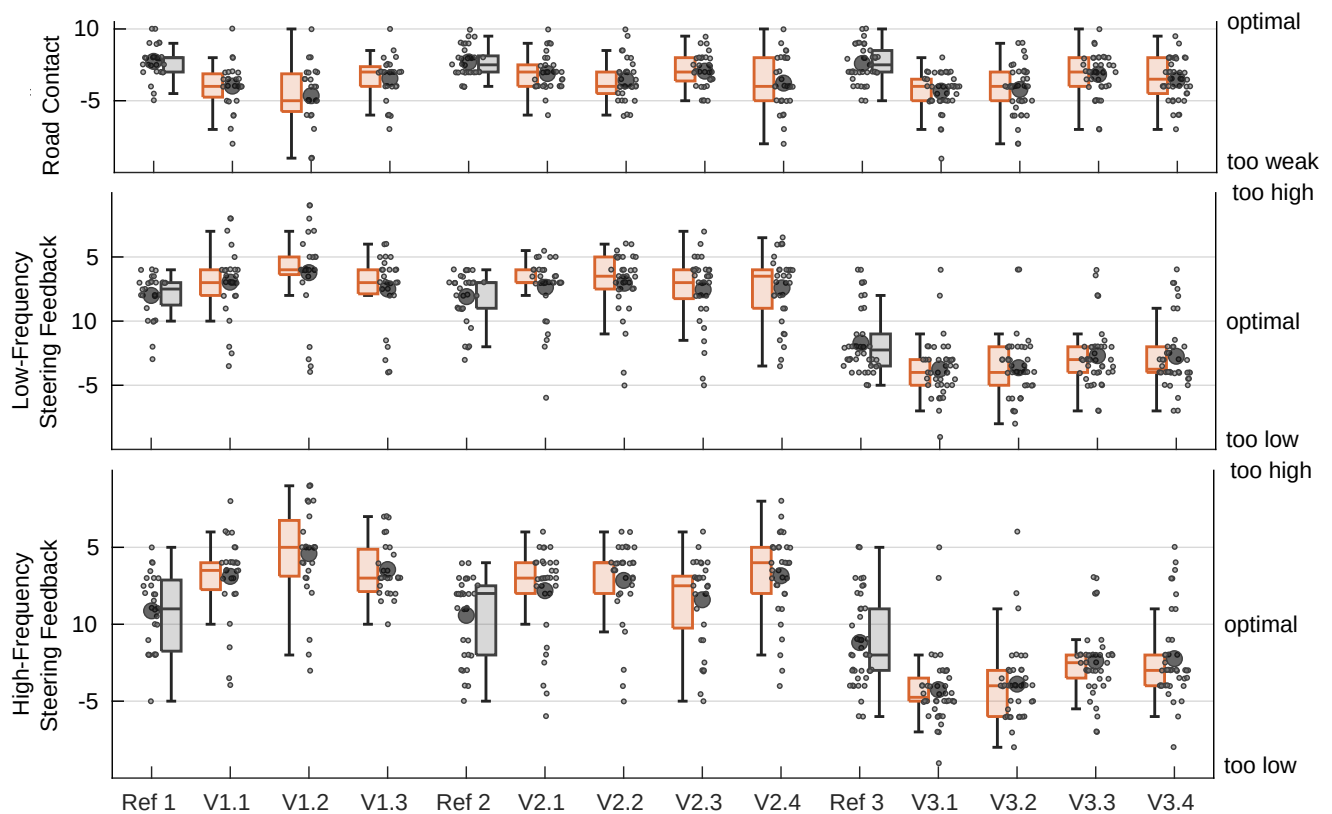


Figure 3: BI ratings of road-feedback-related characteristics for reference vehicle drives (gray) and simulator variants (orange) across all subject studies. Higher absolute values indicate better ratings. Large markers indicate mean values, and medians are marked by the horizontal lines inside the boxes.

with overall levels of road feedback being rated as lower than optimal. Across all dependent variables considered in this meta-analysis, the reference vehicle consistently received higher ratings than any simulator configuration. Consequently, within the following analyses, an improvement in BI ratings corresponds to an increase in behavioral validity.

Effects of Vibrations

The first subject study (Böhle, Schick, and Müller, 2024) identified a significant impact of road-induced excitations transmitted via the motion platform on the subjective perception of steering feedback. In particular, chassis excitations between 10 and 50 Hz exhibited an effect on road-feedback-related steering characteristics that was comparable in magnitude to the effect of an augmentation of rotational steering wheel excitations up to 30 Hz via rack force data measured during the reference drives. Improved validity of subjective ratings and an increase in perceived realism were associated with an increase in the physical validity of high-frequency steering wheel excitations (RFFA), and a reduction of high-frequency feedback through the motion platform (MCA-LP).

Effects of Harshness

In the second subject study (Böhle, Schick, and Müller, 2025b), non-road-induced excitations in the frequency range from 30 to 100 Hz (SST-NRI) were found to significantly influence the perception of both

steering and chassis feedback. Furthermore, the implementation of an FTire model resulted in an improvement in perceived overall realism compared to an MF-Tyre model. However, its influence on specific steering feedback characteristics depended strongly on participants' prior experience. Non-expert participants tended to rate simulator variants with the simplified empirical tire model (MF-Tyre) as slightly more realistic, whereas expert participants assigned substantially higher ratings to configurations employing the physical tire model (FTire).

Effects of Noise

The third subject study examined the role of auditory road feedback and its interaction with tactile excitations up to 500 Hz. While no significant effects of acoustic road feedback (AA-RI) were observed, the results emphasized the pronounced positive contribution of high-frequency tactile road feedback generated by a high-fidelity physical tire model when combined with high-resolution road data. Significant positive effects of road-induced high-frequency excitations via SSTs (SST-RI) on the validity of subjective evaluations and perceived realism were found. Furthermore, all simulator runs in this study were performed using the physical FTire model, since preliminary tests with experts had determined that an implementation of road-induced high-frequency excitations via SSTs using the empirical MF-Tyre model was not feasible in this setup. Due to the significantly less representative tire-road interaction and tire structure dynamics, the MF-Tyre model did not provide suffi-

cient high-frequency vehicle excitations beyond stationary driving conditions. Furthermore, when combined with the high feedback bandwidth of the SSTs, the MF-Tyre model introduced significant structure-borne and airborne noise contributions above 200 Hz that were perceived as highly unrealistic and distracting despite the use of headphones with active noise cancellation.

Figure 4 shows a comparison of driver seat rail acceleration measurements obtained from a drive in the reference vehicle and recordings from the driving simulator running the MF-Tyre and the FTire model. The analysis was performed on sections with the engine running at idle speed in neutral gear. Simulator runs were performed in closed-loop operation with longitudinal control enabled. As a result, longitudinal vehicle velocities were identical throughout the length of the reference road, whereas the lateral position was subject to driver influence.

Effects of Tire Models

The following section presents the results of the meta-analysis of subjective ratings using LMMs. An overview of the resulting contributions of fixed factors and random terms is provided in Table 4.

Table 4: Model fit characteristics of the LMMs used to analyze the BI ratings from all three subject studies

Model Fit	Road Contact	Steering Feedback	
		Low-Freq.	High-Freq.
Marginal R^2	.171	.605	.622
Conditional R^2	.671	.776	.746

The model fit indicates that both fixed and random effects contributed substantially to the representation of subjective data. For Road Contact, only a small share of the explained variance was the result of fixed effects, suggesting a strong influence of between-subject variability. In contrast, LF-Steering Feedback and HF-Steering Feedback exhibit a much larger contribution from fixed effects, with only moderate influence from random components.

Fixed effects omnibus tests revealed a highly significant main effect of Tire Model on the ratings of LF-Steering Feedback, $F(1, 335.0) = 51.123$, $p < .001$, and HF-Steering Feedback, $F(1, 335.0) = 62.489$, $p < .001$. Variants using the FTire model received higher ratings and were perceived as more realistic than those using the MF-Tyre model. A significant interaction between the main effects RFFA and Tire Model was observed for both LF-Steering Feedback, $F(1, 335.0) = 63.754$, $p < .001$ and HF-Steering Feedback, $F(1, 335.0) = 75.768$, $p < .001$. The effect size associated with Tire Model was significantly attenuated in configurations with active RFFA.

Furthermore, a significant main effect of RFFA was found for Road Contact, $F(1, 324.0) = 4.632$, $p = .032$, LF-Steering Feedback, $F(1, 332.0) = 33.579$, $p < .001$, and HF-Steering Feedback, $F(1, 333.0) = 35.363$, $p < .001$. In the presence of active RFFA, Road Contact was rated more positively and perceived as more realistic compared to configurations without RFFA. However, both LF-Steering Feedback and HF-Steering Feedback were

evaluated as “too high” with active RFFA, whereas they were perceived as “too low” when RFFA was inactive.

The variable User Level did not exhibit a significant main effect. However, a significant interaction between User Level and RFFA was observed for Road Contact $F(2, 302.0) = 4.056$, $p = .018$. Generally, participants from the experts subgroup (User Levels 1 and 2) demonstrated larger effect sizes for both main effects compared to non-expert participants, while differences between the User Levels within the expert subgroup were comparatively small. A visual representation of the main effects and their interaction with participants' User Level is shown in Figure 5

No significant effects of Repetition were identified in the subjective evaluations. Despite being included in only one study each, both SST-NRI and SST-RI showed significant contributions to the observed effects. The parameter estimates of all fixed effect terms for all analyzed road-feedback-related DVs are displayed in Table 5 in the appendix.

Discussion

Overall, enhancing the validity of tactile feedback, both via an augmentation of steering rack forces and the implementation of the physical tire model, led to a significant improvement in drivers' perception of road feedback. Behavioral validity, defined as the agreement between subjective evaluations obtained in simulator drives and participants' ratings of the reference vehicle, was consistently higher in configurations employing the FTire model. This underscores the general importance of accurately representing both the steering feedback transfer path and the tire-road interaction for achieving reliable subjective driver feedback in high-fidelity driving simulators.

The combined analysis of the datasets from all three studies provided further insights into the role of different frequency components. The observed interaction between RFFA and Tire Model suggests that a substantial proportion of the FTire model's contribution to perceived road feedback originates from frequency components below 30 Hz. In vehicles with conventional steering systems, road-induced wheel excitations in this frequency range strongly affect the steering feedback path, inducing both rotational and non-rotational steering wheel excitations. Given that the augmentation of steering feedback using reference measurement data, as realized via RFFA, is not a practical solution for general driver-in-the-loop simulations, these findings emphasize the necessity of a high-fidelity representation of the tire-road interaction, even at moderate frequencies.

Furthermore, despite being implemented in only one subject study each, both modifications introducing structure-borne noise above 30 Hz (SST-NRI and SST-RI) exhibited substantial effects on subjective ratings of simulator variants. This finding highlights the critical role of tactile feedback system bandwidth in shaping drivers' subjective perception within dynamic driving simulators. The challenges highlighted in accurately representing high-frequency structure-borne noise input using empirical tire models outside their range of validity underscore the need for more sophisticated tire, subsystem, and vehicle models in order to achieve realistic road feedback at higher frequencies.

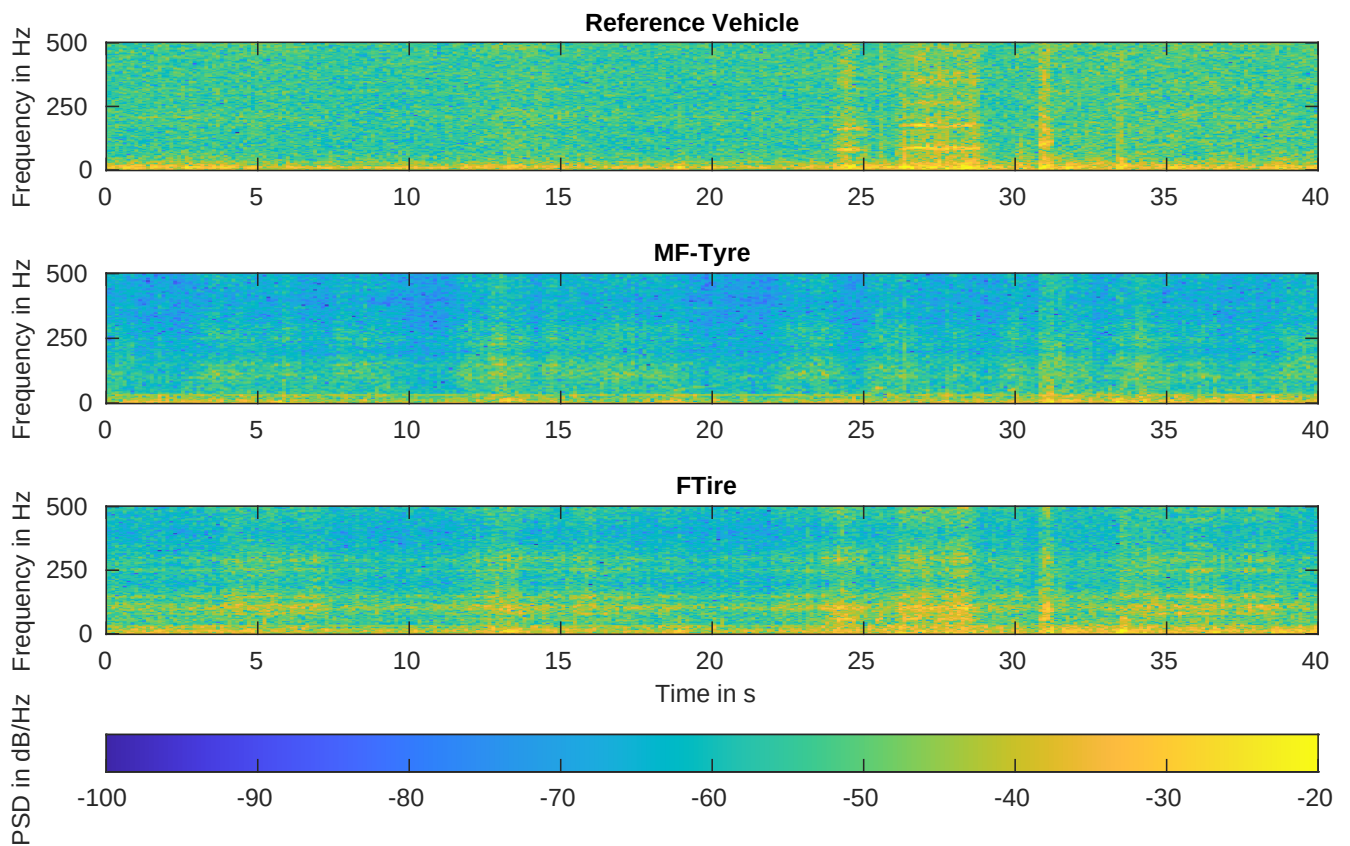


Figure 4: Spectrograms of vertical accelerations measured in the reference vehicle and the driving simulator without engine load

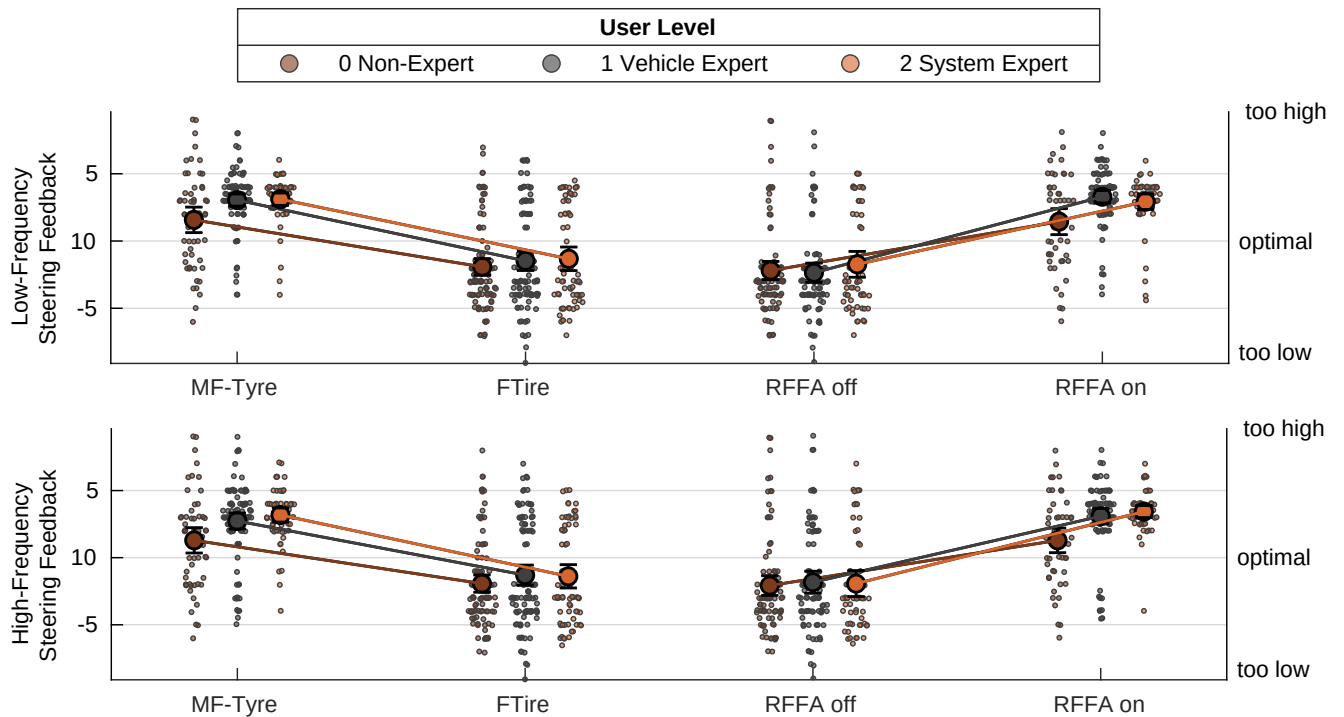


Figure 5: Main effects of Tire Model and RFFA on LF-Steering Feedback and HF-Steering Feedback across all subject studies. Means and standard errors of BI values (higher absolute value means better, more realistic rating).

Across all simulator configurations using the empirical MF-Tyre model, both low-frequency and high-frequency steering feedback were consistently rated as “too high” and as exceeding the levels observed

in the reference vehicle, regardless of RFFA status. In contrast, configurations with the physics-based tire model were generally perceived as “too low” without RFFA and as “too high” with active RFFA, relative

to the reference vehicle. One possible explanation for this observation is an amplitude mismatch arising from unscaled high-frequency steering wheel excitations in combination with vehicle body excitations that were both scaled by the motion cueing algorithm and attenuated by the motion platform's low-pass characteristics. Although this does not affect the overall interpretation of the primary effects, it should be considered when interpreting effect directions and comparing effect sizes. Additionally, given the complexity of the experimental design, potential masking effects from other high-frequency components cannot be excluded.

The significant interaction between participants' simulator experience and the main effects indicates that the ability to perceive and appropriately evaluate high-fidelity road feedback is strongly influenced by prior training. The clear distinction between non-experts and experts, combined with the absence of observable training effects over the course of the study, suggests that such expertise must be developed through extensive prior experience in simulator-based evaluations. These results are consistent with the findings from the meta-analysis of objective driver performance metrics reported in Böhle, Schick, and Müller, 2025a. While increased steering feedback validity through RFFA reduced mental workload and improved lateral vehicle control, the effects of the FTire model varied with experience level: non-experts exhibited increased workload and reduced driving performance, whereas experts demonstrated improved performance despite experiencing higher workload.

Conclusion

As commercially available driving simulators continue to achieve higher levels of fidelity in both tactile and auditory feedback, the accurate representation of high-frequency vehicle model outputs becomes increasingly relevant. The findings of the subject studies presented in this work underscore the importance of highly accurate tire models for ensuring the validity of subjective evaluations in dynamic driving simulators, particularly when real-world road data are applied. In this work, simulator configurations employing the physical FTire model yielded significantly higher validity in subjective assessments of road feedback compared to those based on the empirical MF-Tyre model. However, the positive effects associated with enhanced representations of steering feedback and tire-road interaction were strongly influenced by participants' levels of relevant evaluation experience, with substantially greater benefits observed among more experienced individuals. This suggests that the selection of an appropriate simulator configuration for development tasks should account for the specific characteristics of the intended evaluator population.

The subject studies presented in this work investigated the coupling of both an empirical and a state-of-the-art physical tire model with a simplified rigid-body vehicle model. Despite this modeling constraint, the results demonstrate considerable improvements associated with the use of a physical tire model. Consequently, further research is recommended to investigate the combined application of such tire models with state-of-the-art multibody vehicle models in order to enhance simulation fidelity.

References

- Böhle, M., Schick, B., and Müller, S., Dec. 2024. Steering Feedback in Dynamic Driving Simulators: The Influence of Steering Wheel Vibration and Vehicle Motion Frequency. *IEEE Transactions on Intelligent Vehicles*, 9(12), 7781–7796. <https://doi.org/10.1109/tiv.2024.3401868>.
- Sept. 24, 2025a. Effects of Steering Feedback via Noise, Vibration, and Harshness on Lateral Driver Performance in Dynamic Driving Simulators. In: A. Kemeny, J.-R. Chardonnet, F. Colombet, and S. Espié eds. *Proceedings of the Driving Simulation Conference 2025 Europe XR*. Driving Simulation Association. Stuttgart, Germany, pp. 37–44. <https://doi.org/10.82157/dsa/2025/4>.
- 2025b. Steering Feedback in Dynamic Driving Simulators: Road-Induced and Non-Road-Induced Harshness. *arXiv preprint arXiv:2504.11097*.
- Gallrein, A., Baecker, M., Burger, M., and Gizatullin, A., 2014. *An Advanced Flexible Realtime Tire Model and its Integration Into Fraunhofer's Driving Simulator*. Tech. rep. SAE Technical Paper.
- Gipser, M., 2007. FTire – the tire simulation model for all applications related to vehicle dynamics. *Vehicle System Dynamics*, 45(sup1), pp. 139–151. <https://doi.org/10.1080/00423110801899960>.
- Heißing, B. and Brandl, H. J., 2002. *Subjektive Beurteilung des Fahrverhaltens*. Vogel Buchverlag.
- Hieronimus, K., 1992. Anforderungen an Schwingungs- und Akustikberechnungen aus Sicht der Fahrzeugentwicklung. In: *Rechenmethoden in der Fahrzeugentwicklung. Fortschritte der Fahrzeugtechnik*. Ed. by W. Dirschmid. Vol. 12. Wiesbaden, Germany: Vieweg+Teubner Verlag, pp. 47–71.
- ISO, 2014. *Passenger cars — Validation of vehicle dynamics simulation — Lateral transient response test methods*. International Organization for Standardization.
- 2021. *Passenger cars — Validation of vehicle dynamics simulation — Steady-state circular driving behaviour*. International Organization for Standardization.
- Pacejka, H., 2005. *Tire and vehicle dynamics*. Elsevier.
- Parduzi, A., 2021. Bewertung der Validität von Fahrsimulatoren anhand vibro-akustischer Fahrzeugschwingungen. PhD. Dissertation, Technische Universität Berlin, 2021.
- R Core Team, 2021. *R: A Language and Environment for Statistical Computing*. Version 4.1.2. R Foundation for Statistical Computing. Vienna, Austria. Available at: <https://www.R-project.org/>.
- Rieff, B., Böhle, M., Vorderobermeier, S., and Schick, B., 2022. Comfort evaluation on a dynamic driving simulator with advanced tire, road and vehicle models. In: *International Munich Chassis Symposium*. Springer, pp. 130–143.
- Rieff, B., Joerg, K., Becker, M., and Thaler, K., 2024. A Novel Simulator Setup: Combining Full FTire Dynamics With A High Accurate Steering Feel Emulation. In: *International Munich Chassis Symposium*. Springer, pp. 107–118.
- Rupp, P. and Schick, B., Jan. 2025. FTire vs MFTire: A comparison of the evaluability of vehicle dynamics on a dynamic driving simulator depending on the tire model type. In: pp. 219–232. <https://doi.org/10.51202/9783181024614-219>.
- Schurmann, R., Harendza, N., Rieff, B., and Lion, A., 2024. Comparison of real-time capable multi-body simulation models and map-based models on the dynamic driving simulator. In: *International Munich Chassis Symposium*. Springer, pp. 61–76.
- Shyrokau, B., De Winter, J., Stroosma, O., Dijksterhuis, C., Loof, J., van Paassen, R., and Happee, R., 2018. The effect of steering-system linearity, simulator motion, and truck driving experience on steering of an articulated tractor-semitrailer combination. *Applied Ergonomics*, 71, pp. 17–28. <https://doi.org/10.1016/j.apergo.2018.03.018>.
- The jamovi project, 2024. *jamovi version: 2.3.28 [Software]*. Sydney, Australia. Available at: <https://www.jamovi.org>.
- Xue, H., Prevati, G., Gobbi, M., Mastinu, G., et al., 2023. Research and Development on Noise, Vibration, and Harshness of Road Vehicles Using Driving Simulators-A Review. *SAE International journal of vehicle dynamics, stability, and NVH*, 7(4), pp. 555–577.

Appendix

Table 5: LMM parameter estimates for fixed effects (Parameter $\pm$ SE) of BI ratings across all subject studies

Fixed Effect	Level	Road Contact		LF-Steering Feedback		HF-Steering Feedback	
		Estimate	p	Estimate	p	Estimate	p
Fixed Intercept	–	–2.896 $\pm$ 0.273	< .001	1.130 $\pm$ 0.506	.026	1.992 $\pm$ 0.531	< .001
RFFA	off–on	–0.564 $\pm$ 0.262	.032	–2.790 $\pm$ 0.481	< .001	–2.983 $\pm$ 0.502	< .001
MCA-LP	off–on	–0.527 $\pm$ 0.454	.247	1.343 $\pm$ 0.849	.115	0.116 $\pm$ 0.888	.896
SST-NRI	off–on	–0.736 $\pm$ 0.224	.001	0.241 $\pm$ 0.443	.586	1.207 $\pm$ 0.478	.012
Tire Model	MF–FT	0.161 $\pm$ 0.283	.570	3.784 $\pm$ 0.529	< .001	4.394 $\pm$ 0.556	< .001
SST-RI	off–on	–0.994 $\pm$ 0.154	< .001	–0.987 $\pm$ 0.310	.002	–1.803 $\pm$ 0.339	< .001
AA-RI	off–on	0.093 $\pm$ 0.153	.544	0.010 $\pm$ 0.309	.974	–0.184 $\pm$ 0.338	.587
User Level	1 – 0	0.109 $\pm$ 0.432	.801	0.369 $\pm$ 0.760	.627	0.258 $\pm$ 0.781	.741
	2 – 0	0.493 $\pm$ 0.482	.308	1.017 $\pm$ 0.848	.231	0.972 $\pm$ 0.871	.265
Repetition	2 – 1	0.402 $\pm$ 0.390	.303	0.593 $\pm$ 0.757	.434	0.162 $\pm$ 0.806	.841
	3 – 1	0.564 $\pm$ 0.370	.128	0.370 $\pm$ 0.728	.611	0.550 $\pm$ 0.781	.482
	4 – 1	0.669 $\pm$ 0.349	.056	0.873 $\pm$ 0.671	.194	0.824 $\pm$ 0.711	.248
SST-RI x AA-RI	off–on x off–on	–0.475 $\pm$ 0.307	.124	0.030 $\pm$ 0.620	.961	0.007 $\pm$ 0.678	.992
RFFA x Tire Model	off–on x MF–FT	–0.190 $\pm$ 0.511	.711	7.578 $\pm$ 0.949	< .001	8.666 $\pm$ 0.996	< .001
RFFA x User Level	off–on x 1 – 0	1.206 $\pm$ 0.498	.016	–1.570 $\pm$ 0.911	.086	–1.364 $\pm$ 0.952	.153
	off–on x 2 – 0	0.046 $\pm$ 0.563	.935	–0.658 $\pm$ 1.031	.524	–1.009 $\pm$ 1.077	.350
SST-NRI x User Level	off–on x 1 – 0	–0.525 $\pm$ 0.403	.194	–0.256 $\pm$ 0.797	.748	0.622 $\pm$ 0.860	.470
	off–on x 2 – 0	–0.450 $\pm$ 0.455	.324	–0.429 $\pm$ 0.904	.636	–0.966 $\pm$ 0.979	.324
Tire Model x User Level	MF–FT x 1 – 0	0.638 $\pm$ 0.497	.200	0.521 $\pm$ 0.902	.564	0.192 $\pm$ 0.934	.837
	MF–FT x 2 – 0	0.156 $\pm$ 0.554	.778	1.339 $\pm$ 1.005	.184	0.398 $\pm$ 1.041	.702
SST-RI x User Level	off–on x 1 – 0	–0.163 $\pm$ 0.369	.659	0.626 $\pm$ 0.741	.399	0.415 $\pm$ 0.808	.608
	off–on x 2 – 0	–0.871 $\pm$ 0.407	.034	–1.016 $\pm$ 0.820	.217	–0.338 $\pm$ 0.895	.706
AA-RI x User Level	off–on x 1 – 0	–0.399 $\pm$ 0.366	.276	0.879 $\pm$ 0.736	.234	1.681 $\pm$ 0.804	.038
	off–on x 2 – 0	–0.716 $\pm$ 0.406	.079	0.853 $\pm$ 0.818	.298	1.164 $\pm$ 0.893	.193
RFFA x Repetition	off–on x 2 – 1	–0.429 $\pm$ 0.608	.480	–1.670 $\pm$ 1.156	.150	–1.522 $\pm$ 1.215	.211
	off–on x 3 – 1	–0.427 $\pm$ 0.667	.522	–1.130 $\pm$ 1.311	.390	–0.298 $\pm$ 1.405	.832
	off–on x 4 – 1	–1.012 $\pm$ 0.738	.171	1.214 $\pm$ 1.405	.388	1.204 $\pm$ 1.481	.417
SST-NRI x Repetition	off–on x 2 – 1	–0.029 $\pm$ 0.746	.969	0.364 $\pm$ 1.375	.792	0.152 $\pm$ 1.416	.915
	off–on x 3 – 1	0.112 $\pm$ 0.499	.823	1.654 $\pm$ 0.997	.099	0.670 $\pm$ 1.084	.537
	off–on x 4 – 1	1.140 $\pm$ 0.864	.188	–0.565 $\pm$ 1.603	.725	0.068 $\pm$ 1.658	.967
Tire Model x Repetition	MF–FT x 2 – 1	0.285 $\pm$ 0.452	.529	–0.140 $\pm$ 0.900	.876	–0.626 $\pm$ 0.976	.522
	MF–FT x 3 – 1	–0.380 $\pm$ 0.657	.564	1.185 $\pm$ 1.240	.340	1.152 $\pm$ 1.295	.374
	MF–FT x 4 – 1	0.058 $\pm$ 0.875	.947	0.484 $\pm$ 1.643	.769	0.837 $\pm$ 1.712	.625
SST-RI x Repetition	off–on x 2 – 1	0.518 $\pm$ 0.501	.302	–1.628 $\pm$ 0.984	.099	–2.018 $\pm$ 1.053	.056
	off–on x 3 – 1	0.473 $\pm$ 0.543	.385	–1.824 $\pm$ 1.064	.087	–2.322 $\pm$ 1.134	.041
	off–on x 4 – 1	0.157 $\pm$ 0.508	.758	0.143 $\pm$ 1.002	.887	–1.527 $\pm$ 1.076	.157
AA-RI x Repetition	off–on x 2 – 1	0.725 $\pm$ 0.496	.145	–0.361 $\pm$ 0.977	.712	–0.331 $\pm$ 1.046	.752
	off–on x 3 – 1	–0.064 $\pm$ 0.508	.900	0.766 $\pm$ 1.006	.447	–0.480 $\pm$ 1.083	.658
	off–on x 4 – 1	–0.107 $\pm$ 0.519	.837	–0.832 $\pm$ 1.020	.415	0.250 $\pm$ 1.092	.819
SST-RI x AA-RI x User Level	off–on x off–on x 1 – 0	–0.223 $\pm$ 0.736	.762	0.084 $\pm$ 1.481	.955	–0.490 $\pm$ 1.616	.762
	off–on x off–on x 2 – 0	0.524 $\pm$ 0.812	.519	0.986 $\pm$ 1.635	.547	2.067 $\pm$ 1.785	.248
RFFA x Tire Model x User Level	off–on x MF–FT x 1 – 0	0.554 $\pm$ 0.919	.547	0.667 $\pm$ 1.670	.690	0.796 $\pm$ 1.745	.649
	off–on x MF–FT x 2 – 0	1.796 $\pm$ 1.022	.080	0.481 $\pm$ 1.857	.796	0.085 $\pm$ 1.940	.965
SST-RI x AA-RI x Repetition	off–on x off–on x 2 – 1	–1.536 $\pm$ 0.976	.117	–1.228 $\pm$ 1.927	.524	–0.886 $\pm$ 2.069	.669
	off–on x off–on x 3 – 1	–0.606 $\pm$ 1.055	.567	–2.654 $\pm$ 2.076	.202	0.011 $\pm$ 2.223	.996
	off–on x off–on x 4 – 1	–1.718 $\pm$ 1.031	.097	0.614 $\pm$ 2.030	.762	0.854 $\pm$ 2.175	.695

Occupant behaviour, cognition, and Human-Machine Interaction in Level 3 Autonomous Vehicle driving simulator

Ali Olayan^{1,2}, Jean-Rémy Chardonnet², Baptiste Sandoz¹

(1) Arts et Metiers Institute of Technology, IBHGC, Paris, France : {ali.olayan, baptiste.sandoz}@ensam.eu

(2) Arts et Metiers Institute of Technology, LISPEN, Chalon-sur-Saône, France : jean-remy.chardonnet@ensam.eu

Abstract – This study investigates occupant kinematics, physiological, cognitive, and behavioral responses during braking events in a Level 3 autonomous vehicle (AV) driving simulator. The main objective is to determine how auditory alerts, braking intensity, and occupant activities affect head acceleration, head pitch rotation, and the presence of sternocleidomastoid (SCM) muscle activation. Twelve participants experienced simulated AV rides involving normal and strong braking, with and without auditory alerts. Kinematics and physiological data were collected using inertial measurement units (IMUs) and electromyography (EMG). Results showed a significant interaction between auditory alerts and braking intensity on head acceleration. Strong braking events preceded by auditory alerts produced higher head accelerations than equivalent events without alerts. Conversely, normal braking events with alerts showed reduced head accelerations. Occupants who redirected attention from devices to the road during braking exhibited the highest head accelerations when no alert was provided, suggesting reduced preparedness under these conditions. Head pitch rotation did not differ significantly across conditions, although trends suggested greater rotation when attention shifted from devices to the road. SCM muscle activation varied across individuals and was significantly affected by auditory alerts. Some participants showed double activations, consistent with anticipatory bracing or startle responses, however high EMG noise limited detailed interpretation. Overall, auditory alerts appear to influence occupant preparedness, kinematics, and physiological responses during emergency braking, supporting the need for adaptive AV safety systems.

Keywords: Driving Simulator, Biomechanics, Human-Machine Interaction, Autonomous Vehicles, Occupant Behaviour & Cognition

Copyright: © 2026 by the authors.

Licensee Driving Simulation Association.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.82157/dsa/2026/2>

Introduction

Advanced driver assistance systems (ADAS) and higher levels of vehicle automation are expected to reduce crash frequency and severity by limiting human error; nevertheless, collision scenarios and emergency maneuvers remain relevant safety concerns, while new occupant configurations and behaviors simultaneously challenge conventional restraint and Human-Machine Interaction (HMI) design (Filatov et al., 2019; Litman, 2014; ONISR, 2020). Because occupants are partially relieved from continuous driving tasks and control of the vehicle, they may adopt non-driving activities such as reading, working, or phone use, which can lead to out-of-position postures (OOP) potentially compromising their protection in emergency manoeuvres (Grébonval et al., 2021a, 2021b; Östling et al., 2017; Subit et al., 2017). At the same time, HMI design is recognized as a key safety enabler, as AVs must clearly communicate their status, limits, and requests to intervene using visual,

auditory and haptic sensory feedback (Akamatsu et al., 2013; Kern and Schmidt, 2009; SAE J3016, 2021; Wood et al., 2019; Zhang et al., 2019). To study HMI in SAE Level 3 and above, driving simulators are often used due to their low-risk, cost-effective, and controlled environments (Kemeny et al., 2020). These simulators enable personalized cognitive exploration of occupants, aiming to understand their behaviour, interaction strategies, and cognitive responses in dynamic simulated environments. The flexibility of virtual environments allows testing various scenarios, such as emergency braking or audible alerts, crucial for evaluating HMI effectiveness (Schöner and Morys, 2016). When designing AV experiments, it is essential for subjects to become accustomed to the HMI. Simulator HMI should ensure security, reduce stress through instructive and entertaining features, provide timely alerts, and facilitate environmental awareness. Additionally, simulator sickness, caused by sensory conflict between vestibular and visual systems, must be considered due to potential impacts on behaviour and cognition (Bos et al., 2021; Kemeny et al., 2020;

Rangelova and Andre, 2018; Valk et al., 2010). Studies indicate that once subjects feel secure, familiar, and comprehend the task, their behaviour and cognition closely reflect reality, enhancing simulation validity (Kemeny et al., 2020; Lucas et al., 2020; Riegler et al., 2021; Williams et al., 2020).

This research is part of a major evolution towards increased vehicle automation, addressing crucial road safety issues and the associated risks, costs, and acceptance of autonomous technologies. It focuses on understanding, through simulation and virtual reality, the cognitive mechanisms at play in interactions between a Level 3 AV and its driver/occupant. This includes determining the influence on strategies to stabilize the human body in critical situations, such as emergency braking, and identifying suitable HMI that can adapt in real-time to ensure optimum safety and minimize injuries. It is essential to validate simulation studies against real-world conditions to avoid perceptual biases. Currently, safety systems are evaluated using impact dummies representing average human morphologies, which have improved vehicle safety but do not reflect the diversity of occupants (Grébonval et al., 2021a). These dummies, seated in standardized positions, fail to account for behavioural aspects, OOP movements, and muscle contractions in Level 3+ AVs (Grébonval et al., 2021b; Subit et al., 2017). Understanding the HMI between the occupant and the AV is crucial for improving occupant safety by studying their behaviour and cognition. We expect that occupant activity preceding AV braking will influence head-neck kinematics and physiological responses, including head acceleration, head movement, and the presence of sternocleidomastoid (SCM) muscle activation. Electromyography (EMG) was included alongside head-neck kinematics to detect SCM activity during braking, as muscle recruitment often accompanies bracing or stabilization (Fuentes Del Toro et al., 2020; Homayounpour et al., 2021). However, due to susceptibility to motion artifacts, crosstalk, electrical noise, and the difficulty of distinguishing between voluntary, startle, and reflexive responses, surface EMG requires cautious interpretation (Homayounpour et al., 2021; Roberts and Gabaldon, 2008). Consequently, EMG served here as a conservative indicator of the presence of neuromuscular activation, rather than a precise measure of timing, magnitude, or physiological origin. Additionally, we hypothesize that HMI systems providing timely and clear auditory alerts will influence occupant head-neck dynamics during braking maneuvers.

The main objective of the presented paper is to better understand HMI, behaviour, and cognition of occupants in a Level 3 AV driving simulator during emergency automatic braking. To achieve the objectives, we examined the impact of AV braking maneuvers on kinematics and physiological

parameters, including head movement and acceleration, muscle activation, as well as occupant activity. We expect that individual participants will exhibit significantly different kinematics and physiological responses, despite being subjected to uniform experimental conditions. Moreover, based on prior findings on AV occupant activities, anticipatory postural adjustments, and auditory cueing (Liang et al., 2020; Östling and Larsson, 2019; Santos-Cuadros et al., 2021), we predict that the occupant activities performed by participants during the experiment will align with those documented in the literature for short duration drives in SAE Level 3 autonomous vehicles.

Methodology

Participants

12 participants (8 women, 4 men; mean age = 33.5 ±12.2 years; mean weight = 66.2 ±10.7kg; mean height 1.69 ±0.07m; mean BMI = 23 ±2.2kg/m²) took part in the study. All but one held a driver's license for over 5 years, with weekly driving experience ranging from zero to five hours. Participants had normal or corrected vision, no vestibular/ocular abnormalities, and no prior experience with the current simulator setup. Procedures adhered to the Declaration of Helsinki, with all participants providing written informed consent after being instructed to abstain from caffeine, alcohol, and recreational drugs for 24 hours pre-experiment.

Apparatus

The experiment used the SAAM (Simulateur Automobile Arts et Métiers) dynamic driving simulator (Figure 1), based on an instrumented Renault Twingo II cockpit that preserved the original vehicle appearance (Colombet et al., 2011). The cockpit was mounted on a MOOG 6DOF 2000E platform with six degrees of freedom and electric actuators operating at 30–60 Hz. Visual stimuli were displayed on a hemicylindrical screen (height: 1.20 m; field angle: 150°) using three DLP projectors (Projection Design F10SX+; 1400 × 1050 pixels, 60 Hz). A six-computer cluster controlled the real-time graphics, driving environment, and visual, auditory, and haptic interfaces. Platform motion constraints are reported in Table 1. A classical Stewart motion cueing algorithm (Stewart, 1965) converted horizontal accelerations into vertical forces within simulator limits, using pre-recorded simulation data to reproduce g-forces in the vertical plane. Participants wore two inertial measurement units (Xsens MTx), one on the head and one on the first thoracic vertebra (T1), while a reference IMU was fixed to the vehicle floor. Orientation and acceleration were sampled at 100 Hz. Right SCM activity was recorded at 2000 Hz using a wireless EMG sensor (BioNomadix). Participant activities

were also recorded with a GoPro Hero3+ camera at 720p and 30 fps.



Figure 1: The SAAM dynamic driving simulator

Table 1: The SAAM dynamic driving simulator Kinematic and Dynamic Constraints

D.O.F	Displacement	Velocity	Acceleration
Pitch	±22 deg	±30 deg/s	±500 deg/s ²
Roll	±21 deg	±30 deg/s	±500 deg/s ²
Yaw	±22 deg	±40 deg/s	±400 deg/s ²
Heave	±0.178 m	±0.30 m/s	±0.5 g
Surge	+0.259,-0.241 m	±0.5 m/s	±0.6 g
Sway	±0.259 m	±0.5 m/s	±0.6 g

Simulator Scenario and Protocols

To assess occupant behaviour and physiology, a 13-minute ride in a Level 3 AV simulator was conducted. The virtual environment was built using Unreal Engine and CARLA (Town 4), an open-source autonomous driving simulator built to provide environments, vehicles, and scenario data (Dosovitskiy et al., 2017). The simulated environment comprises a compact urban area amidst lakes and mountainous terrain, featuring a network of streets and intersections surrounded by commercial buildings and residential houses. A ring road, resembling a "shape of 8," encircles the town. The central intersection incorporates underpass and overpass elements, along with circular access roads (Figure 2). CARLA was also used to extract virtual vehicle telemetry data, including horizontal acceleration during the simulated route. acceleration data were provided to the simulator motion-cueing algorithm, which reproduced acceleration and braking cues using the available 6DOF platform motions. Since the platform has limited horizontal travel, horizontal acceleration was represented through a combination of limited translation and pitch-based tilt coordination, rather than by directly reproducing the full horizontal acceleration. This provided physical cues consistent with the visual braking events while respecting the simulator's mechanical constraints. The AV journey began within the town, traversed a highway section, and returned to the urban environment, while the route configuration conveyed the impression of entering a different town. The braking events occurred during this automated route in traffic-relevant situations, including interactions with the

road environment, surrounding traffic, and potential hazards such as pedestrian-crossing situations. These events were classified according to their acceleration level as normal braking (0.2–0.4 g) or strong braking (0.4–0.6 g), with a subset preceded by an auditory alert. These operational ranges were consistent with values reported in the literature (Fitch et al., 2010; Wang et al., 2005). This traffic context is relevant because the visual scene, the expectedness of the event, and the presence or absence of an alert may influence participant preparedness and therefore affect head acceleration, head pitch rotation, and SCM muscle response. The AV operated at Level 3 autonomy, independently perceiving surroundings and making decisions, such as lane changing, without participant intervention throughout the experiment. Participants could freely engage in typical occupant activities, including using their phones or the smart tablet in the vehicle, or listening to the radio, which was maintained at a constant volume. Table 2 details the recorded 16 braking events (8 normal, 8 strong), some accompanied by audible alerts. Before the experiment, participants were equipped with the previously described physiological sensors (Figure 3). Upon completing the experiment, participants exited the driving simulator and proceeded to a desk to complete a Simulator Sickness Questionnaire (Kennedy et al., 1993), which was used to assess their levels of simulator sickness following the simulation.

Table 2: Summary of Braking Events Recorded During the Experiment

Braking event count	Acceleration Range (g)	Designation type	Including Audible Alert
8	0.2 – 0.4	Normal	2
8	0.4 – 0.6	Strong	1



Figure 2: Top-down view of the virtual road network in the CARLA Town 4 Map using Unreal Engine

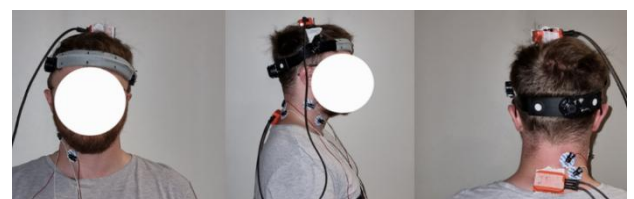


Figure 3: Participant outfitted with kinematics and physiological sensors

Studied Parameters

The selected parameters were intended to capture complementary aspects of occupant response during automated braking events in the Level 3 AV driving simulator. Head acceleration quantified the magnitude of braking-induced head motion, while head pitch rotation characterized angular head-neck motion in the sagittal plane. SCM muscle activation was included to determine whether cervical muscle activity occurred during braking, providing a physiological measure alongside the kinematic variables. Event-specific participant activity was also considered, as the task being performed at the time of braking may influence posture, attention, and preparedness, thereby affecting occupant responses. The analysed variables were head acceleration, head pitch rotation, SCM muscle activation, and participant activity. IMU-derived acceleration and pitch rotation data were left unfiltered. The IMU reference frames were numerically aligned with the simulator coordinate system (Figure 1), with +X directed forward, +Y to the participant's right, and +Z toward the floor. These measures were used to examine the effects of braking intensity and audible alerts on occupant kinematic, physiological, and behavioural responses, including their interaction with participant activity. Detailed processing and calculation procedures for all variables are provided in Table 3.

Table 3: Detailed Summary of data processing parameters and classification criteria during the experiment

Parameter	Description
Head Acceleration	Difference between highest and lowest values during each braking event
Head Pitch Rotation	Difference between highest and lowest pitch values of the head movement
SCM Muscle Contraction	0 for no activation, 1 for one or double muscle activation (once after hearing the alert and again during the braking event).
Participant Activities	A: observing the road, B: using devices, C: using devices before the braking event and then looking at the road when event starts or when audible alert is activated
Braking Intensity	Normal or strong braking, categorized by audible-alert condition (with alert vs. without alert)

Statistics

N-way ANOVA tests were performed in MATLAB using “anovan” to assess the effects of participant, auditory alert, braking intensity, and participant activity on head acceleration, head pitch rotation, and SCM muscle activation. Participant was included as a categorical factor to quantify inter-individual variability in this exploratory event-level analysis. The ANOVA models included main effects and two-way interaction terms only. Because activity was self-selected, the design was unbalanced; therefore, MATLAB's default Type III sums of

squares were used. Since each participant contributed multiple braking events, the data had a repeated-measures structure; however, the analysis was conducted as an event-level factorial ANOVA with participant included as a categorical factor, so the findings should be interpreted as exploratory condition-level comparisons. The significance threshold was set to 0.05, and no post-hoc test was performed because the aim was to assess the presence of significant overall effects.

Results

For all categories, three types of non-driving activities were identified: looking at the road ahead (A), using devices such as a phone or tablet (B), and using devices before the braking event and then looking at the road when event starts or when audible alert is activated (C). In the following parts, a number appended to the activity letter signifies the observed SCM activity change. It is good to note that 52% of the EMG data exhibited excessive noise, leading to its exclusion from detailed analysis. Excessive EMG noise was defined as artefactual or simulator-related interference that prevented reliable SCM interpretation, including signals exceeding the participant's maximum voluntary response. No clear systematic pattern of noisy recordings was observed across participants. However, for the purpose of calculating the percentage of event occurrences, the total event count still included these instances to ensure an accurate representation of the overall event distribution, despite the exclusion of noisy data from further analysis. For normal and strong braking with alert, Activity B (engaged with devices at braking) never occurred since all participants looked at the road when an alert was active.

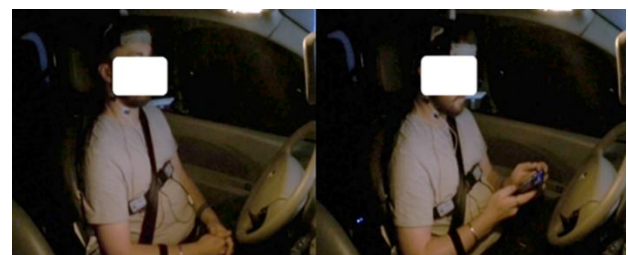


Figure 4: Occupant engaged in Activity: A (left) and B (right)

Head Acceleration

Figure 5 illustrates occupants' head acceleration (g) during three activities: Activity A, Activity B, and Activity C, with and without an audible alert. Each panel shows the distribution of head acceleration values with overlaid individual data points. The coloured regions indicate confidence intervals: red for “with alert” and yellow for “without alert.” The black line and asterisk mark the median and mean, respectively. Figure 5(a) shows strong-braking data, and Figure 5(b) shows normal-braking data. Under strong braking, head acceleration was higher with an

audible alert than without. Without an alert, acceleration varied by activity: it was slightly higher when occupants continuously watched the road (Activity A) than when they used devices at braking onset (Activity B). The highest acceleration without an alert occurred when occupants initially used devices but shifted attention to the road as braking began or an audible alert activated (Activity C). In normal braking, head acceleration was consistent for Activity A with or without an alert; however, in Activity C, acceleration was lower with an alert. Under normal braking with alerts, Activities A and C showed similar acceleration values, whereas without alerts, Activity C produced higher acceleration than Activities A and B. ANOVA showed a significant effect of braking intensity on head acceleration ($p < 10^{-5}$) and that individual participant influences head acceleration under uniform conditions ($p = 0.002$). The interaction between audible alert and braking intensity was also significant ($p = 0.021$), indicating that the effect of braking intensity on head acceleration depended on whether an audible alert preceded braking.

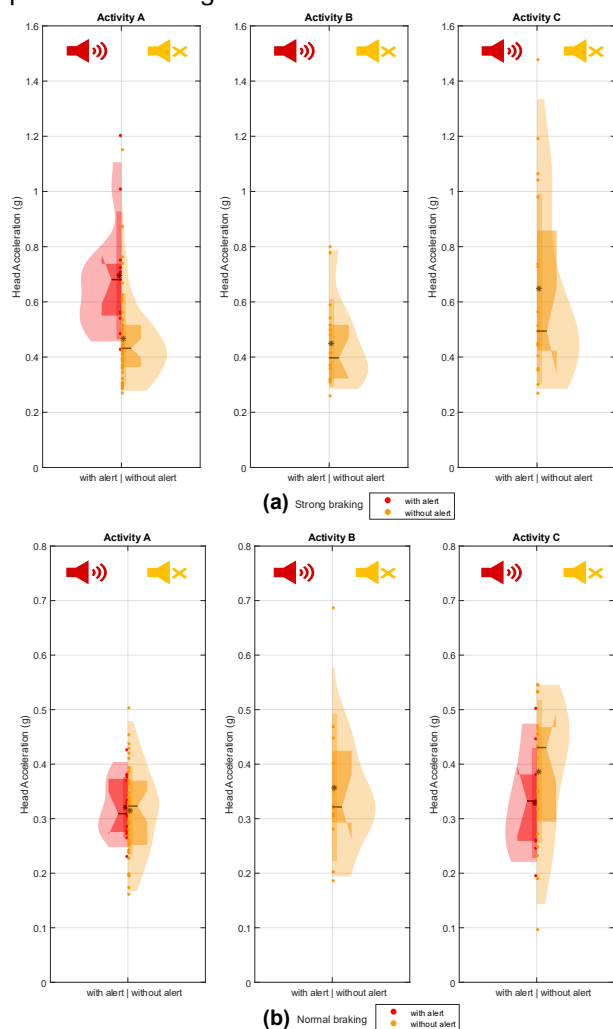


Figure 5: Head acceleration during strong braking (a) and normal braking (b) across different activities with and without audible alert. Activity: (A) looking at the road; (B) using a device; (C) using a device before braking or alert, then looking at the road when it starts.

Head Pitch Rotation

Figure 6 shows occupants' head pitch rotation (degrees) during Activities A, B, and C, with and without an audible alert. Each panel presents the distribution with overlaid individual data points. The coloured regions show confidence intervals: red for "with alert" and yellow for "without alert." The black line and asterisk indicate the median and mean, respectively. Figure 6(a) shows strong braking, and Figure 6(b) shows normal braking. Under strong braking, head pitch rotation was greater with an alert than without in Activity A. Without alerts, rotation was similar in Activity A (road-focused) and Activity B (device-focused), while Activity C, involving an attention shift from device to road during braking or alert, showed the highest rotation. During normal braking, head pitch rotation in Activities A and C was slightly lower with an alert. With alerts, Activity C exceeded Activity A. Without alerts, rotation was highest in Activity C and lowest in Activity B. N-way ANOVA found no statistically significant group differences.

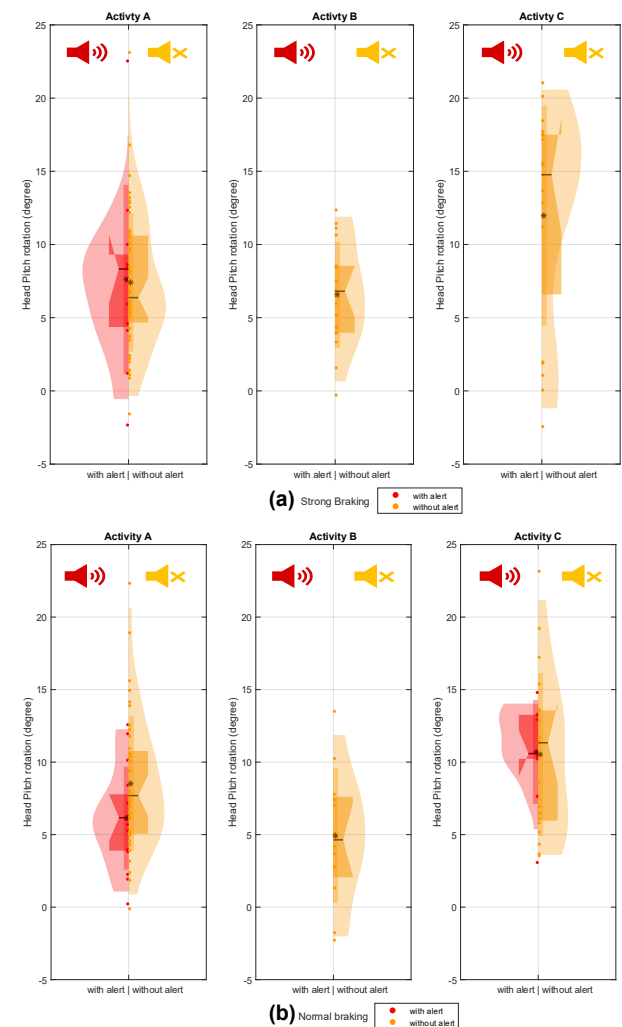


Figure 6: Head pitch rotation during strong braking (a) and normal braking (b) across different activities with and without audible alert. Activity: (A) looking at the road; (B) using a device; (C) using a device before braking or alert, then looking at the road when it starts.

SCM Muscle Activation

Table 4 systematically compares the occurrence rates of muscle activation across different braking scenarios and conditions, highlighting the relationship between occupant activity and muscle activation patterns observed through EMG recordings. On the first column, the first letter of each cell denotes the occupant's activity during the recorded event: (A) looking at the road; (B) using a device; (C) using a device before braking or alert, then looking at the road when it starts. The number following the activity letter represents the observed physiological response: _0 indicates the absence of muscle activation, and _1 denotes muscle activation during or after the event. Table 4 highlights the impact of alert presence on the occurrence rates during strong, and normal braking scenarios under various experimental conditions. Occurrence counts are reported relative to the full set of 192 braking events, including events with noisy EMG data

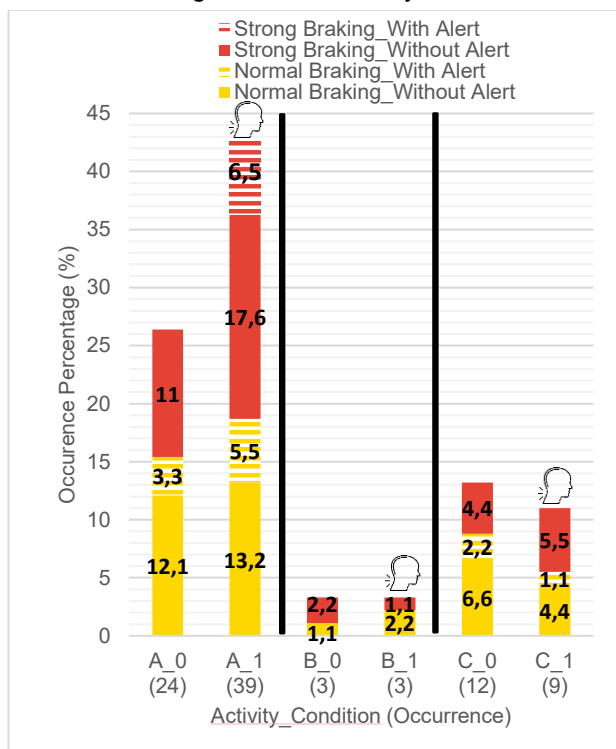


Figure 7: Proportion of SCM Muscle Activation Occurrences Categorized by Braking Intensity and Presence of Audible Alert. Colors indicate braking intensity: Strong (red) and Normal (yellow), with alert (striped) or without (solid). On the x-axis, each bar's first letter shows the occupant's activity: (A) looking at the road, (B) using a device, or (C) using a device before braking or alert, then looking at the road when it starts. A number pair represents the physiological response: _0 for no muscle activation, _1 for activation during or after the event. The number in parentheses shows how often each condition occurred.

Figure 7 displays the percentage distribution of SCM muscle activation occurrences across different braking intensities (Strong, Normal) and the presence or absence of an audible alert. It presents proportions calculated only from the 91 events with interpretable, non-noisy EMG recordings. The x-axis

corresponds to the first column of Table 4, with each bar representing a specific combination of occupant activity and the associated muscle activation response. The most occurrent condition observed was A_1, where occupant experienced muscle activation during or after the event when looking at the road. ANOVA analysis shows that the combined effect of participant and the presence of audible alert on SCM muscle activation is statistically significant ($p = 0.008$). This suggests that the influence of alert on muscle activation depends on individual occupants.

Table 4: Occupant activity and muscle response in relation to braking intensity (with and without an alert): occurrence distribution of muscle activation patterns across various braking scenarios and occupant activity. In the first column, each cell's first letter indicates the occupant's activity: (A) looking at the road, (B) using a device, or (C) using a device before braking or alert, then looking at the road when it starts. A number pair represents the physiological response: _0 for no muscle activation, _1 for activation during or after the event.

Occupant activity & Muscle Response over the total occurrence (192)	Occurrence	Strong braking with alert	Normal braking with alert	Strong braking without alert	Normal braking without alert
A_0	24	0	3	10	11
A_1	39	6	5	16	12
B_0	3	0	0	2	1
B_1	3	0	0	1	2
C_0	12	0	2	4	6
C_1	10	0	1	5	4
Noisy data	101	6	13	46	36

Simulator Sickness

Out of the 12 participants, 11 experienced minimal or no simulator sickness. Only 1 participant reported mild simulator sickness, indicating the presence of symptoms, though mild, which did not significantly impair performance or results (Stewart, 1965).

Discussion

This study investigates the multi-layered nature of occupant responses within Level 3 AV simulator. Kinematics and physiological measures, in conjunction with observed occupant behaviour, provide valuable insights into how individuals respond to various driving events. A multi-faceted approach, incorporating IMU measurements, and SCM muscle activity, provides a picture of occupant kinematics and physiological responses during various braking events. EMG was included to complement kinematic measurements by providing a physiological indicator of neck muscle recruitment during emergency braking, thereby helping to characterize occupant braking responses beyond head motion alone. We sought to explore whether pre-braking alerts could induce a preparatory response, as some participants exhibited double muscle activations that may reflect anticipatory braking or startle-related responses when warned

before braking. Although surface EMG is inherently susceptible to motion artefacts and noise, and distinguishing reflexive from voluntary activation remains difficult in this context, the signal was considered valuable for exploratory interpretation; however, excessive EMG noise ultimately limited detailed conclusions. A complex interplay emerges between external cues, such as audible alerts, the intensity of AV braking, occupant's activity, occupant's head-neck kinematics, and muscle activity. Individual variations in these responses underscore the importance of understanding the degrees of occupant experiences. The occupants own activities during this short drive were similar to what was observed in a previous study by Östling and Larsson, 2019. To the best of our knowledge, this is one of the first studies where it simultaneously examines the impact of Level 3 AV braking on a range of kinematics and physiological parameters, while also considering individual variability and occupant activity. Relevant prior work by Happee and colleagues addressed related dimensions of occupant response, but not within a single empirical Level 3 AV braking study combining head kinematics, SCM activation, and self-selected occupant activity (He et al., 2022). This multi-dimensional approach is crucial for gaining a general understanding of occupant HMI, behaviour, cognition, and safety within AVs, while answering the research question: How do pre-brake conditions and occupant activities in an AV influence the body's response to emergency braking (Strong braking) events? The traffic scenario should also be considered when interpreting the head-motion results, since braking events embedded in urban, intersection, or hazard-related contexts may provide different levels of visual anticipation and preparedness than braking events occurring in less complex road sections.

Our results demonstrated that the effect of braking intensity on the head acceleration is affected by the presence of an audible alert. Specifically, during instances of strong intensity braking events where alerts preceded the braking events, participants exhibited an increased head acceleration compared to the events without a preceding alert. However, during instances of normal braking intensity, an inverse effect was observed, which is more in line with previous studies performed in a real car and level 3 environment (Mackenzie et al., 2022). When looking at the road, head acceleration remained unchanged, whereas when looking at a device then returning attention to the road, it was 0.1 g lower when an alert proceeded the braking event. This suggests that in the event of strong braking either the preparatory effect triggered by audible alerts might lead to anticipatory postural changes potentially influencing head movement, that participants were unprepared for intense braking due to an expectation of normal braking, or a combination of both factors.

This is in accordance with Liang et al., 2020, where they demonstrated that through repetitive exposure to external perturbations, participants were able to rely solely on auditory cues to generate adequate anticipatory postural adjustments. Conversely, the absence of alerts before braking resulted in similar head acceleration levels when looking at the road and looking at a device across both braking intensities, while looking at a device then returning attention to the road exhibited higher head acceleration. This implies that occupants engaged in this latter activity may exhibit a more passive response, indicating lesser preparedness for the force of the braking maneuver. In this sense, the higher acceleration observed with alerts may reflect a different response strategy to the perturbation rather than reduced preparedness. The effect of alerts on SCM muscle activation was influenced by individual differences. This further supports the finding by Santos-Cuadros et al., 2021, where they showed that during emergency braking with a warning, there was a notable difference observed in the SCM muscle response according to gender and age. Some occupants exhibited double SCM activation when alerts were present, regardless of braking intensity. This may represent a heightened startle response or anticipatory bracing. This double muscle activation was similarly observed in a study by Krašna and Đorđević, 2020, which measured upper trapezius muscle activity during a SLED low-severity impact experiment. The study found that in trials where participants were aware of the impending impact, a startle reflex or anticipatory bracing response was evident. Nevertheless, most occupants in our experiment demonstrated SCM activation primarily in response to high-intensity events, even without alerts. These findings highlight the importance of understanding individual variability in muscle responses to AV events, especially with regards to potential comfort and safety implications. EMG findings should be interpreted with caution, the high degree of noise in our EMG data (52%), unfortunately, limited robust conclusions about muscle activation patterns. This could be attributed to factors such as sensor placement, individual differences in skin conductivity, electrical noise, or movement artifacts. Future research must prioritize optimizing EMG signal quality to gain clearer insights into the muscular responses of AV occupants. Possible strategies include refined sensor placement techniques less prone to motion artifacts and reducing electrical noise. Our N-way ANOVA identified a statistically significant effect of the 'participant' variable on head acceleration. This emphasizes the inherent variability among individuals in their responses to identical AV events. Understanding this variability is crucial for developing AV systems that adapt to diverse occupant needs and preferences. Factors such as prior driving experience, simulator susceptibility, and

baseline physiological differences could all contribute to the observed variability. Our findings suggest that the design of auditory alert systems in AV may impact occupant experience and could have safety implications. The interplay between alerts, braking intensity, and individual responses warrants further investigation to develop adaptive systems that cater to diverse occupant preferences and sensitivities. Furthermore, the observed variations in head movement raise questions about potential risk factors in different AV scenarios.

Limitations

Several limitations encountered in this study should be addressed in future experiments. First, electrical noise from the driving simulator affected EMG data accuracy; this could be mitigated by using advanced EMG sensors with better noise reduction. Moreover, covering the simulator's side windows to maintain immersion limited environmental realism. This could be improved by employing modern mixed-reality headsets, such as the Apple Vision Pro or VARJO XR-4, or using a more advanced simulator with a 360-degree screen. Another limitation was the uneven distribution of activity data, which hindered the ANOVA analysis of occupant activity effects. Because activities were self-selected, some activity-by-condition combinations were rare or absent. This limits interpretation of interactions involving alerts, device use, and attention shifts. Furthermore, although participant was included as a factor in the ANOVA, the event-level analysis did not explicitly model within-subject dependence across repeated braking events, which should be considered when interpreting the statistical significance of the findings. Conducting follow-up experiments with clear instructions for specific activities could address this. Moreover, participant was included as a categorical factor to test whether inter-individual differences contributed to the measured responses under the experimental conditions. However, we acknowledge that treating participant as a random effect in a mixed-effects model would be more appropriate for fully accounting for the repeated-measures structure of the data. Furthermore, only 12 experiments were performed due to time and logistical constraints; expanding the sample size in future studies would improve reliability and generalizability. Additionally, the analysis focused on peak-to-peak head acceleration and pitch rotation, which quantify the magnitude of the response but does not fully capture its temporal evolution. Future work should examine the correlation between head acceleration and pitch rotation, as well as their timing relative to auditory alert onset and braking onset. Such analysis would help determine whether observed head responses reflect anticipatory postural adjustments, startle-related reactions, or passive head motion induced by the braking maneuver. Finally, the validity of driving simulators in this study is critical to ensure

kinematics, physiological, and behavioural responses accurately reflect real-world driving scenarios. While simulators offer a controlled, low-risk environment with repeatable events, challenges exist in replicating real-world conditions, such as fidelity of motion cues, visual immersion, and simulator sickness. In particular, simulator decelerations and inertial cues do not fully reproduce on-road braking and may affect the measured kinematic and physiological responses (Bos et al., 2021; Kemeny et al., 2020). These findings therefore require validation in real-vehicle studies. Previous research indicates simulators can mimic real-world responses like braking, but individual variability, sensory mismatches, and environmental artificiality might introduce perceptual biases. This study attempted to align the simulator with reality using pre-recorded braking maneuvers and standardized alerts, but full validation requires further refinement of visual and haptic feedback mechanisms.

Conclusion

This study presents an inclusive analysis of the head-neck kinematics, SCM muscle activity, and behavioural responses of occupants in AVs during braking maneuvers. Our findings underscore the impact of auditory alerts, braking intensity, and individual differences on head acceleration, head rotation, and SCM muscle activity. The interaction between auditory alerts and braking intensity notably influences head acceleration and SCM muscle activation, suggesting that such alerts could enhance preparedness and potentially improve safety by enabling anticipatory postural adjustments. However, the variability in responses highlights the need for AV systems to account for individual occupant differences. Additionally, our results indicate that individual variability affects kinematics and physiological responses during AV braking. This variability can be attributed to factors such as prior driving experience, susceptibility to simulator sickness, and baseline physiological states. Therefore, designing AV systems that can adapt to this individual variability is crucial. Implementing personalized settings or adaptive features that cater to the unique needs of each occupant could optimize comfort and safety. The limitations encountered with EMG data noise underline the necessity for improved data acquisition methods. Future research should focus on enhancing sensor placement and signal processing techniques to obtain more reliable measures of muscle activation, which are essential for understanding the full scope of physiological responses in AV environments. In conclusion, our study contributes to the growing body of knowledge on HMI within AVs. It highlights the importance of considering both universal and individual factors in the design of AV systems to enhance occupant experience and safety. Continued research in this

area is imperative to refine AV technologies, making them more adaptable and responsive to the diverse needs of occupants, thereby facilitating wider acceptance and use of autonomous transportation solutions. These findings underscore the complexity of occupant head-neck kinematics and SCM muscle responses during AV braking, highlighting the need for adaptive systems that can accommodate individual variability in response to auditory alerts and driving conditions. While auditory cues and braking intensity were shown to influence head acceleration and muscle activation, the variability in occupant responses calls for more personalized approaches to AV safety systems. These insights provide a foundation for further investigation into how specific occupant activities influence kinematics, physiological, and cognitive responses during critical driving scenarios, such as emergency braking.

References

- Akamatsu, M., Green, P., Bengler, K., 2013. Automotive Technology and Human Factors Research: Past, Present, and Future. *International Journal of Vehicular Technology* 2013, 1–27.
- Bos, J.E., Nooij, S.A.E., Souman, J.L., 2021. (Im)possibilities of studying carsickness in a driving simulator.
- Colombet, F., Paillot, D., Mérienne, F., Kemeny, A., 2011. Visual Scale Factor for Speed Perception. *Journal of Computing and Information Science in Engineering* 11, 041010.
- Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., Koltun, V., 2017. CARLA: An Open Urban Driving Simulator. 1st Annual Conference on Robot Learning 16.
- Filatov, A., Scanlon, J.M., Bruno, A., Danthuri, S.S.K., Fisher, J., 2019. Effects of Innovation in Automated Vehicles on Occupant Compartment Designs, Evaluation, and Safety: A Review of Public Marketing, Literature, and Standards. Presented at the WCX SAE World Congress Experience, pp. 2019-01-1223.
- Fitch, G.M., Blanco, M., Morgan, J.F., Rice, J.C., Wharton, A., Wierwille, W.W., Hanowski, R.J., 2010. Human performance evaluation of light vehicle brake assist systems: Final report. National Highway Traffic Safety Administration.
- Fuentes Del Toro, S., Santos-Cuadros, S., Olmeda, E., San Román, J.L., 2020. Study of the Emergency Braking Test with an Autonomous Bus and the sEMG Neck Response by Means of a Low-Cost System. *Micromachines* 11, 931.
- Grébonval, C., Beillas, P., Wang, X., 2021a. Experimental investigation of preferred seating positions and postures in reclined seating configurations 7.
- Grébonval, C., Trosseille, X., Petit, P., Wang, X., Beillas, P., 2021b. Effects of seat pan and pelvis angles on the occupant response in a reclined position during a frontal crash. *PLoS ONE* 16, e0257292.
- He, X., Stapel, J., Wang, M., Happee, R., 2022. Modelling perceived risk and trust in driving automation reacting to merging and braking vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour* 86, 178–195.
- Homayounpour, M., Gomez, N.G., Vasavada, A.N., Merryweather, A.S., 2021. Cervical Muscle Activation Due to an Applied Force in Response to Different Types of Acoustic Warnings. *Ann Biomed Eng* 49, 2260–2272.
- Kemeny, A., Chardonnet, J.-R., Colombet, F., 2020. Getting Rid of Cybersickness: In Virtual Reality, Augmented Reality, and Simulators. Springer International Publishing, Cham.
- Kennedy, R.S., Lane, N.E., Berbaum, K.S., Lilienthal, M.G., 1993. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology* 3, 203–220.
- Kern, D., Schmidt, A., 2009. Design space for driver-based automotive user interfaces. In: *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Presented at the AutomotiveUI '09: International Conference on Automotive User Interfaces and Interactive Vehicular Applications, ACM, Essen Germany, pp. 3–10.
- Krašna, S., Đorđević, S., 2020. Estimating the Effects of Awareness on Neck-Muscle Loading in Frontal Impacts with EMG and MC Sensors. *Sensors* 20, 3942.
- Liang, H., Kaewmanee, T., Aruin, A.S., 2020. The role of an auditory cue in generating anticipatory postural adjustments in response to an external perturbation. *Exp Brain Res* 238, 631–641.
- Litman, T., 2014. A New Transit Safety Narrative. *JPT* 17, 114–135.
- Lucas, G., Kemeny, A., Paillot, D., Colombet, F., 2020. A simulation sickness study on a driving simulator equipped with a vibration platform. *Transportation Research Part F: Traffic Psychology and Behaviour* 68, 15–22.
- Mackenzie, J., Dutschke, J., Di Loreto, C., Forrest, M., Van Den Berg, A., Merienne, F., Chardonnet, J.-R., Sandoz, B., 2022. Human Head and Neck Kinematics during Autonomous and Human Braking in Three Initial Head Positions. *SAE Int. J. Trans. Safety* 10, 09-10-02-0007.
- ONISR, 2020. La sécurité routière en France Bilan de l'accidentalité de l'année 2020.
- Östling, M., Larsson, A., 2019. Occupant Activities And Sitting Positions In Automated Vehicles In China And Sweden 10.
- Östling, M., Sunnevång, C., Svensson, C., Kock, H.-O., 2017. Potential future seating positions and the impact on injury risks in a Learning Intelligent Vehicle (LIV) 16.
- Rangelova, S., Andre, E., 2018. A Survey on Simulation Sickness in Driving Applications with Virtual Reality Head-Mounted Displays. *Presence: Teleoperators and Virtual Environments* 27, 15–31.
- Riegler, A., Riener, A., Holzmann, C., 2021. A Systematic Review of Virtual Reality Applications for Automated Driving: 2009–2020. *Front. Hum. Dyn.* 3, 689856.
- Roberts, T.J., Gabaldon, A.M., 2008. Interpreting muscle function from EMG: lessons learned from direct measurements of muscle force. *Integrative and Comparative Biology* 48, 312–320.
- SAE J3016, 2021. (R) Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. *SAE Intl. J. CAV* 3, 12-03-01-0003.
- Santos-Cuadros, S., Fuentes Del Toro, S., Olmeda, E., San Román, J.L., 2021. Surface Electromyography Study Using a Low-Cost System: Are There Neck Muscles Differences When the Passenger Is Warned during an Emergency Braking Inside an Autonomous Vehicle? *Sensors* 21, 5378.
- Schöner, H.-P., Morys, B., 2016. Dynamic Driving Simulators. In: *Winner, H., Hakuli, S., Lotz, F., Singer,*

- C. (Eds.), Handbook of Driver Assistance Systems. Springer International Publishing, Cham, pp. 177–198.
- Stewart, D., 1965. A Platform with Six Degrees of Freedom. *Proceedings of the Institution of Mechanical Engineers* 180, 371–386.
- Subit, D., Vézin, P., Laporte, S., Sandoz, B., 2017. Will Automated Driving Technologies Make Today's Effective Restraint Systems Obsolete? *American Journal of Public Health (ajph)*.
- Valk, P., Grech, M., Bos, J., 2010. A Multi-Factorial Analysis Of Human Performance During A 9-Day Sea Trial. *HPAS 2010 International Conference On Human Performance at Sea*, Glasgow, Scotland, United Kingdom.
- Wang, J., Dixon, K.K., Li, H., Ogle, J., 2005. Normal Deceleration Behavior of Passenger Vehicles at Stop Sign-Controlled Intersections Evaluated with In-Vehicle Global Positioning System Data. *Transportation Research Record: Journal of the Transportation Research Board*.
- Williams, B., Garton, A.E., Headleand, C.J., 2020. Exploring Visuo-haptic Feedback Congruency in Virtual Reality. In: *2020 International Conference on Cyberworlds (CW)*. Presented at the 2020 International Conference on Cyberworlds (CW), IEEE, Caen, France, pp. 102–109.
- Wood, M., Wittmann, D., Liu, S., Knobel, C., Syguda, S., Wiltshko, T., Garbacik, N., O'Brien, M., Dannebaum, U., Weast, J., Dornieden, B., 2019. Safety First for Automated Driving.
- 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: Traffic Psychology and Behaviour* 64, 285–307.

Effects of Cognitive Load, Age and Sex on Driver Control Parameters

Víctor Nalda-Tárrega¹, Juan-Manuel Belda Lois¹, Begoña Mateo-Martínez¹

(1) Instituto de Biomecánica de Valencia, Universitat Politècnica de València, Camino de Vera s/n, Valencia 46022, Spain, e-mail: {juanma.belda, victor.denalda, begona.mateo}@ibv.org

Abstract – Understanding how cognitive load influences system-level parameters is essential for characterizing behavioral and physiological dynamics in driving models. In this study, we analyzed multiple positive, skewed variables (T , τ , U_D , a_t , j_{acc} , j_{brk} , k_v) across levels of cognitive load (n -back driving scenarios), sex, and age group. Using generalized linear mixed models (Gamma distribution with log link), we evaluated main effects and interactions while accounting for repeated measures at the subject level. Model selection was performed using AIC and likelihood-ratio tests, and estimated marginal means (EMMs) were used for inference. Results indicate that cognitive load is the primary driver of variability, with age modulating this effect in several variables, while sex shows limited influence. Distributional analyses support the use of Gamma-like models. These findings suggest that system dynamics under cognitive load are non-linear and age-dependent.

Keywords: cognitive load, driving models, non-linear, mixed models, spatial domain.

Copyright: © 2026 by the authors.

Licensee Driving Simulation Association.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.82157/dsa/2026/3>

Introduction

Cognitive workload has long been recognized as a factor influencing the driver's **behavioral** and **physiological responses** (Engström Johan, 2017). However, its impact on *interpretable driver control parameters* and its interaction with demographic factors such as **age** and **sex** (Gwyther Holly, 2012) remain incompletely understood. Many parameters derived from dynamic systems are strictly positive and exhibit skewed distributions, requiring appropriate statistical modeling approaches.

The main objective is to assess whether the **mental workload**, treated as a categorical factor, produces changes in driver model parameters, or whether its influence is secondary compared to inter-driver variability.

Specifically, this study aims to characterize the distributional properties of key system variables; evaluate the effects of cognitive load (n -back), sex, and age; determine whether interactions between these factors significantly influence system dynamics and provide robust inference using mixed-effects modelling.

Methodology

Driver Model

The driver model employed in this study is an adaptation of the **UMTRI driver model** proposed by MacAdam (MacAdam, 2001), (MacAdam, 2003). It relies on preview-based control, internal reference generation, and delayed feedback mechanisms, suitable for representing near- and at-limit vehicle handling.

The specific formulation and parameterization formulation is embedded within a digital-twin framework described in (Belda-Lois, 2025). The present study builds upon that contribution by extending the analysis to a statistical evaluation across users and mental workload conditions.

Sample characteristics

The dataset of the experimentation included **N = 10 participants** from IBV Users Database. The age range was 25-65 years ($\mu=45.8$, $\sigma=13.11$, $<45\text{yrs}=5$ and $>45=5$) and sex balanced (male=5, female=5).

Each participant contributed repeated measurements across multiple n -back levels, for a total of **38 observations**. Participants performed lane-change scenarios similar to (Cantin V, 2009). According to (Fuller, 2005), driving performance is not merely a matter of reflexive reaction, but a continuous process of error minimization. By introducing a memory-intensive n -back task, we create a competition for cognitive

resources between the tactical decision-making level and the operational control level.

Age was divided in two groups, to make this factor consist of two levels, as well as sex, easing the analysis in how it affects the model parameters. Participants prone to motion sickness were excluded during screening (Keshavarz B, 2018).

Design of the experimentation

Testing was performed in the Human Autonomous Vehicle (HAV) dynamic simulator developed by IBV (Nalda-Tarrega, et al., 2022), featuring a 6-DOF moving platform and CARLA open-source software (Dosovitskiy A, 2017). The simulator contains different screens -three main screens to show the driver's FOV, rearview screens and an HMI- and steering wheels and pedals. The platform receives as inputs, the longitudinal accelerations and angular velocities of the *ego* vehicle, detected by a virtual IMU spawned in the vehicle, for every frame. Therefore, it can replicate the movements of the vehicle.



Figure 1: The dynamic driving simulator (HAV) used in the study.

In a highway scenario, a number (1–4), corresponding to the lane number, appeared every 20 seconds on the central screen. Drivers had to change to or maintain the lane number shown n turns ago. Achieving a >75% success ratio incremented the value of n for the subsequent scenario (maximum $n = 4$). Each participant attempted 4 driving scenarios (two completed only 3).

Statistical Analysis

For each driving segment, seven driving parameters were identified using a **spatial-domain** Unscented Kalman Filter (**UKF**), minimizing discrepancies between real and simulated tracking profiles. Identification uncertainty was explicitly represented through the UKF covariance, avoiding reliance on point estimates alone.

The identified parameters are:

- **T.** Look ahead time (s). Time horizon used by the controller to anticipate future states of the system. It determines how far ahead the control strategy plans, influencing smoothness and stability of the motion.
- **Tau.** Internal response delay (s). Represents the response time of the control system. It characterizes how quickly the

system reacts to changes or errors. Lower values lead to faster responses, while higher values produce smoother but slower dynamics.

- **U_d.** Desired speed (m/s). Target longitudinal velocity that the controller aims to achieve and maintain. It defines the steady-state operating point of the system.
- **a_t.** Max. transverse acceleration (m/s²). Upper bound on the maximum acceptable transverse acceleration.
- **j_{acc}.** Max. acceleration jerk (m/s³). Maximum rate of change of acceleration during acceleration phases. It controls how smoothly acceleration is applied, avoiding abrupt changes.
- **j_{brk}.** Max. braking jerk (m/s³). Maximum rate of change of deceleration during braking. It ensures smooth braking transitions and avoids aggressive or uncomfortable deceleration profiles.
- **k_v.** Speed control gain. Gain parameter of the velocity controller.

Table 1: Identified parameters using a spatial -domain UKF.

Parameter	Description
T	Look ahead time
Tau	Internal response delay
U_d	Desired speed
a_t	Max. transverse acceleration
j_{acc}	Max. acceleration jerk
j_{brk}	Max. braking jerk
k_v	Speed control gain

Statistical Distribution Fitting

To handle potential skewness, each parameter was tested against five continuous probability distributions (Lognormal, Gamma, Weibull, Inverse Gaussian, and Exponential) using the Akaike Information Criterion (AIC). Three generalized linear mixed models were subsequently compared: **M0** (Main effects only; assumes identical n-back impact across demographics), **M1** (Partial interaction: n-back × Age Group + Sex), and **M2** (Full interaction: n-back × Age Group × Sex). Pairwise comparisons of the 5-level n-back factor were conducted using Estimated Marginal Means (EMMs) with Holm correction.

Results

Distributional Properties

The different parameters of the driving model, evaluated in the study, have been tested with different fitting distributions. The results of these comparisons are the following:

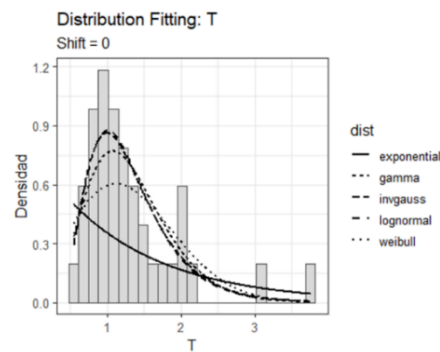


Figure 2: Fitting of the T parameter with different probability distributions.

The following table synthesizes the distribution fitting and mixed-effects model selection results based on the lowest AIC values across all parameters:

Table 2: Fitting models selected for each analysed parameter.

Parameter	Fitting model	AIC	Selected int. model	Model AIC
T	invgauss	58.54	M1_nbackX age	50.5
Tau	weibull	-54.53	M0_main	-37.6
Ud	lognormal	222.78	M0_main	222.9
at	lognormal	67.85	M0_main	74.9
j _{acc}	gamma	48.46	M2_full3way	41.8
j _{brk}	lognormal	41.19	M1_nbackX age	42.8
k _v	invgauss	-65.68	M2_full3way	-54.0

Model Selection

For each parameter, the model that better represented the relationship between the factors was chosen, as showed in the table:

Effects of Cognitive Load, Age and Sex

Once the distribution fitting and the models affecting the correlation of the variables have been chosen, the values for the different parameters have been calculated.

Anticipation/preview time

The parameter T has been analysed considering the cognitive load and the age group, as a strong relationship between these factors was detected. The results obtained on how the value T is affected by these two variables is statistically significant ($p=0.002$).

The anticipation/preview time reduces as the n in the n -back scenarios increases. It also increases for the older age group, who usually anticipate more their own actions when driving.

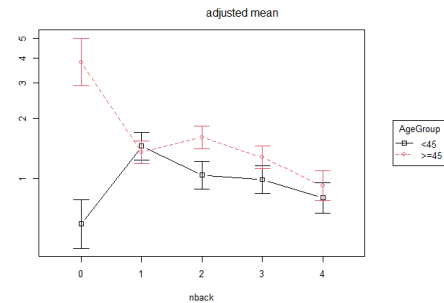


Figure 3: Mean value of T for each n -back scenario and different age groups.

Internal response delay

Tau was analysed considering all of the factors as independent (main effects only), and no statistically significant difference was detected in how the value is affected by the three factors being analysed.

Desired speed

U_d was analysed considering all of the factors as independent (main effects only). The result obtained in how this parameter is affected by the cognitive load n -back is statistically significant ($p=0.0051$).

The difference is visible among the scenario 0-back and the rest of them.

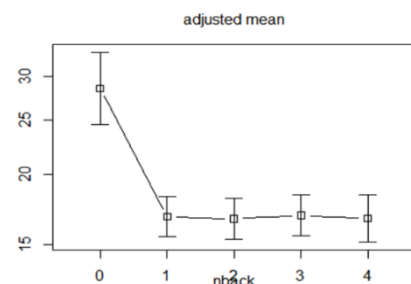


Figure 4: Mean value of U_d for each n -back scenario.

Max. longitudinal acceleration

a_t was analysed considering all of the factors as independent (main effects only), and no statistically significant difference was detected in how the value is affected by the three factors being analysed.

Max. acceleration jerk

j_{acc} was analysed considering a full interaction among the three factors. This parameter is the one that shows a higher statistically significant difference ($p=0.0497$) in the sex factor.

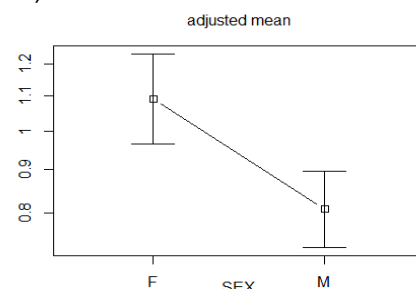


Figure 5: Mean value of j_{acc} for the different sex category.

Max. braking jerk

The parameter j_{brk} has been analysed considering the cognitive load and the age group are related. The results obtained on how the value j_{brk} is affected by these two variables is statistically significant ($p=0.002$).

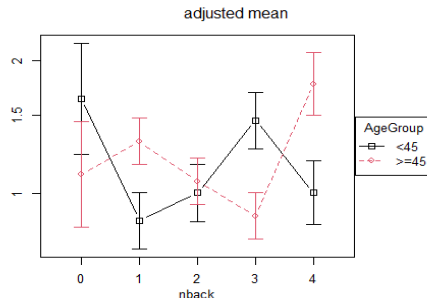


Figure 6: Mean value of j_{brk} for each n -back scenario and different age groups.

Speed control gain

The parameter k_v was analysed considering a full interaction among the three factors. However, the model matrix was found to be rank deficient, indicating that certain factor combinations provided redundant or insufficient information, which reinforces the necessity of focusing on specific interaction contrasts rather than aggregate averages.

Discussion

Main Findings and Interpretation

Cognitive load is the primary determinant of the system behaviour. The n -back factor was the dominant one across the different parameters. There were strong effects observed in T and j_{brk} . More limited or stable effects in Tau , U_d and a_t were detected. Age modulates the response to cognitive load. Sex has limited influence in this dataset. Some variables exhibit complex interaction structure. For the parameters j_{acc} and k_v , full interactions (n -back $\times$ AgeGroup $\times$ Sex) were selected.

The presence of n -back $\times$ GroupAge interactions indicates that aging affects how systems respond to increasing cognitive demand. The selection of full interaction models for j_{acc} and k_v suggests non-linear, multi-factor dependencies.

It is important to also consider the 'learning effect' that the participants acquired when completing the different n -back scenarios. This is clearly reflected in the analysis of the T parameter, where the look ahead time was reduced, independently of the age.

Methodological Implications and Limitations

As variables are not independent, it is likely to reflect shared underlying processes. Gamma-log mixed

models are appropriate for skewed positive variables. Ignoring subject-level variability would bias inference and EMMS are essential for interpreting interaction effects.

The sample $N=10$ participants may not be sufficient to extrapolate the main findings in a larger population. Random slopes were not consistently estimable and age was dichotomized (information loss).

Conclusions

System dynamics **under cognitive load** are **non-linear**, modulated by age and weakly influenced by sex. These findings highlight the **importance** of using **flexible mixed models** and considering **interaction effects** when studying cognitive load.

Acknowledgements

Research activity under BERTHA project (GA101076360) funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

References

- Belda-Lois, J. M., 2025. *A digital twin based driver model for workload aware vehicle control*. s.l., FISITA World Congress.
- Cantin, V, Lavallière, M., Simoneau, M., and Teasdale, N., 2009. Mental workload when driving in a simulator: Effects of age and driving complexity. *Accident Analysis & Prevention*, Volume 41, pp. 763-771.
- Dosovitskiy, A, Ros, G., Codevilla, F., Lopez, A., and Koltun V., 2017. *CARLA: An open urban driving simulator*, Proceedings of the 1st Annual Conference on Robot Learning, PMLR 78:1-16.
- Engström Johan, M. et al., 2017. Effects of Cognitive Load on Driving Performance: The Cognitive Control Hypothesis. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 59(5).
- Fuller, R., 2005. Towards a general theory of driver behaviour. *Accident Analysis and Prevention*, Volume 37, pp. 461-472.
- Gwyther Holly, H. C., 2012. The effect of age, gender and attitudes on self-regulation in driving. *Accident Analysis & Prevention*, Volume 45, pp. 19-28.
- Keshavarz, B, et al., 2018. Comparing simulator sickness in younger and older adults during simulated driving under different multisensory conditions. *Transportation Research Part F: Traffic Psychology and Behavior*, Issue 54, pp. 47-62.
- MacAdam, C. C., 2001. Development of a driver model for near/at limit vehicle handling. *Vehicle System Dynamics*, Issue 35, pp. 311-343.
- MacAdam, C. C., 2003. Understanding and modelling the human driver. *Vehicle System Dynamics*, Issue 40, pp. 101-134.
- Nalda-Tarrega, V. et al., 2022. *Human Autonomous Vehicle (HAV): From sickness prevention to emotional response*. s.l., AHFE: Human Factors in Transportation.

In-car occupant state monitoring using RGB- and IR-images - carsickness as an example

Jelte E. Bos^{1,2}, Susanne Louvenberg¹, Jan L. Souman¹

(1) Integrated Vehicle Safety, TNO Helmond, the Netherlands {jelte.bos, susanne.louvenberg, jan.souman}@tno.nl

(2) Human Movement Sciences, Vrije Universiteit Amsterdam, the Netherlands

Abstract – With growing vehicle automation, occupant state monitoring systems gains interest as well. State monitoring typically concerns multiple states, such as drowsiness, alertness, and impairment. One example of the latter is carsickness, a factor generally considered of great importance for self-driving vehicles. Here we elaborate on that example using RGB images combined with far infrared (IR or thermographic) images of the face. In a lab experiment we induced motion sickness in 14 subjects mapping different regions of interest found in the RGB images onto the matching regions in the IR images. Of all regions considered, the central forehead showed the most significant and consistent effect, with an average estimated reduction of skin temperature of 1.32 °C over the entire range of feeling well up to emesis. Although this study implies an important step forward in monitoring carsickness, showing forehead skin temperature to be a sensitive measure, its specificity should still be determined. The latter, however, requires multiple non-motion stressors to be studied, as well as multiple measures, which poses a major challenge for independent and scientifically sound research that still needs to be carried out.

Keywords: Occupant state monitoring, Motion sickness, Carsickness, Facial images.

Copyright: © 2026 by the authors.

Licensee Driving Simulation Association.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.82157/dsa/2026/4>

Introduction

Now that more and more cameras are being used in cars to monitor the occupants, the question on the validity of what can be measured with these cameras, i.e., their state(s) is becoming increasingly relevant. Occupant state monitoring refers to drivers and passengers and may concern multiple states, such as drowsiness, alertness, and impairment. One example of the latter is carsickness, which typically concerns passengers, considerably more than drivers (Rolnick & Lubow, 1991; Schmidt et al., 2020). This sub-state is of great interest regarding self-driving cars in particular, because over 60% of all occupants, then passengers, are predicted to suffer from carsickness at some moment, while currently, in human-driven cars, this is only about 20%, because less than 40% of all car occupants currently are passengers (Bos et al., 2022).

As an example of state monitoring, we here elaborate on carsickness using RGB images combined with far infrared (IR or thermographic) images of the face. Occupants' faces are generally visible in vehicles by these cameras, which are becoming increasingly common, and deep learning (DL) algo-

ritms allow for determining multiple psychological and physiological measures, here further referred to as biomarkers. Examples are: emotions, blinking, eye movements, pupil size, heart rate (and heart rate variability), oxygenation, skin colour (paleness), and skin temperature. An elaborate review can be found in Louvenberg et al. (2026). As discussed by these authors too, skin temperature in particular is a promising biomarker that can be measured in different facial regions of interest (ROIs).

Our research question accordingly was whether a statistically significant correlation can be shown between facial skin temperature and sickness symptomatology in one or more ROIs.

Study details were published by Louvenberg et al. (2026) and technical details regarding the merging of RGB and IR camera information and image processing by Nijskens et al. (2024). The current paper places these findings within a broader context, i.e., that of sensitivity (the likelihood of finding a true effect), and specificity (the likelihood of not finding a false effect) of camera based biomarkers and in-vehicle occupant state monitoring in general.

Methods

Experimental design

To correlate facial temperatures with sickness ratings, we used a within subjects design with two conditions presented in a fixed order: 20 minutes without motion, the 'stationary condition', and after a 5 min break, 20 minutes of off-vertical axis rotation (OVAR, see below), the 'rotary condition'. During both conditions, subjects were continuously observed by an RGB camera and a far infrared (thermographic) camera (also see below). Sickness was rated every minute in both conditions using the Motion Illness Symptoms Classification scale (MISC, again, see below).

OVAR

Off-vertical axis rotation was used as an analogy for creating carsickness. To that end, the axis of rotation of a rotation chair was tilted by 15 degrees (see Figure 1). As a result, the gravitational acceleration projects onto the subject's 'horizontal' or transverse plane with a magnitude of $9.81 \cdot \sin(15^\circ) = 2.54 \text{ m/s}^2$ with a concomitant reduction of the vertical component to 9.48 m/s^2 . The magnitude of the transverse acceleration is about that of normal car acceleration and deceleration. During rotation, this transverse component rotates about the subject opposite the angular motion. A cabin was mounted around the subject with shields preventing any view on the lab space. As a result the subject's perception of angular motion typically vanishes within a minute, due to the functioning of the semicircular canals effectively acting as high pass-filters. Here we used an angular velocity of $60^\circ/\text{s}$, resulting in a periodical motion with a frequency of 0.17 Hz, close to the peak of sickness incidence, also in the horizontal plane (Bos et al., 2024). Moreover, this stimulus has been used before, showing a gradual increase of symptoms comparable to that of carsickness (Bos, 2015). Active ventilation kept the temperature within the cabin constant.

RGB and IR images

As presented in more detail by Nijskens et al. (2024), RGB images were taken with a 1936×1216 resolution. IR images ($8\text{--}14 \mu\text{m}$ spectral range and a sensitivity of 0.04°C) were taken with 640×512 resolution. Both cameras recorded at 30 fps, which allowed easy synchronisation. The two cameras were housed together in one case with their lenses positioned as close together as possible, and firmly attached to the rotation chair. The cameras were positioned approximately 60 cm from the face, for technical reasons slightly left from and, on purpose below the subject's line of sight at an angle of about 25° upward (see Figure 1). Though not reported on here, the latter allowed for capturing the nostrils to measure respiration from temperature changes re-

lated to air in- and outflow. Both cameras were adjusted such that the face was mapped onto equal parts of the images. Seat cushions were used to adjust the subjects' height centring their faces on the centre of these images. A headrest allowed to reduce additional head motions during the rotary condition in particular. Four white LED strips provided sufficient light for the recordings.

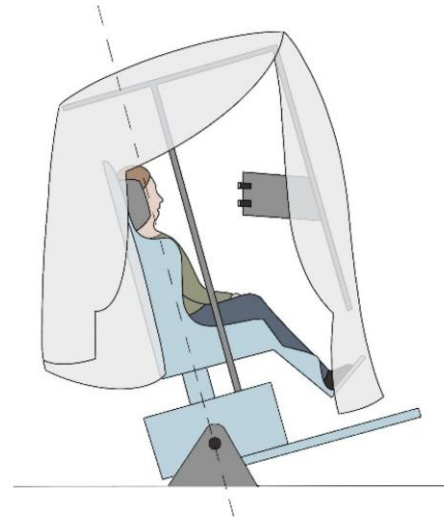


Figure 1: Sketch of the off-vertical axis rotation device and the camera's in front of a subject. Shields were closed during both conditions blocking all view on the lab space. Image adapted from Louvenberg et al. (2026).

RGB images were analysed using DLIB (King, 2009), yielding 68 characteristic facial landmarks. Based on these landmarks, eleven regions of interest (ROIs), shaped as polygons or circles, were determined (see Figure 2 left). A 12th ROI covered the whole face. After mapping these ROIs onto the IR images, the median temperatures over these regions were calculated.

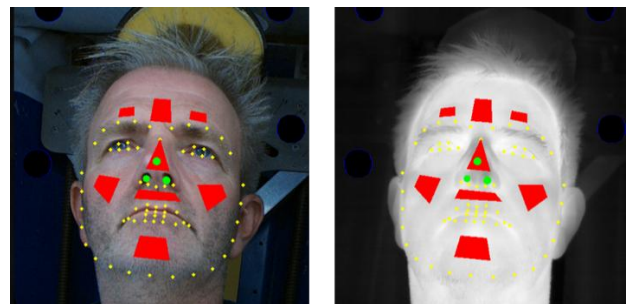


Figure 2: Example pair of RGB (left) and IR (right) images with the mapped facial landmarks (yellow) and ROIs (red polygons, green circles).

MISC

Because we aimed at measuring the progression of sickness during each condition, we used the 11-point Motion Illness Symptoms Classification scale (Reuten et al., 2021; see Table 1), formerly also known

as the Misery Scale (Bos et al. (2005). This scale applies the knowledge that classes of symptoms associated with motion sickness typically occur in a fixed order during accumulation of sickness (Reuten et al., 2021). The MISC consequently provides clear anchor points unambiguously predictive for the endpoint, defined by emesis.

Table 1: Motion Illness Symptoms Classification scale (MISC).

Symptoms		MISC
No problems at all		0
Uneasy (no typical symptoms)		1
Dizziness, warmth, headache, stomach awareness, sweating, ..., but no nausea	vague	2
	little	3
	rather	4
	severe	5
Nausea	little	6
	rather	7
	severe	8
	retching	9
Vomiting		10

Subjects

Fourteen subjects were included in this experiment, seven females and seven males, with an average age of 38 years in the range of 23 to 58. People were included if they had experienced motion sickness at least once in the past five years. They were excluded if they regularly used medication to suppress motion sickness, if they did suffer from (self-known) respiratory or cardiovascular problems, an abnormal blood pressure, head injury, severe headaches, dizziness, fainting, neurological disorders, chronic neck or back problems, were pregnant, or used medication affecting balance or alertness, or reported to have taken alcohol or drugs in the 12 h preceding the experiment. All subjects gave written informed consent and were paid € 30 plus travel expenses. The experiment was approved by the TNO ethical committee for non-medical scientific research.

Procedures

Subjects were first explained about the experiment, including a thorough explanation of the MISC scale, and were given answers to all experiment relevant questions. They were then seated in the rotation chair. This chair was kept tilted throughout the entire experiment. They first sat in the chair for 20 minutes during the stationary condition. After a 5 min. break, the rotary condition was started by increasing chair rotation speed in one minute to the final velocity of 60 °/s. The rotation speed then remained constant for the entire rotary condition (19 minutes). The experiment was stopped either at the end of the rotary condition, or if the subject reported a MISC level of 8 or higher, or when she or he desired to stop, no matter for which reason. To prevent unnec-

essary additional sickness, the chair was always brought to a stand still in one minute too.

During both conditions the MISC was requested every minute, timed to the second, after which the subject returned the rating verbally.

As mental state is known to affect sickness (Bos et al., 2015), in both conditions we aimed to control that state using an auditory 1-back working memory task. In this task one out of four spoken vowels was presented in random order each three seconds via a loudspeaker fixed to the cabin, where subjects had to count the number of incidences in which two equal vowels were presented one after the other.

After the experiments subjects were allowed to leave the lab once their MISC rating was below two.

Apart from the number of correct answers to the 1-back task that might have been considered a performance measure of interest, the experiment included a number of other additional measurements. Here, we only report on those relevant for the present paper, i.e., the MISC and different facial ROI temperatures. For more details, see again Louvenberg et al. (2026) and Nijskens et al. (2024).

Statistical analyses

The temperatures at the different ROIs were taken at a rate of 30 samples per second, while the MISC ratings were only taken every minute. To facilitate their comparison, we determined the median temperature of each ROI over a 50 second interval preceding each MISC rating. Here we excluded 5 seconds before and after each rating, to exclude possible artifacts due the fact that the MISC ratings were given verbally with a variable response time. Because baseline values may differ between different subjects, we normalised skin temperature by subtracting the first median value calculated for each subject and condition from subsequent values for that subject and condition.

Linear Mixed effects Models (LMM) analysis were applied to the observed MISC values as the response variable and each preceding normalised ROI temperature as the predictor variable. Distinct models were created for each individual ROI. To account for variability between subjects, all models included random intercepts and slopes. Differences were considered statistically significant if $p < 0.05$. For more details, again see Louvenberg et al. (2026).

Results

Mainly due to passive head motions made during rotation, several ROIs showed considerably more variability and outliers in temperature than others. We therefore limited further analysis to five ROIs: the entire face, central forehead, nose, left cheek, and chin. The left cheek performed better than the right

cheek because of the fact that the camera was positioned slightly left of the subject's line of sight.

As an example, Figure 3 shows the change in temperature of the central forehead and the matching MISC values for one subject in both conditions. Although during the stationary condition temperature seems to increase, it should be noted that on average this was less than 0.05° and may be explained by some discomfort due to fact that the chair was also tilted in this stationary condition.

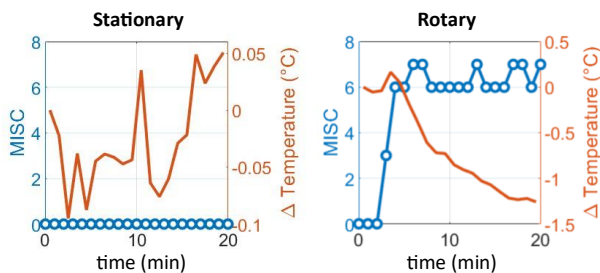


Figure 3: Changes in central forehead temperature (orange) and matching MISC values (blue) for one subject in both conditions tested. Note the difference in temperature scales.

For the five remaining ROIs the p -values of the linear relationships between MISC value and temperature change as determined by the LMM models varied between 0.6 and 0.05 for four ROIs, only the central forehead ROI giving $p < 0.001$.

Figure 4 then shows the change in temperature of the central forehead versus the matching MISC data for all 14 subjects, including a linear regression line determined over all individual data. If extrapolated to the endpoint of emesis (MISC = 10), the total decrease in temperature was calculated to be 1.32°C .

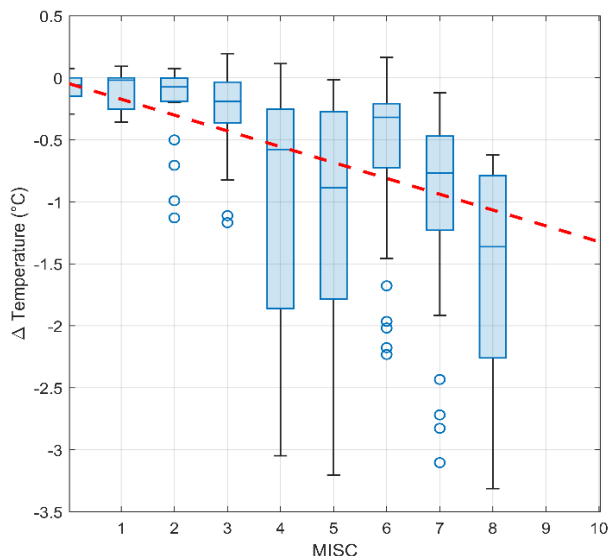


Figure 4: Temperature change of the central forehead of all subjects versus the matching MISC values in terms of the upper and lower quartiles (blue boxes), medians (blue lines inside boxes), whiskers extending to data points within 1.5 time the Interquartile range from the box edges (black), outliers (blue circles) and the linear regression line (red).

Discussion and conclusions

Here we showed that the central forehead is by far the most promising facial region of interest for measuring skin temperature as one of the possible biomarkers for motion sickness. Moreover, Louvenberg et al. (2026), discussed vasodilation and a subsequent sweating as the underlying mechanism explaining this observation.

The average decrease in temperature over the entire range of symptoms up to emesis was 1.32°C . This is a rather small change, and it remains doubtful whether such small changes can be detected using less sensitive (i.e., expensive) camera's as used in our experiment, which would facilitate the application of our finding. We are therefore currently elaborating on the option of using a less sensitive (cheaper) IR camera in combination with the detection of sweating as assumed critical for the drop of forehead temperature.

A main limitation of our study concerns the fact that it was a lab study under highly controlled conditions. Real world conditions while riding in a vehicle with a much larger variability of (partly occluded) faces, head motion and lighting conditions, most likely introduces several challenges still to be solved.

Another limitation concerns the fact that Figure 4 suggests a non-linear relationship, in particular showing a considerable smaller change in temperature at lower sickness ratings as compared to the overall range of sickness symptoms. This may impede the use of biomarkers for the early detection of motion sickness for determining when a counter-measure is required, such as driving more gently. The latter typically does not make sense when occupants are already (on the edge of) emesis. An early warning would therefore be most effective, in this case requiring the detection of an even smaller drop in temperature as concluded on here based on the entire range of symptoms up to emesis. An additional limitation might then be that a drop in skin temperature as suggested by Figure 3 in the rotary condition, lags the symptomatology as rated with the MISC. The above mentioned vasodilation and sweating mechanisms are particularly relevant in explaining this lag. The latter is an additional reason for including the measurement of sweating in our ongoing efforts.

Yet another major critical limitation of our study concerns the fact that we only concluded on sensitivity, i.e., the likelihood of finding a true effect, in this case the temperature drop of the central forehead as a marker for motion sickness. Stressors other than motion, like a raised ambient temperature, or performing a high demanding mental task, such as a 2- or 3-back task, might result in a drop of forehead temperature as well. If not considered, this could lead to adapting the vehicle's motion, instead of, e.g., its interior temperature. It therefore seems critical to

extend the research on this topic to also find the measurement's specificity, i.e., the likelihood of not finding a false effect by also testing such other than motion stressors. However, to increase the chance of obtaining both a high sensitivity and a high specificity of a specific state, most likely multiple biomarkers (a metrics), will be required. When limiting to facial biomarkers obtained by RGB and/or IR images, Figure 5 suggests a number of optional biomarkers that could be included in a suitable metrics to increase both likelihoods. Currently it is unclear which biomarkers could serve this purpose.

In Figure 5, we finally also included multiple states, that might all depend on the same metrics. The mu-

tual weightings of the biomarkers should then make the difference between these different states. The final actions were already exemplified by driving gently in case occupants would get sick. The validation of turning this picture of interrelated states through their underlying biomarkers into practice, will then likely be the most challenging effort of all. This effort might well require a multi partner joint industry and research centre(s) effort or European project. We welcome suggestions.

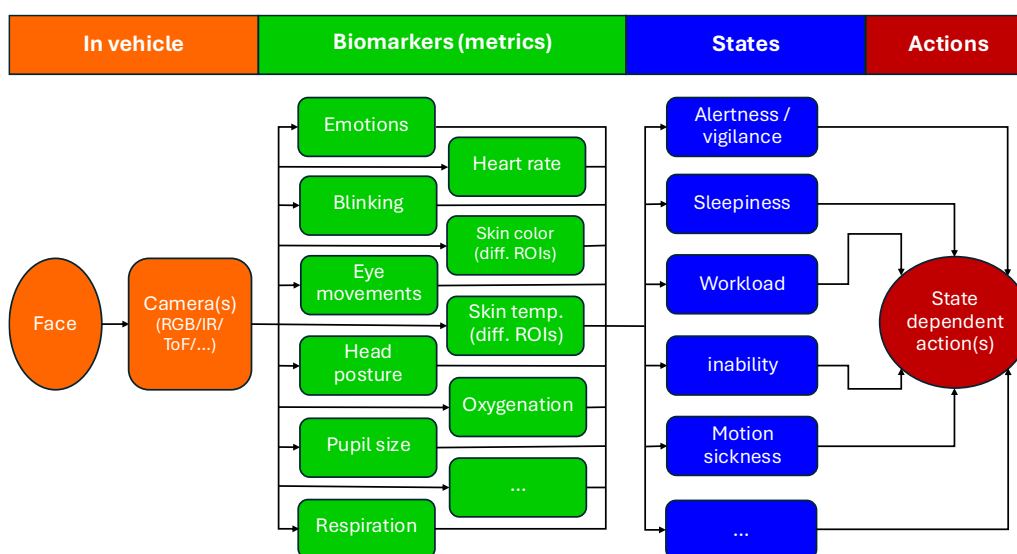


Figure 5. In vehicle occupant multiple state monitoring depending on multiple biomarkers (metrics) as discussed in the text.

References

- Bos, J.E., 2015. Less sickness with more motion and/or mental distraction. *Journal of Vestibular Research*, Vol. 25, pp. 23-33.
- Bos, J.E., Diels, C., and Souman, J.L., 2022. Beyond seasickness: a motivated call for a new motion sickness standard across motion environments. *Vibration*, Vol. 5, pp. 755-769.
- Bos, J.E., MacKinnon, S.N., and Patterson, A., 2005. Motion sickness symptoms in a ship motion simulator: effects of inside, outside, and no view. *Aviation Space and Environmental Medicine*, Vol. 76, pp. 1111-1118.
- Bos, J.E., Souman, J.L., Nooij, S.A.E., and Diels, C., 2024. Advancing ISO 2631-1 considering pre-emesis symptoms in carsickness. *Driving Simulation and Virtual Reality Conference Europe (DSC2024)* Strasbourg, France, September 18-20, pp. 91-95.
- King, D.E., 2009. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, Vol. 10, pp. 1755-1758.
- Louvenberg, S., Bos, J.E., Nijsskens, L., Van der Hurk, S.E., and Souman, J.L., 2026. Facial skin temperature as a biomarker for motion sickness. *Applied Ergonomics*, Vol. 135, pp. 104765,1-10.
- Nijsskens, L., Van der Hurk, S.E., Van der Broek, S.P., Louvenberg, S., Souman, J.L., Bos, J.E., Ter Haar, F.B., 2024. An EO/IR monitoring system for noncontact physiological signal analysis in automated vehicles. In: *Proceedings of the International Society for Optics and Photonics, SPIE 13207, Autonomous Systems for Security and Defence*, Edinburgh, UK, 16-19 November. Dijk, J., and Sanchez-Lopez, J.L. (eds), pp. 13070206.
- Reuten, A.J.C., Nooij, S.A.E., Bos, J.E., and Smeets, J.B.J., 2021. How feelings of unpleasantness develop during the progression of motion sickness symptoms. *Experimental Brain Research*, Vol. 239, pp. 3615-3624.
- Rolnick, A., and Lubow, R.E., 1991. Why is the driver rarely motion sick? The role of controllability in motion sickness. *Ergonomics*, Vol. 34, pp. 867-879.
- Schmidt, E.A., Kuiper, O.X., Wolter, S., Diels, C., and Bos, J.E., 2020. An international survey on the incidence and modulating factors of carsickness. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 71, pp. 76-87.

How We Roll: Validating Cycling Simulation Against Real-Life Riding

My Weidel^{1,2}, Lennart Ochel¹, Jakub Sliacan¹, Katja Kircher¹

(1) Swedish National Road and Transport Research Institute (VTI), Linköping, Sweden, e-mail: my.weidel@vti.se
(2) Linköping University

Abstract - This paper presents a multimodal comparison between real-life and simulated cycling using behavioral, physiological, and subjective measures, including riding dynamics, heart rate, gaze behavior, and perceived workload. A total of 32 participants were included in the study, aged between 18 and 73 years. They performed cycling tasks in both a real-life environment and in a digital replica of it. The simulator aims to preserve natural balance and steering dynamics additionally aided by allowing riders to use their own bicycles. Cycling dynamics can be quantified through measurements of speed, cadence, and longitudinal acceleration, as well as kinematics related to rider and bicycle motion. Participants attended a structured debriefing (interview) and completed the post-experiment questionnaire about their simulator and real-life experiences. Eye tracking was investigated in a controlled indoor environment using a hallway/corridor scenario designed to allow direct comparison between the simulator and cycling in real life. The results show that the simulator seems realistic enough for many research purposes and its design provides opportunities to safely test previously unexplored aspects of cycling.

Keywords: Cycling Simulator, Traffic Safety, Behavioral Research, Human Factors, Mixed Reality

Copyright: © 2026 by the authors.

Licensee Driving Simulation Association.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.82157/dsa/2026/5>

1. Introduction

While simulation technologies for driving and pedestrian behavior have matured considerably over the past decades, cycling simulation remains a nascent research area. Simulators offer a unique opportunity to investigate road user behavior under controlled and repeatable conditions, often unattainable in real-world studies (Carsten and Jamson, 2011). However, the limited development of cycling simulators restricts our ability to systematically examine how cyclists interact with the environment, other road users, and emerging traffic technologies. A dedicated cycling simulator can help address this gap by enabling experiments that would otherwise be difficult or unsafe to conduct in live traffic (Hammami, Borsos, and Sándor, 2025). Nevertheless, many existing systems still face challenges related to physical realism, restricted motion feedback, and limited flexibility. These can reduce ecological validity and restrict the scope of possible investigations.

1.1. Background on Existing Cycling Simulators

A high-end cycling simulator has the potential to address many of the aforementioned challenges (Hammami, Borsos, and Sándor, 2025). However, existing cycling simulation systems often face limitations such

as insufficient physical realism, limited motion feedback, and restricted flexibility in adapting to different experimental setups or bicycle types. A cost-effective and common solution is an instrumented bicycle that has its rear wheel mounted to a bicycle trainer of some kind. Sometimes with a tilt function (Wintersberger, et al., 2022), other times not (Hammami, Borsos, and Sándor, 2025; Keler, et al., 2018; Meulenens, Fraser, and Roberts, 2023; O'Hern, Oxley, and Stevenson, 2017; Sasaki, Fujiwara, and Mitobe, 2024; Schramka, et al., 2018; Sun and Qing, 2018; Ullmann, et al., 2020). This setup resembles riding a bicycle with training wheels, limiting natural leaning and steering dynamics. Other types of simulators with different purposes also exist, see (Dialynas, Happee, and Schwab, 2019; Guo, et al., n.d.; Shoman and Imine, 2021), or (Haasnoot, et al., 2024) who focus on haptic feedback, and (Berghoefer and Vollrath, 2023) who did not use a headset and had a leaning mechanism. However, none of them reach the level of realism that cycling on your own bike in a familiar environment allows for. Moreover, a simulator needs to be versatile enough to be accessible to all riders if it is going to be suitable for future research.

1.2. Why the HWR Simulator?

The VTI (Swedish National Road and Transport Research Institute) Cycling Simulator “How We Roll”



Figure 1: How We Roll: The VTI cycling simulator, featuring a free roller, a cyclist using a mixed reality headset, and the control computer in the background.

(HWR; [Figure 1](#)) was developed with the existing limitations in mind (Ochel, et al., [2025](#)). The system preserves authentic balance and steering dynamics without requiring extensive motion modeling through construction of complex motion platforms. By allowing participants to use their own bicycles, the simulator accommodates a wide range of users and promotes intuitive interaction, which is particularly valuable for inexperienced cyclists. This design choice fosters natural behavior and thereby enhances the ecological validity of the collected data.

The design of the simulator (see [Figure 1](#)) is inspired by common training rollers. It has three big steel rolls, two for the rear and one for the front bike wheel. The rollers are equipped with motors that make it possible to change the resistance to simulate inclines. The integration of mixed reality allows for realistic environmental representation while retaining direct physical feedback from the bicycle. Furthermore, the simulator is portable, does not require three-phase power, and is relatively cost-efficient compared to more complex systems.

The current validation work builds on (Ochel, et al., [2025](#)), which introduced the design and technical implementation of the VTI Cycling Simulator (HWR). The previous publication focused on the simulator architecture and the extent to which the simulator elicits realistic cycling behavior has not yet been systematically evaluated. Establishing such ecological validity is a critical prerequisite for using simulation as a reliable tool for behavioral research.

The present study addresses this gap through a structured validation study comparing cycling behavior in the simulator with cycling in a corresponding real-life environment. Specifically, this paper makes three contributions. First, it provides one of the first empirical validations of a high-end cycling simulator that allows natural balance and steering dynamics. Second, it presents a multimodal comparison between real and simulated cycling using behavioral, physiological, and subjective measures, including riding dynamics, heart rate, gaze behavior, and perceived workload. Third, it identifies key similarities and differences between the two scenarios, highlighting both the strengths of the simulator and areas where further refinement is required.

2. Research question

This study investigates the ecological validity of the cycling simulator through addressing the following research question:

To what extent does the cycling simulator reproduce realistic cycling behaviour and user experience, in terms of riding dynamics, perceived control, motion realism, and physical effort, compared to real-life cycling?

3. Methodology

A validation study was conducted in autumn 2024 to assess the realism, usability, and research applicability of the HWR simulator. Participants performed cycling tasks in both a real-life environment and a digital replica of it administered through the HWR simulator.

3.1. Procedure

Participants were recruited through a Facebook ad. Individuals with varying levels of cycling experience were eligible to participate, ranging from novices to competitive cyclists. The sole exclusion criterion was a predisposition to epilepsy, while individuals prone to motion sickness were not excluded.

All participants provided informed consent prior to participation. The study was ethically reviewed and conducted in accordance with the guidelines of the Swedish Ethical Review Authority (reference number 2024-04594-01, approved 26 August 2024).

Individuals who expressed interest via the survey and met the inclusion criteria were informed about the study procedures in advance and were given the opportunity to schedule a time slot. Each session lasted approximately three hours. Participants received financial compensation of 800 SEK (before tax) for their time, regardless of whether they completed or withdrew from the study, which they could do at any time, without providing a reason. No participant chose to withdraw.

A total of 32 participants (20 men and 12 women), aged between 18 and 73 years, were included in the study. All participants used their own bicycles during the entire study. The only restriction was that the wheelbase must not exceed 117 cm. Two electric bikes, one foldable bike, a few road bikes, many city bikes, and two mountain bikes were used by the participants.

Upon arrival, all participants had to fill out an informed consent form while a test leader instrumented their bike with sensors and adjusted the simulator for the bikes wheelbase. Half of the participants started in the real environment, and half started in the simulator.

In the simulator, participants used a *Varjo XR-4* mixed reality headset with passthrough functionality, enabling them to see their own bicycle and arms while interacting with the virtual environment. The headset includes integrated speakers, offers a wide field of view (120° horizontal × 105° vertical), and weighs 1,021 grams. Participants wore their prescription glasses underneath the headset when necessary (this did not interfere with eye tracking). The

simulator had a handrail on the right that the participant held onto while mounting their bike. When they gained speed and balance, they let go of the handrail and steered freely over the rollers. While wearing the headset, the cyclist could see the handrail in the simulation for the first few meters, and as they got closer to the end, they could see the handrail again and were instructed to go to the right side and grab it. The rail was green, so it would only be seen in the headset when it was programmed into the graphics. Before beginning the simulator scenarios, participants were given the option to cycle without wearing the headset to familiarize themselves with the setup. Most participants chose to do so and subsequently put on the headset once they felt comfortable.

The digital replicas of the environments were done by design students from Linköping University. The simulated scenarios were brought to life using Unreal Engine 5.3 (Epic Games, Cary, NC, USA). The students measured, photographed, and even recorded sounds from the scenarios to ensure realistic and life-like replicas.

3.2. Scenarios

As mentioned above, the validation consisted of two different settings, one conducted in the real-life environment and the other in its simulated replica. Additionally, there was an opportunity of a training session designed to familiarize participants with the simulator. This took place in a virtual underground corridor and included a slalom course to assess steering and control. The corridor was chosen as a location due to its width: it was similar to the width of the simulator rollers. Moreover, the corridor being indoors eliminated various elements like weather, other vehicles, etc. The training in the corridor involved three “rides” and lasted approximately 10–12 minutes in total. In the real-life version, participants stopped halfway to turn around and return to the starting point. In the simulator version, the environment was mirrored at the midpoint, allowing continuous cycling there and back without turning. The simulator corridor scenario was always completed before the other scenario – the campus scenario – to make sure that the participants felt comfortable. Most participants felt comfortable starting the scenario within 3 minutes of practice, while one participant needed almost 15 minutes to feel comfortable.

The study was conducted on the Linköping University campus and involved cycling up and down a 4% slope (sections 2 and 5, see Figure 2). It included an encounter and an overtake of an automated bus. White markings on the ground indicated the width of the simulator. The full distance was about 1 km long. This scenario was repeated twice. In the simulation, there were no moving actors other than the automated bus. There was a recording of birds singing. In the real-life scenario, interactions with people on campus were inevitable.

A counterbalanced within-subject design was applied. All participants completed both the real-life and simulator conditions. The order of conditions was largely randomized, with minor adjustments made to ensure that outdoor data collection was conducted during daylight hours and under conditions with minimal pedestrian traffic.

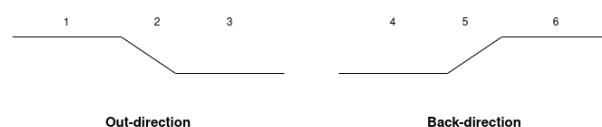


Figure 2: A simplified vertical profile of the route in Scenario 2. On the simulator, these two were concatenated without any gap. In real life, participants turned around after section 3. Note that sections 4,5,6 are sections 3,2,1 in this order but reverse direction.

3.3. Sensors

A multimodal sensor setup was used to capture kinematic, physiological, and environmental data in both the real-life and simulator conditions. All sensor streams were time-synchronized and recorded continuously throughout each trial.

Cycling dynamics were quantified through measurements of speed, cadence, and acceleration. A Wahoo cycling computer was used with a magnet-based speed sensor mounted on the front wheel. Cadence was measured via cadence sensor attached to the crank arm. Acceleration and deceleration were also recorded in both environments. Rider and bicycle motion were further captured using three inertial measurement units (IMUs; Shimmer sensors). The sensors were mounted on the bicycle frame, the handlebars, and the participant's forehead. IMU data was only collected for exploratory purposes. Physiological data were collected in the form of heart rate through a monitor with a chest strap. Lateral positioning relative to surrounding objects was measured using a LiDAR sensor unit, used for assessment of lateral clearance during interactions such as overtaking. The LiDAR unit also incorporated a GNSS (GPS) module, which was meant to provide geospatial positioning data for the real-life trials.

Video data was collected as follows. In the real-life scenario, a forward-facing camera mounted on the handlebar captured the cyclist's field of view. The test leader followed each participant while cycling and recorded the context with the participant in it. In the simulator scenario, a stationary camera was positioned to capture the entire scene.

After each task, subjective workload was assessed using the NASA Task Load Index (NASA-TLX; Hart and Staveland, 1988), which provides ordinal ratings (1–20) of perceived mental and physical demand. Following completion of all their sessions, participants were debriefed in a structured interview and completed a questionnaire (7-point Likert scale) about their experiences in this study.

3.4. Eye Tracking

Glance behavior was investigated in a controlled indoor environment. The corridor session was designed to be as sterile as possible to enable direct comparison between real-life and simulated cycling.

In the real-life corridor scenario, eye movements were recorded using Pupil Labs Invisible eye tracking glasses, 46,9 grams. These glasses are equipped with a binocular pair of infrared eye cameras (850 nm) for capturing eye movements, enabling real-time

gaze estimation from recorded raw data. A detachable scene camera, mounted on the left arm of the glasses, recorded the participant's field of view and included an integrated microphone for audio capture. The system allows for interchangeable prescription lenses, such that participants' visual acuity can be corrected within the range of -3 to +3 diopters.

Gaze data were collected in the straight corridor where participants cycled over five predefined ground markers. These markers, made of tape, were positioned at longitudinal distances of 5, 10, 15, and 20 meters, with a lateral spacing of 70 cm. The virtual tunnel was designed as a photorealistic replica of the real environment, using images of wall paintings and doors to ensure high visual fidelity.

Again, a counterbalanced experimental design was employed: half of the participants completed the real-life scenario first, the other half began in the simulator. In both scenarios, participants cycled back and forth through the tunnel three times. Eye tracking data from the final traversal was extracted for analysis. Two predefined spatial points along the corridor were used to define the start and the end of the analyzed segment. Both directions were included for all participants. The corridor was accessible to others as usual.

Eye tracking data were analyzed using Noldus Observer XT (Wageningen, NL). An observation scheme contained predefined categories, including: forward, empty floor and floor markings, walls, doors, wall painting, ceiling, ceiling signage, and the participant's own equipment (e.g., a wristwatch).

4. Results

4.1. Data preparation

Of the total of 32 participants, both outdoor and indoor data was available for 23 due to bad weather, sensor failures, and other circumstances. Partial data fragments, either heart rate or speed sensor data, were missing or corrupted for further 7. Complete data was available for the remaining 16 participants.

During data cleaning we excluded empty and erroneous output files (test runs and canceled runs) and trimmed files when sensors were left running after the experiment. We standardized missing values. No prolonged periods of stuck sensors values were observed (apart from GPS). Synchronizing timestamps across different sensors was done based on prearranged cues (button presses) and whenever those were missing (5) or incorrect (5), we used value peaks across various sensors to manually identify the matching points and we aligned data based on those.

Given that speed was measured with a magnet-based speedometer mounted on participants' own bicycles (different wheels, tire sizes, tire pressures), we adjusted each individual bicycle speed (as measured by speedometer) based on a discrepancy observed between the bicycle speed and calibrated simulator speed (during the simulator scenario). Knowing the coefficient, we then adjusted bicycle speed accordingly for all calculations.

Virtual cycling included no turnaround at the end of the out-section, and we had precise knowledge where the virtual turnaround happened. Outdoors,

Table 1: Basic speed information for each section, in km/h.

section	XR		RL	
	mean	std	mean	std
1	14.85	2.20	16.10	2.23
2	19.16	2.90	18.61	2.85
3	18.27	3.30	16.26	2.28
4	18.06	3.18	16.58	2.01
5	15.68	2.94	16.92	2.75
6	15.60	3.01	15.90	2.05

however, we did not have that information and the GPS signal was unreliable. Therefore, we used the dip in speed as an indicator of the turnaround. When such a dip was not observable, we excluded an interval of the average length of the turnaround interval from the middle of the flat out-and-back section (sections 3 and 4) between the two incline sections (between 2 and 5).

Outdoors incline sections were identified from the speed profile (decreased/increased for uphill/downhill respectively). Ideally, one would use GPS, and when it was available, we used it to validate and adjust our speed-based detection. Once acceptable performance was reached, we used the speed-based method for all participants.

4.2. Speed

The rollers on the simulator are motor-driven to be able to simulate uphill and downhill. This creates a task of calibrating them realistically. We see from the responses (perception) and speeds themselves (see Figure 3) that the simulator exaggerates the experience of uphill and downhill compared to real life. In Figure 3 this shows up as higher speeds (in blue) in section 2 and lower speeds (in blue) in section 5. However, it is important to note that several participants were slightly more interrupted outdoors than on the simulator, thereby creating speed disruptions which go unaccounted for.

You can find the summary of average speeds over the sections in Table 1. Sections in bold are inclines.

Note that several participants had anomalous speeds on various sections in real-life scenario. This was usually when they needed to dismount or slow down for life on the street. We did not specifically treat these anomalies and they are included in the data (see example in Figure 4).

4.3. Heart rate

The average heart rate is similar in both environments on the first two flat sections, and lower on the downhill slope in XR. After the turnaround, the heart rate is consistently lower outdoors (Table 2). Note, however, that participants made a U-turn in the real life, meaning they slowed down or even stopped.

4.4. Eye Tracking

The distribution of gaze behavior differed markedly between the real-life and simulator scenarios (see Figure 5). In the real-life, participants' gaze was relatively dispersed across several areas of interest. The

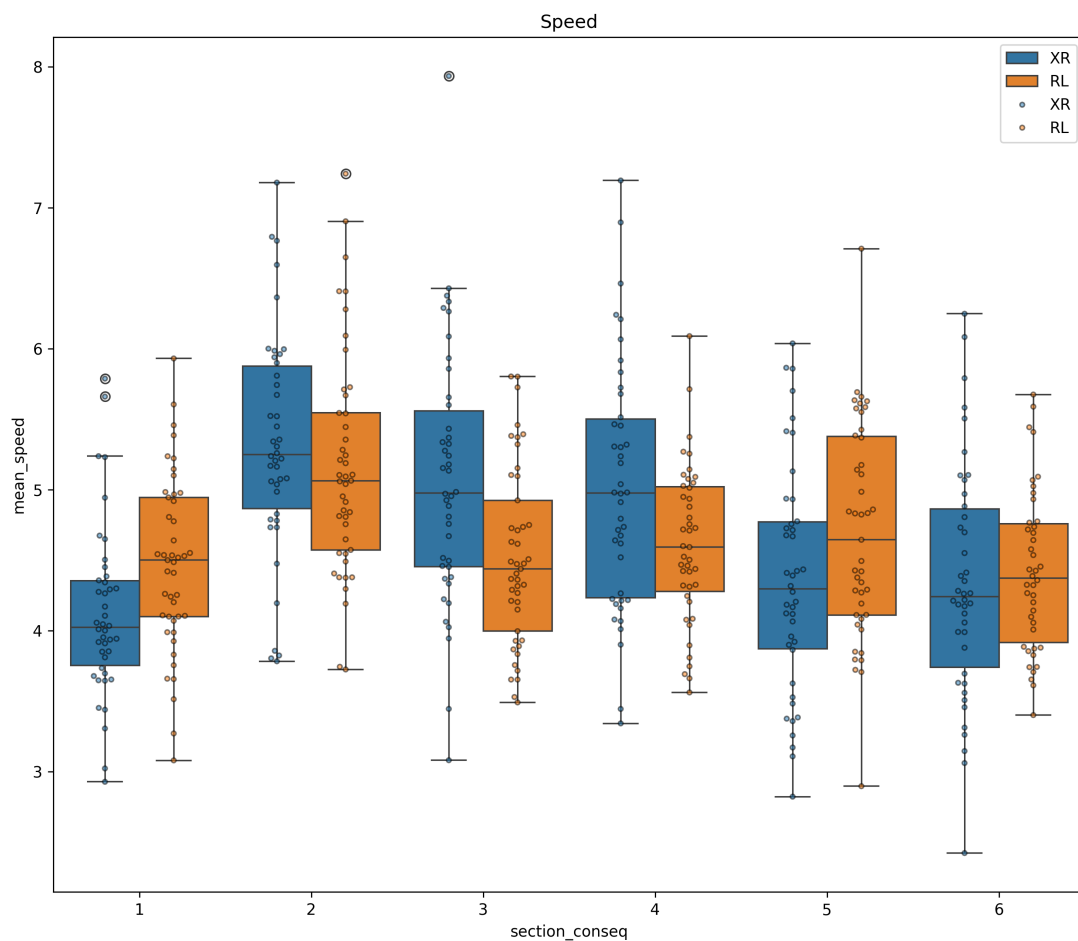


Figure 3: Comparing participants' mean speeds (in m/s) for each section on the simulator (XR) and in real life (RL). Sections are: out-direction flat(1)-downhill(2)-flat(3) and back-direction flat(4)-uphill(5)-flat(6).

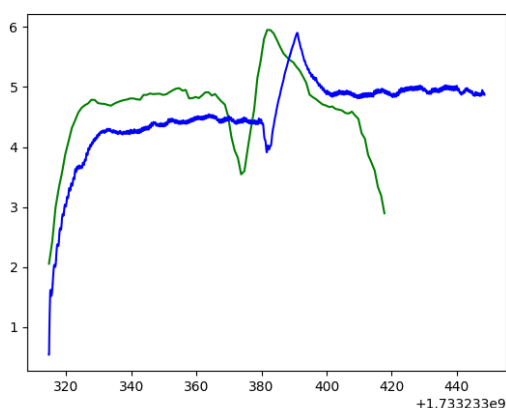


Figure 4: Individual speed comparison between real life and simulator. Unix time on the x -axis.

Table 2: Basic heart rate information for each section, in bpm.

section	XR		RL	
	mean	std	mean	std
1	124.05	19.32	118.36	22.08
2	108.17	22.34	128.03	24.5
3	121.3	19.81	122.47	24.1
4	146.43	21.16	119.78	23.09
5	142.64	22.06	131.38	25.26
6	147.8	20.22	130.79	24.19

target area receiving the largest glance share was the floor 34% in total, (empty floor 30% and floor markings 12%), followed by the right wall (25%) and the forward direction (20%). Additionally, participants glanced at doors (9%), the left wall (3%), cables at the ceiling (1%), and paintings on the wall (0%).

In contrast, eye movements in the simulator scenario were more restricted. Most glances were directed toward the empty floor (60%) and floor markings (38%),

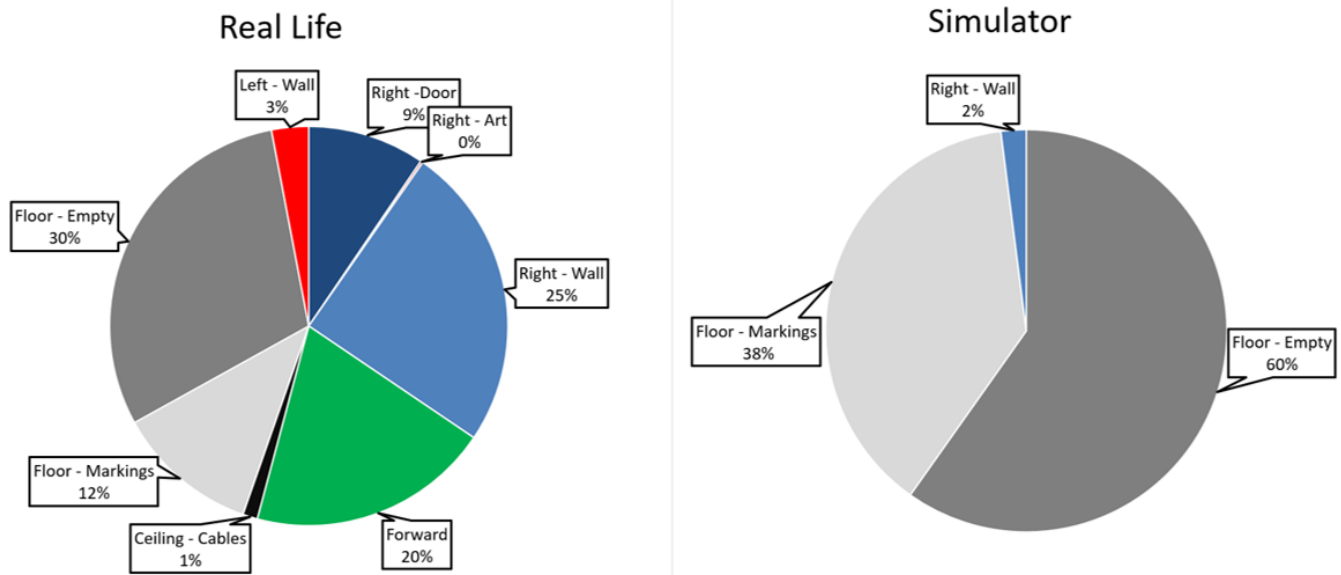


Figure 5: Distribution of gaze behavior in real-life and simulated cycling conditions.

with a small share of glances targeting other elements such as the right wall (2%). Notably, glances to doors, wall art, and ceiling elements were virtually absent in the simulator scenario.

4.5. Subjective Workload

The results from NASA-TLX indicate that both perceived mental and physical workload were higher in the simulator than in real life. The mean mental workload reported for the simulator was 6.63 (SD = 4.14), whereas the corresponding value for real life was 3.18 (SD = 2.12). Similarly, the mean physical workload was higher in the simulator (M = 9.56, SD = 4.39) than in real life (M = 4.61, SD = 2.76).

4.6. Subjective Experience

Participants answered on a 7-point Likert scale where a high value (7) was seen as the most positive and the lowest value (1) was the most negative. Participants' subjective ratings indicate a generally positive perception of the simulator in terms of realism and usability (see also Figure 6). The perceived similarity to real-life cycling was rated moderately high (M = 4.5, SD = 1.2). The virtual environment was also perceived as moderately varied and rather realistic (M = 5.1, SD = 0.9).

Participants reported that it was relatively easy to detect details in the environment in time (M = 5.1, SD = 1.4), and other road users (the automated shuttle in the campus environment) were perceived as realistic (M = 5.4, SD = 1.3). The ability to judge the bicycle's lateral position on the road was rated as easy to accomplish (M = 5.9, SD = 1.2). Lateral and longitudinal control such as steering (M = 4.8, SD = 1.6), speed maintenance (M = 5.2, SD = 1.3) and speed changes (M = 4.5, SD = 1.5) were perceived as relatively easily achievable.

The XR headset was experienced as rather comfortable (M = 5.1, SD = 1.7), and also the performance

of the visual representation while riding was rated as positive (M = 5.3, SD = 1.2), indicating a smooth image flow. The participants found it easy to make use of the railing to aid starting and stopping (M = 5.6, SD = 1.0).

Importantly, the participants did not report much simulator-induced discomfort, such as dizziness or nausea (M = 1.5, SD = 1.2). Note that this scale was reversed, so a lower value indicates less nausea.

4.7. Debriefing interview

Participants' experiences of cycling in the simulator followed a clear progression from initial uncertainty to increasing familiarity and comfort. First impressions were often characterized by mixed emotions, with several participants describing the experience as *"scary but fun"* and *"exciting, especially the first time"*. The simulator was frequently perceived as physically demanding – *"much more strenuous and physical than I thought"* – and somewhat unfamiliar, with comments such as *"it wasn't just to start cycling, you had to learn again"*. Early difficulties related to balance, steering, and control were common, with the system described as *"hard to control at first and a bit wobbly"* and *"slippery, especially during quick lateral movements"*. However, most participants reported a relatively rapid adaptation process. Within minutes, control improved and the experience became more intuitive, with participants noting that *"you got used to it quickly"* and that *"after some test cycling, it worked well"*. Several described how the experience began to feel natural, even *"almost like cycling for real"*, with *"muscle memory kicking in"*.

In terms of realism, the simulator was generally perceived as convincing, particularly after familiarization. Many participants stated that the environment was *"very true to life"* or *"almost like reality"*, although some limitations were noted, including a slightly *"flat impression"* and the absence of wind, sound, and dynamic elements such as other road users. Cycling mechanics were considered reasonably realistic, al-

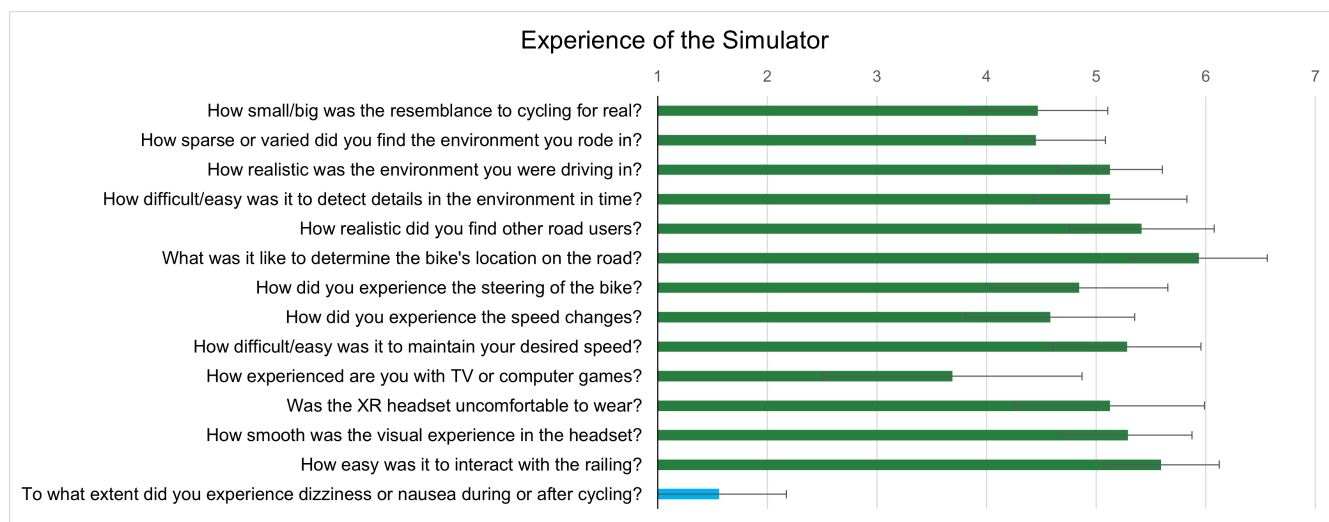


Figure 6: Subjective experience of the cycling simulator. Mean ratings (M) and standard deviation (SD) on a 7-point Likert scale (1 = most negative, 7 = most positive; reversed interpretation for the final item (blue) across questionnaire items.

though often described as *“heavier than in real life”* and *“as if there was a constant slight uphill”* or requiring more effort.

Compared to real life, the simulator experience was described as more physically demanding, less natural in terms of balance, and more cognitively engaging. Participants reported focusing more on the act of cycling itself – *“full focus all the time, higher pulse”* – while the real-life cycling allowed more attention to the surrounding environment and traffic. The simulator was also perceived as more predictable and controlled, lacking the variability and unpredictability of real traffic situations. This influenced behavior, with some participants reporting reduced environmental scanning – *“I checked my surroundings much less”* – and more cautious or constrained movements, such as maintaining a central lane position or avoiding sharp turns. For most, biking in the simulator even felt easier with the headset and helped participants *“get a sense of speed”* more than without it.

5. Discussion

Overall, the simulator was perceived as acceptable to rather realistic, with the cycling dynamics, the absence of simulator sickness and the visual richness leaving positive impressions. The usage of free rollers allows a fully realistic representation of lateral cycling movements, including a correctly mapped lateral position, which likely explains the high tolerance of the simulated situation. At the same time, the construction based on free rollers does not afford making turns, stronger braking, or frequent starts and stops. Therefore, these manoeuvres were not included in the tested scenarios.

The headset with its higher weight and narrower field of view, the safety harness, and the limited lateral movement range are restrictions that are inherent to the setup. While improvements can be made, it will not be possible to fully eliminate them.

The possibility to use one's own bike increases familiarity and realism at the cost of more effort of installing sensors on the bike, and of being limited by

the physical dimensions of the simulator, which currently only allow for a limited wheelbase range. If a complex sensor setup is required, providing a kitted out bike instead of using participants' own bikes might be preferable.

For individual variables, we looked at comparability on different conceptual levels. In the simulator, the slope seems to affect the average speed more, with higher speeds downhill and lower speeds uphill than in reality. This seems to carry over somewhat to the segment following the slope. This effect could be due to differences in friction, air resistance, inaccuracies in the model used or biases when estimating the system weight of cyclist and bike. Additional fine-tuning and calibration is needed to improve these relative differences. This should also include a comparison of the wattage in relation to speed.

Comparable reasoning applies to glance behavior. The visual impression of the static environment was rather similar in both, but in the real corridor people could be present and doors could open. The participants knew that this was not the case in the virtual corridor. Although the distribution of glances differs nominally between the two environments, it can be argued that the functional behavior was comparable – participants tended to focus on the areas relevant for cycling in the specific environment. More similar conditions are likely to also improve nominal comparability, but methodological constraints like the limited field of view, the narrow range in brightness, or the weight of the headset affect validity systematically.

The higher perceived mental workload in the simulator can partially be explained by the novelty of the situation. However, some features of the simulator setup probably systematically increase workload, like the lateral physical constraints and the comparatively low visual resolution. The higher perceived physical workload could be linked to a factual higher power output. This was not measured for practical reasons. However, the slightly higher heart rate in the simulator suggests the same. It could also be connected to the higher temperature in the simulator setting, to the absence of a natural break by turning around halfway,

or to being in a virtual and less stimulating environment (Olsson, et al., 2024).

5.1. Limitations

The construction of the simulator entails limitations, some of which were mentioned above (cycling a curve, braking, starting and stopping). Peripheral vision, the range of luminance and visual resolution is limited by the quality of the headset. The accessible lateral width and the range of bikes that can be accommodated are also limited by construction, but could be easily increased via hardware modifications.

The power/speed ratio appears to have been not well calibrated for the simulator, as participants rather consistently reported a higher perceived physical exertion in the simulator than in the same situation outdoors. An improvement of the resistance modeling taking additional factors into account could ameliorate this. Fast accelerations and decelerations are a problem due to inertia of the rollers. Additional research and development are needed to investigate to what extent this can be addressed.

The data collection suffered from substantial data loss and challenges to direct comparability, due to failing sensors, unpredictable occurrences in the environments, too heavy precipitation outdoors, etc. Also, for practical reasons, the environments were not fully comparable. The concatenation of the sections in XR to avoid stopping affected the speed and heart rate profile and may have affected perceived workload.

Only a subset of the available data is analyzed so far to provide examples. More fine-grained analyses to also incorporate tactical behavior could be performed, and potential interaction effects between variables could be investigated. Additional analyses can be made to prepare for future data collections and fine-tune the relevant sensors, but with some reservation for the data quality.

6. Conclusions

The study shows that the HWR cycling simulator has the potential to be a reliable research tool for certain research questions. After fine tuning, one might expect to draw fairly realistic conclusions e.g. in a study on safety critical traffic situations not involving excessive bike manoeuvring or speed changes. Being aware of the limitations will help assess the outcomes. On the other hand, there exist physical, logistical, and design constraints which one needs to be aware of while designing a study. For instance, a scenario where a cyclist turns 90° onto a side road cannot be implemented in the HWR simulator. However, and perhaps most importantly, fundamental yet readily implementable modifications can often lead to extended use cases. For example, studying cycling dynamics of people cycling two abreast is possible after technically trivial modifications. Similarly, conducting a study about children riding bikes (short wheelbase) is feasible after very little adjustments to the simulator construction.

The HWR simulator seems to be realistic enough up to fixable issues and its design provides opportunities to safely test previously unexplored aspects of cycling.

References

- Berghoefer, F. L. and Vollrath, M., 2023. Prefer what you like? Evaluation and preference of cycling infrastructures in a bicycle simulator. *Journal of Safety Research*, 87, pp. 157–167. ISSN: 0022-4375. <https://doi.org/10.1016/j.jsr.2023.09.013>.
- Carsten, O. and Jamson, A. H., 2011. Chapter 7 - Driving Simulators as Research Tools in Traffic Psychology. In: B. E. Porter ed. *Handbook of Traffic Psychology*. San Diego: Academic Press, pp. 87–96. ISBN: 978-0-12-381984-0. <https://doi.org/10.1016/B978-0-12-381984-0.10007-4>.
- Dialynas, G., Happee, R., and Schwab, A. L., 2019. Design and hardware selection for a bicycle simulator. *Mechanical Sciences*, 10(1), pp. 1–10. <https://doi.org/10.5194/ms-10-1-2019>.
- Guo, X., Angulo, A., Robartes, E., Chen, T. D., and Heydarian, A., n.d. ORCLSim: A System Architecture for Studying Bicyclist and Pedestrian Physiological Behavior through Immersive Virtual Environments. *Journal of Advanced Transportation*, 2022(1), p. 2750369.
- Haasnoot, J., Happee, R., Wijk, V. van der, and Schwab, A. L., 2024. Validation of a novel bicycle simulator with realistic lateral and roll motion. *Vehicle System Dynamics*, 62(7), pp. 1802–1826. <https://doi.org/10.1080/00423114.2023.2264418>.
- Hammami, A., Borsos, A., and Sándor, A. P., 2025. How realistic a bicycle simulator can be? - A validation study. *Multimodal Transportation*, 4(1), p. 100193. <https://doi.org/10.1016/j.multra.2025.100193>.
- Hart, S. G. and Staveland, L. E., 1988. *NASA-Task Load Index (NASA-TLX): Paper and pencil package (TLX.pappen.manual)*. Tech. rep. NASA Ames Research Center.
- Keler, A., Kath, J., Chucholowski, F., Chucholowski, M., Grigoropoulos, G., Spangler, M., Kath, H., and Busch, F., 2018. A bicycle simulator for experiencing microscopic traffic flow simulation in urban environments. In: *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, pp. 3020–3023.
- Meuleners, L., Fraser, M., and Roberts, P., 2023. Improving cycling safety through infrastructure design: A bicycle simulator study. *Transportation Research Interdisciplinary Perspectives*, 18, p. 100768. ISSN: 2590-1982.
- Ochel, L., Saparia, S., Weidel, M., Johansson, A., and Kircher, K., 2025. How We Roll - Realistic Cycling Simulator for Behavioural Research. In: *Proceedings of the Driving Simulation Conference 2025 Europe XR*. Stuttgart, Germany: Driving Simulation Association, pp. 53–58.
- Olsson, K. S. E., Ceci, R., Wahlgren, L., Rosdahl, H., and Schantz, P., 2024. Perceived exertion can be lower when exercising in field versus indoors. *PLOS ONE*, 19(5), e0300776.
- O'Hern, S., Oxley, J., and Stevenson, M., 2017. Validation of a bicycle simulator for road safety research. *Accident Analysis & Prevention*, 100, pp. 53–58. ISSN: 0001-4575. <https://doi.org/10.1016/j.aap.2017.01.002>.
- Sasaki, Y., Fujiwara, K., and Mitobe, K., 2024. Risks that induce bicycle accidents: Measurement and analysis of bicyclist behavior while going straight and turning right using a bicycle simulator. *Accident Analysis & Prevention*, 194, p. 107338. ISSN: 0001-4575. <https://doi.org/10.1016/j.aap.2023.107338>.
- Schramka, F., Arisóna, S., Joos, M., and Erath, A., 2018. Development of virtual reality cycling simulator. *Journal of Computers*, 13(6), pp. 603–615. <https://doi.org/10.3929/ethz-b-000207175>.
- Shoman, M. M. and Imine, H., 2021. Bicycle simulator improvement and validation. *IEEE Access*, 9, pp. 55063–55076. <https://doi.org/10.1109/ACCESS.2021.3071214>.
- Sun, C. and Qing, Z., 2018. Design and Construction of a Virtual Bicycle Simulator for Evaluating Sustainable Facilities Design. *Advances in Civil Engineering*, 2018(1), p. 5735820. <https://doi.org/10.1155/2018/5735820>.
- Ullmann, D., Kreimeier, J., Götzelmann, T., and Kipke, H., 2020. BikeVR: A virtual reality bicycle simulator towards sustainable urban space and traffic planning. In: *Proceedings of Mensch und Computer 2020*, pp. 511–514.
- Wintersberger, P., Matviienko, A., Schweidler, A., and Michahelles, F., 2022. Development and Evaluation of a Motion-based VR Bicycle Simulator. *Proceedings of the ACM on Human-Computer Interaction*, 6(MHCI), pp. 1–19. <https://doi.org/10.1145/3546745>.

Simulation-Based Human-in-the-Loop Assessment of Shifting Comfort in Two Speed Electric Vehicles

Omar Fahmy¹, Safa Kaskaya¹, Stephan Rinderknecht¹

(1) Institute for Mechatronic Systems in Mechanical Engineering, Otto-Berndt-Straße 2, 64287 Darmstadt, Germany, e-mail: omar.fahmy@tu-darmstadt.de, {safa.kaskaya, rinderknecht}@ims.tu-darmstadt.de

Abstract – This paper presents a simulation-based human-in-the-loop methodology for the early-stage assessment of longitudinal shift comfort in two-speed battery electric vehicles. Representative upshift profiles are generated in a longitudinal MATLAB/Simulink model and reproduced in a dynamic driving simulator dedicated to longitudinal dynamics studies. Objective evaluation is based on vehicle acceleration and jerk criteria, while subjective assessment uses five-point Likert ratings of daily use acceptance, comfort, perceived acceleration interruption, and shift noticeability. Statistical analysis of the results confirms clear and systematic differences between the investigated configurations. Interruption-related metrics align particularly well with subjective ratings of perceived interruption and comfort, confirming traction continuity as the dominant contributor to shift comfort perception, whereas jerk contributes mainly to the noticeability of the shift event. The proposed approach therefore enables a development-oriented assessment of shift comfort before physical drivetrain prototypes are available and provides a basis for future work on a mathematical correlation model and on validation against a physical drivetrain.

Keywords: Battery electric vehicle, Two-speed-transmission, Shift comfort, Objective-subjective evaluation, Human-in-the-loop.

Copyright: © 2026 by the authors.

Licensee Driving Simulation Association.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.82157/dsa/2026/6>

Introduction

Two-speed electric powertrains are increasingly investigated for battery electric vehicles, as they allow more effective use of the electric machine and inverter operating range and thereby improve efficiency and performance (Ahssan, et al., 2018; Machado, et al., 2021). Relative to single-speed drivetrains, however, they introduce additional challenges for longitudinal drivability: torque interruption, acceleration deficits, and increased jerk during gear shifts can negatively affect perceived shift comfort and overall driving quality (Yu, et al., 2012; Liang, et al., 2013). Prior work has concentrated primarily on the technical realization and control of shift processes, with the aim of reducing torque interruption and improving traction continuity (Nguyen, et al., 2019); shift quality in these studies is typically assessed objectively through technical parameters, without detailed results on perceived quality (Sornioti, et al., 2011).

Subjective evaluation, by contrast, is resource- and time-intensive and is particularly difficult during early development phases and proves to be a challenging obstacle, especially when no physical prototype is available. Possible approaches to address this limitation include testing with mockups and also establishing a correlation model between objective and subjective evaluations, allowing virtual measurements to substitute for physical prototype testing. Such models have been applied to holistically perceived qualities, including ride comfort and steering feel (Kim, et al., 2019).

Driving simulators offer a useful complement in this context, as they enable reproducible test conditions, rapid variation of test cases, and subjective assessment prior to hardware availability. Longitudinal dynamics perception has been investigated in a structured manner for powertrain shuffle (Erler, et al., 2018) and drive-off procedures, with findings supported by real-vehicle validation (Kraft, et al., 2023). Gear shifting scenarios have likewise been implemented in motion-cueing

simulators and evaluated subjectively (Ghasemi, et al., 2020); however, no significant effect on driver behavior was identified using coarse behavioral measures, indicating that dedicated signal-based comfort metrics are required to capture perceptual differences during shifting.

A methodology that combines simulation-based shift profile generation, objective signal analysis, and subjective simulator-based assessment for two-speed electric vehicles remains absent from the literature. This paper addresses this gap by presenting a human-in-the-loop approach that enables early, reproducible evaluation of longitudinal shift comfort prior to the availability of physical drivetrain hardware. The evaluation combines objective and subjective variables, with statistical and correlation analysis supporting the results.

Research Question

The research question is which objective metrics derived from the longitudinal acceleration profile are suitable to describe shift comfort during upshifts in two-speed battery electric vehicles, and how these metrics correlate with the acceleration-related sensations subjectively perceived by test participants in a dynamic driving simulator.

The working hypothesis is that participants can perceive objective metrics derived from the longitudinal acceleration profile during a shift event, so these metrics should correlate with subjective ratings. Interruption-related metrics are expected to reflect perceived comfort, while jerk-related metrics are expected to reflect how noticeable the shift event is.

Methodology

The methodology proceeds in four stages. First, a longitudinal vehicle simulation implemented in MATLAB/Simulink generates representative upshift profiles: shift events are simulated under defined conditions to produce longitudinal acceleration responses that serve as test profiles for the subsequent simulator study. Second, these profiles are evaluated objectively, whereby acceleration continuity, acceleration loss, and jerk are extracted from the acceleration traces as signal-based descriptors of the transient longitudinal behavior during shifting. Third, the shift profiles are reproduced in a dynamic driving simulator designed for longitudinal dynamics studies, in which test participants experience the different shift events under controlled and repeatable conditions and rate them using a standardized questionnaire. Finally, the subjective and objective results are analyzed jointly:

statistical methods are applied to compare subjective ratings across configurations and to investigate relationships between perceived comfort and the derived objective parameters. This workflow enables the assessment of shift comfort prior to physical prototype availability.

Vehicle Simulation and Profile Generation

Vehicle Simulation

The investigated shift profiles are generated using a longitudinal vehicle simulation environment implemented in MATLAB/Simulink. The simulation framework represents different drivetrain topologies within one common longitudinal modelling approach, including a two-motor two-speed topology with support torque during shifting and a conventional automatic two-speed topology based on a planetary transmission with friction clutches. For the parameterization of the model, relevant vehicle and drivetrain parameters are defined via an input script in MATLAB, including vehicle mass, frontal area, drag coefficient, rolling resistance, wheel radius, electric machine characteristics, and transmission ratios.

Overall, during a shift event, the longitudinal acceleration follows a characteristic sequence. The torque is first ramped down at a defined rate, after which the transmission is disengaged, and the motor is synchronized to the target gear ratio. Once synchronization is complete, the torque is ramped back up and normal drive operation resumes. The longitudinal output of interest is therefore the transient acceleration response during this sequence, which depends on both the ramp rate and the degree of traction support available during the disengagement and synchronization phase.

For the two-motor two-speed topology, the present work adopts the general principle of a Two-Drive-Transmission-type architecture, in which two electric machines and two transmission paths are combined to improve efficiency and performance (Viehmann, et al., 2019). Traction support during shifting is achieved by transferring torque from one motor to the other, thereby allowing the wheel-torque disturbance to be partially or fully compensated. This is consistent with the shifting concept described for the Speed4E powertrain (Schöneberger, et al., 2020) considered in this study, in which support torque is used to reduce the wheel-torque disturbance during the shift, while the gear change itself consists of dog-clutch disengagement, synchronization, and re-engagement of the target gear.

For the conventional automatic topology with a planetary gearbox and friction clutches, the shift is realized as a clutch-to-clutch transition. Although the internal actuation mechanisms differ fundamentally from the dog-clutch-based multi-motor concept, both topologies produce a characteristic transient longitudinal acceleration response that can be analysed objectively and evaluated subjectively within the same methodological framework.

Test Profile Generation

Based on the simulation environment, representative upshift profiles are generated by systematically varying shift-relevant characteristics; the resulting acceleration traces form the basis for both the objective signal analysis and the subjective simulator evaluation. To enable a compact and systematic description of the generated variants, a DXSX notation is introduced in this study. The first index, D, characterizes the shift dynamics, specifically the rate at which the torque, and hence the longitudinal acceleration, is ramped down before the gear change and ramped up after synchronization. Three levels are considered: D1 for a slow ramp, D2 for a medium ramp, and D3 for a fast ramp. The second index, S, describes the degree of traction support during the shifting and synchronization phase, reflecting the proportion of acceleration retained relative to the pre-shift level while the transmission is disengaged and the motor is synchronized to the target gear ratio. Five levels are defined, S1 to S5, covering the range from 0% to 100% of the pre-shift acceleration in equal steps of 25%, excluding powertrain mechanical losses and driving resistances.

The resulting profile set spans a broad range of longitudinal shift characteristics, encompassing clearly perceivable differences in both transient dynamic and interruption behavior. Figure 1 illustrates a group of profiles generated within the set. Profile D3S5 represents the planetary gearbox cases, for which no discrete interruption-related acceleration loss occurs by definition; the evaluation of these configurations is therefore governed primarily by the continuous transient behavior and the jerk-related profile shape.

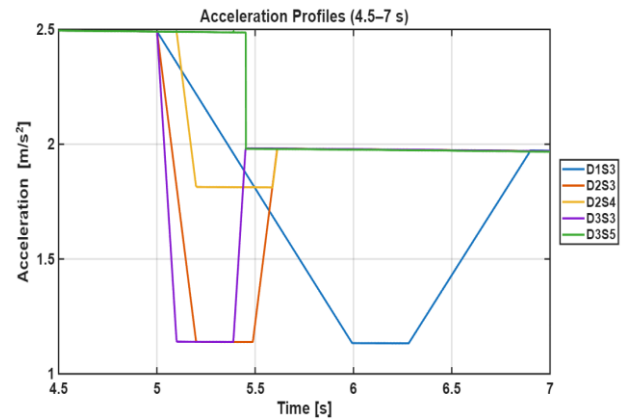


Figure 1: Group of generated Profiles

Experimental Setup

Dynamic Driving Simulator

The subjective evaluation was conducted in a dynamic driving simulator specifically designed for longitudinal dynamics studies. The simulator, as shown in Figure 2, provides a translational workspace of 10 m, with the neutral position set at 0.2 m from the rail end to allow a safety buffer, yielding 9.8 m of usable travel, a maximum translational acceleration of $\pm 8 \text{ m/s}^2$, and cabin tilt up to $\pm 25^\circ$. In previous studies the simulator was able to reproduce longitudinal acceleration cues relevant to maneuvers such as drive-off procedures, powertrain shuffle, and gear shifting (Kraft, et al., 2021; Kraft, et al., 2023; Erler, et al., 2018)). Beyond the motion system, the simulator provides a visual environment and further human-machine-interface components supporting longitudinal perception studies; prior work using the platform has combined a virtual reality environment with noise-cancelling headphones that provide both acoustic cue presentation and noise suppression, alongside haptic elements and realistic driver controls. In the present study, the simulator serves as a controlled platform for reproducing predefined longitudinal shift transients in a repeatable manner.



Figure 2: Dynamic Driving Simulator used for the Study

Objective and Subjective Criteria

The objective evaluation is based on the longitudinal acceleration signal $a(t)$, since it directly reflects the traction behavior perceived by the participant and therefore represents the physical interface between powertrain behavior and longitudinal comfort perception. The longitudinal acceleration is the mass-normalized consequence of the resulting traction force at the vehicle. When traction is lost during shifting, this manifests as a drop in $a(t)$; when the shift is particularly abrupt, it appears in the rate of change of acceleration, that is, in the jerk. The selected variables were chosen to describe three main comfort dimensions of the shift event: interruption, dynamics, and duration.

Each shift event is analyzed in a defined shift window with the characteristic form plateau–ramp–down–hold–ramp–up–plateau, as shown in Figure 3. The event times t_1 to t_4 define the beginning of the systematic deviation from the pre-shift acceleration level, the end of the ramp–down, the end of the minimum region, and the end of the stabilized recovery respectively. Based on this window, the segment times and the total shift duration T_{shift} are calculated as:

$$T_{shift} = t_4 - t_1 \quad (1)$$

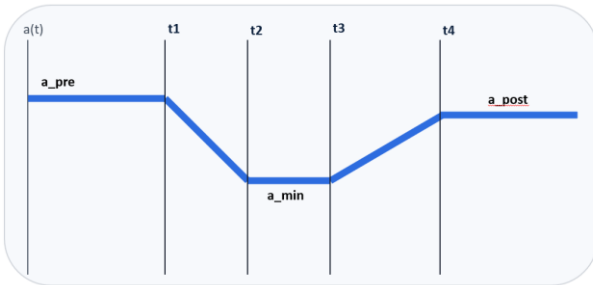


Figure 3: Acceleration Levels and Time Markers during Shifting

Three signal-derived variables are used to describe different aspects of longitudinal shift quality. The RMS value of the jerk j_{rms} characterizes the overall dynamic sharpness of the shift event and is defined as:

$$j_{rms} = \sqrt{\frac{1}{T_{shift}} \int_{t_1}^{t_4} \left(\frac{da(t)}{dt} \right)^2 dt} \quad (2)$$

The relative acceleration loss $a_{loss,rel}$ quantifies the reduction in longitudinal acceleration during the shift relative to the pre-shift level:

$$a_{loss,rel} = \frac{a_{pre} - a_{min}}{a_{pre}} \quad (3)$$

where a_{pre} is the acceleration level before the shift and a_{min} is the minimum acceleration reached during the shift. The area-based metric $A_{between}$ captures the overall deviation of the realized acceleration profile from the reference progression during the shift event, combining both the depth and the duration of the interruption:

$$A_{between} = \int_{t_1}^{t_4} |a_{ref}(t) - a(t)| dt \quad (4)$$

with a_{ref} representing the reference acceleration profile representing the ideal, uninterrupted acceleration trajectory without a shift event. Together, $a_{loss,rel}$ and $A_{between}$ describe the interruption-related comfort dimensions, while j_{rms} describes the dynamic sharpness of the transition.

The subjective evaluation is based on a standardized questionnaire using a five-point Likert scale. After each simulator run, participants rate whether the shift was noticeable, whether it was perceived as dynamic, whether it was perceived as comfortable, whether the acceleration was interrupted during shifting, and whether the configuration would be acceptable in everyday use. This selection directly mirrors the objective dimensions: the interruption rating corresponds to $a_{loss,rel}$ and $A_{between}$, the dynamics rating corresponds to j_{rms} , and the comfort and acceptance ratings allow an overall assessment across both dimensions.

Participants and Study Design

The participant study was conducted with adult test subjects recruited from the university environment, including students and staff members, aged between 18 and 65 years. A total of 30 participants were planned for the study. Before participation, subjects were informed about the procedure and screened for simulator or VR incompatibility as well as selected exclusion criteria relevant to safe simulator use. Demographic data and information on driving experience were collected. Participation was voluntary and could be terminated at any time without disadvantages.

The experiment follows a within-subject design, in which each participant experiences all predefined shift configurations in the driving simulator and evaluates them using a standardized questionnaire. The total duration per participant was approximately 40 to 60 minutes. The study procedure consists of three phases. In the briefing phase, participants are introduced to the study, and the simulator provide informed consent, complete a general well-being questionnaire, and receive safety instruction followed by a familiarization run. In the test phase,

participants evaluate the selected shift profiles in the simulator. Each run consists of an acceleration phase, a subsequent upshift event, and a deceleration phase to standstill. After each run, the participant rates the experienced shift event on a five-point Likert scale. The order of the presented profiles is permuted across participants to avoid order effects and reduce the influence of habituation or expectation on the subjective ratings. In the concluding phase, participants complete a final questionnaire on the overall procedure, simulator impression, and post-experiment wellbeing.

The simulator scenario is restricted to a straight virtual road and focuses exclusively on longitudinal maneuvers. The vehicle follows a predefined speed and acceleration profile automatically, so that active steering, braking, or throttle input by the participant is not required.

Results

Objective Evaluation

Table 1: Objective Criteria of Test Profiles

Profile	j_{rms}	$a_{loss,rel}$	$A_{between}$
D1S1	31.34	1.08	3.63
D1S3	1.25	0.54	1.15
D2S1	55.42	1.08	1.26
D2S2	27.91	0.81	0.90
D2S3	4.85	0.54	0.49
D2S4	4.65	0.27	0.16
D2S5	0.01	0	0
D3S1	64.88	1.08	0.7
D3S3	8.06	0.54	0.4
D3S5	8.35	0	0

The simulated test profiles a range of longitudinal shift characteristics, differing with respect to the magnitude of acceleration interruption, the steepness of the transient response, and the overall shift progression. The objective metrics for all investigated configurations are summarized in Table 1. From an objective perspective, the interruption-related metrics $a_{loss,rel}$ and $A_{between}$ follow the S-index directly by design: S1 configurations show the largest acceleration loss and deviation area, while S5 configurations show zero interruption-related loss. The jerk j_{rms} is primarily governed by the D-index, with D3 configurations exhibiting the highest values and D1 the lowest. Notably, D2S5 and D3S5 represent the planetary gearbox cases, for which no discrete interruption-related acceleration loss occurs, and the evaluation is therefore governed by the transient behavior and the jerk-related profile shape.

Subjective Evaluation

The subjective ratings reveal clear differences between the investigated configurations with respect to comfort, daily use acceptance, perceived acceleration interruption, and shift noticeability. The results are summarized in Tables 2 and 3.

The highest comfort ratings were obtained for D3S5, D1S3, and D2S5, while the lowest were observed for D3S1, D1S1, and D2S1. A similar pattern was found for daily use acceptance, with D3S5 and D2S4 receiving the highest scores. The strongest perceived interruption was reported for D2S1, D1S1, and D3S1, whereas D3S5, D2S5, and D2S4 were rated as clearly less interruptive. The same configurations rated as most interruptive were also rated as most noticeable, while D3S5 and D2S4 were among the least conspicuous cases.

The confidence intervals of the mean ratings are small relative to the observed differences between configurations, indicating that the rating patterns reflect consistent perception tendencies across participants rather than random variation.

Table 2: Subjective Comfort and Daily Use results reported as mean $\pm$ 95% confidence interval

Profile	Comfort	Daily Use
D1S1	2.16 $\pm$ 0.49	1.80 $\pm$ 0.40
D1S3	4.31 $\pm$ 0.34	3.69 $\pm$ 0.38
D2S1	2.07 $\pm$ 0.35	1.93 $\pm$ 0.39
D2S2	2.70 $\pm$ 0.36	2.60 $\pm$ 0.39
D2S3	3.43 $\pm$ 0.40	3.46 $\pm$ 0.42
D2S4	4.07 $\pm$ 0.38	4.14 $\pm$ 0.31
D2S5	4.20 $\pm$ 0.27	3.77 $\pm$ 0.44
D3S1	1.87 $\pm$ 0.30	1.94 $\pm$ 0.35
D3S3	2.79 $\pm$ 0.35	2.76 $\pm$ 0.38
D3S5	4.35 $\pm$ 0.31	4.13 $\pm$ 0.39

Table 3: Subjective Shift Perception and Acceleration Interruption results reported as mean $\pm$ 95% confidence interval

Profile	Perception	Acceleration Interruption
D1S1	4.64 $\pm$ 0.26	4.28 $\pm$ 0.44
D1S3	2.97 $\pm$ 0.44	2.66 $\pm$ 0.43
D2S1	4.57 $\pm$ 0.31	4.27 $\pm$ 0.37
D2S2	4.33 $\pm$ 0.34	3.93 $\pm$ 0.40
D2S3	3.89 $\pm$ 0.37	3.25 $\pm$ 0.44
D2S4	2.93 $\pm$ 0.45	2.36 $\pm$ 0.38
D2S5	3.10 $\pm$ 0.41	2.30 $\pm$ 0.28
D3S1	4.68 $\pm$ 0.20	3.87 $\pm$ 0.41
D3S3	4.00 $\pm$ 0.32	3.42 $\pm$ 0.29

The one-way ANOVA confirmed that configuration had a highly significant effect on all four subjective dimensions, as summarized in Table 4. All F-values exceeded 18 and all p-values were smaller than 10^{-23} , indicating that the observed rating differences

across configurations are not attributable to random variation. Comfort showed the strongest configuration effect with $F(9, 258) = 33.43$, reflecting the high sensitivity of comfort ratings to the systematic variation of shift characteristics.

Post-hoc Tukey-Kramer comparisons indicate that significant pairwise differences were concentrated between extreme S-levels: S1 configurations were consistently rated significantly worse than S4 and S5 configurations across all four subjective dimensions, with differences reaching high significance throughout. Adjacent S-levels — for example S1 vs S2, S2 vs S3, and S4 vs S5 — were generally not perceptually distinguishable, suggesting a non-linear perception function with saturation effects at both ends of the support scale. Within the same S-level, D-level effects were rarely significant for comfort and acceptance but emerged more clearly for noticeability and interruption at intermediate S-levels, where D3 configurations tended to be rated as more noticeable and interruptive than D1 configurations at equivalent support levels. A notable exception was the comfort comparison between D1S1 and D3S1, where the slow ramp at zero support was rated marginally more comfortable than the fast ramp, an effect attributable to the reduced jerk associated with the slower transient.

Table 4: One-Way ANOVA Results per Subjective Dimension

Variable	F(9, 258)	p-Value	Significance
Acceptance	23.91	1.22×10^{-26}	***
Comfort	33.43	3.42×10^{-36}	***
Interruption	20.41	3.37×10^{-24}	***
Noticeability	23.97	3.27×10^{-26}	***

The MANOVA further, as shown in Table 5 confirmed that the combined multivariate subjective response differed significantly across configurations. Three significant discriminant dimensions were identified, while Dimension 4 was not significant ($p = 0.2483$). The first and dominant dimension reflects the traction continuity effect, consistent with the strong influence of $a_{loss,rel}$ on comfort and perceived interruption. The second dimension captures a smaller but statistically significant additional source of variation, consistent with the contribution of shift dynamics and jerk to subjective conspicuity. The presence of exactly two significant dimensions, corresponding to the two independently varied factors D and S, supports the validation of the experimental design.

Table 5: MANOVA Results per Subjective Dimension

Dimension	p-value	Significance
1	8.89×10^{-26}	***
2	0.0058	**
3	0.02833	*
4	0.2483	n.s.

The confidence intervals of the mean ratings are small relative to the observed differences between extreme configurations across all dimensions, indicating that the rating patterns reflect stable and consistent perception tendencies across participants rather than random variation.

Correlation and Statistical Analysis

A comparison of the objective and subjective results shows that the selected metrics capture relevant and statistically significant aspects of perceived shift quality across all investigated pairs. The Spearman correlation coefficients and corresponding significance levels are summarized in Tables 5 and 6. Figure 4 visualises the correlation between subjective and objective acceleration interruption.

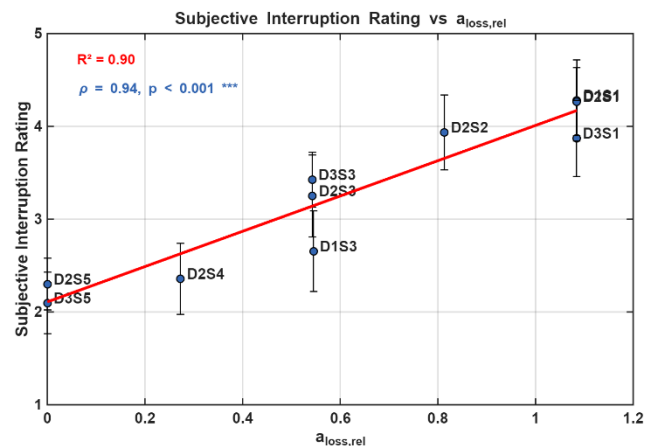


Figure 4: Correlation between felt interruption and $a_{loss,rel}$

Relative acceleration loss showed the strongest and most consistent correlations across all subjective dimensions. The correlation with perceived acceleration interruption was the strongest finding in the entire dataset ($p = 0.94, p < 0.001, R^2 = 0.90$), confirming that $a_{loss,rel}$ directly reflects the traction loss felt by participants during the shift. Strong correlations were also found with comfort ($p = -0.81, p = 0.005, R^2 = 0.83$), acceptance ($p = -0.91, p < 0.001, R^2 = 0.86$), and noticeability ($p = 0.84, p = 0.002, R^2 = 0.84$), indicating that $a_{loss,rel}$ is not only the primary descriptor of perceived interruption but also the dominant contributor to overall comfort and acceptance perception.

$A_{between}$ showed strong Spearman correlations with interruption ($\rho = 0.85$, $p = 0.002$) and noticeability ($\rho = 0.71$, $p = 0.022$), but weaker linear fits ($R^2 = 0.47$ and 0.36 respectively). The divergence between R^2 and ρ is explained by the behavior of D1 profiles: the slow ramp rate inflates the area metric through long shift duration without a proportional increase in perceived discomfort, since participants appear to adapt to gradual transitions more readily than to abrupt ones. This confirms that $A_{between}$ captures a relevant but duration-sensitive dimension of the interruption that is complementary to $a_{loss,rel}$ rather than equivalent to it.

Table 5: Linear regression coefficient of determination R^2 with corresponding p-values.

Obj. Metric	Subj. Metric	p-value	R^2
$a_{loss,rel}$	Interruption	< 0.001	0.90
$a_{loss,rel}$	Acceptance	< 0.001	0.86
$a_{loss,rel}$	Comfort	0.005	0.83
$a_{loss,rel}$	Noticeability	0.002	0.84
$A_{between}$	Interruption	0.002	0.47
$A_{between}$	Noticeability	0.022	0.36
$A_{between}$	Acceptance	0.003	0.48
$A_{between}$	Comfort	0.042	0.33
j_{rms}	Acceptance	0.024	0.65
j_{rms}	Comfort	0.018	0.66
j_{rms}	Noticeability	0.016	0.50
j_{rms}	Interruption	0.073	0.49

Table 6: Spearman rank correlation coefficients ρ .

Obj. Metric	Subj. Metric	Spearman ρ
$a_{loss,rel}$	Interruption	0.94
$a_{loss,rel}$	Acceptance	-0.91
$a_{loss,rel}$	Comfort	-0.81
$a_{loss,rel}$	Noticeability	0.84
$A_{between}$	Interruption	0.85
$A_{between}$	Noticeability	0.71
$A_{between}$	Acceptance	-0.83
$A_{between}$	Comfort	-0.65
j_{rms}	Acceptance	-0.72
j_{rms}	Comfort	-0.75
j_{rms}	Noticeability	0.76
j_{rms}	Interruption	0.60

The jerk metric showed significant correlations with noticeability ($\rho = 0.76$, $p = 0.016$, $R^2 = 0.50$), comfort ($\rho = -0.75$, $p = 0.018$, $R^2 = 0.66$) and acceptance ($\rho = -0.72$, $p = 0.024$, $R^2 = 0.65$). The correlation with noticeability is consistent with the hypothesis that jerk reflects the sharpness of the acceleration transient perceived by participants. The correlation with comfort and acceptance, while significant, is partly attributable to the inherent confound between the D and S factors in the investigated profile set — configurations with high D-index also tend to have lower S-index, making it difficult to fully isolate the

jerk contribution from the interruption contribution in these dimensions. The correlation with interruption was borderline significant ($\rho = 0.60$, $p = 0.073$), further supporting the interpretation of jerk as a secondary rather than primary comfort descriptor. Overall, no single objective metric fully explains the complete subjective response. However, the results show clear and physically plausible relationships across all investigated pairs, with $a_{loss,rel}$ emerging as the dominant descriptor of perceived shift comfort and j_{rms} contributing primarily to shift conspicuity. The combined use of both metrics captures the two perceptual dimensions identified in the MANOVA and provides a physically motivated basis for the development-oriented assessment of longitudinal shift comfort.

Discussion & Conclusion

The objective of this work was to determine which objective metrics derived from the longitudinal acceleration profile are suitable for describing shift comfort during upshifts in two-speed battery electric vehicles, and how these metrics correlate with the acceleration-related sensations perceived by test participants in a dynamic driving simulator.

The results support the working hypothesis. In the human-in-the-loop simulator study, the one-way ANOVA confirmed that shift configuration exerted a highly significant effect across all four subjective dimensions, while the MANOVA corroborated this at the multivariate level, identifying two significant discriminant dimensions corresponding to the two independently varied experimental factors, shift dynamics and traction support. This convergence indicates that the investigated configurations produced clear, systematic differences in subjective perception rather than random variation. The correlation analysis further showed that the selected objective metrics are suitable for describing the relevant comfort dimensions of the shift event. The interruption-related metrics $a_{loss,rel}$ and $A_{between}$ exhibited the strongest correspondence with perceived interruption and comfort, confirming traction continuity as the dominant contributor to shift comfort; the weaker linear fit obtained for $A_{between}$ relative to its rank correlation indicates a monotonic but non-linear relationship. By contrast, j_{rms} correlated most strongly with shift noticeability and only marginally with perceived interruption and is therefore better interpreted as a complementary descriptor of shift conspicuity than as a primary comfort variable. Acceptance ratings reflected the combined influence of both, consistent with everyday usability being governed by multiple perceptual dimensions simultaneously.

The independent variation of the D and S factors proved methodologically valuable, since a shift event may be perceived as uncomfortable owing to a deep traction interruption, an abrupt transient, or a combination of both — a separation that would otherwise have been difficult to isolate in the subjective responses, and one supported by the identification of exactly two significant MANOVA dimensions. The planetary gearbox profiles D2S5 and D3S5, which produce a largely continuous acceleration transition, consistently received favorable comfort and acceptance ratings even at higher jerk levels, reinforcing that traction continuity has a stronger influence on comfort perception than transient sharpness alone. Overall, the presented methodology enables shift comfort to be assessed prior to physical drivetrain hardware availability and provides a basis for perception-oriented development of shift strategies in multi-speed electric vehicles. Future work will focus on formulating a mathematical correlation model from the findings presented here, and on transferring the methodology to a real vehicle environment, where selected shift profiles will be reproduced on a physical drivetrain and evaluated using the same objective metrics and subjective procedure to enable direct comparison with the simulator-based results.

Acknowledgements

The presented contents are part of the Opt4E (Opt4E.de) research project (01MV23031), funded by the Federal Ministry for Economic Affairs and Energy (BMWE) and supervised by the Project Management Agency DLR-PT. The authors thank them for the sponsorship and support from the BMW and the Project Management Agency DLR-PT. Many thanks go to all funding bodies, project partners from industry, and universities for the great collaboration and support during the application for the project and the ongoing partnership. Thanks to the Research Association for Drive Technology e. V. (FVA) for supporting the project.

Supported by:



Federal Ministry
for Economic Affairs
and Energy

Opt4E.de

on the basis of a decision
by the German Bundestag

References

- Ahssan, M., Ektesabi, M. & Gorji, S., 2018. Electric Vehicle with Multi-Speed Transmission: A Review on Performances and Complexities. *SAE International Journal of Alternative Powertrains*, 7(2), p. 169–181.
- Erler, P. et al., 2018. Investigating the perception of powertrain shuffle with a longitudinal dynamic driving simulator. *2018 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, pp. 1427–1431.
- Ghasemi, N. et al., 2020. Driver response to gear shifting system in motion cueing driving simulator. *VEHICULAR 2020: The Ninth International Conference on Advances in Vehicular Systems*.
- Kim, D., Jeong, M., Bae, B. & Ahn, C., 2019. Design of a Human Evaluator Model for the Ride Comfort of Vehicle on a Speed Bump Using a Neural Artistic Style Extraction. *Sensors*, 19(24).
- Kraft, E., He, P. & Rinderknecht, S., 2023. Evaluation of drive-off procedures in dynamic driving simulators: Multisensory cues, interaction and sense of presence. *Transportation Research Part F: Traffic Psychology and Behaviour*, pp. 129–142.
- Kraft, E., Viehmann, A., Erler, P. & Rinderknecht, S., 2021. Virtual Driving Tests of Powertrain Systems in the Driving Simulator. *ATZ Worldw*, pp. 40–45.
- Liang, Q., Xiang, Y., Guo, L. & Chen, H., 2013. The seamless gear shifting control for pure electric vehicle with 2-speed inverse-AMT. *IFAC Proceedings Volumes*, pp. 603–610.
- Machado, F., F. N. & Fraser, R., 2021. Multi-speed gearboxes for battery electric vehicles. *International Journal of Vehicle Technology*, pp. 419–434.
- Nguyen, T. S. et al., 2019. Design and Development of a Real-Time Simulation and Testing Platform for a Novel Seamless Two-Speed Transmission for Electric Vehicles. *ASME. J. Dyn. Sys., Meas., Control*, 141(2).
- Schöneberger, D., Rinderknecht, S., Reitmeier, D. & Mertens, A., 2020. Development of a transmission control for an innovative high-speed powertrain using motor-related control: Simulative approach to develop an angular-position-controlled engagement of dog clutches in a two-drive transmission. *Drivetrain for Vehicles 2020*, pp. 195–208.
- Sorniotti, A. et al., 2011. A Novel Seamless 2-Speed Transmission System for Electric Vehicles: Principles and Simulation Results. *SAE Int. J. Engines*, 4(2), pp. 2671–2685.
- Viehmann, A., Schleiffer, J.-E., König, R. & Rinderknecht, S., 2019. Innovative powertrain concept for the smart vehicle of the future. *ATZ worldwide*, pp. 70–75.
- Yu, Z. et al., 2012. Research on an integrated control strategy to improve electric vehicle shift quality. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, pp. 1404–1414.

Usability of a mixed reality head mounted display in a static driving simulator

Eva Morgand¹, Florent Colombet¹, Javier Posselt¹, Nicolas Mignotte¹

(1) Renault Group, Technocentre, 1 avenue du Golf 78280 Guyancourt, e-mail: saceva78@gmail.com, {florent.colombet, javier.posselt, nicolas.mignotte}@renault.com

Abstract – This paper presents the setup of a head-mounted display (HMD) in a static driving simulation mean of Renault Group and a user behavioral comparison. The differences were studied between a three-screens display and an HMD. We first investigated how to validate a driving simulator, observing that there is no standardized method nowadays, and therefore chose to measure the most common variables used in the state-of-the-art research. The measures include subjective questionnaires, driving performances and head rotation; and were taken from a total of 21 participants. Similar driving performances were observed but driving with an HMD have led to a greater mental load. Screens and HMD induced the same feeling of presence, but the participants felt more comfortable when driving with the screens. In general, future work on the HMD visual quality should enhance the driving experience.

Keywords: Mixed Reality, Driving Simulation, Motion Sickness, Presence, Driving Performance

Copyright: © 2026 by the authors.

Licensee Driving Simulation Association.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.82157/dsa/2026/7>

Introduction and related work

Driving simulators are increasingly used in automotive development to reduce physical prototypes, evaluate more design alternatives, and expose drivers to potentially risky situations without real danger. Immersive, driver-in-the-loop simulation is particularly useful when subjective feedback is required. Conventional driving simulators generally rely on large projection systems or screen displays, whereas head-mounted displays (HMDs) offer a more compact and potentially more immersive alternative. However, HMD use can affect motion perception and may increase simulator sickness, especially in static simulators where visual and vestibular cues may conflict (Reason and Brand, 1975; Kemeny, Chardonnet and Colombet, 2020). Simulator sickness, cybersickness, and motion sickness share related symptoms but differ according to context and sensory conditions (Stanney, Kennedy and Drexler, 1997; LaViola Jr, 2000; de Winkel, Talsma and Happee, 2022).

Previous work has reported several limitations of HMD-based driving simulation, including restricted field of view, latency, visual resolution, and discomfort (Blissing and Bruzelius, 2018). Parduzi, Venrooij and Marker (Parduzi, Venrooij and Marker, 2020) compared an HTC Vive Pro with three LCD screens and observed that drivers made more efforts

to stabilize their position with the HMD, without significant simulator sickness differences. Heiderich et al. (Heiderich et al., 2025) compared a Varjo XR3 with a projection system and reported no simulator sickness difference, although participants felt more present with projection. In these studies, the absence or quality of body, hand, or steering-wheel representation was identified as a possible factor affecting behaviour.

Simulated driving with HMD has thus already been compared to simulated driving with regular screens. However, no study seems to have enabled the driver to see his/her hands or his/her body while driving.

This research aims thus to explore the feasibility of using HMDs for immersive simulations in terms of technological usability and user behaviour, with body or hand incarnation. Specifically, this experiment examines whether HMDs lead to similar driving performances, compared to screens display, and a better presence. As user comfort affects performance, it is also a key component of the evaluation.

Presence is considered a subjective experience of being in the virtual environment, while immersion refers to the system's objective capacity to dominate the user's sensory experience (Witmer and Singer, 1998; Slater, 1999). Because no standardized method currently exists for validating driving simulators, this study will focus on commonly used indicators of relative validity, including

questionnaires, driving performance and head rotations (Wynne, Beanland and Salmon, 2019).

Experimental setup and protocol

The experiment was conducted on Renault Group's SEI static driving simulator, used as a reference platform for Human-Machine Interfaces (HMI) studies (Figure 1). Two display conditions were compared using the same driving simulation software: a three-screen setup and a Varjo XR3 HMD. The screen configuration provided a 120° field of view with three 4K LCD screens, additional mirror and dashboard displays, four speakers, and two VIVE base stations. The positions of the mirrors relatively to the driver were similar as those of a Renault R5 (Figure 2). The Varjo XR3 provided a 115° horizontal field of view and a 90 Hz refresh rate, with focal and peripheral display zones. In the HMD condition, a Renault R5 virtual cockpit was geometrically aligned with the real steering wheel, and depth occlusion was used to allow real hands to remain visible (Figure 3 & Figure 4).



Figure 1: SEI Driving Simulator.

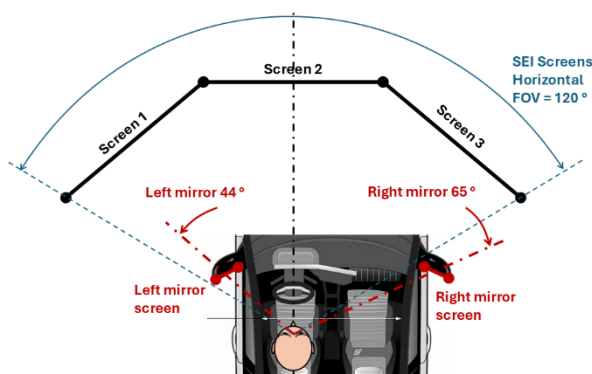


Figure 2: Field of view covered by SEI screens



Figure 3: Display in the HMD with the Renault R5 cockpit, real hands and the virtual environment.

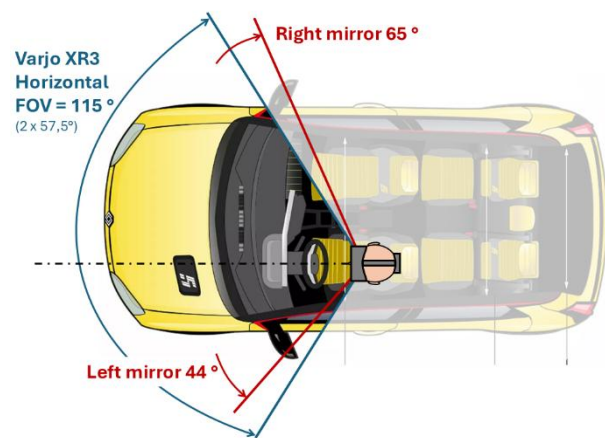


Figure 4: Field of view and virtual cockpit with the Varjo XR3 HMD

The simulated environment was a French national highway, the N104, with 10 to 15 surrounding vehicles continuously present in traffic (Figure 5). The route was selected to include a potentially recognizable highway context while avoiding difficult turns and manoeuvres that may increase simulator sickness (Mourant et al., 2007; Roe, Brown and Watson, 2007; Verwey, 2000). The scenario included overtaking, being overtaken, and merging situations, with acceleration, braking and merging triggers.

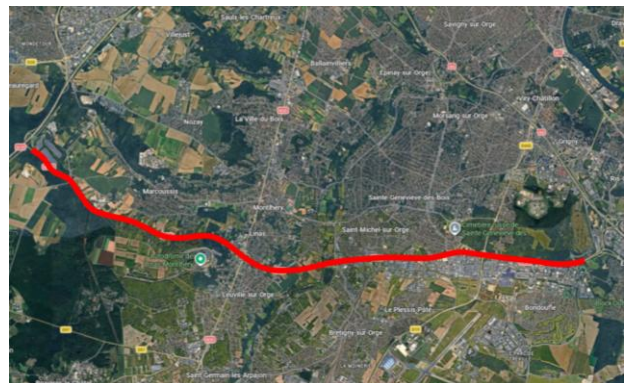


Figure 5: Part of the N104 highway used for the experiment.

The study used a within-subject design with 21 licensed participants. The sample included 2 women and 19 men, with a mean age of 44.6 years and a mean driving experience of 26 years. Display conditions were randomized and counter-balanced. Participants completed an initial Simulator Sickness Questionnaire, followed by a 2-minute training session and two main driving sessions of 8 to 10 minutes each. After each driving session, they completed questionnaires assessing simulator sickness, presence, workload and subjective driving performance.

Simulator sickness was assessed with the Simulator Sickness Questionnaire (Kennedy *et al.*, 1993). Presence was assessed with the Igroup Presence Questionnaire (Schubert, Friedmann and Regenbrecht, 2001). Mental workload was assessed with the NASA Task Load Index (Hart and Staveland, 1988), as workload may influence visual attention to dashboards and mirrors (Ostlund and others, 2006; Tsai *et al.*, 2007). Driving metrics were selected from validation literature (Wynne, Beanland and Salmon, 2019), including speed, lane gap, lane-position variability and steering-wheel direction changes. Head rotations were measured in pitch and yaw using either HMD sensors or a Vive Tracker mounted on a cap (Wells and Venturino, 1989; Kollenberg *et al.*, 2010).

Results

Oral feedback

Participants reported that speed and distance perception were difficult for both visual rendering conditions, mainly because of missing haptic feedback and inertial restitution. Preference results are summarized in Figure 6.

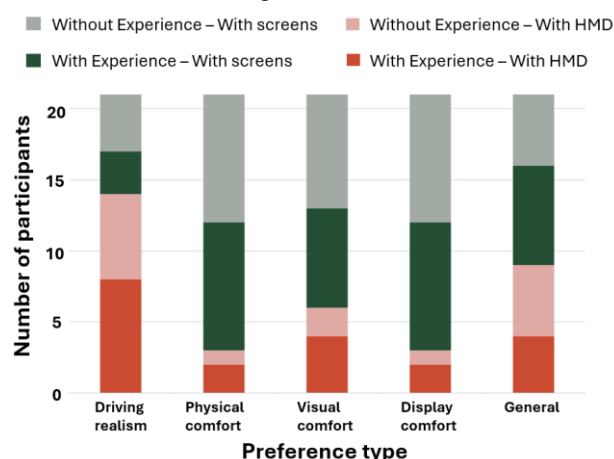


Figure 6: Participants' choices for each preference.

Screen visual quality was judged better, with better angular resolution and a higher frame rate. The HMD was considered more immersive by some participants, especially because it displayed the

complete vehicle cockpit, but blurred mirrors, depth-occlusion artifacts, steering-wheel superposition and reduced field of view affected visual quality and comfort. More participants preferred driving with screens, even though the HMD was perceived by some as providing a more realistic driving situation.

Statistical analysis

10 in 21 participants scored less than 2 for experience in VR and/or video games so we decided to test whether experience in VR and/or video games had an impact on the user's behaviour changes due to display. We did a Shapiro-Wilk test for each variable to know if the data were normally distributed and, in the end, only 8 pairings in 31 variables were normally distributed both for screens and HMD display conditions. As we knew that our sample size was small, we preferred considering that our data was not normally distributed.

We made an Aligned Rank Transform (ART) for each variable and then assigned these ranks in a mixed ANOVA to counteract the non-normal distribution. No p-values of interaction effect were inferior to 0.05, meaning that experience in VR and/or video games had no influence on the behaviour changes dependent on the display condition in our experiment. We can say that the behaviour differences between driving with screens and driving with an HMD are not significantly different according to the user's experience. For each significant difference, we ran a t-test according to the source of this difference (experience or display).

Questionnaires

Participants felt more simulator sickness symptoms when driving with the HMD (see Figure 7). The nausea, disorientation and oculomotor scores are all significantly higher when driving with an HMD (respectively a p-value of 0.023, 0.012 and 0.04).

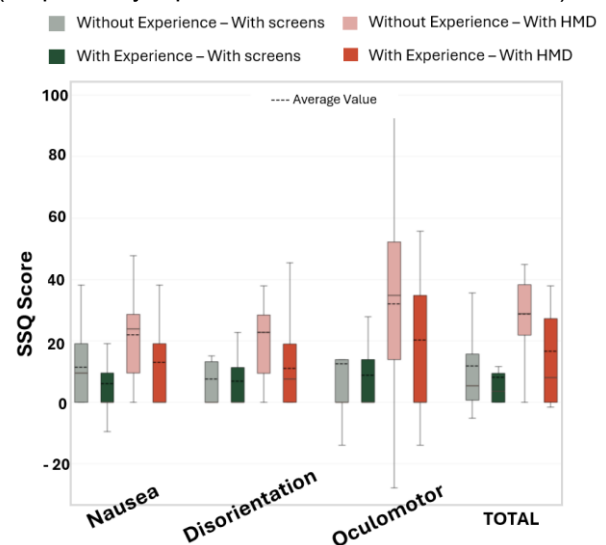


Figure 7: Differences sub-scores and total score of the SSQ between pre and post driving states

For presence, the IPQ revealed no difference in presence (see Figure 8). There is nevertheless a significant difference between the display conditions only for the involvement subscore, participants feeling more involved when driving with an HMD (p -value < 0.01). Involvement means here that the stimuli from the real world were ignored.

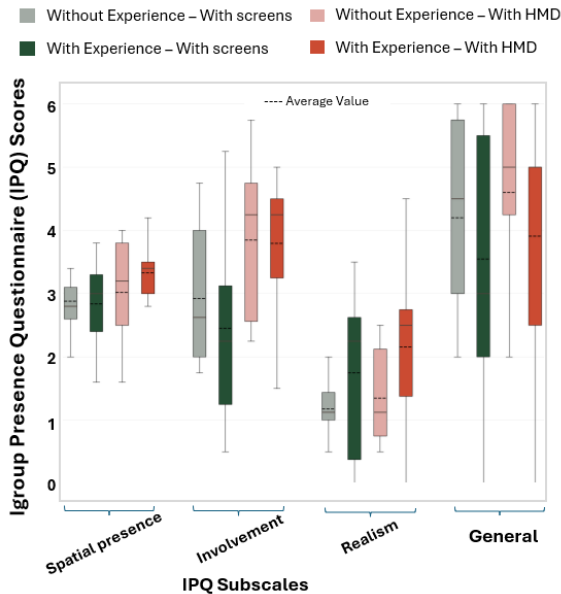


Figure 8: Sub-scores and total scores of the Igroup Presence Questionnaire (IPQ)

Results of the NASA-TLX are shown in Figure 9. Participants considered making more head-neck and visual efforts when driving with an HMD (p -values < 0.01).

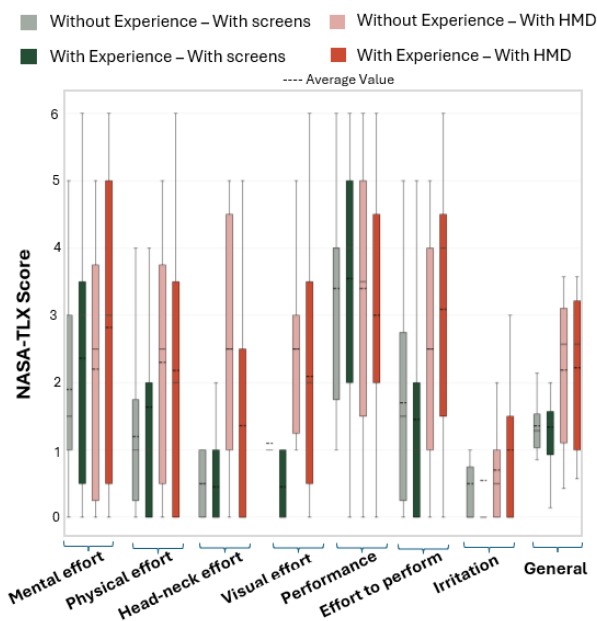


Figure 9: Subscales and general score of the NASA -TLX

The visual artifacts of the HMD were qualified as disturbing by half of the participants (10 for the focal

zone and depth occlusion, 13 for the blurred mirrors). The weight of the HMD is certainly one of the reasons for the head-neck efforts. It is also important to note that participants claimed to make more efforts with the HMD to perform at their level (p -value of 0.002). Concerning mental workload, the general score of the NASA-TLX is significantly higher with an HMD (p -value of 0.001).

Driving performance

In terms of subjective driving performance, the HMD brings a better speed perception according to subjective questionnaires (p -value of 0.04). There is no interaction effect between experience in VR and/or video games, and the display condition.

As can be seen on Figure 10, the cumulative number of steering wheel direction changes over the scenario is significantly higher when driving with an HMD (p -value < 0.001). Keeping their lane position is objectively more difficult with the HMD.

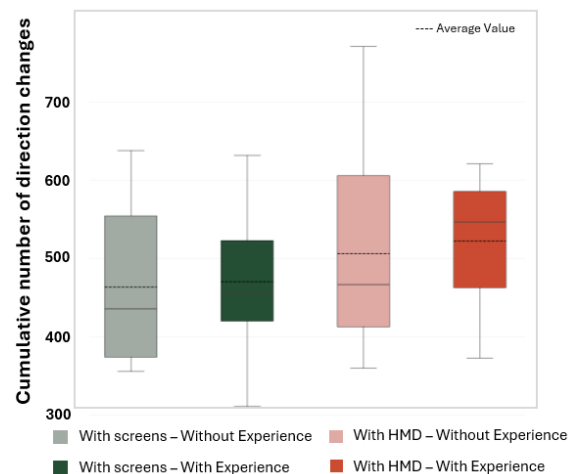


Figure 10: Cumulative number of direction changes for the steering wheel.

Lane gap, i.e. the gap between the vehicle lateral position with regards to the center of the lane, is close to zero for both groups and both display conditions, meaning that the HMD has no influence on the lane position. This level of performance is due to an increased effort to maintain the lane position, according to oral feedback. However, the participants with no or few experience had a less consistent lane position (p -value of 0.008).

Head rotations

Concerning head rotations, when we compare head angular direction changes, the amount between screens and HMD is similar around vertical axis (yaw movements), and higher with screens around lateral axis (pitch movements, p -value < 0.001).

In terms of amplitude (see Figure 11), pitch movements are significantly less wide with the HMD (p -value < 0.001), whereas yaw movements are significantly wider with the HMD (p -value = 0.004).

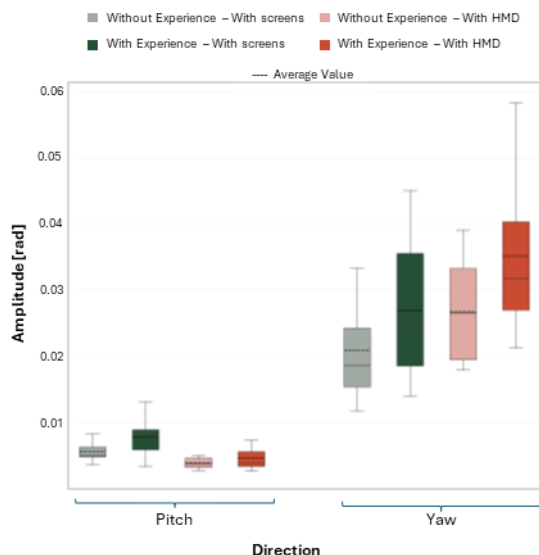


Figure 11: Head angular amplitudes

Discussion

Stereoscopy

A difference between the 2 visual rendering conditions lies in the fact that vision is in stereoscopy with the HMD, whereas it is in monoscopy on the screens. Some studies show an impact of stereoscopy on perceived speed for instance (Drouet, Colombet and Kemeny, 2022). But the closest elements of the vehicle cockpit (steering wheel, dashboard, mirrors, etc.) were physical and seen in stereoscopy in the screen condition. Thus, only the road and the virtual environment outside the car were seen in monoscopy and this should not have induced a bias in the experiment. However, this could also be an explanation of participants reporting a better speed perception with HMD.

Simulator sickness & Presence

The difference in simulator sickness symptoms between the 2 visual rendering conditions could come from the reduced frame rate (around 60 fps) in the HMD. In the projection-HMD comparison by (Heiderich *et al.*, 2025), the frame rate for both conditions was 90 fps and no difference in simulator sickness was noticed. Besides, a reduced framerate is known to create more simulator sickness (Kemeny, Chardonnet and Colombet, 2020). Also, in Heiderich experiment, participants had motion cues from the simulator, which is known to reduce simulator sickness.

We can also add that participants with no or few experience tend to show more symptoms of disorientation when driving with an HMD, but not significantly (p -value = 0.07). The general discomfort is then caused more by a visual expectation of movement than a lack of motion rendering. Literature

shows that experience in VR and/or video games slightly helps the user have fewer symptoms (Rosa *et al.*, 2016; Dużmańska, Strojny and Strojny, 2018).

We would have expected to obtain more spatial presence thanks to the HMD, especially since participants that preferred driving with the HMD said orally that they felt more present. But difference was not significative (p -value = 0.07). We think that the reduced visual quality with the HMD counterbalances the possible higher presence.

Steering wheel reversals & head movements

The significative difference in the number of steering wheel reversal shows that keeping their lane position was objectively more difficult with the HMD. This mainly explains why participants felt like making more efforts to do the same task with the HMD.

Heiderich *et al.* (Heiderich *et al.*, 2025) also observed more steering wheel reversals, but they thought that the lack of hand representation was the cause of this difference. In our experiment, the participants could see their own hands, but the superposition of the virtual and physical steering wheels maybe added a layer of difficulty rather than helped the users. Moreover, some participants said during the driving session with the HMD that the car deviates from its lane position when they look at one of the mirrors and, therefore, when they turn their head. It is maybe due to an instinctive movement of turning the steering wheel where we look, rather than keeping it at its initial position. The reduced FOV of the HMD makes the stabilization on the lane more difficult.

A possible explanation for the higher number of head pitch movements, is that the dashboard is more hidden by the steering wheel when driving with screens. The better speed perception perceived with the HMD may also have led participants to look at the dashboard not as often as with screens.

For yaw movements, we saw that the difference was only for amplitude. This can be explained by the fact that participants need to have the mirror in their focal zone to see it clearly with the HMD. Whereas in the screen conditions, mirrors are always in the field of view of the driver (see Figure 2) and need less amount of yaw movement to be seen clearly. Moreover, a greater head amplitude leads to an increased physical load for the head and the neck, and an unstable lane position as their eyes stop looking at the road.

Conclusion and future work

Conclusion

More symptoms of simulator sickness when driving with the HMD can be noticed, and experience in VR or video games has a positive effect on these

symptoms. Unlike our hypothesis, participants did not feel more present with the HMD. They however notice less stimuli from the real world with the HMD. More efforts, especially head-neck and visual, are needed for the participants when driving with the HMD.

Participants reported a better speed perception with the HMD, but the produced speeds showed no significant difference regarding the display condition or the experience. The higher number of steering wheel direction changes when driving with the HMD reveals a difficulty to maintain the lane position but participants succeeded in maintaining it. The reduced (physical) FOV of the HMD makes participants turn their head to look at their mirrors, making them lose sight of the road and hence possibly explaining the difficulty to keep their position on the road. Yaw movements were wider with the HMD, though with the same quantity between the 2 conditions. These differences can be explained by the reduced FOV of the HMD.

HMD requires participants to make more efforts to perform at the same level. This higher mental load explains why more people preferred driving with screens. But despite this, participants succeeded in driving similarly with screens and with the HMD.

Future work

This experiment showed similar results than those of (Parduzi, Venrooij and Marker, 2020) and (Heiderich *et al.*, 2025), except for simulator sickness. We know now that a similar driving behavior is possible with the HMD.

One step further in our work would be to incorporate an HMD in one of the Renault's dynamic simulators. This would allow us to have a more high-level synthesis simulation driving, testing both dynamic behavior of the vehicle, embedded software & systems and interior design such as future HMIs. This also rises new questions about incarnation methods or HMD choice criterias. Especially in a dynamic simulator with high accelerations up to 1G, it is not clear if a HMD lighter but with a smaller FOV would be preferable to a heavier HMD with a large FOV. The weight of the HMD could be in this case a safety criteria in terms of neck pain and damages.

References

- Blissing, B. and Bruzelius, F., 2018. Exploring the suitability of virtual reality for driving simulation, *Driving Simulation Conference 2018*, pp. 163–166.
- Drouet, T., Colombet, F. and Kemeny, A., 2022. Contribution of stereoscopy and motion parallax for speed perception in driving simulator experiments, *Proceedings of the DSC Europe VR 2022*, pp. 93–100.
- Dużmańska, N., Strojny, P. and Strojny, A., 2018. Can simulator sickness be avoided? A review on temporal aspects of simulator sickness, *Frontiers in psychology*, 9, p. 2132.
- Hart, S.G. and Staveland, L.E., 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research, in P.A. Hancock and N. Meshkati (eds.) *Human Mental Workload*. Amsterdam: North Holland, pp. 139–183.
- Heiderich, M. *et al.*, 2025. Validation of the usability of head-mounted displays in a highly dynamic driving simulator, *Proceedings of the DSC Europe XR 2025*.
- Kemeny, A., Chardonnet, J.-R. and Colombet, F., 2020. *Getting Rid of Cybersickness*. Springer. Springer. <https://doi.org/10.1007/978-3-030-59342-1>.
- Kennedy, R.S. *et al.*, 1993. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness, *The International Journal of Aviation Psychology*, 3(3), pp. 203–220.
- Kollenberg, T. *et al.*, 2010. Visual search in the (un) real world: how head-mounted displays affect eye movements, head movements and target detection, *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, pp. 121–124.
- LaViola Jr, J.J., 2000. A discussion of cybersickness in virtual environments, *ACM Sigchi Bulletin*, 32(1), pp. 47–56.
- Ostlund, J. *et al.*, 2006. Effects of cognitive and visual load in real and simulated driving, *Transportation Research Part F: Traffic Psychology and Behaviour*, 9(4), pp. 249–258. <https://doi.org/10.1016/j.trf.2006.06.003>.
- Parduzi, A., Venrooij, J. and Marker, S., 2020. The effect of head-mounted displays on the behavioural validity of driving simulators, *Proceedings of the Driving Simulation Conference Europe VR 2020*, pp. 125–132.
- Reason, J.T. and Brand, J.J., 1975. *Motion Sickness*. Academic Press.
- Rosa, P.J. *et al.*, 2016. The immersive virtual reality experience: A typology of users revealed through multiple correspondence analysis combined with cluster analysis technique, *Cyberpsychology, behavior, and social networking*, 19(3), pp. 209–216.
- Schubert, T., Friedmann, F. and Regenbrecht, H., 2001. The Experience of Presence: Factor Analytic Insights, *Presence: Teleoperators and Virtual Environments*, 10(3), pp. 266–281.
- Slater, M., 1999. Measuring Presence: A Response to the Witmer and Singer Presence Questionnaire, *Presence: Teleoperators and Virtual Environments*, 8(5), pp. 560–565.
- Stanney, K.M., Kennedy, R.S. and Drexler, J.M., 1997. Cybersickness is not simulator sickness, *Proceedings of the Human Factors and Ergonomics Society annual meeting*, 41(2), pp. 1138–1142.
- Tsai, J. *et al.*, 2007. Task Performance and Eye Activity: Predicting Behavior Relating to Cognitive Workload, *Aviation, Space, and Environmental Medicine*, 78(5 Suppl), pp. B176–B185.
- Wells, M. and Venturino, M., 1989. Head movements as a function of field-of-view size on a helmet-mounted display, *Proceedings of the Human Factors Society 33rd Annual Meeting. Santa Monica, CA: Human Factors Society, Inc*, p. 9.
- de Winkel, K.N., Talsma, T.M. and Happee, R., 2022. A meta-analysis of simulator sickness as a function of simulator fidelity, *Experimental Brain Research*, 240(12), pp. 3089–3105.
- Witmer, B.G. and Singer, M.J., 1998. Measuring presence in virtual environments: A presence questionnaire, *Presence*, 7(3), pp. 225–240.
- Wynne, R.A., Beanland, D.J. and Salmon, M.A., 2019. Systematic Review of Driving Simulator Validation Studies, *Safety Science*, 117, pp. 138–151.

Human Vision Driven Standard for Automotive Lit Surfaces

Victor Loya Garnica¹

(1) Ansys, part of Synopsys, Inc., 2600 Ansys Drive, Canonsburg, Pennsylvania, 15317, USA

Abstract – The increasing volume of light-emitting devices such as signaling lamps and displays in automobiles is the product of stricter safety regulations, the need for enhanced communication, and a higher demand for customization. Therefore, there is critical importance in ensuring the proper validation of any new product that emits light from the user vision perspective using quantitative metrics.

Contextually, lit uniformity, brightness, colorimetry, and activation time, to mention a few, must coexist to achieve an acceptable concept before considering mass production or even some cost-prohibitive physical prototypes. Every company, particularly OEMs, specifies its own acceptance criteria based on subjective judgement most of the time. Truly, there is no single standard definition for this multi-factor metric across the automotive industry. For this reason, a new guideline for assessing illuminated surfaces is proposed. This novel method takes shape as a graph showing a relationship between viewing position and a mixed performance rating, illustrating the dependency and impact of the beforementioned parameters. It should be then possible to use this graphical score to evaluate a vehicle lamp, an infotainment screen, a head-up display, and any other similar device.

As a result, the adopters of this new human vision driven standard for automotive illuminated surfaces should benefit by replacing ambiguous terms for numerical rules. Ultimately, the development and validation processes in this field will become clearer and easier to follow, enabling designers and engineers to speak the same language.

Keywords: light uniformity, automotive lighting, human vision, light simulation, displays

Copyright: © 2026 by the authors.

Licensee Driving Simulation Association.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.82157/dsa/2026/8>

Introduction

The assessment of the lit appearance of engineered automotive parts is most of the time considered a non-objective process, resulting in unacceptable-looking products that are hard to describe with quantitative metrics. Because of the nature of the human eye's behavior against light, this evaluation must involve not only photometric measurements but also psychological and perception factors. This implies that the use of a just numerical analysis that ignores the human vision parameters is not sufficient.

Clearly, physical luminance is linear, but the human perception of brightness is not; Luminance is measured in cd/m^2 , but brightness is a subjective perceptual magnitude (Wysecki & Stiles, 1982). In fact, brightness follows logarithmic or power-law behavior. Humans perceive around twelve orders of magnitude of luminance (that is between seeing an object using starlight on a moonless night and staring at the solar disk at horizon).

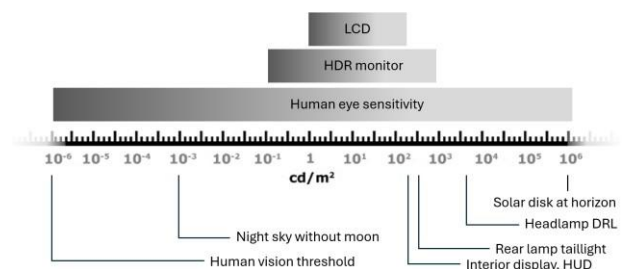


Figure 1: Luminance orders of magnitude

More importantly, the human eye is sensitive to relative contrast, not absolute light levels. It encodes differences. The eye retina responds to changes in luminance across space, making our vision more sensitive to gradients and contrasting edges. Per Weber's Law, the smallest noticeable difference in luminance ΔL is proportional to the background luminance L :

$$\Delta L/L = \text{constant} \quad (1)$$

This means that the eye is extremely sensitive to small changes at low levels of luminance and, in

contrast, it needs much larger increments to notice a difference at high luminance. This could explain why the mura defects are more evident in displays in dark ambient conditions.

Similarly, the concept of *Contrast Sensitivity Function* (CSF) must be considered. CSF quantifies the human eye's ability to differentiate light and dark areas (contrast) across different object sizes (spatial frequencies). A lit surface might be physically non-uniform by having small variations in luminance, but if those variations fall into spatial frequencies where the human eye is less sensitive, they won't be noticeable.

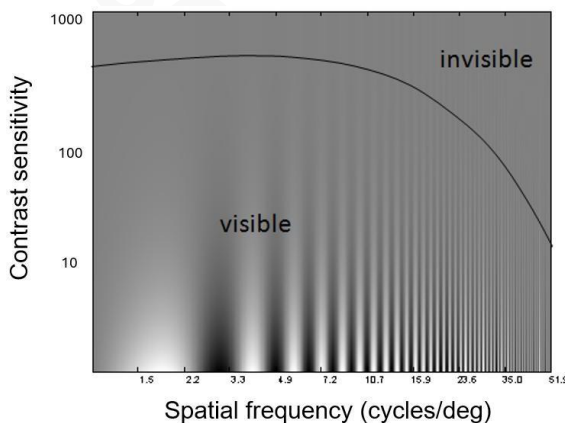


Figure 2: CSF peaks at 2-5 cycles per degree

As for color, CIE ΔE (Delta E) explains noticeable color change by expressing how far two colors (XYZ values) are separated in a perceptually uniform color space that models human vision.

Further concepts were considered but not mentioned in this paper for brevity.

Current Industry Solution

Today's vehicles are equipped with *Interactive Social Displays* (ISD) in China, *Signal Road Projectors* (SRP) in Japan, and lit emblems in various markets. All these illuminated parts are designed, in theory, to be ultimately perceived by the human eye. Or at least, they should be. Achieving perfect light homogeneity is difficult. OEMs need to ensure quality through a control method in the design phases of the different products.

Traditional light uniformity metrics including maximum-to-minimum (Max:Min), average-to-minimum (Avg:Min), and coefficient of variation (CV), treat all luminance variations equally, regardless of whether they are visible to humans. As a result, a smooth gradient and a checkerboard can score identical, which is unsatisfactory. Additionally, the viewing positions used for analysis in early stages of development of lit devices have no connection or are impractical to the end user interacting with the device

in most cases. For instance, measuring the light homogeneity of a low-mounted lamp (e.g. on the bumper, close to the ground) from an even lower location can be deemed as unnecessary.

Most uniformity metrics are derived, in fact, from instrumental pixel-wise measurements, while human perception integrates luminance and color spatially and contextually. This leads to cases where measured non-uniformities in production are missed because they fall below pixel-level thresholds.

It is true that engineers and studio designers struggle to interact because the latter describe "nice" aesthetics in subjective, emotional language, while engineers must realize them through precise, measurable requirements—creating a gap between creative intent and quantifiable technical constraints.

Proposed New Method

Given the complexity to associate engineering metrics with human vision performance and the current industry limitations around automotive lit surfaces, the following new method is proposed.

Analysis Workflow

Optical simulation and data processing tools were adopted for the creation of this workflow. The necessary steps were automated and follow the order shown in Figure 3.

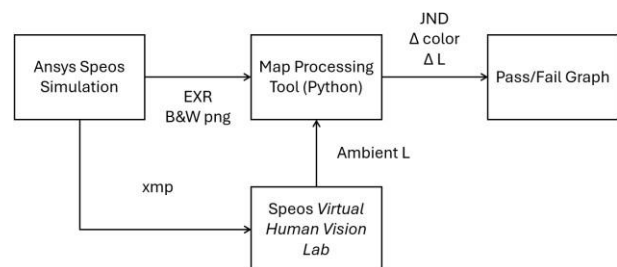


Figure 3: Analysis workflow for lit surface uniformity

Ansys Speos Simulation

The adoption of a simulation tool like Ansys Speos became relevant to address the testing. Thanks to its physics-based capabilities combined with its CIE-approved virtual human vision features, it is possible to quantify the quality of a 'virtual' product that is illuminated. A simulation of this kind results in multi-view luminance maps that provide enough data to create the applicable correlation.

Speos allowed us to compare different optical designs of the same external topology of lit surfaces (e.g. same vehicle lamp with different internal optics). The additional benefit of using a luminance extended map from Speos (xmp format) is its large spatial field that closely resembles the human vision's field of view. In fact, these luminance maps

are 2D representations of the eye's spherical projection.

In contrast to the traditional viewpoint locations, different positions are proposed. For exterior lighting, the viewing points are located at a constant height from the ground based on the assumption that the observer's eyes remain at a fixed distance from the floor (a person sitting in a car or a pedestrian crossing a street) as shown in Figure 4. The value of the 'constant height' is tentatively set to the average population height of the region of interest. The different positions of the viewing points parallel to the ground and pivoting around the lit object are set in a grid manner (Figure 5). These proposed points of view are more relevant and represent more realistic scenarios.

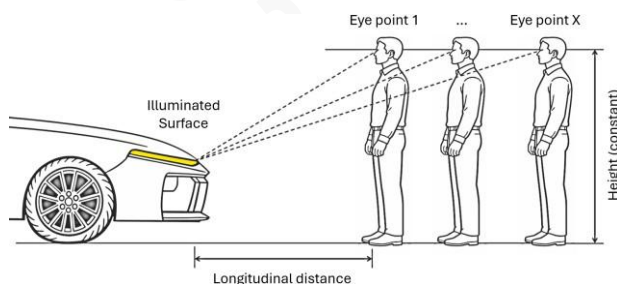


Figure 4: Viewpoint locations for exterior, side view

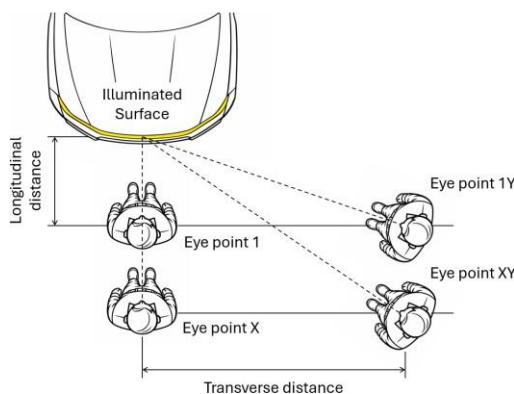


Figure 5: Viewpoint locations for exterior, top view

A similar approach can be applied to interior applications such head-up displays (HUD), infotainment screens, and interior lighting using an eye box distribution. In such case, the dimensions must be within the occupant local reach inside the cabin.

In either case, the viewpoints are defined as *Radiance Sensors* in Speos. Within the CAD environment of this light simulation tool, it is possible to extract luminance maps using these photometric camera-like sensors. Figure 6 shows an example of a luminance extended map (xmp) resulting from the simulation of an automotive position lamp plus contextual lighting from the environment in false color. A second map is also created using an idealized light source with no ambient lighting. This results in a B&W image that

will be used in a later stage of the analysis as a light filtering mask.

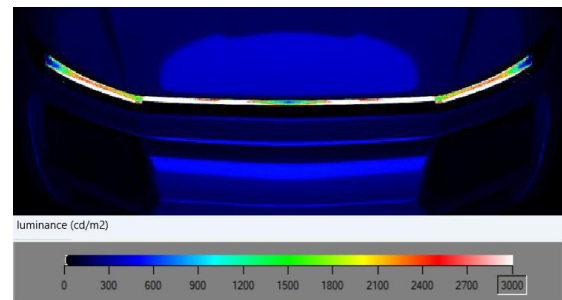


Figure 6: Photometric luminance map

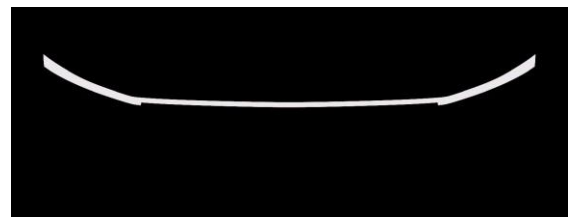


Figure 7: B&W mask

Speos Virtual Human Vision Lab

Since most of the monitors (laptops, phones, tablets, etc.) used to display luminance maps obtained using Speos do not have a dynamic range and brightness as the simulated products, with *Virtual Human Vision Lab*, you can mimic the visual appearance of modeled scenes on this media. This feature restores contrasts perceived by an observer placed in the scene and adds the contributions of glare, eye adaptation, observation time, and more. It provides a human eye-biased version of any photometric map as shown in Figure 8.

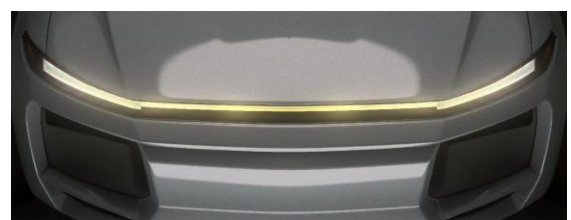


Figure 8: Human vision rendering from Speos showing a front Position light activated

Map Processing Tool

Clearly, Speos provides both quantitative and qualitative analysis tools. The remaining piece is a common language to describe them together. A Python-based map processing tool performs a perceptually based analysis of automotive lighting or displays luminance maps, focused on how bright and noticeable different regions appear to a human observer.

At a high level, it:

- Takes a luminance map turned into an HDR image (EXR format) and a region-of-interest mask (B&W png)
- Maps scene values to real-world luminance (cd/m^2) based on a specified peak output (e.g., DRL, display, etc.)
- Accounts for viewing conditions, including ambient luminance (night, overcast day, bright daylight) and image field of view (how large the source appears to the eye)
- Estimates human visibility using *Just Noticeable Difference* (JND, Figure 9a) based perceptual models, producing a luminance visibility map (brightness salience) and a CIE Delta E map (perceived color/tint changes, Figure 9b)

The output is a set of spatial maps that show where brightness or color differences are likely to be perceptible, under realistic automotive viewing conditions.

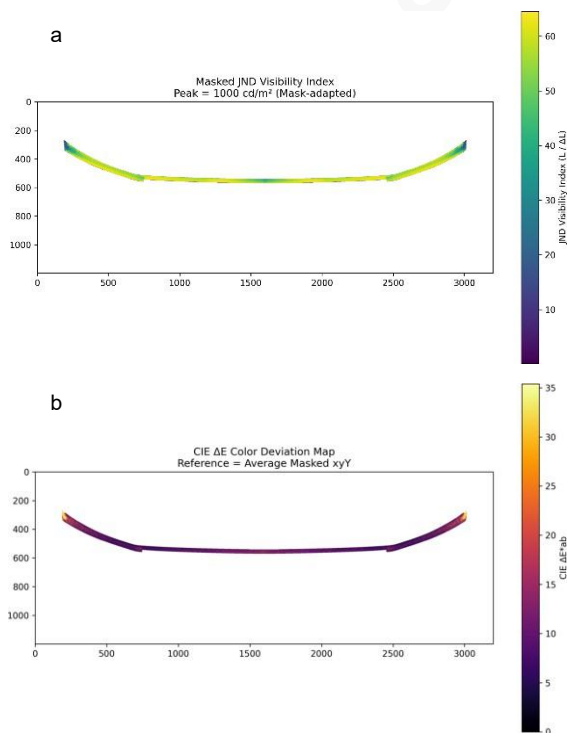


Figure 9: (a) JND visibility index, (b) CIE ΔE deviation map

Pass/Fail Graph

Three metrics per map define uniformity in a standardized pass/fail graph:

1. Relative JND – It answers how noticeable a pixel is compared to the average of the region. Values equal to 1 or less are below visibility thresholds (not noticeable). Higher values represent obvious differences.
2. CIE Delta E – This score ensures color compliance using a reference CIE XYZ value and

assesses the color shift within the lit surface. The lower the value the better (scores depend on specific application).

3. Delta Luminance – It measures the change in luminance from viewpoint to viewpoint using as a reference the object's front view average luminance (assumed to be maximum overall). Human detection of luminance changes follows Weber–Fechner behavior at photopic levels, with just noticeable differences of approximately 1–2% of background luminance (Barten, 1999).

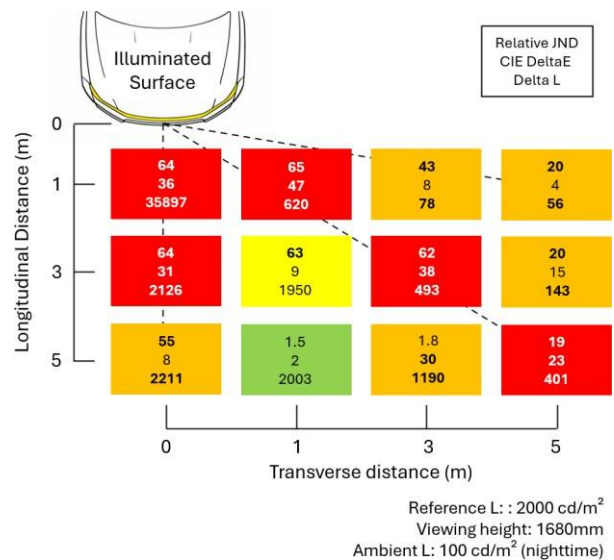


Figure 10: Lit uniformity pass/fail graph

A passing (green) score means that defects are not present from the human eye perspective. Specifically, nonuniformities, change in color, and change in average luminance are nonexistent. Yellow means one out of three scores failed, orange means two out of three failed, and red means all three scores failed.

By adopting this method, a light-emitting object that needs to be aesthetically pleasant can now be described with this graphical standard.

References

- Abboushi, B., Irvin, L., Rodriguez-Feo Bermudez, E., Royer, M., 2023. *Evaluating luminance uniformity metrics using online experiments*, pp. 2-30.
- Barten, P., 1999. Contrast Sensitivity of the Human Eye and Its Effects on Image Quality, pp. 39-71.
- Blackwell, R., 1946. *Contrast Thresholds of the Human Eye*, pp. 2-24.
- Wagner, L., 2025. *Detectability of Nonuniformities in Automotive Exterior Displays : A Study for Model Validation*, ISAL Conference.
- Wyszecki, G., & Stiles, W. S., 1982. *Color science: Concepts and methods, quantitative data and formulae* (2nd ed.), Wiley–Interscience.

MOTION CUEING & CONTROL & AI

A Model Predictive Control Algorithm for Matched Lateral/Yaw Motion

Maurice Kolff¹, Daan M. Pool², Vincenzo Ricciardi³, Martin Heiderich³,
Frank Zantner³, Barys Shyrokau¹

(1) Delft University of Technology, Faculty of Mechanical Engineering, section of Intelligent Vehicles, 2628 CD Delft, e-mail : {m.j.c.kolff, b.shyrokau}@tudelft.nl.

(2) Delft University of Technology, Faculty of Aerospace Engineering, section of Control & Simulation, 2629 HS Delft, e-mail : d.m.pool@tudelft.nl.

(3) Honda R&D Europe (Deutschland) GmbH, Vehicle Technology Research, 63073 Offenbach, Germany.

Abstract - This paper describes a novel model-predictive control motion cueing approach for driving simulation. The simulator's lateral and yaw motions are explicitly matched through known relations from vehicle dynamics, rather than the state-of-the-art approach of optimizing for the lateral and yaw motion channels separately. To evaluate the implementation, three representative example motion systems are considered: (MS 1) a hexapod, (MS 2) a hexapod with a yaw-drive underneath, and (MS 3) a hexapod with a y-drive underneath. Offline, open-loop simulations for a driving scenario were compared objectively through the root-mean square difference. The results show the effectiveness of the matched approach compared to the state-of-the-art baseline approach without explicit matching. By purposefully reducing the motion in either the lateral or yaw direction, an improved motion reproduction compared to the state-of-the-art approach is achieved, which was confirmed by four drivers in a pilot study. While the presented approach requires further experimental validation, the matching of vehicle lateral and yaw motion may enable a leap towards vehicle dynamics-based motion cueing.

Keywords: Motion Cueing Algorithm, Model Predictive Control, Vehicle Dynamics

Copyright: © 2026 by the authors.

Licensee Driving Simulation Association.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.82157/dsa/2026/9>

1. Introduction

Over the last twenty years, Model-Predictive Control (MPC) Motion Cueing Algorithms (MCAs) (Dagdelen, et al., 2004; Jacumet, et al., 2026; Khusro, et al., 2020) have dominated motion cueing research and development in the field of driving simulation. The working principle of such algorithms is to optimize the simulator motion through a cost function, while respecting the workspace limitations through explicit optimization constraints. The typical approach in such algorithms is to optimize for the motion of the six motion channels (i.e., three specific forces and three rotational rates) separately. These channels are therefore considered independently in the optimization, such that the simulator motion is generally brought to these individual reference signals of the vehicle as closely as possible. Furthermore, different and independent weights are typically set for the different motion channels. This means that one typically only has limited control over the extent to which the relative magnitude between different motion channels, as present in the vehicle, is retained in the simulator.

In real-world driving, however, the six motion chan-

nels of the vehicle are not independent of each other. Most notably, in curve driving, the lateral acceleration and yaw rate are directly related through the vehicle's velocity. For motion cueing that is representative of real driving, drivers may therefore expect the same match between lateral acceleration and yaw rate as present in the real vehicle, even if the magnitudes of both motions are reduced. Especially in the expert evaluation of vehicles, ensuring motion magnitudes across different channels that match the real vehicle dynamics is crucial for valid simulator-based assessments.

Current MPC MCAs typically consider the lateral and yaw rate motion to be independent optimization channels (Hvitfeldt, Drugge, and Jerrelind, 2025; Jacumet, et al., 2026), such that this relative matching is not provided at all. If the lateral and yaw motion channels are independently optimized, there may thus be a resulting mismatch in the representation of the simulated vehicle dynamics. This can occur especially in a case where a motion system is able to provide a significant lateral specific force without the corresponding yaw motion, resulting in the sensation of understeering (too little yaw). In that case, it may be more beneficial to deliberately limit the lateral

specific forces provided in the simulator to match in magnitude with the provided yaw rate motion, and as a result, better correlation with the original vehicle's dynamics. Vice versa, simulators with large yaw motion and small lateral motion capabilities may create the sensation of oversteering (too much yaw), where it would be beneficial to purposefully reduce the yaw motion to match the lateral motion. Existing work has investigated additional control on sideslip dynamics of a vehicle, (Bruschetta, et al., 2018), and using a motion scaling approach (Hvitfeldt, Drugge, and Jerrelind, 2025). To the best of our knowledge, however, a general MPC implementation of magnitude-matched control between the lateral acceleration and yaw rate, without the need of explicit scaling relations, does not yet exist.

The contribution of this paper is to present a method to explicitly implement the matching between the lateral acceleration and yaw rate in an MPC cost function. The well-known vehicle dynamics relation between lateral acceleration and yaw is rewritten in terms of the lateral specific force rather than acceleration. It is then implemented as a reference tracking term in the MPC cost function. The paper provides an offline open-loop comparison of a state-of-the-art MPC (as the baseline) and the novel approach using the explicit vehicle dynamics "matching". For the evaluation, three motion systems are considered (MS 1: hexapod, MS 2: hexapod + yaw drive, MS 3: hexapod + y-drive) to demonstrate its functionality across a range of typical driving simulator motion systems with varying lateral and yaw motion workspaces. Apart from its successful implementation in curve driving (for which it was designed), the method is furthermore demonstrated to work on straight and inclined roads.

The structure of the paper is as follows. Section 2 derives the cost function formulation for the magnitude-matched lateral/yaw motion based on vehicle dynamics. Section 3 presents the simulation results for the different motion systems. A discussion is provided in Section 4. Finally, the paper is concluded in Section 5.

2. Method

2.1. Vehicle Dynamics

Considering vehicle dynamics, see Figure 1, the linear acceleration in the body frame B in the center of gravity can be calculated as (Jazar, 2025):

$$\begin{aligned} \mathbf{a}^B &= \dot{\mathbf{u}} + \boldsymbol{\omega} \times \mathbf{u} = \begin{bmatrix} \dot{u}_x \\ \dot{u}_y \\ \dot{u}_z \end{bmatrix} + \begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix} \times \begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix} \\ &= \begin{bmatrix} \dot{u}_x + u_z\omega_y - u_y\omega_z \\ \dot{u}_y - u_z\omega_x + u_x\omega_z \\ \dot{u}_z + u_y\omega_x - u_x\omega_y \end{bmatrix}, \end{aligned} \quad (1)$$

where $\mathbf{u}$ are the linear velocities of the body frame and $\boldsymbol{\omega}$ are its rotational rates. Hence, the vehicle lateral acceleration equals:

$$a_y^B = \dot{u}_y - u_z\omega_x + u_x\omega_z \approx \dot{u}_y + u_x\omega_z, \quad (2)$$

where the right-most expression holds for small u_z , i.e., for an even surface. The equation relates the lateral acceleration a_y and the yaw rate ω_z through the longitudinal velocity u_x and $\dot{u}_y$ (the time derivative of

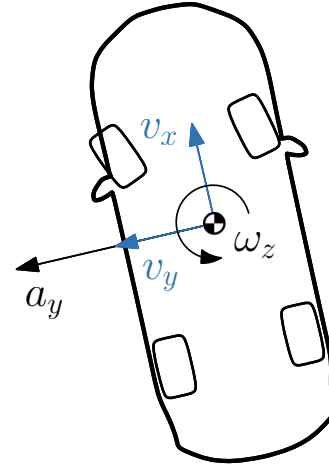


Figure 1: The vehicle system seen from the top, showing the relevant accelerations and velocities in the body reference frame around the center of gravity.

the lateral velocity u_y). These quantities apply to the vehicle's center of gravity, as indicated in Figure 1.

To provide fully matching motion between a_y^B and ω_z in the simulator, the relationship given by Eq. (2) must be respected. However, Eq. (2) cannot be directly incorporated into the MCA architecture. This is because the otoliths in the human vestibular system, located in the inner ear of the human body, are sensitive to specific forces, rather than accelerations. This makes the specific forces the relevant quantity for an MCA to reproduce. The specific forces may be considered as the combination of the accelerations and the reaction forces caused by gravity that acts on the driver via the seat (Stroosma, 2025). As a result, a tilting of the vehicle over a roll angle φ will thus also result in a gravity component in the body lateral specific force f_y^B .

In a real vehicle, a driver will thus perceive the linked dynamics of Eq. (2) between the lateral acceleration and yaw rate indirectly via the specific forces. Thus, in the simulator, the goal is to reproduce specific force and rotational rate motion that is matching via the vehicle relation of Eq. (2). Our approach is therefore to rewrite the vehicle dynamics relation of Eq. (2) to an equivalent form in terms of the specific forces and rotational rates, which may then be implemented in the MCA. We first consider the 3-element specific force vector of the vehicle, which may be calculated as:

$$\mathbf{f}^B = \mathbf{T}(\boldsymbol{\beta})(\mathbf{a}^I + \mathbf{g}^I) = \mathbf{a}^B + \mathbf{T}(\boldsymbol{\beta})\mathbf{g}^I, \quad (3)$$

with B the body system, I the inertial system, and where $\boldsymbol{\beta} = [\varphi, \theta, \psi]$, i.e., the Euler angles. The vector $\mathbf{g}^I = [0, 0, g]^T$ represents the contribution of gravity. Finally, the rotation matrix represents the attitude of the vehicle:

$$\mathbf{T}(\boldsymbol{\beta}) = \mathbf{R}_x^T(\varphi) \mathbf{R}_y^T(\theta) \mathbf{R}_z^T(\psi), \quad (4)$$

with each rotation matrix $\mathbf{R}$ representing pure rotations around the given axis. The lateral specific force then equals:

$$f_y = \mathbf{c}_2 \mathbf{f} = a_y^B + \mathbf{c}_2 \mathbf{T}(\boldsymbol{\beta})\mathbf{g}^I, \quad (5)$$

where the role of the selector vector $\mathbf{c}_2 = [0, 1, 0]$ is

to select only the lateral specific force. We may now substitute Eq. (2) into Eq. (5):

$$f_y = \dot{u}_y + u_x \omega_z + c_2 T(\beta) g^I. \quad (6)$$

This resulting equation is thus the specific force equivalent of the vehicle dynamics equation in Eq. (2). Note that this relation also remains true for a curved surface (non-zero roll φ and pitch θ angles) due to its definition in terms of the specific force rather than acceleration.

2.2. Simulator Control

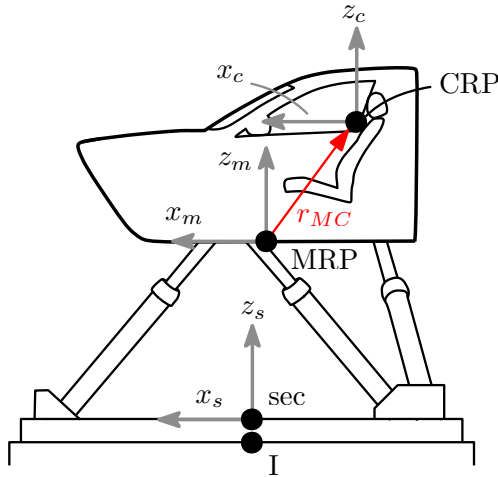


Figure 2: Simulator side view showing the relevant reference points and coordinate systems, adapted from (Kolff, et al., 2023).

Providing “matched” motion in terms of the vehicle dynamics relationship of Eq. (2) in the simulator thus implies constraining the simulator lateral specific force and yaw rate according to the relation of Eq. (6). The specific force cueing enabled by the simulator depends strongly on its kinematic configuration (Kolff, et al., 2023). In this paper, three relevant and representative driving simulator motion systems are considered:

- MS 1: A hexapod, the most widely adopted motion system, with a 12 cm stroke.
- MS 2: The same hexapod as MS 1, mounted on top of an additional $\pm 20^\circ$ yaw drive (e.g., Jacumet, et al., 2025).
- MS 3: The same hexapod as MS 1, mounted on an additional ± 5 m y-drive, similar to the systems available at VTI (Jansson, et al., 2014), IKA (Wagener, et al., 2023), and Daimler (Cleij, et al., 2019).

These different motion systems are used to demonstrate the functionality of the lateral/yaw motion matching for different available workspaces. A schematic side view of the considered hexapod system is shown in Figure 2, showing the relevant reference points. Note that in the case of MS 2 and MS 3 the additional motion axes under the hexapod are represented by the movement of the reference point “sec” (secondary), to which the hexapod motion of the Motion Reference Point (MRP) is relative. This point is where the motion signals are defined. Relevant for the driver’s perception is the Cueing Reference Point (CRP), approximating the location of the

human vestibular system in the inner ear. The difference between the two reference points is denoted by the position vector r_{MC} . In further analysis, the simplifying case of $r_{MC} = [0, 0, 0]^T$ is considered.

The specific forces generated by the motion system may be calculated as (Kolff, et al., 2023):

$$\begin{bmatrix} f_x^s \\ f_y^s \\ f_z^s \end{bmatrix} = \mathbf{R}_x^\top(\varphi^s) \mathbf{R}_y^\top(\theta^s) \mathbf{R}_z^\top(\psi^s) \left(\begin{bmatrix} a_x^s \\ a_y^s \\ a_z^s \end{bmatrix}_{hex} + \dots \right. \\ \left. \underbrace{\mathbf{R}_z^\top(\psi_d^s) \begin{bmatrix} 0 \\ a_y^s \\ 0 \end{bmatrix}}_{\text{MS 3 only}} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ g \end{bmatrix}}_{\text{MS 2 only}} \right). \quad (7)$$

The superscript $(\cdot)^s$ here indicates that these are “simulator” signals. Note that Eq. (7) again highlights why the matching can indeed not be implemented directly on the simulator acceleration, as the relevant quantity for the reference tracking is the specific force, which may be reproduced as a combination of the simulator acceleration and cabin tilting (tilt-coordination, Stratulat, et al., 2011).

Similarly, the simulator rotational rates are calculated as (Kolff, et al., 2023):

$$\begin{bmatrix} \dot{\omega}_x^s \\ \dot{\omega}_y^s \\ \dot{\omega}_z^s \end{bmatrix} = \begin{bmatrix} 1 & 0 & -\sin \theta^s \\ 0 & \cos \varphi^s & \sin \varphi^s \cos \theta^s \\ 0 & -\sin \varphi^s & \cos \varphi^s \cos \theta^s \end{bmatrix} \begin{bmatrix} \dot{\varphi}^s \\ \dot{\theta}^s \\ \dot{\psi}^s \end{bmatrix} + \underbrace{\dot{\psi}_{sec}^s}_{\text{MS 2 only}}. \quad (8)$$

2.3. Cost Function Formulation

To integrate the vehicle dynamics relation, a typical MPC cost function formulation is used:

$$\mathcal{J}_{(k)} = \sum_{i=1}^{N_p} (\mathbf{y}_{(i,k)}^v - \mathbf{y}_{(i,k)}^s)^\top \mathbf{Q} (\mathbf{y}_{(i,k)}^v - \mathbf{y}_{(i,k)}^s) \\ + \sum_{i=1}^{N_p} \mathbf{x}_{(i,k)}^\top \mathbf{R} \mathbf{x}_{(i,k)}. \quad (9)$$

where the first term represents the reference tracking, with $\mathbf{y}_{(i,k)} \in [f_x, f_y, f_z, \omega_x, \omega_y, \omega_z]_{(i,k)}^\top$ and its weighting matrix $\mathbf{Q}$. The superscript $(\cdot)^v$ is added to indicate the reference vehicle signals. The second term, with the weight $\mathbf{R}$, penalizes the states $\mathbf{x}$ for washout behavior and added cost function convexity (Katliar, 2020). The MPC performs the optimization at each time step k , over elements i to a prediction horizon N_p . Especially in the real-time application of MPC, additional terms on the optimization input and the change in input are common (Jacumet, et al., 2026). However, these are not applied in the present analysis due to the offline optimization.

For the driver in the simulator to perceive motion that corresponds to the vehicle dynamics, we again consider and slightly rewrite Eq. (6):

$$-f_y^v + \dot{u}_y^v + u_x^v \omega_z^v + c_2 T(\beta^v) g^I = 0. \quad (10)$$

The yaw rate and lateral specific force are now replaced by those of the simulator:

$$-f_y^s + \dot{u}_y^v + u_x^v \omega_z^s + c_2 T(\beta^v) g^I = 0. \quad (11)$$

f_y^s and ω_z^s are physical motion signals produced by

the simulator, whereas the signals u_x^v and $\dot{u}_y^v$, as well as the vehicle attitude β^v , are visual motion signals, which thus remain in the vehicle system. Eq. (11) can be directly implemented in the reference tracking term of Eq. (9), to replace the independent baseline f_y^v tracking term:

$$\mathbf{y}_{(i,k)}^v - \mathbf{y}_{(i,k)}^s = \begin{bmatrix} \dot{u}_y^v + u_x^v \omega_z^s + c_2 \mathbf{T}(\beta^v) \mathbf{g}^I \\ f_x^v \\ f_y^v \\ \omega_x^v \\ \omega_y^v \\ \omega_z^v \end{bmatrix}_{(i,k)} - \begin{bmatrix} f_x^s \\ f_y^s \\ f_z^s \\ \omega_x^s \\ \omega_y^s \\ \omega_z^s \end{bmatrix}_{(i,k)} \quad (12)$$

Thus, the MCA still optimizes the specific forces in Eq. (7) and rotational rates in Eq. (8), but now aims to track the relation of Eq. (11) that matches the yaw rate simulator motion rather than the full lateral specific force of the vehicle reference data.

2.4. Solver

The optimization results presented in this paper used a pre-recorded drive, such that the optimization was performed offline. Hence, to analyze a best-case MPC result, the optimization therefore had full knowledge of all future states (Ellensohn, et al., 2019). The optimization problem was programmed in CasADi Python (Andersson, et al., 2019) and solved using Ipopt (Wächter and Biegler, 2006) at $\Delta t = 0.01$ s.

2.5. Scenario

The tested scenario involves representative handling with a sedan car on a two-lane test track. It first considers a left-right-left curve sequence (5 s to 15 s), followed by a straight section, and then slalom motion (30 s to 55 s). The total drive length was 60 s. The velocity remained constant around 80 km/h. In practice, the simulator performs fading in and out at the beginning and end of the simulation, respectively. Therefore, the novel matching reference tracking was not active in the fading in until $t_s = 5$ s, and the fading out after $t_e = 55$ s.

2.6. Verification

Considering the implementation of an additional tracking term rather than a constraint, the optimization is still allowed to find solutions that do not exactly satisfy the matching relation. This can especially occur when a (too) low weight on the f_y^v tracking term is used. Therefore, we quantify the difference between the target lateral specific force term to the actual generated specific force via the Root Mean Square Difference:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i_s}^{N-i_e} [-f_y^s + \dot{u}_y^v + u_x^v \omega_z^s + c_2 \mathbf{T}(\beta^v) \mathbf{g}^I]^2} \quad (13)$$

with N the number of samples, $i_s = t_s/\Delta t$, and $i_e = t_e/\Delta t$. Considering the present case of open-loop optimization, we may set $N = N_p$.

3. Results

3.1. Objective Evaluation

The results for the three motion systems are shown in Figure 3a-3f. The reference vehicle data (blue) are tracked by both the “baseline” and “matched” MPC variants, which are not able to fully reproduce the reference data, as expected due to the limited available workspaces. The orange lines represent the baseline MPC, i.e., without the matching tracking term active. Furthermore, the orange dashed lines in the lateral specific force plots (left column) represent what the lateral specific force *should have been* based on the yaw rate (right column), if matching motion were provided. From the difference between the solid and dashed lines it is evident that the baseline method does not provide matching motion.

The green lines show the outputs of the MPC *with* the matching tracking term. The term was not active in the grey shaded areas (see Section 2.5). Again the dashed lines are plotted to represent the matching target, but they are mostly not visible as they correspond exactly to the matched MPC’s actual lateral specific force motions. Qualitatively, this thus shows that matching motion is indeed provided.

The matching method is also seen to be effective across the considered variety in motion systems. For example, for MS 2 with the additional yaw drive, the system strongly reduces the yaw motion (Figure 3d) below what the workspace actually allows to match the limited available lateral motion (Figure 3c). Furthermore, for MS 3 with the additional y-drive for lateral motion, the optimization deliberately reduces the available lateral motion (Figure 3e) to match with the limited available yaw motion (Figure 3f).

An explicit quantitative comparison of the achieved lateral specific force/yaw motion match is shown in Figure 4, with the actual lateral specific force f_y^s on the y-axis and the matching target on the x-axis. The blue data represents the vehicle data, which indeed indicates perfectly matched motion between the two quantities, as expected. If matched simulator motion is provided, it should thus also lie on this straight 45° line. Here the efficacy of the MPC with matching tracking (green) is clearly visible compared to the baseline MPC (orange). Furthermore, the RMSD (see figure legends) strongly decreases from 1.02 to $5 \cdot 10^{-4}$ (MS 1), from 0.89 to $1.23 \cdot 10^{-3}$ (MS 2), and from 1.13 to $5.11 \cdot 10^{-4}$ (MS 3), when including the matching term. This confirms that the novel approach effectively achieves matching motion. The RMSD of the vehicle reference data is shown for completeness; it equals $2.98 \cdot 10^{-3}$.

A fourth case is considered where the vehicle drives the same slalom road, but with a constant 10° inclination of the road, which leads to a constant roll angle. This creates a constant offset in the vehicle specific force, as shown in Figure 3g, while not affecting the yaw rate (Figure 3h). As the matching relation in Eq. (11) accounts for any effects of the vehicle orientation due to the term $c_2 \mathbf{T}(\beta^v) \mathbf{g}^I$, the proposed matched MPC is effectively able to handle the additional road inclination and will still provide matching motion. This is further shown in Figure 4d, where the data are shifted to the right by $c_2 \mathbf{T}(\beta^v) \mathbf{g}^I$ due to the road inclination, and by the same amount upwards in f_y^s due

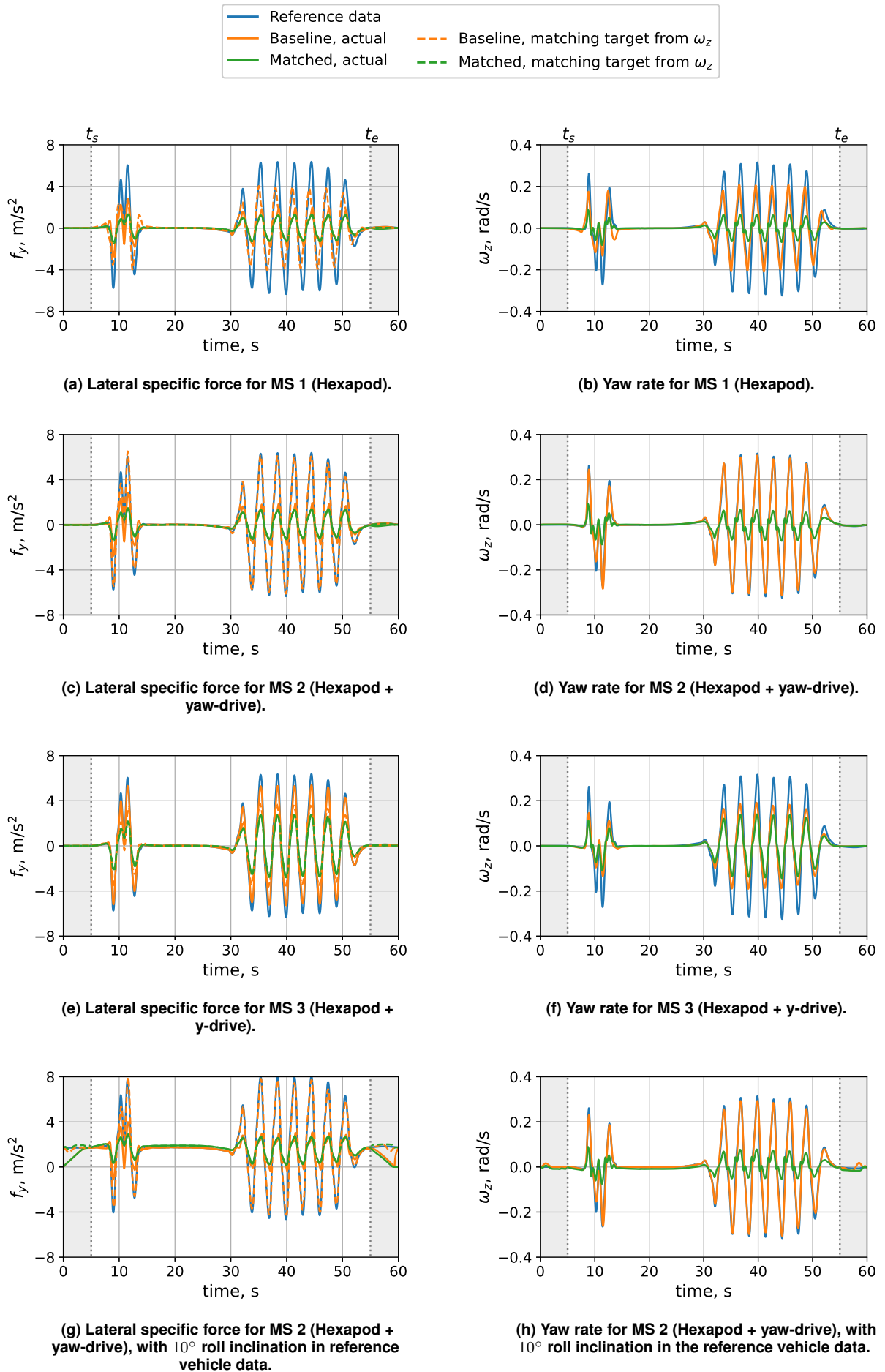
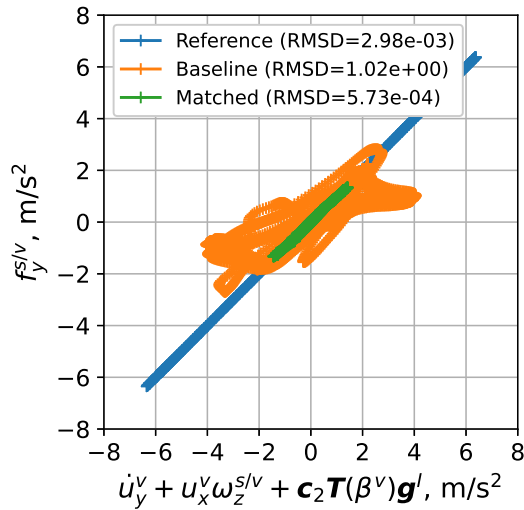
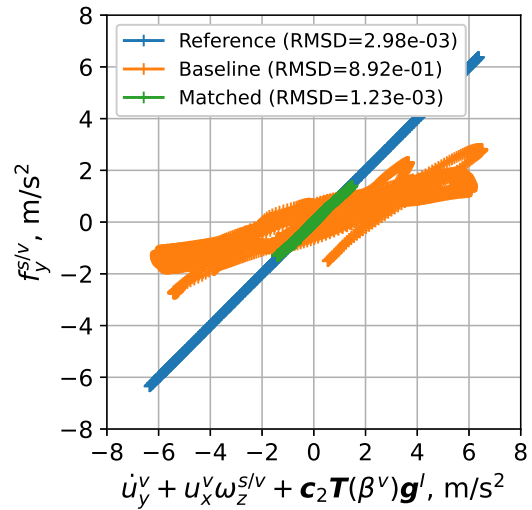


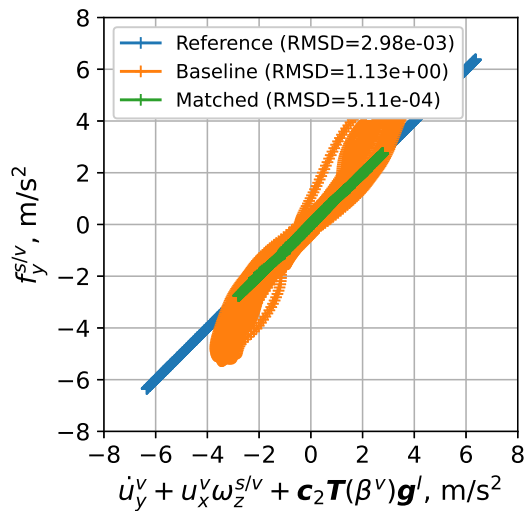
Figure 3: Optimization outputs for the baseline MPC (orange solid lines) and the MPC with the matching cost function (green solid lines), compared to the reference vehicle data (blue solid lines). Dashed lines indicate the matching lateral motion corresponding on the yaw rate. The grey areas were excluded from further analysis.



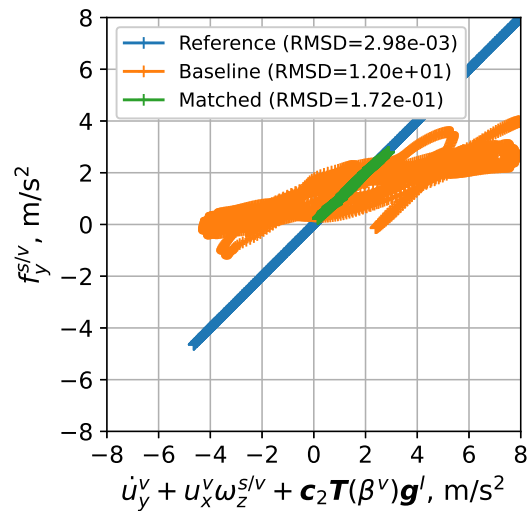
(a) Hexapod (MS 1).



(b) Hexapod + yaw-drive (MS 2).



(c) Hexapod + y-drive (MS 3).



(d) Hexapod + yaw-drive (MS 2), with 10° roll inclination in reference data.

Figure 4: Verification results of the optimization for the baseline MPC (orange) and the MPC with the matching tracking (green), compared to the baseline data (blue).

to the simulator rolling to reproduce the resulting specific force. The matching relation is thus valid, visible by the 45° blue and green lines. This shows that the matching MPC method is not restricted to fully flat road surfaces, but can directly handle a large variety of realistic road geometries.

3.2. Workspace Management

Figure 5 shows the time traces of the motion subsystems for the four conditions. In all four cases, the added matching provides a different workspace utilization in the hexapod. The magnitude of the hexapod stroke utilization is, however, similar. Larger differences are present in the yaw drive signals (Figures 5c and 5g), which are strongly reduced in magnitude in the matched approach. Overall, this shows

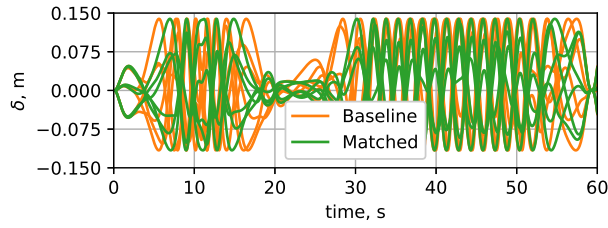
an improved workspace management of the novel matched approach.

3.3. Subjective Evaluation

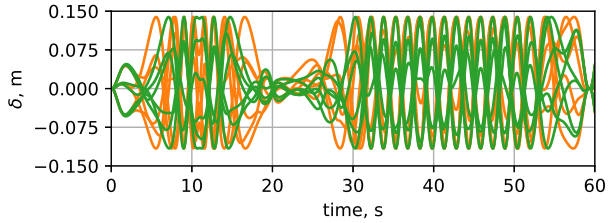
The baseline and matching MPCs were tested in a pilot study. All four drivers from TU Delft subjectively preferred the matching MPC over the baseline version. A larger controlled experiment is planned to systematically investigate the subjective differences between the two methods.

4. Discussion

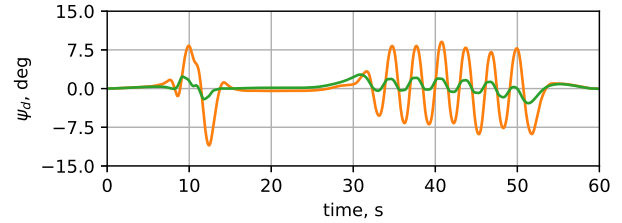
The main contribution of this paper is to propose a method that can enable matching lateral specific



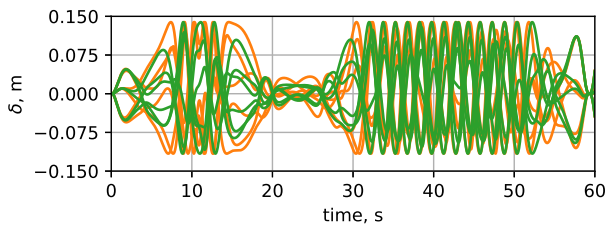
(a) MS 1: hexapod



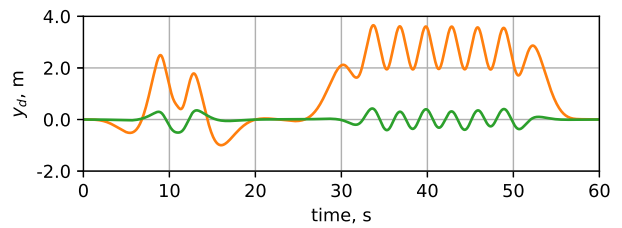
(b) MS 2: hexapod



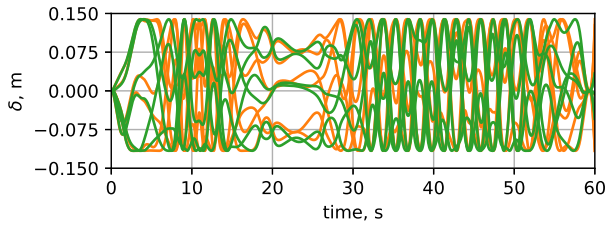
(c) MS 2: yaw-drive



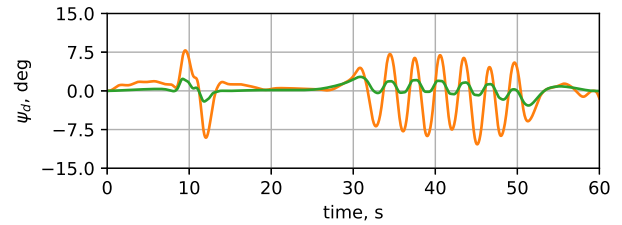
(d) MS 3: hexapod



(e) MS 3: y-drive



(f) MS 2, with 10° roll inclination in reference data: hexapod.



(g) MS 2, with 10° roll inclination in reference data: yaw-drive.

Figure 5: Workspace utilization of the four tested conditions.

force and yaw rate motion in MPC MCAs. The purpose of this is to provide motion that is more representative of real vehicle dynamics, where both motions are naturally linked. While providing more realistic motion cues to the driver, the workspace usage is reduced. The implemented methods are not only valid in curve driving; its general formulation describes the lateral motion at all times, also for straight driving. In that case, the lateral specific force will simply equal the term $c_2 \mathbf{T}(\beta^v) \mathbf{g}^I$, which is then only dependent on the inclination of the road. This provides a novel MPC approach that, without any added computational cost, remains closer to the vehicle dynamics than state-of-the-art MPC MCAs, while reducing the workspace usage. The proposed methods also suggest that always using the full workspace may not be beneficial if, as a result, the motion does not match anymore. The matching MPC thus speaks in favor of

simulator systems that can provide additional lateral *and* yaw motion, rather than only emphasizing the motion in one of these.

No direct limitations were observed in the implementation of the matching MPC; the amount of tunable parameters remains the same and the optimization time was unaffected. Further adaptations of the reference data (e.g., due to the reference point shift between the CRP and MRP, as well as vestibular system model integration) can be relatively easily obtained by expressing Eq. (11) accordingly. However, significant future work is required in several directions. Whereas the presented methods have been developed within an MPC framework, the extension to real-time MPC with matched motion remains to be made. Furthermore, while the matching MPC was found to be preferred in small pilot study, the presented methods require a more thorough subjective

evaluation to prove the preference of the matched MPC.

Current proxy methods of subjective evaluations through objective metrics (e.g., Grottolli, et al., 2019; Kolff, et al., 2024) also follow the same paradigm as the baseline MPC in this paper, where each channel is treated separately and fully independent of all others. Such metrics would thus, also in our presented results, favor the baseline MPC, as it will always provide objectively better tracking in the lateral specific force and yaw rate channels. The lack of considering the matching in current quality metrics may explain the disputed role regarding the beneficial effect of yaw motion on subjectively rated motion quality (Hogema, Wentink, and Bertolini, 2012; Kolff, et al., 2024). Yaw motion may be beneficial, but *only* if it matches in shape and magnitude with the perceived lateral motion. To be able to then show the true benefits of the matching MPC proposed in this paper, an improvement in the existing metrics that can account for the quality of the matching would be necessary.

5. Conclusion

This paper presented a novel method to explicitly match the lateral and yaw motion in driving simulators based on the well-known vehicle dynamics relation between these dynamic states. This relation was integrated in an MPC architecture and simulated offline (playback) in a combined slalom and double lane change scenario. The simulation results, for three representative driving simulator motion system architectures (MS 1-3), show the effective implementation of the matching reference tracking. Verifiably matching motion as a result is obtained, while also reducing the workspace usage. The matching MPC is especially relevant for simulators where one of the relevant motion channels (lateral or yaw) is strongly limited compared to the other (i.e., our MS 2 and MS 3). The matched MPC then purposefully reduces the motion in the other channel to provide a less strong, but matching, combination of lateral and yaw motion compared to the baseline MPC approach. Furthermore, by implementing the matching term on the level of the specific force, the MPC can effectively provide matching motion even on inclined roads. An informal pilot study revealed the preference of the matching MPC of four drivers, although further and more extensive subjective testing is necessary to fully prove its efficacy. Overall, the proposed approach provides a first key step towards vehicle dynamics-based motion cueing, rather than optimization of all motion channels separately. This may greatly improve the motion cueing especially for expert evaluations and lead to more effective expert testing of vehicle dynamics in motion-based driving simulators.

References

- Andersson, J. A. E., Gillis, J., Horn, G., Rawlings, J. B., and Diehl, M., 2019. CasADi – A software framework for nonlinear optimization and optimal control. *Mathematical Programming Computation*, 11(1), pp. 1–36. doi: 10.1007/s12532-018-0139-4.
- Bruschetta, M., Mendola, D. L., Beghi, A., and Minen, D., 2018. An MPC based Motion Cueing Algorithm with side slip dynamics. In: *Proceedings of the Driving Simulation Conference 2018 Europe VR*. Antibes, France, pp. 55–59.
- Cleij, D., Venrooij, J., Pretto, P., Katliar, M., Buelthoff, H. H., Steffen, D., Hoffmeyer, F., and Schöner, H.-P., 2019. Comparison between filter- and optimization-based motion cueing algorithms for driving simulation. *Transportation Research Part F: Traffic Psychology and Behaviour*, 61, pp. 53–68. doi: 10.1016/j.trf.2017.04.005.
- Dagdelen, M., Reymond, G., Kemeny, A., Bordier, M., and Maïzi, N., 2004. MPC Based Motion Cueing Algorithm: Development and Application to the ULTIMATE Driving Simulator. In: *Proceedings of the Driving Simulation Conference 2004 Europe*. Paris, France, pp. 221–233.
- Ellensohn, F., Venrooij, J., Schwenbacher, M., and Rixen, D., 2019. Experimental evaluation of an optimization-based motion cueing algorithm. *Transportation Research Part F: Traffic Psychology and Behaviour*, 62, pp. 115–125. doi: 10.1016/j.trf.2018.12.004.
- Grottolli, M., Cleij, D., Pretto, P., Lemmens, Y., Happee, R., and Bülthoff, H. H., 2019. Objective evaluation of prediction strategies for optimization-based motion cueing. *Simulation*, 95(8), pp. 707–724. doi: 10.1177/0037549718815972.
- Hogema, J., Wentink, M., and Bertolini, G., 2012. Effects of Yaw Motion on Driving Behaviour, Comfort and Realism. In: *Proceedings of the Driving Simulation Conference 2012 Europe*. Paris, France, pp. 149–158.
- Hvitfeldt, H., Drugge, L., and Jerrelind, J., 2025. Motion cueing scaling for planar vehicle dynamics and body slip feedback. *Vehicle System Dynamics*, pp. 1–23. doi: 10.1080/00423114.2025.2497849.
- Jacumet, R., Kolff, M., Venrooij, J., Schwenbacher, M., Wagner, S., Wollherr, D., Leibold, M., Pool, D. M., and Mulder, M., 2026. A Review on Optimization-Based Motion Cueing Algorithms for Driving Simulation. *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–25. doi: 10.1109/TITS.2026.3661819.
- Jacumet, R., Wetzel, J., Wagner, S., Schwenbacher, M., Wollherr, D., and Leibold, M., 2025. Optimization-based Motion Cueing Algorithm for Worst Case Maneuvers in Driving Simulation. In: *Proceedings of the Driving Simulation Conference 2025 Europe*. Stuttgart, Germany, pp. 163–168.
- Jansson, J., Sandin, J., Augusto, B., Fischer, M., Blissing, B., and Källgren, L., 2014. Design and performance of the VTI Sim IV. In: *Proceedings of the Driving Simulation Conference 2014 Europe*. Paris, France, pp. 128–138.
- Jazar, R. N., 2025. *Vehicle Dynamics: Theory and Application*. 4th ed. Springer New York, NY.
- Katliar, M., 2020. Optimal control of motion simulators. PhD Dissertation. Albert-Ludwigs-Universität Freiburg.
- Khusro, Y. R., Zheng, Y., Grottolli, M., and Shyrokau, B., 2020. MPC-based motion-cueing algorithm for a 6-DOF driving simulator with actuator constraints. *Vehicles*, 2(4), pp. 625–647.
- Kolff, M., Venrooij, J., Schwenbacher, M., Pool, D. M., and Mulder, M., 2023. The Importance of Kinematic Configurations for Motion Control of Driving Simulators. In: *IEEE 26th International Conference on Intelligent Transportation Systems*. Bilbao, Spain, pp. 1000–1006. doi: 10.1109/ITSC57777.2023.10422127.
- Kolff, M., Venrooij, J., Schwenbacher, M., Pool, D. M., and Mulder, M., 2024. Reliability and Models of Subjective Motion Incongruence Ratings in Urban Driving Simulations. *IEEE Transactions on Human-Machine Systems*, 54(6), pp. 634–645. doi: 10.1109/THMS.2024.3450831.
- Stratulat, A., Roussarie, V., Vercher, J., and Bourdin, C., 2011. Does tilt/translation ratio affect perception of deceleration in driving simulators? *Journal of Vestibular Research: Equilibrium & Orientation*, 21(3), pp. 127–139. doi: 10.3233/VES-2011-0399.
- Stroosma, O., 2025. Understanding Motion Cueing: Two Sides of Newton's Second Law. In: *AIAA SCITECH 2025 Forum*. Orlando, FL, pp. 1–11. doi: 10.2514/6.2025-0977.
- Wächter, A. and Biegler, L., Mar. 2006. On the Implementation of an Interior-Point Filter Line-Search Algorithm for Large-Scale Nonlinear Programming. *Mathematical programming*, 106, pp. 25–57. doi: 10.1007/s10107-004-0559-y.
- Wagener, N., Russ, F., Beckmann, J., Legran, P., and Eckstein, L., 2023. ika's highly dynamic driving simulator. In: *Product Solution Book of the Driving Simulation Conference 2023 Europe*. Antibes, France, pp. 103–104.

Learning-Based Acceleration Prediction for Real-Time Model Predictive Motion Cueing

Keno Pape¹, Kevin Malena², Julia Timmermann¹, Ansgar Trächtler^{1,3}

(1) Chair of Control Engineering and Mechatronics, Heinz Nixdorf Institute, Paderborn University, Germany, e-mail: {keno.pape, kevin.malena, julia.timmermann, ansgar.traechtler}@hni.uni-paderborn.de

(2) Private University of Applied Sciences for Business and Technology (PHWT), Rombergstrasse 40, 49377 Vechta, Germany;

(3) Fraunhofer Institute for Mechatronic Systems Design IEM, Zukunftsmeile 1, 33102 Paderborn, Germany.

Abstract – Interactive driving simulation relies on dynamic simulators whose motion systems are controlled by motion cueing algorithms. Model predictive motion cueing algorithms are particularly promising, but exploiting their full potential requires future motion references. These are not directly available because future vehicle motion depends on unknown driver intent. This paper introduces a hybrid predictor for longitudinal and lateral acceleration references that combines structured physics-based preview information with data-driven, driver-specific prediction. The framework comprises a rule- and physics-based SpeedPreviewer, physics-guided feature construction, and a long short-term memory (LSTM) predictor. It is evaluated in simulation regarding prediction accuracy, acceleration reproduction within the model predictive motion cueing algorithm, and real-time capability. The results show good prediction accuracy, consistent transfer to unseen road, and promising transfer to different driver types. Compared with a constant reference, the predictor improves acceleration reproduction, particularly in longitudinal dynamics, while meeting real-time requirements.

Keywords: Interactive Driving Simulation, Motion Cueing, Reference Prediction, Physics-Guided Learning, LSTM

Copyright: © 2026 by the authors.

Licensee Driving Simulation Association.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.82157/dsa/2026/10>

Introduction

Interactive driving simulation is an important tool for the development and evaluation of modern vehicle systems, particularly when realistic driver behavior and perception must be considered early in the design process. By integrating the human as a driver-in-the-loop, it enables the investigation of vehicle dynamics, human-machine interaction, and subjective driver experience under realistic but reproducible conditions. In applications where motion perception is of central relevance, this requires dynamic driving simulators that reproduce vehicle motion as perceptible accelerations and angular rates.

To generate these motion cues, the simulator motion system is controlled by motion cueing algorithms (MCA). Compared with conventional filter-based approaches, model predictive motion cueing algorithms (MP MCA) explicitly account for future system behavior and the physical limitations of the simulator in motion planning (Dagdelen, et al., 2009). Their performance, however, depends critically on suitable future references, in particular longitudinal and lateral accelerations, which are central to driver perception.

This is precisely where a central challenge arises in interactive driving simulation: future vehicle motion depends strongly on driver intent, which is not known at the current time. A reference predictor is therefore required that provides realistic and driver-specific future acceleration profiles while remaining consistent with vehicle dynamics, road preview information, and real-time constraints. In this setting, reference generation becomes a methodological problem rather than merely a downstream component of motion cueing (Mohammadi, et al., 2016).

Existing research in motion cueing has focused primarily on improving classical MCA and on the further development of motion cueing approaches (Scheidel, et al., 2024; Huang, et al., 2025; Kolff, et al., 2025). In contrast, reference generation for MP MCA is only rarely addressed as a distinct methodological challenge. Existing approaches range from constant or linearly extrapolated references (Bruschetta, et al., 2018) to model-based methods (Biemelt, et al., 2020; Lamprecht, et al., 2022) and data-driven approaches (Qazani, et al., 2021; Paparusso, Melzi and Braghin, 2023). Model-based approaches are interpretable and physically consistent, but they capture individual driving styles only to a

limited extent. Purely data-driven methods can represent driver-specific patterns; however, without appropriate structuring, they are often difficult to interpret, do not generalize reliably, and make it challenging to meet real-time requirements.

This highlights a central research gap: there is currently no sufficiently developed approach that combines physically motivated preview information, driver-specific behavior, and data-driven acceleration prediction to support robust, interpretable, and real-time-capable reference generation. This paper introduces and investigates an initial hybrid reference prediction approach intended to address this gap. The proposed formulation combines physics-guided preview and feature construction with a neural-network-based predictor for future longitudinal and lateral accelerations. The core idea is to derive informative and interpretable features from available signals and prior physical knowledge, such that the network does not have to learn the full problem structure directly from raw data, but can instead operate on a physically pre-structured signal representation. Its suitability and practical potential are assessed in the application context of MP MCA.

The main contributions of this work are:

- (i) an initial hybrid framework for driver-specific acceleration prediction
- (ii) a physics- and rule-based SpeedPreviewer
- (iii) a physics-guided feature construction for LSTM-based acceleration prediction, and
- (iv) a simulation-based evaluation with respect to prediction accuracy, acceleration reproduction in the MP MCA, and real-time capability

Application Context

This section introduces the ATMOS dynamic driving simulator, the role of the MP MCA, and the resulting requirements for acceleration prediction.

ATMOS and MP MCA

The ATMOS (Atlas Motion System) dynamic driving simulator at the Heinz Nixdorf Institute in Paderborn, Germany, serves as the application example in this work. In ATMOS, vehicle and environment simulation, driver inputs, the motion system, and visual and acoustic feedback are coupled in real time. Figure 1 illustrates the closed loop of driver, environment, vehicle, and motion system. Based on visual and acoustic cues (C_{env}) from the virtual environment and motion cues (a_{sim} , ω_{sim}), the driver determines the desired vehicle motion r_{des} . The predictor estimates the resulting future reference trajectory $\hat{r}_{veh}$, from which the MP MCA computes the motor commands u_{motor} for the motion system. An in-depth system description can be found in (Biemelt, et al., 2018).

Longitudinal and lateral accelerations as well as rotational rates provide essential motion cues and

influence driver behavior. These simulated vehicle motions must be reproduced by the motion system as realistically as possible. In ATMOS, this is achieved by an MP MCA using horizon-based optimization, enabling proactive planning of braking, acceleration, and cornering maneuvers while efficiently utilizing the limited simulator workspace. At this point, the central reference problem arises: future vehicle motion is not directly available in interactive simulation, as it results from driver decisions, vehicle dynamics, and road geometry. MP MCA performance therefore depends not only on the controller but also on the quality of the predicted future references.

Predictor Requirements

Providing these references requires a predictor that acts as an interface between the driver, the simulated vehicle, and our MP MCA, and approximates the future motion of the virtual vehicle. Such a reference predictor must account for both the unknown components of individual driver decisions and the known effects of vehicle dynamics and road geometry. The prediction must be physically plausible, comply with vehicle-dynamic and road constraints, generalize across drivers and roads, and operate in real time. Constant references cannot anticipate maneuvers or capture driver-specific behavior, while purely model-based approaches represent individual driving behavior only to a limited extent.

For the ATMOS system, these requirements translate into constraints. The primary target variables are the longitudinal and lateral accelerations a_x and a_y , which should be reproduced as accurately as possible. Rotational degrees of freedom, by contrast, are used mainly for tilt coordination, such that no angular-rate reference prediction is required in the present work. Accordingly, $\hat{r}_{veh}$ denotes the future a_x and a_y profiles. Based on the MP MCA implementation, the relevant horizon is typically 1 s; however, the predictor should also remain applicable to longer horizons of up to 3 s. In addition, real-time capability is mandatory: with a control cycle of 25 ms, a maximum of 10 ms is available per prediction. The following section therefore introduces a hybrid reference prediction approach designed to address these requirements by combining physically motivated preview information with data-driven, driver-specific prediction.

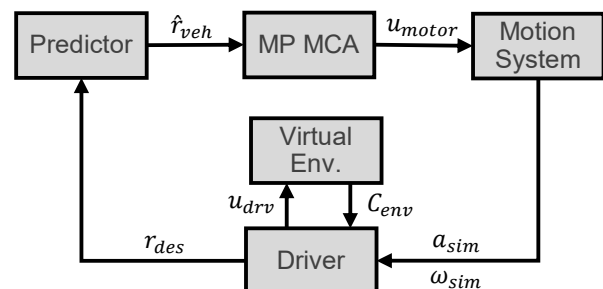


Figure 1: Signal flow of the ATMOS driving simulator components and integration of the predictor.

Prediction Framework

The proposed hybrid framework combines physics-based preview information with data-driven prediction of future longitudinal and lateral accelerations a_x and a_y , while considering physical prior knowledge, driver behavior, and real-time requirements.

Figure 2 illustrates the overall process. In contrast to a purely data-driven black-box predictor, the task is pre-structured by integrating prior knowledge at three levels. First, the SpeedPreviewer condenses road-related information over the prediction horizon into a speed preview comprising admissible speed limits v_{cap} and a reference speed v_{ref} . During feature construction, history features are derived from recorded past driver-vehicle signals, whereas preview features are obtained mainly from this speed preview together with selected additional preview quantities. The predictor therefore operates on a structured sequence of measured past signals and approximated future context derived from prior knowledge. Preview and prediction span the same future horizon, with the preview providing the context for the predicted accelerations. This history-preview sequence is mapped to a fixed time grid and processed by an LSTM network. The LSTM refines the SpeedPreviewer output using driver history and preview context to account for deviations not captured by a predefined empirical model. In addition, the speed preview guides training through behavior-oriented, rule- and physics-based auxiliary loss terms. Thus, physically defined aspects of driving are incorporated before learning, allowing the network to focus on driver-specific and context-dependent acceleration behavior.

Speed Preview Generation

The SpeedPreviewer forms the first stage of hybrid reference prediction. Here, road preview information, vehicle-dynamic limits, and driver characteristics are combined into the speed preview. The generation of this speed preview is carried out in three steps: (i) determination of physical and driver-specific speed limits along the preview, (ii) generation of a longitudinally consistent reference speed trajectory using a forward-backward pass, and (iii) conversion into a time-equidistant preview signal for the reference predictor. The forward-backward methodology is based on established procedures for deriving valid speed profiles and is extended here by an explicit driver-oriented formulation (Velenis and Tsotras, 2005).

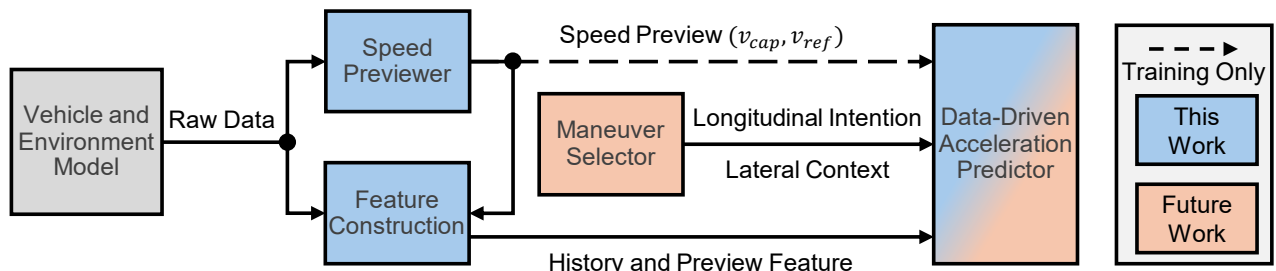


Figure 2: The overall hybrid architecture of the acceleration predictor.

Physical and driver-specific speed limit

The speed preview is defined over a velocity-dependent distance s_{prev} , which is discretized into D equidistant sample points

$$s_{prev} = \frac{v_{x,0}^2}{2|a_{x,min}|}, \quad s_i = \frac{i}{D} s_{prev}, \quad (1)$$

where $i = 1, \dots, D$, $v_{x,0}$ denotes the current speed, and $a_{x,min} < 0$ is the maximum deceleration. For each segment s_i , the road curvature $\kappa_y(s_i)$, the vertical profile curvature $\kappa_z(s_i)$, and a local speed limit $v_{lim}(s_i)$ are given. In addition, the dynamic limits $a_{y,max}$, $a_{x,max}$, $a_{x,min}$ and the vertical comfort limit $a_{z,max}$ are considered. First, a physical speed limit $v_{cap,phys}$ is defined as the minimum of the lateral, vertical, and road speed limits

$$v_{lat}(s_i) = \sqrt{\frac{a_{y,max}}{|\kappa_y(s_i)| + \epsilon_y}},$$

$$v_{vert}(s_i) = \sqrt{\frac{a_{z,max} \cos \beta(s_i)}{|\kappa_z(s_i)| + \epsilon_z}}, \quad (2)$$

$$v_{cap,phys}(s_i) = \min[v_{lat}(s_i), v_{vert}(s_i), v_{lim}(s_i)],$$

where $\epsilon > 0$ is a small regularization term for numerical stability and $\beta(s_i)$ denotes the local road slope.

Since drivers generally maintain a safety margin and anticipate upcoming restrictions, $v_{cap,phys}$ is transformed into a driver-specific speed limit. The anticipation distance is defined as

$$d_{shift}(s_i) = T_{ant} v_{cap,phys}(s_i), \quad (3)$$

and a conservative target limit is obtained using the driver-specific braking acceleration $a_{brk} > 0$:

$$v_{cap,tar}(s_i) = \sqrt{\max[0, v_{cap,phys}^2(s_i) - 2 a_{brk} d_{shift}(s_i)]}. \quad (4)$$

Anticipation is introduced by propagating the most restrictive target limit upstream over the corresponding look-ahead interval:

$$v_{cap,drv}(s_i) = \min_k v_{cap,tar}(s_k), \quad s_k \in [s_i, s_i + d_{shift}(s_i)] \quad (5)$$

Thus, T_{ant} determines how early upcoming restrictions are considered, whereas a_{brk} controls the safety margin below the physical limit.

Forward-Backward Pass

Based on $v_{cap,drv}$, a longitudinally consistent reference speed v_{ref} is generated that satisfies local speed limits and remains feasible with respect to the available acceleration and braking capability. Since lateral acceleration demand and road slope reduce the remaining longitudinal capability, their coupling is represented by a driver-specific longitudinal-lateral acceleration envelope (g-g diagram). From this envelope, a utilization factor $\gamma(s_i) \in [0,1]$ is derived. This factor scales the available longitudinal acceleration and braking capability at each preview position, while the road-slope effect is introduced through an additional gravity term $a_{grav}(s_i)$

$$\begin{aligned} a_{eff}^{acc}(s_i) &= a_{x,max}\gamma(s_i) + a_{grav}(s_i), \\ a_{eff}^{brk}(s_i) &= a_{x,min}\gamma(s_i) + a_{grav}(s_i). \end{aligned} \quad (6)$$

A forward pass determines the maximum possible speed profile given the acceleration limit, where Δs_i denotes the distance between two segments

$$\begin{aligned} \tilde{v}_{fwd}(s_{i+1}) &= \sqrt{\max\left[0, \left(v_{fwd}(s_i)\right)^2 + 2a_{eff}^{acc}(s_i)\Delta s_i\right]}, \\ v_{fwd}(s_{i+1}) &= \min[\tilde{v}_{fwd}(s_{i+1}), v_{cap,drv}(s_{i+1})]. \end{aligned} \quad (7)$$

The backward pass ensures that desired speeds remain feasible under the available braking capacity

$$\begin{aligned} \tilde{v}_{bwd}(s_{i-1}) &= \sqrt{\max\left[0, \left(v_{bwd}(s_i)\right)^2 - 2a_{eff}^{brk}(s_{i-1})\Delta s_{i-1}\right]}, \\ v_{bwd}(s_{i-1}) &= \min[\tilde{v}_{bwd}(s_{i-1}), v_{cap,drv}(s_{i-1})]. \end{aligned} \quad (8)$$

The raw speed profile is given by

$$v_{raw}(s_i) = \min[v_{fwd}(s_i), v_{bwd}(s_i)]. \quad (9)$$

To reduce terminal effects over short horizons, the preview is extended and the terminal velocity is selected online from the current driving situation. Since v_{raw} may still contain sudden local changes, a jerk limiter is applied to the segment accelerations. First, acceleration and related segment time are computed

$$a_x(s_i) = \frac{v_{raw}(s_{i+1}) - v_{raw}(s_i)}{2\Delta s_i}, \quad (10)$$

$$\Delta t(s_i) \approx \frac{2\Delta s_i}{v_{raw}(s_i) + v_{raw}(s_{i+1}) + \epsilon_v}. \quad (11)$$

Given the maximum acceptable longitudinal jerk $j_{x,max}$, the following applies

$$|a_x(s_i) - a_x(s_{i-1})| \leq j_{x,max} \Delta t(s_{i-1}). \quad (12)$$

This constraint is enforced using a second forward-backward pass, in which the segment accelerations are clipped to admissible values. This results in a smooth preview $a_{ref}(s_i)$, used to reintegrate a consistent velocity profile

$$\begin{aligned} \tilde{v}_{ref}(s_{i+1}) &= \sqrt{\max\left[0, v_{ref}^2(s_i) + 2a_{ref}(s_i)\Delta s_i\right]}, \\ v_{ref}(s_{i+1}) &= \min[\tilde{v}_{ref}(s_{i+1}), v_{cap,drv}(s_{i+1})]. \end{aligned} \quad (13)$$

The result is a smooth, limit-compliant v_{ref} and the corresponding segment acceleration a_{ref} . In addition, a state machine is used as a heuristic stabilization layer for the driving states of acceleration, braking, and cruise. In particular, it limits or suppresses positive longitudinal acceleration in curves and thereby reflects typical cornering behavior.

Time-Domain Representation

The SpeedPreviewer initially provides a spatial speed-preview representation, e.g., $v_{cap,drv}(s)$ and $v_{ref}(s)$, while road-preview signals such as $\kappa_y(s)$ are available in the same spatial domain. Since the reference predictor operates on a fixed time grid, these spatial preview quantities are transformed into the time domain. To this end, segment times are derived from the discrete spatial nodes s_i and the corresponding reference velocities and accumulated to obtain a sequence of monotone time nodes $t_s(s_i)$. For the prediction horizon, an equidistant time grid $t_h = h\Delta t_h$, $h = 1, \dots, H$, is used. The spatial preview signals are interpolated onto this grid, yielding uniformly sampled signals such as $v_{cap,drv}(t_h)$ and $\kappa_y(t_h)$. The resulting preview representation is thus available in a form that can be used directly for feature construction and acceleration prediction.

As shown in Figure 3, the physical speed limit $v_{cap,phys}$ defines the maximum admissible speed imposed by road geometry and vehicle-dynamic constraints. The driver-specific limit $v_{cap,drv}$ is lower than $v_{cap,phys}$, reflecting a safety margin. Moreover, upcoming restrictions are shifted to earlier time instants, thereby capturing anticipatory driving behavior, as visible before the speed reduction around $t \approx 130$ s. Consequently, the reference speed v_{ref} is adapted earlier and more strongly and provides a closer approximation of the simulated vehicle speed v_{veh} . In this way, the resulting speed preview serves as an intermediate representation for subsequent feature construction and learning-based prediction.

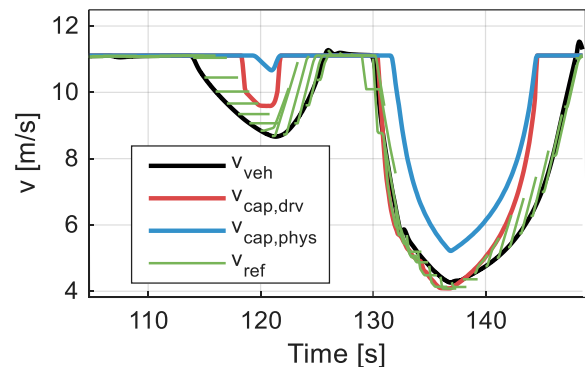


Figure 3: Time-domain speed preview with speed limits v_{cap} , reference speed v_{ref} , and vehicle speed v_{veh} .

Feature Construction

Feature construction aims at a compact and physically interpretable input representation rather than the direct use of raw signals. To this end, prior knowledge is used to process and combine available signals and the speed preview into quantitative features that capture simplified relations of vehicle dynamics and driver behavior, such as braking-related, limit-related, or preview-based lateral-demand quantities. In this way, the learning task is pre-structured, so that the network does not have to infer these basic relations from raw data alone, but can focus on driver-specific and context-dependent deviations.

The feature design distinguishes explicitly between longitudinal and lateral dynamics. Longitudinal acceleration is mainly influenced by driver intention, anticipation, and conflicts between current speed and upcoming road demands. Accordingly, longitudinal features are designed to describe urgency, conflicting demands, and required speed adaptation. Lateral acceleration, in contrast, is determined primarily by future speed and road geometry; lateral features therefore describe road curvature and the resulting lateral demands.

The input representation is organized into history features and preview features. History features are derived from known past driver-vehicle signals and describe past vehicle states together with observable driver responses, e.g., a_x , a_y , v_x , pedal position, and steering angle δ . Preview features describe the future preview situation over the prediction horizon and are derived mainly from the speed preview. This partitioning reflects both the distinction between known past signals and approximated future information and the input-output structure of the sequential LSTM predictor described in the next section.

Representative longitudinal preview features are the required average braking deceleration,

$$a_{brk,req}(t_h) = \frac{v_{x,0}^2 - v_{cap,phys}^2(t_h)}{2 \max(s(t_h), \epsilon_s)}, \quad (14)$$

and the utilization of the driver-specific reference speed relative to the physical limit,

$$v_{util}(t_h) = \frac{v_{ref}(t_h)}{v_{cap,phys}(t_h) + \epsilon_v}. \quad (15)$$

For lateral dynamics, a key preview quantity is the lateral acceleration implied by the driver-specific speed limit and road curvature

$$a_{y,cap}(t_h) = v_{cap,drv}^2(t_h) \kappa_y(t_h). \quad (16)$$

Overall, each time step is represented by a feature vector consisting of 8 history features and 9 preview features. The LSTM thus operates on a sequence of 17-dimensional feature vectors that systematically links past driver-vehicle response with future preview information and provides the basis for robust acceleration prediction.

LSTM Predictor

A sequential LSTM network is used to predict future acceleration profiles and forms the learning-based stage of the hybrid predictor. The SpeedPreviewer already provides a simplified physics- and rule-based representation of future driving dynamics. Rather than replacing this structure, the LSTM combines the resulting preview features with recent driver-vehicle history to account for driver- and context-dependent deviations not captured by the pre-defined SpeedPreviewer parameterization.

The input is arranged as a combined history-preview sequence of length $T_{seq} = T_{hist} + T_{prev}$, using the features introduced in the previous section. The history part describes the recent driver-vehicle response and provides the temporal and behavioral context required to interpret the preview. The preview part contains approximated future road- and speed-related information over the prediction horizon, but no future accelerations. Since the LSTM processes the complete sequence sequentially, it produces an output at each sequence position and thus returns an output sequence of equal length. Only the T_{prev} outputs associated with the future horizon are used as acceleration predictions; the T_{hist} outputs corresponding to the observed history are removed by a binary mask and excluded from the loss.

The network comprises two LSTM layers with 64 hidden units each, dropout, and a fully connected output head. This architecture balances model capacity, generalization, and real-time performance. It operates on a 0.1 s time grid, while the predicted trajectories are interpolated to 0.025 s for integration into the MP MCA. Instead of absolute accelerations, the network predicts acceleration increments, from which the trajectories are reconstructed by cumulative summation. This reduces the learning scale and supports smoother and more consistent multi-step predictions.

An auxiliary output head predicts the longitudinal velocity v_x . Since a_x and v_x are directly related, this secondary task promotes a coherent and physically plausible representation of longitudinal motion. Velocity is not treated as a primary target but enters the loss with a lower weight, as described in the following section.

Training Objective and Loss Design

The loss function constitutes the third physics-based component of the proposed hybrid approach by introducing behavior-oriented and physically consistent regularization into the learning process. In this context, a_x is given greater weight than a_y , since longitudinal dynamics are more difficult to learn. The following weighted total loss function is used

$$L = L_{pred} + \lambda_v L_v + \lambda_{phys} L_{phys} + \lambda_{ax,s} L_{ax,s} + \lambda_{evt} L_{evt}. \quad (17)$$

Here, L_{pred} denotes the masked multi-step prediction error over the horizon with respect to acceleration increments. It forms the core prediction loss and would be the only term in a non-extended formulation. The auxiliary term L_v uses the prediction of v_x as a longitudinal consistency term. L_{phys} penalizes violations of vehicle-dynamics limits, while $L_{ax,s}$ smooths a_x and improves temporal consistency. The event term L_{evt} is a rule-based, one-sided term derived from the generated speed preview. In contrast to the preview features used as contextual inputs to the network, this speed preview is used only during training to identify situations such as cruise, acceleration, and braking phases and to increase the weight of the corresponding prediction errors. It is therefore not imposed as a target trajectory, but acts only as a soft guide within the loss. The weights λ_i are determined using gradient-based loss balancing (Chen, et al., 2018) by scaling the contribution of each term relative to L_{pred} .

Experimental Setup

The proposed hybrid reference predictor is evaluated in terms of prediction accuracy, acceleration reproduction in the MP MCA, and real-time capability. All data generation, training, validation, and MP MCA integration are performed in simulation. Acceleration reproduction is evaluated using a validated ATMOS model in Simulink, while driving data are generated in the commercial simulation environment used with ATMOS.

Driver-specific trajectories are generated by a parameterizable driver model comprising longitudinal and lateral controllers. Three driver types (*slow*, *moderate*, and *fast*) are considered, differing primarily in their longitudinal and lateral acceleration limits and thus in their dynamic behavior (Table 1). Five roads with different geometries and speed limits of up to 70 km/h represent urban and rural driving without surrounding traffic. For each road, 600 s of data are recorded at a simulation step size of 0.025 s.

Four roads are used for training, while the fifth is excluded entirely and serves as an unknown road. Evaluation on known roads uses separate samples not included in training. Each input sequence spans 6 s, comprising 3 s of history and 3 s of preview. Training uses Adam with a learning rate of 10^{-4} , a batch size of 128, gradient clipping at 1, early stopping with a patience of 20 epochs, and a maximum of 100 epochs on a GPU.

For each driver type, the SpeedPreviewer is parameterized using the corresponding dynamic limits to generate the driver-specific preview inputs. The LSTM is trained exclusively on slow-driver data and remains unchanged in all test scenarios. Driver generalization therefore refers to transferring the trained network to unseen driver types while using their

respective SpeedPreviewer parameterizations. Five scenarios of increasing difficulty are considered:

- S1: known driver (slow), known road
- S2: known driver (slow), unknown road
- S3: unknown driver (moderate), known road
- S4: unknown driver (moderate), unknown road
- S5: unknown driver (fast), unknown road

Prediction accuracy and MP MCA acceleration reproduction are quantified by the root mean square error (RMSE). For the latter, simulator accelerations obtained using an ideal reference with full future knowledge, a constant reference, and the proposed LSTM prediction are compared with the simulated vehicle accelerations. Since RMSE may mask brief or directionally incorrect deviations and does not capture perceived motion fidelity, the trajectories are additionally inspected qualitatively; perceptual evaluation remains outside the scope of this study.

Table 1: Driver-specific acceleration limits [m/s^2]

Driver	$a_{x,min}$	$a_{x,max}$	$a_{y,max}$
Slow	-2	1.5	2
Moderate	-3	2	3
Fast	-4.5	3	4.5

Results

This section presents the results obtained for the proposed reference prediction approach with respect to predictive accuracy, acceleration reproduction in the MP MCA, and real-time capability.

Prediction Accuracy

To assess the suitability of the reference predictor itself, its ability to predict future accelerations is evaluated both quantitatively and qualitatively. Table 2 reports the RMSE for a_x and a_y across all scenarios and prediction horizons.

As expected for multi-step prediction, the error increases with horizon length. The lowest errors are observed in the training-related scenarios with a known driver (S1, S2), while road transfer remains comparatively robust. Transfer to unknown drivers has a clearly larger impact: errors remain moderate for S3 and S4, but increase noticeably for the highly dynamic driver in S5. The proposed approach thus shows promising accuracy, particularly for short-term (0.5 s) and application-relevant (1 s) horizons.

These quantitative findings are consistent with the qualitative behavior shown in Figure 4 for scenario S4. Both a_x and a_y exhibit good shape consistency, with braking and acceleration maneuvers detected early and trajectories remaining mostly smooth. For longer horizons, a_x shows occasional drift and slight oscillations, while a_y tends to flatten and become nearly constant. These effects intensify for the highly

dynamic driver, indicating that driver transfer, rather than road transfer, is the main source of uncertainty. Overall, the results suggest that the proposed hybrid predictor makes meaningful use of the physics-based preview and feature representation, supporting its suitability as a promising basis for further development.

Table 2: LSTM prediction RMSE for longitudinal and lateral acceleration at different horizons [m/s²]

Scenario	0.5 s RMSE	1.0 s RMSE	3.0 s RMSE
	$a_x \parallel a_y$	$a_x \parallel a_y$	$a_x \parallel a_y$
S1	0.17 // 0.12	0.23 // 0.16	0.43 // 0.24
S2	0.18 // 0.11	0.26 // 0.16	0.38 // 0.25
S3	0.28 // 0.20	0.41 // 0.30	0.52 // 0.51
S4	0.21 // 0.20	0.29 // 0.31	0.36 // 0.51
S5	0.31 // 0.29	0.43 // 0.48	0.46 // 0.98

Acceleration Reproduction

To assess its effect on acceleration reproduction, the proposed hybrid reference predictor is integrated into the MP MCA and evaluated based on the resulting simulator accelerations. As shown in Table 3, the ideal reference yields the lowest errors in all scenarios, whereas the constant reference produces the largest deviations. Compared with the constant reference, the hybrid predictor consistently reduces the reproduction error, with the greatest improvement for longitudinal acceleration a_x and a smaller but consistent benefit for lateral acceleration a_y . Although performance decreases with increasing scenario complexity, the predicted reference remains more accurate than the constant reference in all transfer cases.

Table 3: Acceleration reproduction RMSE for the evaluated reference approaches [m/s²]

Scenario		Ideal	Constant	LSTM
S1	a_x	0.037	0.118	0.051
	a_y	0.010	0.032	0.022
S2	a_x	0.041	0.085	0.045
	a_y	0.014	0.029	0.021
S3	a_x	0.037	0.125	0.063
	a_y	0.015	0.054	0.043
S4	a_x	0.036	0.077	0.041
	a_y	0.021	0.047	0.042
S5	a_x	0.035	0.095	0.063
	a_y	0.046	0.102	0.089

Since a_y is reproduced more robustly overall, the qualitative analysis focuses on a_x . Figure 5 shows scenario S4, with the ideal reference serving as an upper-bound benchmark under complete future knowledge. Compared with the constant reference, the LSTM-based prediction captures peaks,

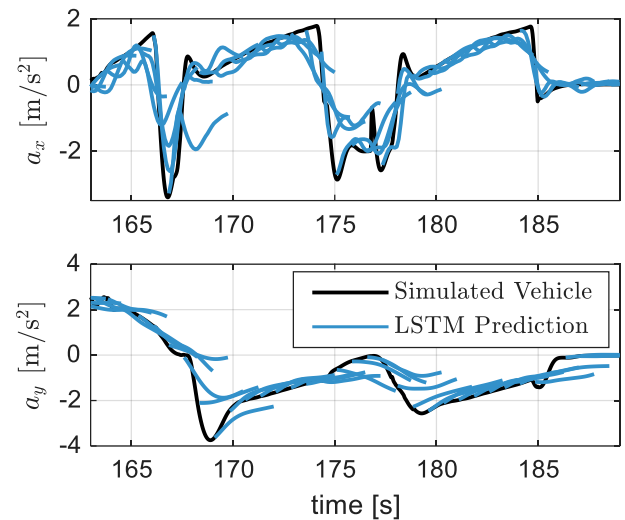


Figure 4: LSTM-based prediction for scenario S4 with a moderate driver and road transfer over a 3 s horizon.

transitions, and sudden changes more accurately, enabling earlier and closer tracking of the desired acceleration. Remaining deviations include slight overshoot at the onset of acceleration maneuvers and small fluctuations or offsets during nearly constant phases, which become more pronounced for the highly dynamic unknown driver.

Overall, the results demonstrate a consistent quantitative benefit over the constant-reference baseline, particularly for longitudinal acceleration. The remaining gap to the ideal reference indicates further potential for improving reference prediction. The present results provide an initial quantitative assessment, while broader comparisons with model-based predictors and perceptual motion-fidelity evaluations remain for future work.

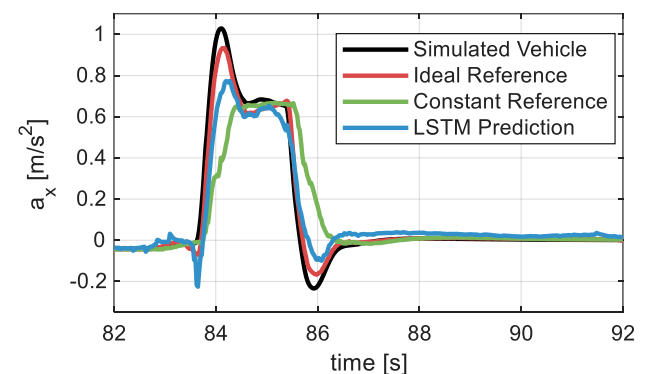


Figure 5: Longitudinal acceleration reproduction for an example maneuver in scenario S4.

Real-Time Capability

The real-time capability of the proposed approach is assessed from the execution time of the complete prediction pipeline. Using more than 11500 samples, Table 4 shows that computation is dominated by network prediction (65.11%) and feature construction (22.12%), while the SpeedPreviewer contributes

only a small part. The total runtime is 2.92 ms at the median and 3.85 ms at the 95th percentile, remaining well below the available real-time budget of 10 ms. These results indicate that the proposed approach is suitable for online use in the MP MCA and that sufficient margin for further extension of the framework is available.

Table 4: Computation time of the prediction pipeline (Setup: Intel Core i7-13700K, 32 GB RAM, MATLAB R2025b)

Step	Median [ms]	P95 [ms]	Share [%]
SpeedPreviewer	0.093	0.145	3.76
Data Processing	0.264	0.347	9.01
Features	0.724	0.852	22.12
Prediction	1.837	2.508	65.11
Total runtime	2.918	3.852	100

Conclusion

The results demonstrate that the proposed hybrid reference prediction approach is effective for predicting future longitudinal and lateral accelerations in interactive driving simulation. Its main contribution is a physics-guided methodology that combines a structured speed preview, physics-guided feature construction, and learning-based acceleration prediction within a modular and real-time-capable framework.

When integrated into the MP MCA, the approach consistently improves acceleration reproduction over the constant-reference baseline, with particularly clear benefits in longitudinal dynamics, while maintaining sufficient computational margin for online operation. Moreover, the LSTM trained exclusively on slow-driver data is transferred unchanged to more dynamic driver types, supported only by driver-specific SpeedPreviewer parameterization. The promising results for the moderate driver demonstrate meaningful transfer capability under this demanding setting, while the fast-driver case identifies targeted opportunities for further adaptation.

Overall, the study demonstrates the potential of hybrid reference prediction to improve MP MCA acceleration reproduction and provides a solid basis for further development. Future work should strengthen this foundation through comparisons with model-based predictors, feature and preview ablations, perceptual motion-fidelity evaluation, and validation with real driving data. Further potential lies in online driver classification for adaptive SpeedPreviewer parameterization and maneuver-specific prediction models.

Acknowledgments

This work was funded by the Ministerium für Kultur und Wissenschaft NRW under grant no. NW21-059D (SAIL).

References

- Biemelt, P., Henning, S., Rüdtenklau, N., Gausemeier, S. and Trächtler, A., 2018. A Model Predictive Motion Cueing Strategy for a 5-Degree-of-Freedom Driving Simulator with Hybrid Kinematics. *Proceedings of the Driving Simulation Conference 2018*, pp. 79–85.
- Biemelt, P., Link, C., Gausemeier, S. and Trächtler, A., 2020. A Model-Based Online Reference Prediction Strategy for Model Predictive Motion Cueing Algorithms. *IFAC-PapersOnLine*, 53(2), pp. 6082–6088.
- Bruschetta, M., Cenedese, C., Beghi, A. and Maran, F., 2018. A Motion Cueing Algorithm With Look-Ahead and Driver Characterization: Application to Vertical Car Dynamics. *IEEE Transactions on Human-Machine Systems*, 48(1), pp. 6–16.
- Chen, Z., Badrinarayanan, V., Lee, C.-Y. and Rabinovich, A., 2018. GradNorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In: *PMLR. International conference on machine learning*, pp. 794–803.
- Dagdelen, M., Reymond, G., Kemeny, A., Bordier, M. and Maïzi, N., 2009. Model-based predictive motion cueing strategy for vehicle driving simulators. *Control Engineering Practice*, 17(9), pp. 995–1003.
- Huang, X., Shi, X., Wang, P., Xu, H., Tang, X. and Zhang, G., 2025. MPC-Augmented Reinforcement Learning for Motion Cueing Algorithms in Driving Simulators. *Advanced Intelligent Computing Technology and Applications*. Springer Nature Singapore, pp. 499–510.
- Kolff, M., Jacumet, R., Wagner, S., Wollherr, D. and Leibold, M., 2025. A Frequency-Aware Model Predictive Control Motion Cueing Algorithm. *Proceedings of the Driving Simulation Conference 2025*, pp. 131–138.
- Lamprecht, A., Steffen, D., Nagel, K., Haecker, J. and Graichen, K., 2022. Online Model Predictive Motion Cueing With Real-Time Driver Prediction. *IEEE Transactions on Intelligent Transportation Systems*, 23(8), pp. 12414–12428.
- Mohammadi, A., Asadi, H., Nelson, K. and Nahavandi, S., 2016. Future reference prediction in model predictive control based driving simulators. *Australasian Conference on Robotics and Automation*.
- Paparusso, L., Melzi, S. and Braghin, F., 2023. Real-time forecasting of driver-vehicle dynamics on 3D roads: A deep-learning framework leveraging Bayesian optimisation. *Transportation Research Part C: Emerging Technologies*, 156.
- Qazani, M. R. C., Asadi, H., Lim, C. P., Mohamed, S. and Nahavandi, S., 2021. Prediction of Motion Simulator Signals Using Time-Series Neural Networks. *IEEE Transactions on Aerospace and Electronic Systems*, 57(5), pp. 3383–3392.
- Scheidel, H., Asadi, H., Bellmann, T., Seefried, A., Mohamed, S. and Nahavandi, S., 2024. A Deep Reinforcement Learning Based Motion Cueing Algorithm for Vehicle Driving Simulation. *IEEE Transactions on Vehicular Technology*, pp. 1–11.
- Velenis, E. and Tsiotras, P., 2005. Optimal velocity profile generation for given acceleration limits: theoretical analysis. *American Control Conference, 2005: IEEE*, pp. 1478–1483.

Comparison of Cost Functions for Automatic Motion Cueing Algorithm Parametrisation

Craig Brown¹, Marco Schlueter², Christian Holzapfel²

(1) ProCue Ltd, London, UK, e-mail: craig.brown@procue.xyz

(2) FKFS, Pfaffenwaldring 12, 70569 Stuttgart, Germany, e-mail: marco.schlueter, christian.holzapfel@fkfs.de

Abstract - Optimal configuration of Motion Cueing Algorithm (MCA) parameters is critical for motion fidelity and the success of simulator-based development. However, it can be a difficult and time-consuming task, even for an experienced simulator operator. Therefore, automatic parametrisation is an attractive alternative to manual tuning.

Automatic parametrisation requires “motion error” (as subjectively perceived) to be formally quantified. There are many relevant contributions in the literature to this question, though many of these proposed metrics are applied post-hoc rather than as part of a cost function. That is to say, they are not properly tested for the purposes of automatic parametrisation. Or, they are heavily customised to a specific simulator and use-case. Or, constrained in their design by the choice of optimiser—quadratic, convex, continuous etc...

In this paper, it is attempted to make a general statement about what metrics work best as cost functions. To achieve this, a flexible adaptive MCA that can produce different cueing “styles” is applied to a state-of-the-art simulator with redundant degrees of freedom. Combined with a powerful black-box optimiser, the overall setup isolates the cost function as primary contributor to cueing quality.

Parametrisations are rated subjectively by a number of experienced drivers. It is found that frequency-domain approaches are more promising than time-domain approaches for automatic MCA parametrisation.

Keywords: optimisation, optimization, parameterisation, parametrisation, tuning

Copyright: © 2026 by the authors.

Licensee Driving Simulation Association.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.82157/dsa/2026/11>

Introduction

Motion cueing can be achieved in many ways, including but not limited to Classical Washout Filter (CWF) (Conrad and Schmidt, 1970), Adaptive Filters (AF) (Parrish, et al., 1975), Model Predictive Control (MPC) (Dagdelen, et al., 2009) and Linear Quadratic Regulators (LQR) (Reid and Nahon, 1985). These Motion Cueing Algorithms (MCAs) have free parameters that must be set correctly to achieve sufficient motion cueing fidelity. Especially in approaches where the setting of said free parameters is unintuitive and time-consuming, automatic parametrisation is an attractive alternative to manual tuning (Arango, et al., 2022; Casas, et al., 2016; Jones, 2018; Pham, 2021). The number of free parameters varies, but could be as high as 30 per axis.

Automatic motion cueing parametrisation may be formulated as an optimisation problem. In the most general case, the optimisation problem can be expressed as in Equation 1

$$\text{minimise } f(\mathbf{x}), \mathbf{x} \in \mathbb{R}^D \quad (1)$$

where $f(\mathbf{x})$ is a measure of ‘motion cueing error’ and

$\mathbf{x}$ is a vector of D decision variables. Typically, there are at least decision variable boundary constraints as in Equation 2.

$$L_j < x_j < U_j, j \in [1, D] \quad (2)$$

where $\mathbf{L} \in \mathbb{R}^D$ and $\mathbf{U} \in \mathbb{R}^D$ are set by an experienced operator. For example, scaling or gain, normally denoted K , may fall into the range $[0, 1]$.

The optimisation problem is considered a black box. That is to say, no specific properties to the optimisation problem are presupposed, such as convexity, differentiability or monomodality. Fortunately, black box optimisers have made significant advancements in recent years, even showing convergence guarantees under certain conditions, or otherwise showing very high probability of finding a solution in the neighbourhood of the true solution for realistic problems (Melo and Carosio, 2012; Opara and Arabas, 2019).

In past work on automatic parameter tuning, cost functions are engineered to achieve specific outcomes with a specific simulator. Unfortunately this does not fully alleviate the burden of tuning, instead moving it from MCA parameters to weighting factors



Figure 1: Total view of the Stuttgart Driving Simulator

in a cost function. In this work, in order to provide a more general overview to practitioners, such “meta tuning” is avoided. Then, a clearer picture of what could work well in the general case is acquired.

Methodology

Inaugurated in 2012 the Stuttgart Driving Simulator, shown in Figure 1, focuses on reproducing the vehicle behaviour in normal driving situations. A full-size mock-up based on a real vehicle is used to accommodate the passengers. The mock-up is mounted in a dome that moves with the vehicle and contains the optical environment simulation with projectors for the virtual reality for a 360° view as well as the acoustic simulation for the sound of own and third-party vehicles and sound from the environment (Baumann, 2012).

The dome is mounted on a hexapod with 6 Degrees of Freedom (DoFs), which in turn stands on an XY-table (2 DoFs), totalling 8 DoFs. The physical properties are shown in Table 1

Using an intercom system, experts provide their subjective assessments in real time during the driving task. The respective driving scenarios can be repeated as often as required, allowing experts to refine and validate their judgments. Drivers will also be asked to summarize their experience post-hoc.

Table 1: Motion system specifications

System	DoF	Stroke [m/°]	Speed	Accel
XY-table	X	±5 m	2 ms ⁻¹	±5 ms ⁻²
XY-table	Y	±3.5 m	3 ms ⁻¹	±5 ms ⁻²
Hexapod	Surge	[−0.45, 0.54]	0.5 ms ⁻¹	±5 ms ⁻²
Hexapod	Sway	[−0.45, 0.45]	0.5 ms ⁻¹	±5 ms ⁻²
Hexapod	Heave	[−0.39, 0.37]	0.5 ms ⁻¹	±6 ms ⁻²
Hexapod	Roll	±18°	30° s ⁻¹	±90° s ⁻²
Hexapod	Pitch	±18°	30° s ⁻¹	±90° s ⁻²
Hexapod	Yaw	±21°	30° s ⁻¹	±120° s ⁻²

The experimental setup is based on a full-scale truck mock-up used as the driver interaction platform. The truck cabin remains fully functional and is equipped with the production mirror camera system, enabling a continuous 360° visualization of the surrounding environment.

Vehicle dynamics are modelled using a multi-body

simulation of a truck implemented in Simpack, which represents the chassis behaviour during dynamic maneuvers. The powertrain is modelled in Simulink using a truck-specific model that includes an internal combustion engine as the propulsion source. The evaluated driving scenario represents an urban environment with 5 curves, characterized by turning angles between 60° and 90°. The vehicle performs these maneuvers at a nominal speed of 50 kmh⁻¹, while one selected turn is executed at a reduced speed of 30 kmh⁻¹ in order to represent a tighter urban maneuver.

Here, the role of the MCA is to facilitate a breadth of different cueing approaches. With this in mind, an adaptive MCA derived from Miunske (2020) is used. With its high parameter count it can produce a range of cueing ‘styles’. With nonseparable parameters, it is difficult to tune manually.

In order to draw clear conclusions about the cost functions under study, it is decided to focus on just the lateral axis. The cueing of long corners is challenging even to simulators with large workspaces, requiring careful management of the same.

The treatment of perceived acceleration due to tilt in any cost function relative to true translational acceleration is a complex problem. Therefore, it is excluded from optimisation and set by an experienced operator.

The MCA is based on a high-pass filter structure with an additional damping element, which automatically adapts during runtime. Depending on the occurring lateral accelerations and jerk levels, the algorithm switches between four different filter parameter sets, each defined by three characteristic parameters. In addition to optimizing the filter parameters themselves, the optimizer also determines the state conditions under which transitions between the parameter sets are triggered. Furthermore, the transition behaviour between the parameter sets is tuned to ensure smooth parameter switching and stable system performance. The number of individual blocks of parameters and the total number of parameters to be optimized can be found in Table 2.

Table 2: MCA Parameters for automatic optimization

Parameter category	N ^o
Filter parameters: frequency, gain and damping	12
Switching conditions	11
Transition characteristics	6
Sum of parameters for optimization	29

The optimiser is based on Brest, et al. (2009), Gong, Cai, and Wang (2014) and Brown, et al. (2016) as implemented in ProCue 1.1.20 (Brown, 2024). It is a global, derivative free, population based direct search method, that is extremely performant on various benchmarks, such as those described in Suganthan, et al. (2005) and is well suited to engineering applications (Melo and Carosio, 2012).

As the offline cueing optimisation problem does not have a strict time constraint (as in online-optimisation) a population size of 10D is chosen, where D is the dimensionality of the optimisation problem. This defensively large population size helps prevent unwanted convergence to a non-optimal local minima at the expense of convergence speed (Piotrowski, 2017).

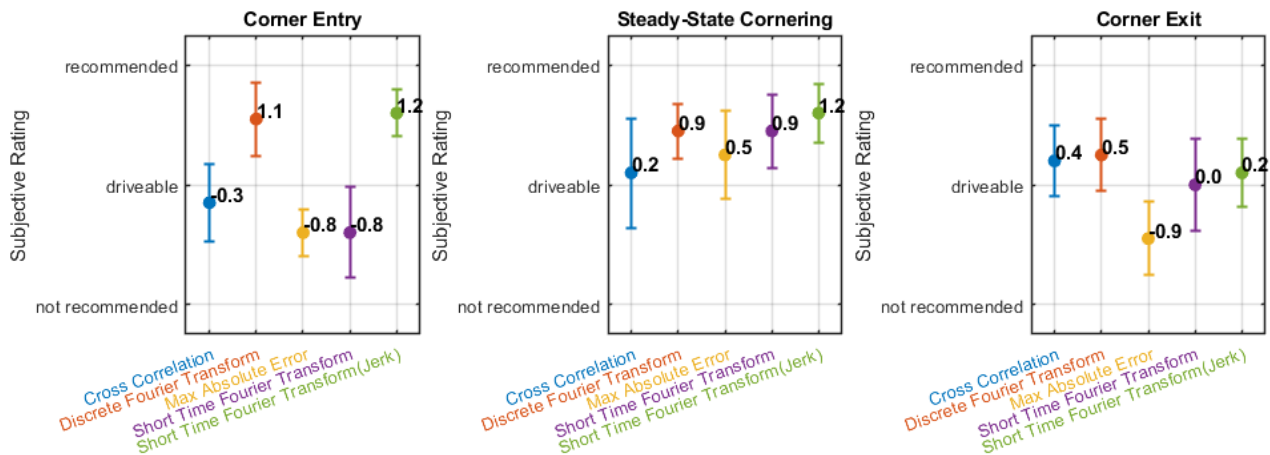


Figure 2: Subjective Rating of test persons for the 3 parts of cornering

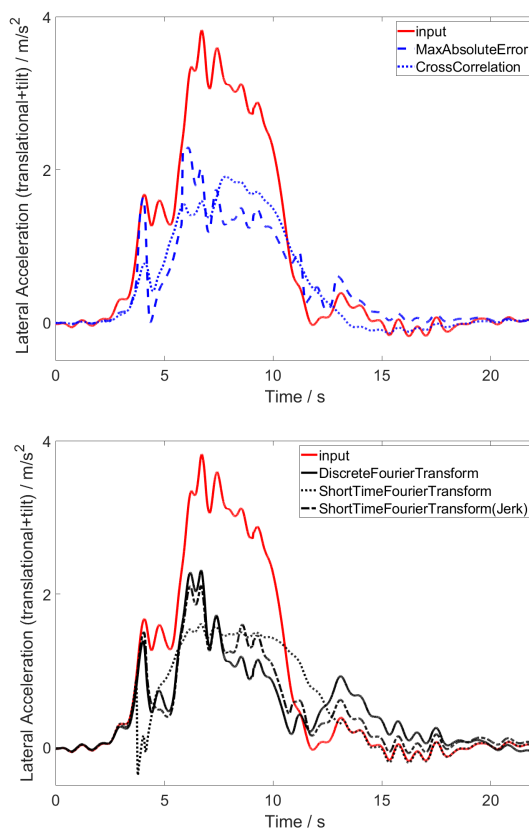


Figure 3: Optimization results for the selected cost functions: top row—MxAE and Pearson correlation coefficient; bottom row—DFT, STFT, and STFT(jerk).

The software implementation of the optimiser exploits Central Processing Unit (CPU) parallelism. Optimisations are carried out on a 20 core Intel(R) Xeon(R) w5-2445 (3.10 GHz) for a period of 12 hours. It was found convergence occurs well within this time budget, despite an ‘oversized’ population.

As a representative segment of the urban scenario, a 90° corner with the smallest radius and a maximum allowable speed of 30 kmh⁻¹ is considered. This section constitutes the worst-case scenario, as it produces the highest lateral accelerations along the route. *Accordingly, the optimization is explicitly performed for this segment, while the resulting MCA pa-*

rameters are applied to the entire urban course. It should be noted that only the translational component of lateral acceleration is optimized. The rotational component displayed by tilt coordination is determined by the road curvature and the current vehicle speed.

Arbitrary workspace constraints in engineering optimisation problems may be achieved using penalty functions (Melo and Carosio, 2012). Here, workspace constraints are implied by a penalty function developed by an expert operator. The penalty function is also treated as a black box.

Cost Functions

The cost functions described here are a curated subset from a broader literature review. Cost functions were rejected for either producing motion too similar to other (simpler) approaches, or, for not producing a feasible parametrisation, as judged by an experienced operator.

Unless otherwise stated, these cost functions are based on the simple absolute difference between the real outputs of each. For example, the sum of the absolute difference between the real outputs of FFT was used. Needless to say, there is a scope to treat these outputs in a more sophisticated way. For example, reproduction of high frequency content should probably be prioritised over low frequency content to reflect their relative contributions to workspace usage. Phase is also something that could be considered explicitly.

Each cost function returns a scalar value ≥ 0 , where 0 is perfect cueing.

Results & Discussion

Figure 3 shows the resulting lateral acceleration profiles of the optimized translational movement Hexapod and XY-Table and the rotational movement of the Hexapod (in ms⁻²) together with the corresponding reference input.

Ten participants took part in the study. Half of them have more than ten years of experience working with the simulator and are familiar with a wide range of motion cueing strategies across various scenarios.

The other half have been involved in multiple projects on vehicle dynamics using the Stuttgart Driving Simulator. Consequently, all participants possess expertise well beyond that of non-experts.

To quantify subjective perception during cornering, each participant completed at least one, and if necessary multiple, runs per optimized parameter set on the urban course. For each parameter set, ratings were provided for three phases of cornering: corner entry, steady-state cornering and corner exit.

A five-point bipolar rating scale ranging in ± 2 is used. This allows the five cost functions to be compared quantitatively across the three cornering phases, as in Figure 2.

An issue with time domain based cost functions is the loss of high frequency information, as also experienced in the context of MPC by Kolff, et al. (2025). Cost functions that do not consider jerk explicitly risk extreme maneuvers in the resulting motion.

On the other hand, frequency-domain cost functions consider higher derivatives of acceleration implicitly, due to the relationship between rate of change and the same. In the case of STFT (jerk), the additional differentiation step places more emphasis on high frequencies compared to STFT, which can clearly be seen visually. However, STFT could likely be improved by better configuration, especially by increasing the window size (from $256 \approx 0.25$ s used here).

Frequency domain approaches do not explicitly penalise miscues, especially sign or washout miscues. Future work could try to understand how miscue penalties can be properly implemented with frequency-domain approaches. Then, there is the further problem of properly incorporating the effect of tilt cueing into the cost function.

Conclusion

In this work, a set of cost functions are applied to offline parameter optimisation of a sophisticated, adaptive MPC on a highly performant, proven motion simulator with redundant axes.

Of the shortlisted functions, Discrete Fourier Transform and Short Time Fourier Transform appear to be the most promising. The output of these are *competitive with expert tuned parametrisation*.

Regarding impact on the broader literature. One problem with comparing MCAs is ensuring that the preference made is due to the MCA itself, and not merely superior parametrisation. Future comparisons of MCAs could feature an optimisation step to ensure each MCA under test is actually at its best. By also including an expertly tuned manual parametrisation, the 'difficulty' of parametrisation, which is important for (industry) adoption of a technique, could be approximated as, for example, the Euclidean distance between the two parameter sets.

With increasing automation, "hard to tune" but highly effective MCAs should become more commonplace. Especially with the availability of specialist tools, such as ProCue (Brown, 2024).

References

- Arango, C. G., Asadi, H., Qazani, M. R. C., Mohamed, S., and Nahavandi, S., 2022. Automatic Tuning of Adaptive Gradient Descent Based Motion Cueing Algorithm Using Particle Swarm Optimisation. In: *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 85–92.
- Baumann, G., 2012. *The New Driving Simulator of Stuttgart University*. Stuttgart: Springer Vieweg Wiesbaden.
- Brest, J., Zamuda, A., Boskovic, B., Maucec, M. S., and Zumer, V., 2009. Dynamic optimization using Self-Adaptive Differential Evolution. In: *2009 IEEE Congress on Evolutionary Computation*, pp. 415–422.
- Brown, C., 2024. ProCue: State-of-the-art Motion Cueing For All. In: A. Kemeny, J.-R. Chardonnet, F. Colombet, S. Espié, and F. Merienne eds. *Proceedings of the Driving Simulation Conference 2024 Europe VR*. Driving Simulation Association. Strasbourg, France.
- Brown, C., Jin, Y., Leach, M., and Hodgson, M., 2016. μ JADE: adaptive differential evolution with a small population. *Soft Computing*, 20, pp. 4111–4120.
- Casas, S., Coma, I., Portales, C., and Fernandez, M., 2016. Towards a simulation-based tuning of motion cueing algorithms. *Simulation Modelling Practice and Theory*, 67, pp. 137–154. ISSN: 1569-190X.
- Conrad, B. and Schmidt, S., 1970. *Motion drive signals for piloted flight simulators*. Tech. rep. National Aeronautics and Space Administration.
- Dagdelen, M., Raymond, G., Kemeny, A., Bordier, M., and Maïzi, N., 2009. Model-based predictive motion cueing strategy for vehicle driving simulators. *Control Engineering Practice*, 17(9), pp. 995–1003.
- Gong, W., Cai, Z., and Wang, Y., 2014. Repairing the crossover rate in adaptive differential evolution. *Applied Soft Computing*, 15, pp. 149–168.
- Jones, M., 2018. Enhancing motion cueing using an optimisation technique. *The Aeronautical Journal*, 122(1249), pp. 487–518.
- Kolff, M., Jacumet, R., Wagner, S., Wollherr, D., and Leibold, M., Sept. 2025. A Frequency-Aware Model Predictive Control Motion Cueing Algorithm. In: A. Kemeny, J.-R. Chardonnet, F. Colombet, and S. Espié eds. *Proceedings of DSC 2025 VR*. Antibes.
- Melo, V. V. de and Carosio, G. L. C., 2012. Evaluating differential evolution with penalty function to solve constrained engineering problems. *Expert Systems with Applications*, 39(9), pp. 7860–7863.
- Miunske, T., 2020. *Ein szenarienadaptiver Bewegungsalgorithmus für die Längsbewegung eines vollbeweglichen Fahrsimulators*. Stuttgart: Springer Vieweg Wiesbaden.
- Opara, K. R. and Arabas, J., 2019. Differential Evolution: A survey of theoretical analyses. *Swarm and Evolutionary Computation*, 44, pp. 546–558.
- Parrish, R. V., Dieudonne, J. E., Bowles, R. L., and Martin, D. J., 1975. Coordinated Adaptive Washout for Motion Simulators. *Journal of Aircraft*, 12(1), pp. 44–50.
- Pham, D.-A., 2021. Mean-Variance Mapping Optimization for Auto-tuning Parameters of Classical Motion Cueing Algorithm. In: B. T. Long, Y.-H. Kim, K. Ishizaki, N. D. Toan, I. A. Parinov, and N. P. Vu eds. *Proceedings of the 2nd Annual International Conference on Material, Machines and Methods for Sustainable Development (MMMS2020)*. Cham: Springer International Publishing, pp. 952–957. ISBN: 978-3-030-69610-8.
- Piotrowski, A. P., 2017. Review of Differential Evolution population size. *Swarm and Evolutionary Computation*, 32, pp. 1–24.
- Reid, L. and Nahon, M., 1985. *Flight Simulation Motion-Base Drive Algorithms: Part 1. Developing and Testing Equations*. Tech. rep. University of Toronto Institute for Aerospace Studies.
- Suganthan, P., Hansen, N., Liang, J., Deb, K., Chen, Y.-p., Auger, A., and Tiwari, S., Jan. 2005. Problem Definitions and Evaluation Criteria for the CEC 2005 Special Session on Real-Parameter Optimization. *Natural Computing*, 341–357.

Effects of motion cueing and adaptive procedure on acceleration JND estimation

Erik Gustaf Lilljebjörn^{1,2}, Sogol Kharrazi^{1,2}, Jan Åslund²

(1) Swedish National Road and Transport Research Institute (VTI), e-mail: {erik.gustaf.lilljebjorn, sogol.kharrazi}@vti.se

(2) Linköping University, Sweden, e-mail: jan.aslund@liu.se

Abstract - To assess driveability in a driving simulator, it is important that test participants can notice subtle differences in acceleration in the simulator. In a moving-base driving simulator, the just-noticeable differences (JNDs) in longitudinal acceleration were estimated for nine professional truck test drivers using three variants of the classical motion-cueing algorithm (MCA). No statistically significant differences in JNDs were observed across MCA variants, possibly due to the low statistical power of the study. The mean JND across all participants and MCA variants was 5.4%.

Motivated by the low statistical power of the driving simulator study, standard deviations and biases of JND estimators were subsequently estimated using Monte Carlo simulations of three adaptive procedures. The stationary distributions of the procedures provide an explanation for why JND estimates obtained with some procedures are asymptotically biased. The results further show that good statistical power can be achieved with the proposed procedures.

Keywords: Driveability, motion-cueing algorithm, acceleration perception, just-noticeable difference (JND), Monte Carlo simulation

Copyright: © 2026 by the authors.

Licensee Driving Simulation Association.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.82157/dsa/2026/12>

Introduction

Early in the product development of a new driveline, when physical prototypes are not available, the driveability cannot be evaluated using conventional methods. Testing the driveability using a virtual prototype of the driveline in a moving-base driving simulator could therefore be an attractive alternative. Driveline-related driveability has been tested in moving-base driving simulators in several previous studies (e.g. Baumgartner, et al., 2019b; Erler, et al., 2018; Kraft, et al., 2021).

Because the workspace of a moving-base driving simulator is limited, a motion-cueing algorithm (MCA) maps the motions of the vehicle model to feasible simulator motions. The classical MCA recreates transient accelerations using translational acceleration and recreates sustained accelerations by tilting the cabin – commonly known as *tilt coordination* (Reid and Nahon, 1985). The classical MCA typically scales down accelerations, and as a result, they may be less noticeable in the simulator. To assess driveability in a driving simulator, it is important that test participants can notice subtle differences in acceleration. One goal of the present study is therefore to investigate if small differences in longitudinal acceleration are more noticeable with a classical MCA tuned specifically for drive-off tests than with a general one.

The just-noticeable differences (JNDs) in acceleration across motion-cueing variants can be estimated using psychophysical methods. Psychophysics is a branch of experimental psychology that studies the relationship between stimuli and sensations (Gescheider, 2013).

Adaptive test procedures, where the stimulus intensity is adjusted based on participant responses in previous trials, are commonly used to estimate JNDs in driving simulator studies (e.g. Baumgartner, et al., 2019a; 2019b; Jamson, 2010; Menig, 2025; Nesti, et al., 2012; 2016). Several adaptive procedures have been proposed (Leek, 2001), and the choice of procedure may have a considerable effect on the variability and biases of JND estimators (García-Pérez, 1998), but this is rarely discussed in previous studies. Therefore, this paper also investigates how the rules and settings of adaptive test procedures influence the mean and variability of JND estimates in psychophysical experiments.

Estimating JNDs

To estimate a test participant's JND in acceleration, pairs of accelerations are presented to the participant successively. One stimulus is always the same and is called the *standard stimulus*, whereas the *compar-*

ison stimulus is varied from trial to trial. Rather than asking participants if they perceived a difference or not, the two alternative forced-choice (2AFC) method is commonly used in JND experiments (Gescheider, 2013). In a 2AFC task, participants are forced to choose which of two stimuli that they perceived as being more intense.

A staircase procedure is an adaptive method in which the stimulus intensity (or stimulus difference) is decreased after one or more correct responses and increased after an incorrect response. A reversal in a staircase procedure is a trial in which the direction of the staircase changes (i.e., the stimulus intensity is increased after a number of trials with decreasing stimulus intensity or vice versa). A sequence of trials between two reversals is referred to as a *run*. When a staircase procedure is used to estimate absolute or differential sensitivity, a threshold is usually estimated as the mean of the reversals, as shown in the top subplot of Figure 1. This method is equivalent to taking the average of the mid-points of every second run (Levitt, 1971), and these estimates are therefore referred to as *mid-run estimates*.

In this work, a hybrid method (Hall, 1981) for estimating the JND from staircase data is compared to the standard mid-run estimates. In a 2AFC task, the *psychometric function* describes the relationship between stimulus difference and the probability of responding that the comparison is higher. In the hybrid method, the participant's unknown psychometric function is approximated by fitting a sigmoid (e.g., logistic) function to the binary responses. The fitted model is then used to calculate the JND.

Figure 1 illustrates how the JND is estimated using the hybrid method. Responses of a hypothetical participant in the top subplot were generated using an assumed true logistic psychometric function shown in the bottom subplot. The fitted psychometric function, also shown in the bottom subplot, was obtained by fitting a logistic model to the responses using maximum-likelihood estimation. The circles represent the proportion of trials in which the participant judged the comparison as higher.

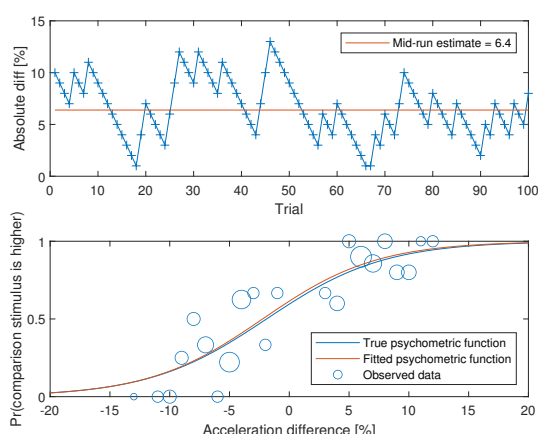


Figure 1: Example of an adaptive staircase procedure. Circle size reflects the total number of trials at each acceleration difference.

The point of subjective equality (PSE) is defined as the stimulus difference at which the probability of responding "comparison stimulus is higher" equals



Figure 2: Driving simulator used in the study.

0.5. A high JND estimate can be calculated as the difference between $X_{0.75}$, i.e., the difference where $p = 0.75$, and the PSE. Similarly, a low JND estimate is the difference between the PSE and $X_{0.25}$ (Gescheider, 2013). A single JND estimate is obtained by calculating the mean of the low and high JND estimates.

Driving simulator study

To investigate the effects of the MCA variants on the JNDs in acceleration, a driving simulator study was conducted with nine professional drivers that routinely assess driveability of heavy truck drivelines.

A moving-base driving simulator with a seven-meter-long translational workspace, shown in Figure 2, was used in the study. The simulator can recreate accelerations of up to 8.0 m/s^2 using translational motion and can simulate sustained longitudinal accelerations of up to approximately 1.5 m/s^2 by tilting the cabin.

Two variants of the classical MCA that are tuned specifically for short drive-off tests and one more general variant were considered. The Unscaled Tip-In (UTI) MCA variant does not scale down the acceleration, whereas the Downscaled Tip-In (DTI) and Downscaled Generic (DG) variants use a scale factor of 0.7. The UTI and DTI variants can recreate a 1.4 m/s^2 step input for three and five seconds, respectively, before the simulator reaches the end of its linear rail. Therefore, the manoeuvre must be stopped automatically after three seconds, and the simulator is returned to the starting position. The Downscaled Generic (DG) variant is the most general and can reproduce a 1.4 m/s^2 step input indefinitely by using cabin tilt more than the other variants.

The simulator is pre-positioned before the manoeuvre so that the variants that are tuned specifically for the drive-off manoeuvre can take full advantage of the long linear rail. Some of the most important parameters of the MCAs are shown in Table 1 and the outputs of the MCAs are shown in Figure 3.

The simulator study used a within-subjects design, i.e., all participants tested all MCA variants, which was the only factor of the study.

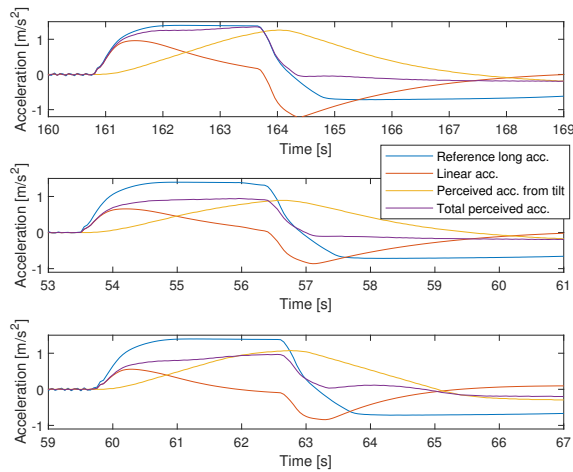


Figure 3: Outputs of UTI (top subplot), DTI (middle subplot) and DG (bottom subplot).

Motion cueing	Scale factor	Cut-off frequency [rad/s]	Max tilt speed [deg/s]
UTI	1	0.3	3
DTI	0.7	0.32	3
DG	0.7	0.55	3

Table 1: Summary of the motion cueing variants.

The participants were asked if a modified acceleration was lower or higher than a 1.4 m/s^2 reference acceleration in a 2AFC task. Stimulus intensities around participants' JNDs in acceleration were obtained by using an adaptive up/down staircase with an up-step twice the size of the down-step, as shown in Figure 4. A test with an MCA variant ended after 20 trials, and the signs of the acceleration differences were randomized.

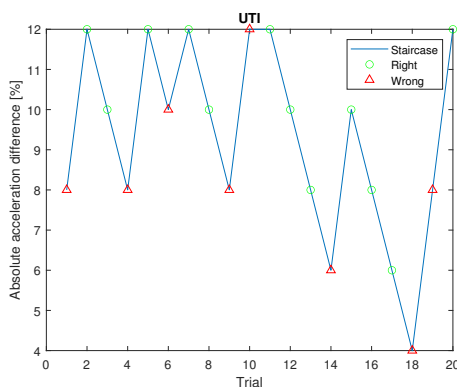


Figure 4: Example staircase of one of the participants in the simulator study.

Simulator study results

Psychometric functions fitted to the data, such as those shown in Figure 5, were used to estimate participants' JNDs using the hybrid method.

The median values, 25th and 75th percentiles and

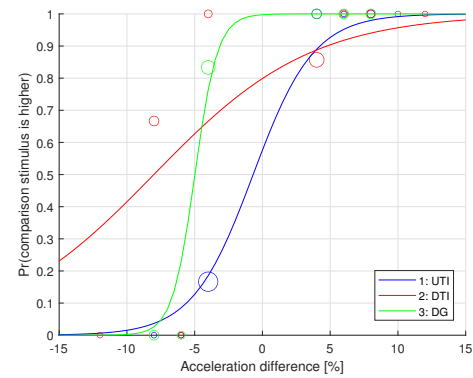


Figure 5: Fitted psychometric functions for each MCA variant for one participant.

extreme values of the JND estimates across MCA variants are shown in Figure 6. The median JND estimate with the DTI variant is higher than that of the UTI and DG variants, but the distributions of the UTI and DG variants are right-skewed and the mean value of all three variants are very similar.

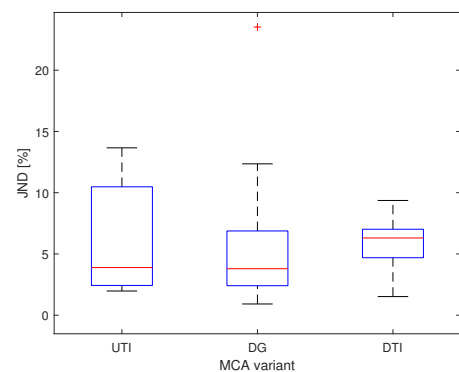


Figure 6: Medians, 25th and 75th percentiles and extreme values of the JND estimates. Estimates judged to be outliers are shown as red crosses.

Although the variability between participants is high in Figure 6, the within-subjects design allowed paired t-tests to be used to compare mean JNDs. Specifically, we compared the mean JNDs of the MCA variants that scale down accelerations – which are expected to yield higher JND estimates – to the UTI variant, based on within-participant differences.

Using a one-tailed paired t-test with significance level $\alpha = 0.05$, we tested the hypothesis that the mean JND of the DTI variant is higher than that of the UTI variant. The null hypothesis is that the mean JNDs are equal. The resulting p-value was 0.55, which exceeds α , and the null hypothesis can therefore not be rejected. Similarly, the hypothesis that the mean JND of the DG variant is higher than that of the UTI variant was tested using a one-tailed paired t-test. The p-value was 0.44, which is also greater than α , and the null hypothesis cannot be rejected. In other words, no statistically significant differences in the mean JNDs of the tested MCA variants were observed.

The participants only did 20 trials with each motion cueing variant and as a consequence, the standard deviations of the JND estimates are high. Baumgart-

ner et al. (Baumgartner, et al., 2019a) reported that the JND in acceleration is related to the reference acceleration according to the model

$$a_{JND} = 0.02 + 0.0275a_{ref}.$$

Using this model, the predicted JND at a 1.4 m/s² reference acceleration is 4.2%. If the reference acceleration is scaled down to 0.7, the predicted JND, using the same model, is 4.8%. The 0.7 scale factor is thus expected to have a relatively small effect on the mean JND. Because of the high standard deviations, small expected differences across motion cueing variants, and the small number of participants in the study, the tests have low statistical power. This could explain why we did not see the expected difference in mean JND between variants.

Because the mean JNDs were found to not differ significantly across MCA variants, and to avoid overfitting, the responses with all MCA variants were used to estimate a single JND for each participant. In Figure 7, a more reliable fit – compared to the models shown in Figure 5 – is obtained by fitting a single psychometric function to all responses. The circles represent the proportion of trials in which the participant answered that the comparison stimulus was higher, at different acceleration differences. The solid curve is the fitted psychometric function. As shown in Figure 8, the variability in the responses of some participants is high.

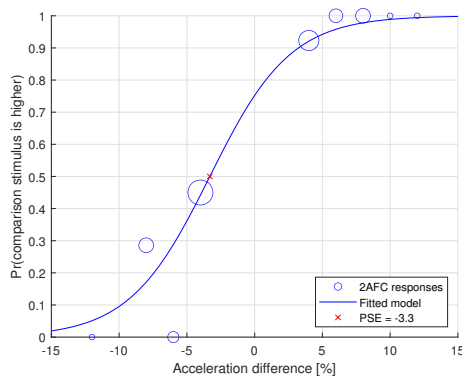


Figure 7: Fitted psychometric function of one of the participants.

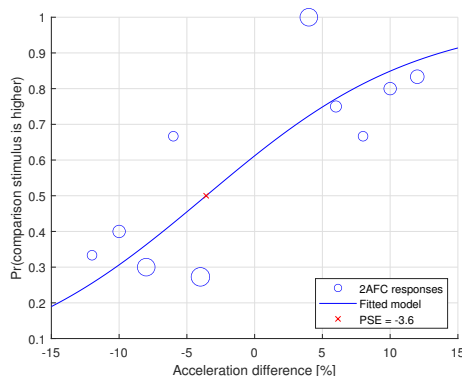


Figure 8: Fitted psychometric function of one participant with high variability in responses.

Mean	Median	Min value	25th percentile	75th percentile	Max value
5.4	5.4	1.7	3.1	7.9	8.6

Table 2: Summary statistics for the JNDs (rounded to one decimal).

Mean	Median	Min value	25th percentile	75th percentile	Max value
-1.9	-2.1	-3.6	-3.4	-0.8	1.4

Table 3: Summary statistics for the PSEs (rounded to one decimal).

Summary statistics for the JNDs and PSEs are shown in Tables 2 and 3, respectively. Notably, the mean PSE was -1.9%. To test the hypothesis that the mean PSE is not equal to zero, a two-tailed one-sample t-test with significance level $\alpha = 0.05$ was used. The null hypothesis is that the mean PSE is equal to zero. The resulting p-value is 0.009, which is less than α , and the null hypothesis is rejected. In the performed simulator study, the comparison stimulus always preceded the standard stimulus, and the negative mean PSE may be the result of a so-called *time error*. It has been reported before that when pairs of stimuli are presented to test participants successively, the memory image of the first stimulus may fade rapidly and participants may therefore perceive the second stimulus as being stronger than the first (Gescheider, 2013).

More detailed information on the statistical analysis of the simulator study results can be found in (Lilljebjörn, et al., 2026).

Effects of the adaptive procedure on JND estimates

The low statistical power of the simulator study highlights the need for an adaptive procedure with low variability for estimating JNDs. To determine how different adaptive procedures affect the variability and bias of JND estimators, a separate Monte Carlo simulation study was conducted. This analysis is independent of the simulator study and is not intended to validate its results, but rather to investigate general properties of the considered procedures. Typical psychometric functions, informed by the simulator study results, were used in the simulations.

Three adaptive test procedures were compared, each targeting difference thresholds at a correct response probability of 75-79%. The first procedure is a weighted up/down staircase (Kaernbach, 1991) in which the up-step is three times the size of the down-step, as shown in Figure 9. The second procedure is a transformed up/down staircase (Levitt, 1971) requiring three consecutive correct answers to take a step down, as shown in Figure 10. The third procedure is a combination of a transformed and weighted up/down staircase in which two consecutive correct answers are required. To target the 75% threshold with this procedure, the up-step is – perhaps somewhat unconventionally – 1.286 times larger than the down-step, as shown in Figure 11.

An adaptive procedure oscillates around a target

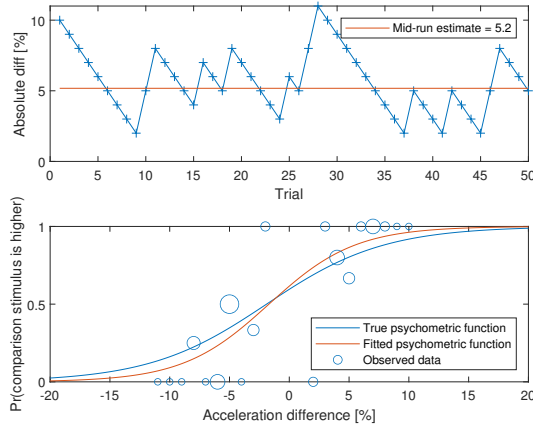


Figure 9: Example simulation of a JND experiment with procedure 1 and a logistic PF with JND = 5.4 and PSE = -1.9.

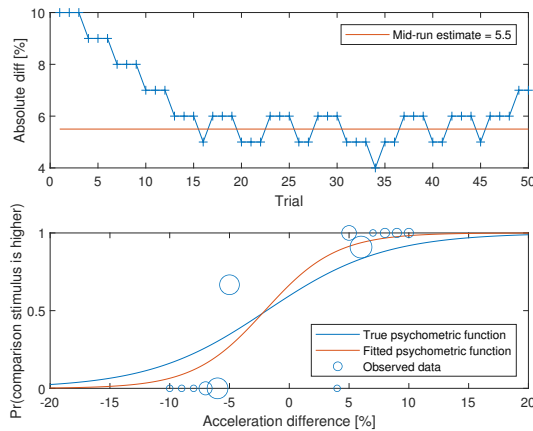


Figure 10: Example simulation of a JND experiment with procedure 2 and a logistic PF with JND = 5.4 and PSE = -1.9.

threshold because the step size and the number of correct responses required to take a step down are chosen so that the expected value of the step, $E(\Delta)$, after a trial with the stimulus intensity corresponding to the target threshold is zero.

$$E(\Delta) = \Delta^+(1 - p_{\text{down}}) - \Delta^- p_{\text{down}} = 0, \quad (1)$$

where Δ^+ is the size of the up-step, Δ^- is the size of the down-step and p_{down} is the probability of a down-step. It follows from equation 1 that

$$\frac{\Delta^+}{\Delta^-} = \frac{p_{\text{down}}}{1 - p_{\text{down}}}. \quad (2)$$

With procedure 1, only one right answer is required to take a step down, so $p_{\text{down}} = p_{\text{right}}$. To target the stimulus intensity where the probability of a right answer is 75%, $X(p_{\text{right}} = 0.75) = X_{0.75}$, it follows from equation 2 that

$$\frac{\Delta^+}{\Delta^-} = \frac{p_{\text{right}}}{1 - p_{\text{right}}} = \frac{0.75}{0.25} = 3.$$

With procedure 2, $\Delta^- = \Delta^+$, so it follows from equation 1 that

$$E(\Delta) = \Delta^+(1 - 2p_{\text{down}}) = 0 \Leftrightarrow p_{\text{down}} = 0.5.$$

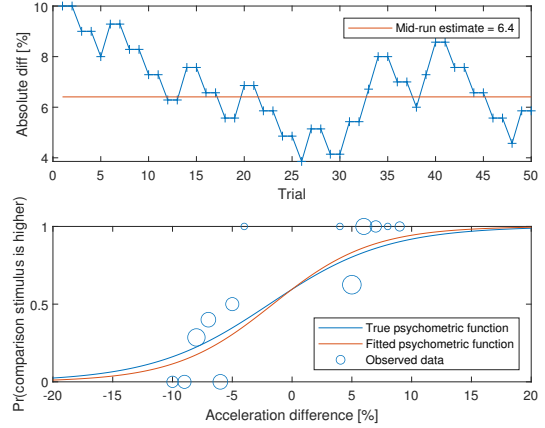


Figure 11: Example simulation of a JND experiment with procedure 3 and a logistic PF with JND = 5.4 and PSE = -1.9.

If three consecutive right answers are required to take a step down, then

$$p_{\text{right}} = p_{\text{down}}^{1/3} = 0.5^{1/3} = 0.79.$$

With procedure 3, $p_{\text{right}} = 0.75$ and two right answers are required to take a step down. With $p_{\text{down}} = 0.75^2 = 0.5625$, it follows from equation 2 that

$$\frac{\Delta^+}{\Delta^-} = \frac{p_{\text{down}}}{1 - p_{\text{down}}} = 1.2857.$$

Monte Carlo simulations in which an adaptive procedure is simulated 10000 times are done to calculate the mean and standard deviation of the JND estimators resulting from each procedure. One main result is that, with many (100+) trials, the hybrid method is hard to beat. As seen in Figure 12, the standard deviation of the hybrid estimators matches that of the mid-run estimators for 100 or more trials, while the bias is virtually zero in Figure 13. However, with fewer trials, the hybrid estimators show substantially higher variability than the mid-run estimators.

The bias of a JND estimator is defined as the difference between the expected value and the theoretical JND of the estimator, i.e., $\text{Bias}(\widehat{JND}) = E[\widehat{JND}] - JND$. The theoretical JND of mid-run estimators resulting from procedures 1 and 3 is the acceleration difference corresponding to a 75% correct response rate, while it is the 79% point for procedure 2.

In Figure 13, the bias of the first procedure stands out. When the starting stimulus in an adaptive procedure is far from the true threshold, the mid-run estimate is typically biased toward the starting value. However, here the procedure starts at an acceleration difference of 5.4%, i.e., the true 75% JND. The large bias of procedure 1 can therefore not be explained by a poor starting value, but rather by the estimator being asymptotically biased. This will be discussed further in the section on the asymptotic properties of the procedures below.

Note that since procedure 2 targets the 79% threshold, it goes from having a negative bias after 30 trials to a slightly positive bias after 200 trials in Figure 13.

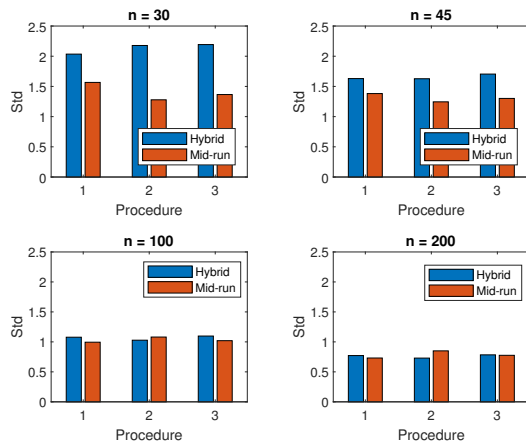


Figure 12: Standard deviations of JND estimators with 30, 45, 100 and 200 trials. The Starting acceleration difference is 5.4%.

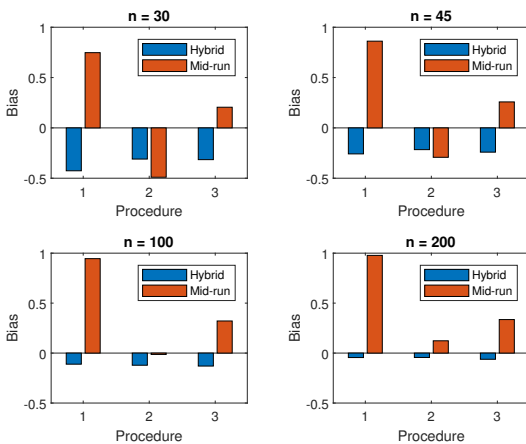


Figure 13: Biases of JND estimators with 30, 45, 100 and 200 trials. The starting acceleration difference is 5.4%.

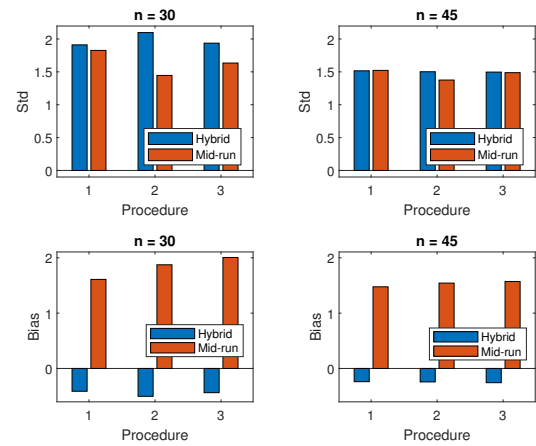


Figure 14: Standard deviations and biases of JND estimators with 10% starting acceleration difference and 1% base step size.

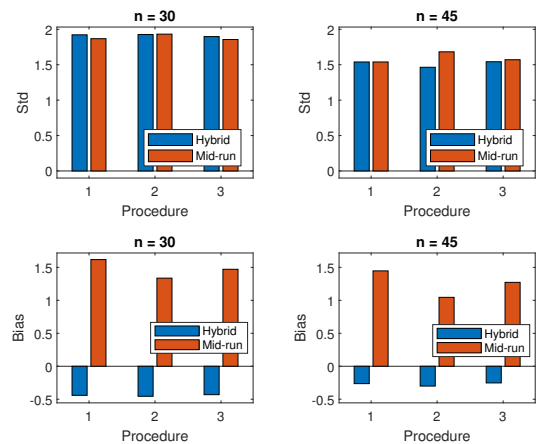


Figure 15: Standard deviations and biases of JND estimators with 10% starting acceleration difference. The base step size is 1% for procedure 1 and 2% for procedures 2 and 3.

Effects of starting stimulus and step size

In the previous section, the procedure started at the target JND. In a real experiment, this would rarely happen. In Figure 14, the starting acceleration difference is 10% instead of 5.4%, and the biases of procedures 2 and 3 are more affected by the starting stimulus than that of procedure 1. The reason for this is that procedures 2 and 3 require more trials than procedure 1 to reach the target threshold when starting from a high stimulus level. As shown in Figure 15, this can be mitigated to an extent by increasing the step size. The bias of procedure 1 is not improved by increasing the base step size, so it is kept at 1%.

Effects of the psychometric function

Procedure 2, which requires three consecutive correct responses to take a step down, has emerged as the best procedure based on simulations using a psychometric function whose JND and PSE were set to the mean JND and PSE obtained in the simulator study. Procedure 2 is nearly asymptotically unbiased, and its variability is comparable to that of procedure

1. In the previous section, the starting acceleration difference was shown to affect the bias of procedure 2 considerably. However, this effect can be mitigated to an extent by increasing the step size. Procedure 2 is also more practical than procedure 3 with its unconventional up-step/down-step ratio of 1.286.

To verify that these conclusions also hold for participants with JNDs below and above the average value of 5.4%, simulations were done with the psychometric functions shown in Figure 16. The red curve represents a participant with a below-average JND of 4%, and the yellow curve a participant with an above-average JND of 8%. The effect of the point of subjective equality is also examined in this section by using a PSE of 0%, instead of the average PSE of -1.9% used previously.

In Figure 17, a psychometric function with a 75% JND of 4% and a PSE of 0% was used, and procedure 2 is considerably less biased than procedure 1. Figure 18 shows results obtained using a psychometric function with a 75% JND of 8% and a PSE of 0%. Under these conditions, the bias of procedure 2 is comparable to, although somewhat larger than, that of procedure 1. The reason that procedure 2 is neg-

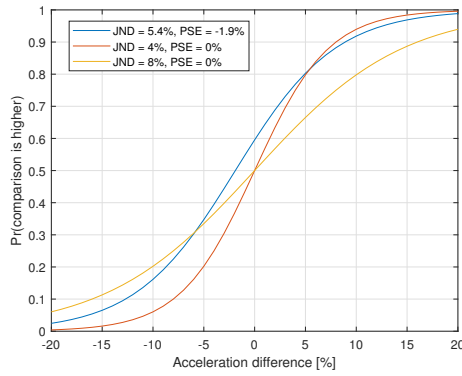


Figure 16: Three different psychometric functions used in simulations.

actively biased with 30 trials under these conditions is that it targets $X_{0.79}$, whereas the starting acceleration difference in the simulation is $X_{0.75} = 8\%$.

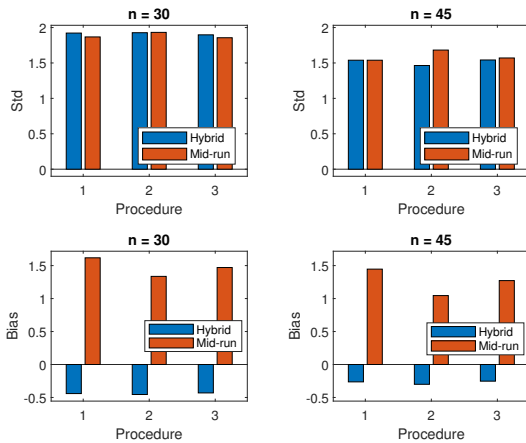


Figure 17: Standard deviation and bias of JND estimators. JND = 4% and PSE = 0%.

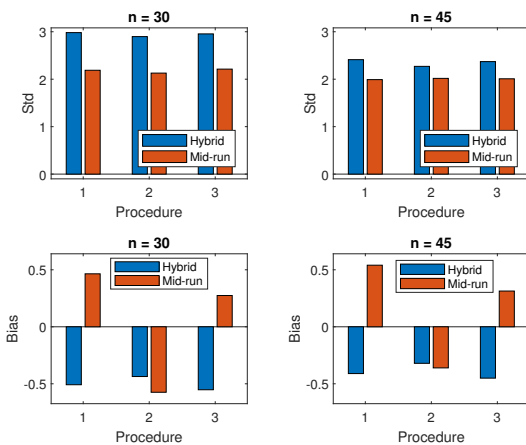


Figure 18: Standard deviation and bias of JND estimators. JND = 8% and PSE = 0%.

Asymptotic properties of the adaptive procedures

In a fixed-step size adaptive procedure, the probability of transitioning to any other state depends only on the current state, so it is, by definition, a Markov chain. The states of an adaptive procedure are the stimulus intensities. The state probabilities after the i 'th transition is

$$\pi_i = \pi_{i-1}P,$$

where π_{i-1} are the state probabilities before the i 'th transition and P is the state transition matrix. As the number of transitions tends to infinity, the vector of state probabilities tends to the stationary distribution of the Markov chain

$$\pi = \pi P.$$

The stationary distribution, π , is thus an eigenvector to the state transition matrix with eigenvalue 1.

The stationary distributions of the first and second adaptive procedures with a psychometric function with $JND = 5\%$ and $PSE = 0\%$ are shown in Figure 19 and in Figure 20, respectively. It is easy to see why the mid-run JND estimator that uses procedure 1 is biased high. In Figure 19, the variability of the stimulus intensities is high, and the distribution is right-skewed. The histogram is the distribution of the stimulus intensities of a Monte Carlo simulation of the procedure and the solid red line is the theoretical stationary distribution of the procedure.

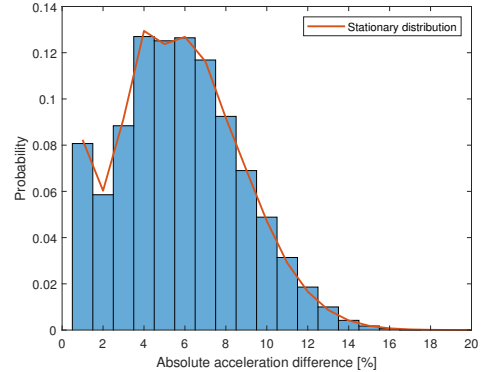


Figure 19: Stationary distribution of procedure 1.

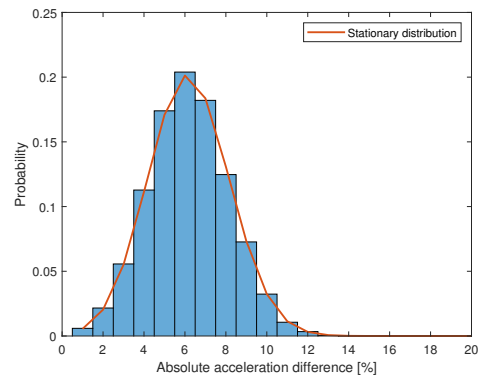


Figure 20: Stationary distribution of procedure 2.

Statistical power

To compare whether 30 or 45 trials per condition yield sufficiently low variability, we estimated the statistical power of the best JND estimators in a hypothetical study.

For practical reasons, each participant can perform around 90 trials. Thus, participants can either test three conditions with 30 trials each, or two conditions with 45 trials each. Mean JNDs (in %) are compared using paired t-tests.

Assuming equal variance of single JND estimates across conditions, the variance of within-participant differences is $\sigma_d^2 = 2\sigma_{JND}^2$, and the variance of the mean of all differences in a sample of N differences is σ_d^2/N .

Expected effect size was set to 1.2% (absolute) based on (Baumgartner, et al., 2019a). With 40 participants, the resulting sample sizes are $N = 40$ for 30 trials and $N = 20$ for 45 trials. Using a one-tailed paired t-test with $\alpha = 0.025$, statistical power of the best JND estimators with 30 and 45 trials was computed and is shown in Table 4. Under these assumptions, statistical power was consistently higher for 30 trials than for 45 trials across all JND estimators.

Table 4: Statistical power (%) of the best JND estimators

Trials	Procedure 1		Procedure 2	
	Staircase	Hybrid	Staircase	Hybrid
30	93	75	91	79
45	79	64	69	68

Conclusions

In a study with nine professional truck test drivers in a moving-base driving simulator, we compared the mean JNDs with three variants of the classical motion cueing algorithm. The mean JND of the generic variants with a scale factor of 0.7 did not differ significantly from that of the MCA variant tuned specifically for tip-in tests. The uncertainty of the JND estimates was high, and the sample size was low. As a result, the power of the statistical tests was low, and we believe this to be the reason why we did not see the expected effect of the scale factor on the JNDs.

Motivated by the low statistical power of the driving simulator study, Monte Carlo simulations were done to compare the standard deviations and biases of JND estimators using different adaptive procedures. We demonstrated that a target threshold around 75-79% can be estimated using a weighted or a transformed adaptive staircase procedure, or by combining ideas from both procedures, as we did in procedure 3. We also demonstrated that the JND can be estimated by either calculating the mean of the reversals in the staircase, which is the standard method, or with a hybrid method in which the JND is estimated using a fitted logistic psychometric function. With less than 50 trials, the standard deviation of the mid-run estimators is considerably lower than that of the hybrid estimators, but the hybrid estimators are generally less biased.

In our next study, we plan to use either procedure 1 with a base step size of 1%, or procedure 2 with a base step size of 2%, as both yield low standard deviations, which is desirable when studying ef-

fect on JNDs. To reduce bias from the starting stimulus difference, the experiment should begin near the expected mean JND. The negative mean PSE in the simulator study indicates that participants detect smaller differences when the higher stimulus is presented last. We therefore recommend using comparison stimuli only below or only above the standard stimulus, with the comparison presented last in half of the trials and first in the other half. Finally, a within-subject design is preferred to maximize statistical power. Learning and fatigue effects will be managed through frequent breaks and a balanced order of conditions.

References

- Baumgartner, E., Ronellenfisch, A., Reuss, H.-C., and Schramm, D., 2019a. A perceptual approach for evaluating vehicle drivability in a dynamic driving simulator. *Transportation research part F: traffic psychology and behaviour*, 63, pp. 83–92.
- Feb. 2019b. Using a dynamic driving simulator for perception-based powertrain development. *Transportation Research Part F: Traffic Psychology and Behaviour*, 61, pp. 281–290. ISSN: 13698478. <https://doi.org/10.1016/j.trf.2017.08.012>.
- Erlor, P., Menig, A., Uphaus, F., Makosi, C. A. M., Rinderknecht, S., and Vogt, J., July 2018. Investigating the perception of powertrain shuffle with a longitudinal dynamic driving simulator. In: *IEEE*, pp. 1427–1431. ISBN: 978-1-5386-1854-7. <https://doi.org/10.1109/AIM.2018.8452255>.
- García-Pérez, M. A., June 1998. Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties. *Vision Research*, 38, 12, pp. 1861–1881. ISSN: 00426989. [https://doi.org/10.1016/S0042-6989\(97\)00340-4](https://doi.org/10.1016/S0042-6989(97)00340-4).
- Gescheider, G. A., 2013. *Psychophysics: the fundamentals*. Psychology Press.
- Hall, J. L., 1981. Hybrid adaptive procedure for estimation of psychometric functions. *The Journal of the Acoustical Society of America*, 69, 6, pp. 1763–1769.
- Jamson, A. H. J., 2010. *Motion cueing in driving simulators for research applications*. University of Leeds.
- Kaernbach, C., May 1991. Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics*, 49, 3, pp. 227–229. ISSN: 0031-5117. <https://doi.org/10.3758/BF03214307>.
- Kraft, E., Viehmann, A., Erlor, P., and Rinderknecht, S., 2021. Virtual Driving Tests of Powertrain Systems in the Driving Simulator. *ATZ worldwide*, 123(3), pp. 40–45.
- Leek, M. R., Nov. 2001. Adaptive procedures in psychophysical research. *Perception & Psychophysics*, 63, 8, pp. 1279–1292. ISSN: 0031-5117. <https://doi.org/10.3758/BF03194543>.
- Levitt, H., Feb. 1971. Transformed Up-Down Methods in Psychoacoustics. *The Journal of the Acoustical Society of America*, 49, 2B, pp. 467–477. ISSN: 0001-4966. <https://doi.org/10.1121/1.1912375>.
- Lilljebjörn, E. G., Kharrazi, S., Åslund, J., and Singull, M., 2026. Effects of motion cueing on longitudinal acceleration perception in a driving simulator. <https://doi.org/10.48550/arXiv.2606.28055>. arXiv: 2606.28055 [eess.SY].
- Menig, A., 2025. Der eben merkbare Unterschied von Längsbeschleunigung im Fahrsimulator—Experimentelle Analysen unter Berücksichtigung verschiedener Einflussfaktoren. PhD.
- Nesti, A., Masone, C., Barnett-Cowan, M., Giordano, P. R., Bülthoff, H. H., and Pretto, P., 2012. Roll rate thresholds and perceived realism in driving simulation. In: *Driving Simulation Conference Europe (DSC 2012)*, pp. 23–31.
- Nesti, A., Nooij, S., Losert, M., Bülthoff, H. H., and Pretto, P., May 2016. Roll rate perceptual thresholds in active and passive curve driving simulation. *SIMULATION*, 92, 5, pp. 417–426. ISSN: 0037-5497. <https://doi.org/10.1177/0037549716637135>.
- Reid, L. D. and Nahon, M. A., 1985. *Flight simulation motion-base drive algorithms: Part 1 - Developing and testing the equations*.

Acceptance of Adaptive Cruise Control on a Dynamic Driving Simulator: the Role of Tilt Coordination

Xabier Carrera Akutain¹, Paolo Cattaneo², Francesco Comolli², Massimiliano Gobbi², Gianpiero Mastinu², Roger Torm Obradors¹

(1) Toyota Motor Europe, Hoge Wei 33, 1930 Zaventem, Belgium, e-mail: {xabier.carrera.akutain, roger.torm}@toyota-europe.com

(2) Politecnico di Milano, Via Candiani 72, 20158 Milan, Italy, e-mail: {paolo.cattaneo, francesco.comolli, massimiliano.gobbi, gianpiero.mastinu}@polimi.it

Abstract – *This study investigates the influence of tilt coordination tuning on driver acceptance of Adaptive Cruise Control (ACC) in a dynamic driving simulator. While motion cueing algorithms (MCAs) are essential for reproducing realistic longitudinal dynamics, improper tuning, especially of tilt coordination, may introduce perceptual conflicts that bias user evaluation of automated driving systems. A within-subject experiment was conducted using a dynamic driving simulator, comparing three tilt gain configurations across representative acceleration and braking manoeuvres. Objective fidelity metrics (e.g., RMS acceleration error, jerk, tilt rate) were combined with subjective measures of realism, comfort, and trust. Results show that increasing tilt gain improves objective fidelity but may degrade user acceptance when perceptual thresholds are exceeded. In particular, high tilt gain produced unrealistic and disturbing pitch sensations, while no tilt reduced the perception of sustained acceleration. A moderate configuration achieved the best balance, enhancing realism without introducing discomfort. Findings highlight that acceptance is not solely determined by physical fidelity, but by perceptual coherence. The study provides practical guidelines for MCA tuning, emphasizing the importance of balancing acceleration reproduction with human sensitivity thresholds to ensure reliable simulator-based evaluation of ACC systems*

Keywords: Adaptive Cruise Control; Driving Simulator; Motion Cueing; Tilt Coordination; User Acceptance.

Copyright: © 2026 by the authors.

Licensee Driving Simulation Association.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.82157/dsa/2026/13>

Introduction

Adaptive Cruise Control (ACC) is one of the most widespread automation functions in production vehicles. Yet the actual benefits of ACC depend on drivers sustained and appropriate use, which is influenced by perceived comfort, safety, and trust (Schakel, et al., 2017). Because physical prototyping and proving ground studies are time consuming and expensive, simulator-based acceptance and human factors testing is common during the development (Stabile, et al, 2023).

In dynamic driving simulators, however, acceptance testing can be biased when longitudinal motion cueing algorithm (MCA) is unrealistic. This issue is particularly relevant for high-fidelity platforms with limited longitudinal workspace. Sustained acceleration and deceleration are typically rendered by a motion cueing algorithm that combines translational motion with tilt coordination. Tilt coordination generates a component of the longitudinal specific force by slowly pitching the cockpit (Beghi, et al., 2012). While this technique can improve cue fidelity for low frequency events, it may also introduce cue conflict, false sensations, and perceptible washout motions if non properly tuned.

Recent contributions in the literature provide valuable context for this trade off. Prepositioning strategies and compensational MCAs have been proposed to better exploit limited workspace (Kraft, et al., 2022). Comparative studies highlight that “common knowledge” tuning rules do not always generalize and that subjective evaluation should be treated as an integral part of MCA development (Kolff, et al., 2024); (Nicolai, et al., 2025). In parallel, frequency-aware and model predictive control (MPC) frameworks are increasingly adopted to formalize tuning and improve robustness across manoeuvres (Gonzalez, et al., 2024), (Kolff, et al., 2025).

The present work investigates how tilt coordination tuning affects ACC acceptance on a dynamic driving simulator. The motivation is twofold: from a scientific perspective, to clarify the link between cue fidelity and psychological metrics such as acceptance and trust; and from an industrial perspective, to provide actionable tuning guidance for simulator-based ACC evaluation campaigns, where inadequate motion cueing could distort judgments of comfort and safety.

Research question and working hypotheses

The main research question is “to what extent does tilt coordination tuning influence drivers’ acceptance and use of ACC in a dynamic driving simulator?”.

Tilt coordination can be described as a controlled “leaning” of the cockpit to mimic the feeling of braking and accelerating without requiring long translations. Used moderately, it increases realism; used too strongly, it can become noticeable as an unnatural pitch motion, potentially reducing immersivity. (Bruschetta, et al., 2017).

The tilt angle θ add a contribution on the perceived longitudinal acceleration $\tilde{a}_x$ as in Eq. 1:

$$\tilde{a}_x = a_x + K g \sin(\theta) \approx a_x + K g \theta \quad (1)$$

where a_x is the real longitudinal acceleration of the driving simulator, g is the gravity, and K is the so-called Tilt gain. We evaluate the following working hypotheses: Well-tuned tilt coordination increases perceived realism and comfort of longitudinal events, thereby improving ACC acceptance (Van Der Laan, et al., 1997) and increasing trust in automation (Lee & See, 2004); (Hoff & Bashir, 2015). Aggressive tilt coordination (excessive pitch magnitude, pitch rate, or insufficient onset/jerk limiting and washout shaping) introduces cue conflict and perceptible “false cues”, reducing acceptance. Changes in acceptance will be associated with objective improvements in longitudinal cue fidelity (e.g., reduced root-mean-square (RMS) error between reference and measured accelerations, reduced longitudinal jerk peaks) to define quantitative fidelity metrics and to test whether improvements in measured fidelity translate into statistically significant increases in ACC acceptance.

Method

The study is conducted at the DriSMi dynamic driving simulator (www.drismi.polimi.it). Proving-ground data recorded on an instrumented vehicle (accelerations filtered out at 5 Hz) are replayed in VI-CarRealTime and followed by the simulator motion platform through the VI-DriveSim motion cueing framework.

Five longitudinal manoeuvres are used as representative building blocks for ACC events: braking 30→0 km/h, 50→0 km/h, 100→50 km/h and 140→70 km/h, and acceleration 0→60 km/h.

A within-subject design compares three motion conditions: baseline motion cueing, and two motion cueing configurations with systematically varied tilt coordination. In this phase of the study, the subjects have not tested the manoeuvres on the real vehicle; their evaluations are based on their driving experience.

The tuning space includes longitudinal high-pass/low-pass filtering, longitudinal acceleration gain, non-linear gain shaping (useful to reduce annoying effects due to high pass filtering), tilt gain and tilt filtering, balance (biasing between acceleration and braking phases), offset (workspace

pre-positioning), and washout/inversion avoidance. (VI-Grade GmbH, 2025).

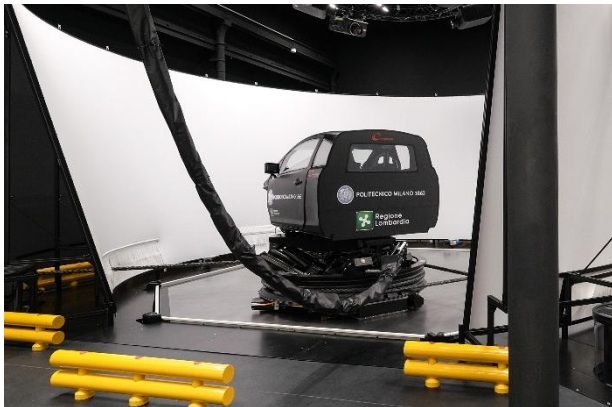


Figure 1: DriSMi Driving Simulator

Within the explored parameter space and considering the set of manoeuvres taken into exam, tilt-coordination tuning (tilt gain and filtering) has a stronger perceptual impact than longitudinal acceleration gain. Cue fidelity will be quantified with metrics such as RMS and peak errors between reference and measured acceleration. Acceptance can be measured using the van der Laan usefulness/satisfaction scale (Van Der Laan, et al., 1997) and trust-in-automation ratings (Lee & See, 2004); (Hoff & Bashir, 2015). Behavioural indicators include time headway during ACC, manual overrides and longitudinal jerk (Xue, et al., 2025).

Results and discussion

A dedicated preliminary tuning and validation campaign explored multiple MCA parameter sets to reproduce the reference longitudinal acceleration profiles while respecting the simulator workspace.

From that base cueing configuration, three different Tilt gain setups are considered to analyse the influence of the tilt on the longitudinal acceleration perception for the ACC acceptance.

The tuning of the cueing algorithm is made to ensure that the acceleration and deceleration sensations generated with the simulator remain coherent with those experienced in real vehicles. In this way drivers will be able to recognise the same differences between different ACC configurations in virtual and in real-road tests. The Tilt Low Pass filter is set to 0.3 Hz to catch only the longitudinal acceleration contribution and filter out the road roughness.

Tilt gains typical range is from 0 (simulator pitch reflects only the vehicle's pitch angle) to 1, which has the target to add the longitudinal acceleration of the reference vehicle. These limits are common for hard handling manoeuvres with strong accelerations. For low-magnitude but sustained accelerations, however, the tilt gain can be increased up to 10. In

this way it is possible to recreate a strong and sustained sensation of acceleration without using platform displacement.

The Tilt gain is set initially to 0 to have a starting configuration without considering tilt. As second attempt, the gain is set to 5 to analyse a medium contribution of the Tilt, and lastly the gain is set to 8 due to the driving simulator working space limits. Figure 2 shows the longitudinal accelerations of the reference vehicle and the three cueing sets for the braking manoeuvre from 30 to 0 km/h.

The Tilt gain 0 cueing has a completely different shape from the reference, except for the onset of deceleration, where the driver expects a bite when starting to brake. The subjects tested report that for this cueing configuration, the acceleration is difficultly perceived. The Tilt gain 8 cueing follows well the reference acceleration shape, showing a consistent initial amplitude in the first part of the manoeuvre. Also, the final part of the manoeuvre is well followed, showing the final bounce expected when the vehicle stops. The subjects reported an exaggerated tilt rate that do not permit them from evaluating the ACC sensation and it is perceived as annoying and far from real vehicle feeling. Finally, the Tilt gain 5 cueing is following the reference shape, showing both the initial jerk and the final bounce, but not reaching the acceleration value of the previous set. The subjects tested report a good feeling. The tilt rate is no more annoying, and the longitudinal acceleration is well perceived. In Table 1 are reported the RMS errors between the reference and the proposed cueing sets accelerations. Higher the Tilt gain is, lower the RMS error.

Figure 3 shows the pitch rates of the three different cueing proposed, considering both the vehicle pitch and the tilt coordination. The vehicle pitch rate reference is the input of the cueing. The tilt gain 0 configuration corresponds to the vehicle pitch rate output without tilt coordination. A delay can be observed because the pitch motion is processed with a gain of 0.5 and a high-pass filter set to 0.1 Hz. The Tilt gain 5 configuration exceeds the value of 0.02 rad/s, reported as the threshold for pitch rate perception (Fitze, et al., 2023), in the first manoeuvre instants. This tilt rate value helps the driver to feel the tilt contribution on the longitudinal acceleration, but it is not perceived as a tilt rotation. In the final part, the pitch rate value is higher, but it is not evaluated as annoying in all cases. The Tilt gain 8 exceeds the threshold of 0.1 rad/s reported in (Bruschetta, et al., 2017) that indicates a recognizable rotation, and it is reported as unnatural, especially in the last manoeuvre part. Due to the very low linear longitudinal acceleration, the tilt rate threshold is determined solely by the tilt coordination.

Figure 4 shows the longitudinal jerk for the cueing sets proposed. Referring to the first braking bite, the Tilt gain 5 and 8 are close, and similar to the

reference, while the Tilt gain 0 is lower. Considering the final part of the braking, both the proposed cueing sets show lower jerk than the reference, but they're close each other. All the analysed sets are sensitively lower than the abrupt braking threshold of 15 m/s^3 proposed in (de Winkel, et al., 2023).

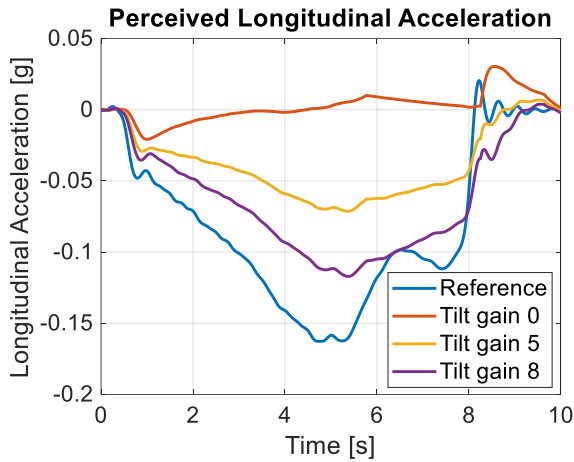


Figure 2: Braking 30 → 0 km/h, longitudinal acceleration in the cockpit and vehicle reference acceleration profile. Comparison of the three cueing configurations analysed.

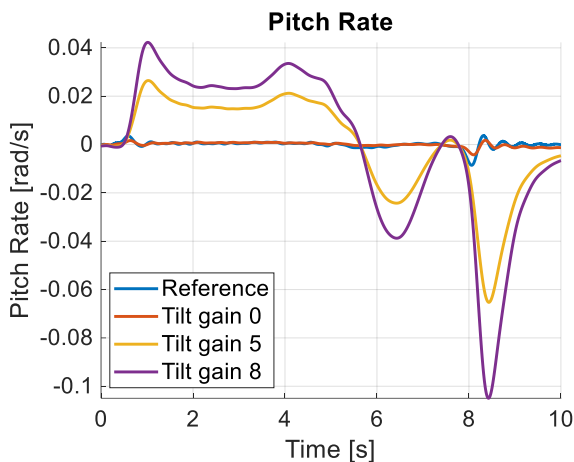


Figure 3: Braking 30 → 0 km/h, Pitch rate of the cueing configurations analysed.

Figure 5 shows the longitudinal accelerations for the other manoeuvres analysed, comparing the results of the three different cueing sets proposed. It is possible to note that the most aggressive manoeuvre is the braking from $50 \rightarrow 0 \text{ km/h}$, which reaches a reference deceleration of about 0.3 g .

Figure 6 shows the tilt rate for the other manoeuvres analysed, from which is possible to note that even in this case, the most aggressive manoeuvre is the braking from $50 \rightarrow 0 \text{ km/h}$, which shows a tilt rate of about 0.15 rad/s , reported as annoying. In the tilt gain 5 case, the 0.1 rad/s threshold is reached only marginally and is not exceeded.

Finally, Figure 7 shows the longitudinal jerk of the manoeuvre analysed. The maximum jerk obtained is for the braking manoeuvre from $50 \rightarrow 0 \text{ km/h}$. The

manoeuvres $100 \rightarrow 50 \text{ km/h}$, $140 \rightarrow 70 \text{ km/h}$ and $0 \rightarrow 60 \text{ km/h}$ are jerky because of the experimental data used as input in the simulator, but the contribution is clear in the initial phase of the manoeuvres.

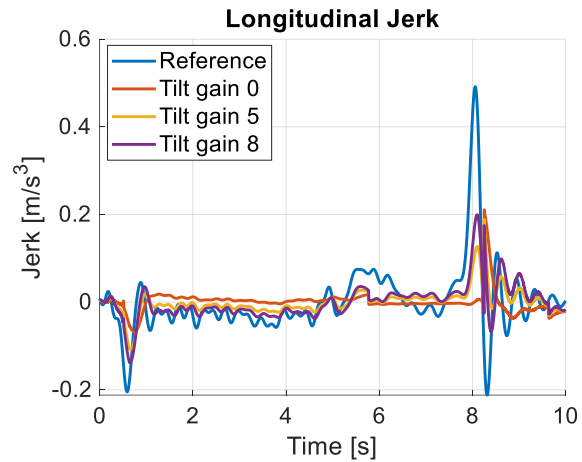


Figure 4: Braking 30 → 0 km/h, Longitudinal jerk of the cueing configuration analysed.

A panel of 12 drivers with different driving skills experienced the different cueing setups in all the manoeuvres proposed (age from 25 to 57, driving experience from 7 to 39 years). A questionnaire (adapted from (Kolff, 2025)) has been submitted to the drivers at the end of each manoeuvre. Each driver tested three different cueing sets. They were asked to rate the realism of the ACC feelings during the test with the three different cueing sets, focusing on the sustained accelerations, the initial braking bites and the final parts of the manoeuvres, when the acceleration/deceleration is going to zero. They were asked also to compare their sensations on the basis of their experience with real-world ACCs.

Drivers report that the accelerations with the Tilt gain 0 set is barely perceived. The initial bite and the final phase of the manoeuvre are well perceived, but there is no feeling for the sustained acceleration.

On the opposite side, drivers report that the initial bite and final washout with Tilt gain 8 set are very annoying, they clearly feel the tilt contribution as unnatural and artificial, especially in the two braking manoeuvres from $30 \rightarrow 0 \text{ km/h}$ and $50 \rightarrow 0 \text{ km/h}$, where the tilt rate is clearly overcoming the annoying threshold (Figure 6). In the other manoeuvres, the initial bite is perceived as good, as the sustained acceleration phase, but it is not sufficient to rate positively the cueing set.

Table 1: RMS errors of the longitudinal acceleration of the analysed cueing sets wrt the reference.

	Tilt gain 0	Tilt gain 5	Tilt gain 8
30→0 km/h	0.044	0.024	0.013
50→0 km/h	0.097	0.053	0.031
100→50 km/h	0.036	0.018	0.009
140→70 km/h	0.061	0.032	0.015
0→60 km/h	0.049	0.025	0.011

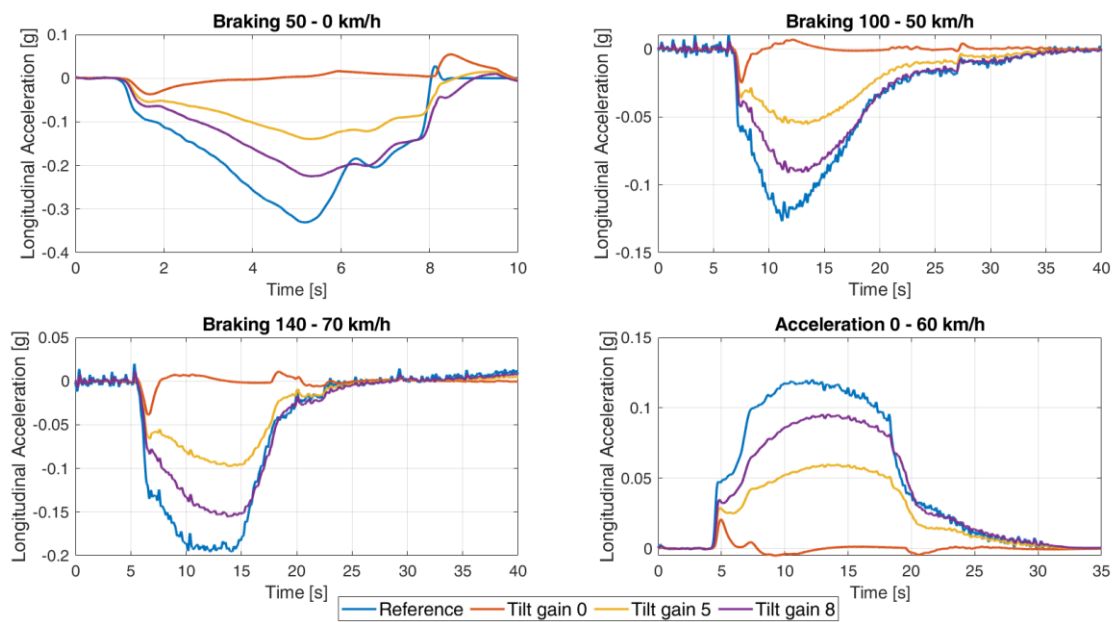


Figure 5: Longitudinal acceleration in the cockpit and vehicle reference acceleration profile. Comparison of the three cueing configurations analysed. Four different manoeuvres.

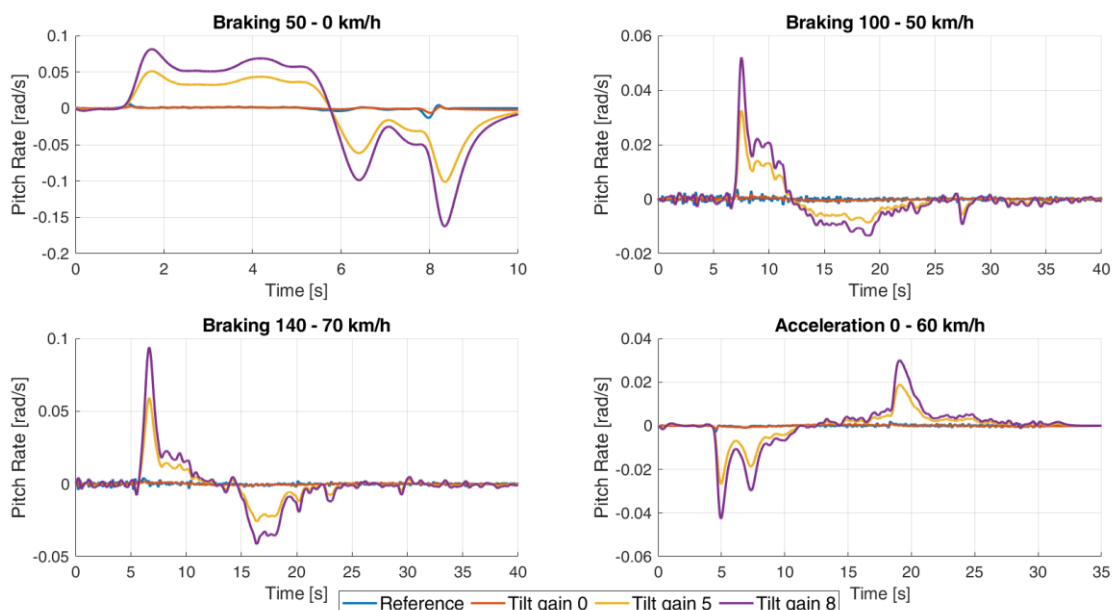


Figure 6: Pitch rate - comparison of the three cueing configurations analysed. Four different manoeuvres.

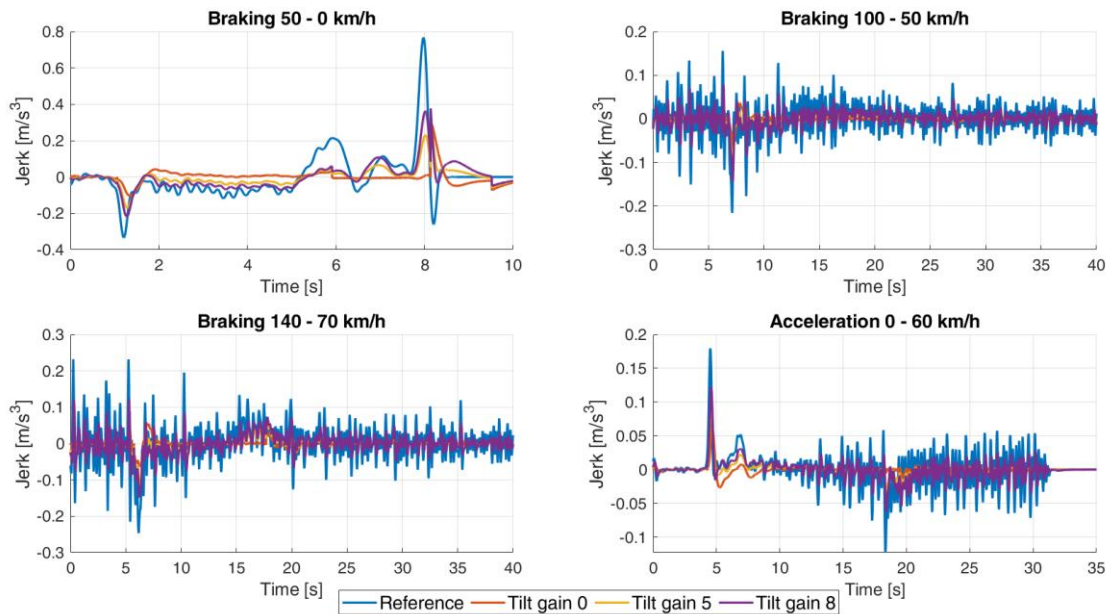


Figure 7: Longitudinal jerk - comparison of the three cueing configurations analysed. Four different manoeuvres.

Finally, the drivers report the Tilt gain 5 cueing set as the best option. The initial bite is clearly perceived for all the manoeuvres, as the final bounce for the two braking to 0 km/h, but they are not annoying and help the driver to feel the acceleration changes (Figure 6). In addition, the sustained acceleration is clearly perceived as quite natural and helpful to evaluate the ACC performances.

Figure 7 shows the longitudinal jerk for all the manoeuvres, from which is possible to see that the contribution of all the cueing sets is similar. Because of the lower tilt contribution, drivers reported that the initial bite jerk is perceived better with the Tilt gain 0 set.

In the following, Figure 8 shows the radar charts of the questionnaire's answers for the three different types of manoeuvres are reported. The drivers are requested to evaluate the realism of the acceleration/deceleration feeling compared to the expectations, differentiating the evaluation for the three manoeuvres phases analysed: initial bite, sustained acceleration, final phase.

In the braking to stop manoeuvres, the participants expressed a clear preference of the Tilt gain 5 set for all the phases. The Tilt gain 8 is quite appreciated in the sustained deceleration phase but judged too much disturbing in the final bounce. Tilt gain 0 is evaluated negatively due to the poor perception of the braking deceleration.

In the braking not to stop manoeuvres, there is the same trend, but the preference is not so marked between Tilt gain 5 and 8, due to the absence of the final bounce of the vehicle stop.

In the acceleration manoeuvre the tilt gain 5 is preferred with respect to tilt gain 8 by the drivers.

The last chart shows the sum of the answers, considering all the manoeuvres together. It appears a clear preference of the drivers for the Tilt gain 5 cueing set, while too high tilt contribution, as in Tilt gain 8, in the final phases of the manoeuvres is judged annoying, mainly in the braking from 50 km/h to 0 manoeuvre. Tilt gain 0 is generally considered as the weakest in terms of deceleration perception.

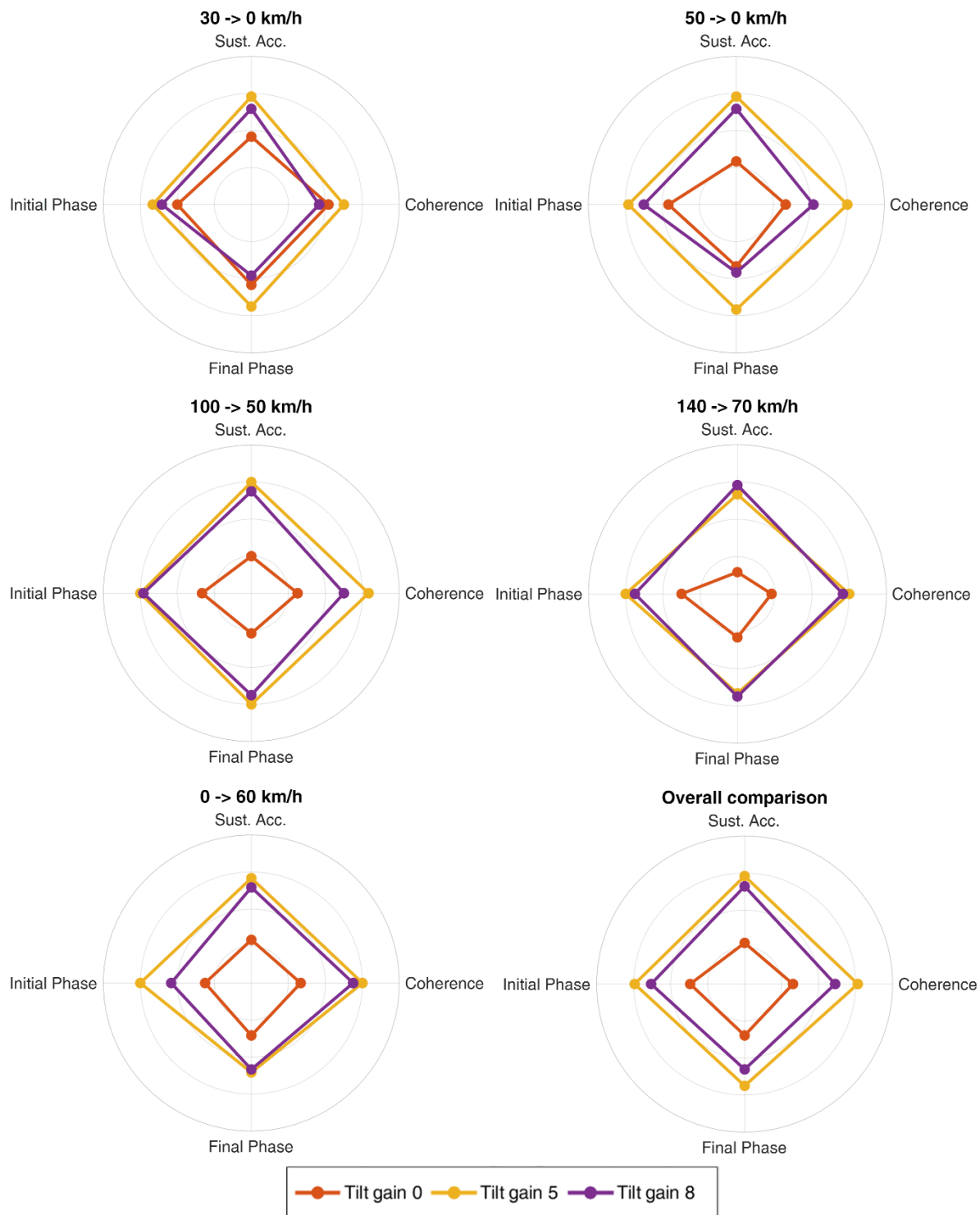


Figure 8: Realism evaluation through questionnaires for all the manoeuvres.

Conclusions

The tilt coordination role has been considered as beneficial to simulate the ACC acceleration deceleration manoeuvres. In particular cases, increasing the contribution of the tilt coordination

beyond 100%, to emphasize the longitudinal acceleration feeling, is positive.

The results indicate that acceptance studies should include acceleration measurements at the occupant location when motion cueing is under evaluation.

The full experiment quantifies, through appropriate statistical testing, the effect size of tilt-coordination

tuning on acceptance scales and translates these findings into guidelines for robust simulator-based ACC evaluation.

It is clear that the RMS error between the reference acceleration and the one obtained by the simulator must guide the choice of the cueing algorithm parameters, but it should not be the only focus to properly set the simulation.

The tilt rate analysis should be one of the main guides: it is important to overcome the minimum sensitivity threshold for the pitch rate, but it is fundamental not to overcome the disturbing threshold.

The same for the jerk: it is useful to evaluate the ACC intervention, but sometimes its contribution is overshadowed by the tilt contribution.

A measurement-driven workflow for MCA tuning can reduce the risk of false negative acceptance outcomes caused by motion-cue artifacts.

The simulator operators have to maximize the cueing output, by using all the tools at their disposal, but they have to consider that not all the cueing parameters values can improve the realism of the simulation.

In the next phases of the study, a larger number of subjects will be involved to obtain a wider set of subjective evaluations in order to identify the most appreciated cueing configuration. Once the cueing set is defined, different ACC logics will be evaluated virtually on the simulator while keeping the cueing parameters fixed. This will allow an initial assessment of the different ACC behaviours before trying them on real vehicle.

References

- Beghi, A., Bruschetta, M., Maran, F. & Minen, D., 2012. A Model-based Motion Cueing strategy for compact driving simulation platforms. *Proceedings of the Driving Simulation Conference 2012*.
- Bruschetta, M., Maran, F. & Beghi, A., 2017. A fast implementation of MPC-based motion cueing algorithms for mid-size road vehicle motion simulators. *Vehicle System Dynamics*, 55(6), pp. 802-826.
- de Winkel, K. N., Irmak, T., Happee, R. & Shyrokau, B., 2023. Standards for passenger comfort in automated vehicles: Acceleration and jerk. *Applied Ergonomics*, 106(103881).
- Fitze, D. C., Mast, F. W. & Ertl, M., 2023. Human vestibular perceptual thresholds — A systematic review of passive motion perception. *Gait & Posture*, Volume 107, pp. 83-95.
- Gonzalez, C., Kooijman, L., Lim, C. P. & Asadi, H., 2024. Downsampling for Efficient Tuning of Nonlinear Model Predictive Control Motion Cueing Algorithm. *Proceedings of the Driving Simulation Conference Europe 2024 VR*, pp. 103-110.
- Hoff, K. A. & Bashir, M., 2015. Trust in automation: integrating empirical evidence on factors that influence trust. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 57(3), pp. 407-434.
- Kolff, M., 2025. *Motion Cueing Quality in Driving Simulation*. s.l.:Doctoral Thesis TU Delft.
- Kolff, M. et al., 2024. Motion Cueing in BMW's Driving Simulation Center: Experiences Versus Common Knowledge. *Proceedings of the Driving Simulation Conference Europe 2024 VR*, pp. 119-125.
- Kolff, M. et al., 2025. A Frequency-Aware Model Predictive Control Motion Cueing Algorithm. *Proceedings of the Driving Simulation Conference Europe 2025 XR*, pp. 131-138.
- Kraft, E., He, P. & Rinderknecht, S., 2022. Application of Prepositioning Strategies in a Compensational Motion Cueing Algorithm. *Proceedings of the Driving Simulation Conference Europe 2022 VR*, pp. 127-134.
- Lee, J. D. & See, K. A., 2004. Trust in automation: designing for appropriate reliance. *Hum Factors*, 46(1), pp. 50-80.
- Nicolai, T. et al., 2025. Subjective Evaluation of Motion Cueing Algorithms: A Comparative Study on Real-World Expectations in a Robot-Based Driving Simulator. *Proceedings of the Driving Simulation Conference Europe 2025 XR*, pp. 125-130.
- Schakel, W. J., Gorter, C. M., de Winter, J. C. F. & van Arem, B., 2017. Driving Characteristics and Adaptive Cruise Control ? A Naturalistic Driving Study. *IEEE Intelligent Transportation Systems Magazine*, 9(2), pp. 17-24.
- Stabile P., Ballo F., Previati G., Mastinu G., Gobbi M. , 2023. Eco-Driving Strategy Implementation for Ultra-Efficient Lightweight Electric Vehicles in Realistic Driving Scenarios, *Energies*, 16 (3).
- Van Der Laan, J. D., Heino, A. & De Waard, D., 1997. A simple procedure for the assessment of acceptance of advanced transport telematic. *Transportation Research Part C: Emerging Technologies*, 5(1), pp. 1-10.
- VI-Grade GmbH, 2025. *VI-DriveSim 2025.2 Documentation*.
- Xue, H. et al., 2025. Investigating comfort-oriented optimization of eco-driving cruising strategy: A driving simulator study. *Transportation Research Interdisciplinary Perspectives*, Volume 34:101733.

Physics-Informed Neural Network-based Motion Cueing Algorithm

Robert Jacumet^{1,2}, Simon Duell¹, Sebastian Wagner², Markus Schwienbacher², Marion Leibold¹

(1) Technical University of Munich, School of Computation, Information and Technology, Chair of Automatic Control Engineering, 82090 Munich, Germany, e-mail : {r.jacumet, simon.duell, marion.leibold}@tum.de

(2) BMW Group, Inventions, Innovation Management, 80788 Munich, Germany, e-mail : {robert.jacumet, sebastian.ws.wagner, markus.schwiebacher}@bmw.de

Abstract - Driving simulator motion systems rely on Motion Cueing Algorithms (MCAs) to generate realistic motion feedback. Current MCAs either require tuning for worst-case maneuvers, thus leading to conservative behavior and inefficient workspace use, or are computationally expensive and therefore struggle to meet real-time requirements in driver-in-the-loop simulations. To address these limitations, this work introduces a physics-informed neural network (PINN)-based MCA framework that, for the first time, explicitly integrates physical knowledge into the training process, combining fast inference with accurate modeling of complex system dynamics and actuator-level constraints. The developed framework combines three NNs, working jointly towards a high motion fidelity, as well as compliance with physical laws and simulator limits. To this end, the NNs learn an efficient approximation of the simulator motion system direct kinematics, a black box model of various aspects of human motion perception, as well as the desired motion control signals resulting from the simulator state, human motion perception, and the reference motion. The applicability of the framework is demonstrated using a large 9-degrees of freedom simulator.

Keywords: Learning-based Motion Cueing Algorithm, Driving Simulation, Physics-Informed Neural Networks

Copyright: © 2026 by the authors.

Licensee Driving Simulation Association.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.82157/dsa/2026/14>

1. Introduction

Driving simulators reduce the need for costly and demanding real-vehicle tests by offering controlled, repeatable, and risk-free experiments. Achieving this requires simulators to create a realistic experience that evokes responses comparable to those in an actual vehicle. To reproduce the motion sensations of real driving, simulators rely on Motion Cueing Algorithms (MCAs), which generate the controls for the motion system of the driving simulator. Due to the motion system being restricted in size, it is not possible to reproduce the vehicle dynamics directly, leading to mismatches between the resulting simulator motion and that of the vehicle. Such mismatches negatively affect the perceived realism for the driver (Bruschetta, Cenedese, and Beghi, 2019). The goal of the MCA is to compute a motion that is perceived as realistic as possible, given the simulator capabilities and limitations (Kolff, et al., 2024a).

Existing MCAs can be grouped into filter-based and optimization-based methods (Jacumet, et al., 2026). Filter-based methods process vehicle motion signals to obtain simulator control inputs by scaling and filtering, offering transparent behavior, low computational cost, and reliable real-time performance (Asadi, et al., 2019). However, because they do not explicitly consider simulator limits, they must be tuned for worst-case scenarios, which results in conservative

motion cues. Optimization-based MCAs, on the other hand, compute platform commands through online optimization, achieving better workspace exploitation and yielding higher cueing quality (Jacumet, et al., 2024). Their drawback is the heavy computational load, which makes real-time operation challenging for driver-in-the-loop studies (Scheidel, et al., 2024). Additionally, optimization-based MCAs heavily depend on the quality of the predicted vehicle motion reference. Incorrect predictions of the reference signal lead to the computation of unrealistic simulator motions.

Aiming to address these disadvantages, learning-based MCAs have emerged, where Neural Networks (NNs) are trained to act as an MCA. Koyuncu, et al. (2020) train a nonlinear autoregressive network to replicate an optimization-based MCA. The resulting fully NN-based MCA runs in real-time and improves cueing over a filter-based benchmark MCA. However, the approach does not generalize well when applied to scenarios not represented in the training data and, using supervised learning, the cueing quality is limited to that of the MCA used in training. Scheidel, et al. (2024) introduce a deep reinforcement learning-based MCA trained using a proximal policy optimization (PPO) actor-critic framework. Building on this, Huang, et al. (2025) improve the training speed by substituting PPO with an analytic policy-gradient

method. Arango, et al. (2025) train a neural controller through analytic policy gradients rather than reward-based exploration. Their method embeds the simulator dynamics directly in the training loop and optimizes a neural control policy offline, resulting in fast online inference. Despite the progress made in learning-based MCAs, the existing approaches handle the NNs solely as black box models. Physical knowledge on the simulator dynamics, human motion perception, and motion system limitations, is only infused into the training process by computing a reward or penalty that takes this knowledge into account. Therefore, learning an input-output mapping that respects physical constraints or platform dynamics is only indirectly encouraged. As a consequence, these methods may violate actuator limits, overlook dynamic feasibility, or generate motions that remain plausible only for the training distribution.

Physics-informed neural networks (PINNs) address this limitation by embedding physical structure into the learning problem itself, rather than treating it as an external evaluation criterion. Instead of relying on reward signals or penalty terms that indirectly incentivize physically plausible behavior, PINNs parameterize the solution space such that the governing differential equations, boundary conditions, and constraints are explicitly enforced during training. As a result, the network is optimized to satisfy both data consistency and physical feasibility, constraining the learned mapping to admissible system dynamics. This formulation reduces reliance on empirical reward design and promotes physically consistent behavior even in regions of the state space that are sparsely sampled or unseen during training.

To the authors' knowledge, this paper presents the first PINN-based MCA. The proposed framework combines three NNs. First, a NN is designed to predict the participant ratings of perceived motion incongruence. The novel learning-based approach is trained on a large dataset collected from multiple participant studies. These ratings are fed to a second NN, trained to compute the simulator control signals which optimize both the participant ratings and the tracking of the vehicle reference motion. The third NN then computes the direct kinematics of the simulator motion system, enabling a physics guard layer to enforce constraints on the actuator level. The applicability of the proposed framework was demonstrated in a participant study on a large 9-degree of freedom simulator with ten participants performing acceleration, braking, and double lane-change maneuvers.

This work is structured as follows: Section 2 presents the preliminaries, i.e., the necessary motion cueing and PINN concepts. The proposed PINN-based MCA is then presented in Sec. 3. An evaluation of the PINN-based MCA with regards to cueing quality, generalization, and runtime is performed in Sec. 4. A discussion is provided in Sec. 5 and the paper is concluded in Sec. 6.

2. Preliminaries

Before presenting the PINN-based MCA, this section provides the necessary motion cueing and subjective rating preliminaries (Jacumet, et al., 2026), followed by an introduction of PINNs.

2.1. Motion Cueing Preliminaries

Motivated by human motion perception, MCAs replicate the angular accelerations ω^V and specific forces f^V that occupants would experience in an actual vehicle. To this end, MCAs compute the motion system control inputs u^S such that the resulting simulator motion acting on the occupants, i.e., $y^S = [\omega^S f^S]^\top$, is perceived as similar as possible to that acting in the real vehicle $y^V = [\omega^V f^V]^\top$. Additionally, the control signals $u^S = [\beta^S a^S]^\top$, i.e., the rotational accelerations β^S and translational accelerations a^S , must ensure that the simulator stays within its limits. That is, the actuator positions q , actuator velocities $\dot{q}$, and actuator accelerations $\ddot{q}$ of each actuator of the motion system must stay within their respective constraint sets $\mathcal{Q}$, $\dot{\mathcal{Q}}$, and $\ddot{\mathcal{Q}}$.

The specific forces and angular velocities acting on the driver are computed from the simulator controls u^S and the simulator state

$$x^S = [\beta^S \ p^S \ \dot{\beta}^S \ v^S]^\top, \quad (1)$$

which contains the translational positions p^S , translational velocities v^S , rotational positions β^S , rotational velocities $\dot{\beta}^S$ of the simulator. To this end, the simulator state and control signals expressed in the inertial frame must be transformed into the body-fixed frame using the joint rotation matrix T . The specific force in the body fixed frame is then (Jacumet, et al., 2026)

$$f^S = T(\beta^S) (a^S + g). \quad (2)$$

Here, it is observed that the gravity vector g contributes to the specific force experienced by the driver, which is known as tilt-coordination. The exact expression (2) depends on the simulator motion system. The angular velocities acting on the driver are obtained from the current rotation β^S , its velocity $\dot{\beta}^S$, and the Jacobian matrix J as (Jacumet, et al., 2026)

$$\omega^S = J(\beta^S) \dot{\beta}^S. \quad (3)$$

Combined, the simulator motion is computed by

$$y^S = \begin{bmatrix} \omega^S \\ f^S \end{bmatrix} = f^y(u^S, x^S) = \begin{bmatrix} J(\beta^S) \dot{\beta}^S \\ T(\beta^S) (a^S + g) \end{bmatrix}. \quad (4)$$

The Cartesian workspace position, velocity, and acceleration are defined as $w = [\beta^S \ p^S]^\top$, $\dot{w} = [\dot{\beta}^S \ v^S]^\top$ and $\ddot{w} = [\ddot{\beta}^S \ a^S]^\top$, respectively. For constraining the simulator actuator space, the actuator positions q , velocities $\dot{q}$, and accelerations $\ddot{q}$ must be computed from these workspace values. In general, the mapping between joint space and workspace is nonlinear for complex motion systems such as a hexapod. Generally, their relation is described by

$$w = f(q), \quad (5a)$$

$$\dot{w} = \frac{\partial f(q)}{\partial q} \dot{q}, \quad (5b)$$

$$\ddot{w} = \frac{\partial f(q)}{\partial q} \ddot{q} + \dot{q}^\top \frac{\partial^2 f(q)}{\partial q^2} \dot{q}. \quad (5c)$$

Depending on the kinematics of the motion system at hand, the transformation from actuator space to workspace, i.e., solving (5) for q , $\dot{q}$, and $\ddot{q}$, is computationally expensive and requires time-intensive,

iterative methods such as the Newton-Raphson method (Ellensohn, et al., 2018). The transformation from actuator space to workspace is often referred to as “direct kinematics”. The hexapod is an example, where the direct kinematics is highly non-linear.

2.2. Participant Ratings of Realism

To assess the performance of an MCA, participants are tasked to provide a rating of their perceived motion realism during studies (Cleij, et al., 2018). As the ultimate goal of motion cueing is to create a motion that is perceived realistic, these rating values are an essential metric. Optimization-based MCAs are designed to minimize the tracking error $\|y^V - y^S\|_Q^2$, where Q is a weighting matrix defining the relative importance of the different motions channels in y . However, a low tracking error does not capture the complex human motion perception, i.e., a better tracking error does not necessarily coincide with an improved rating and vice versa (Cleij, et al., 2018; Jacumet, et al., 2025). This underlines the importance of participant ratings (Jacumet, et al., 2026).

To obtain the participants’ rating of realism with high time resolution, participants give their subjective feedback with a slider or dial, while experiencing a prerecorded vehicle and simulator motion. To obtain such ratings without running cumbersome user studies, previous work (Cleij, 2020; Cleij, et al., 2018; Ellensohn, et al., 2019) proposes models that predict the expected rating directly from the vehicle and simulator motion signals. However, these models showed to generalize insufficiently or are too complex to be used in MCAs. Hence, directly optimizing or training an MCA to achieve better modeled ratings of realism has not been done yet.

2.3. PINN Preliminaries

Deep learning, a subfield of machine learning and artificial intelligence, has gained attention in recent years due to its capacity to model complex, nonlinear relationships in high-dimensional data. In the supervised learning setting, NNs are trained on pairs of inputs and corresponding labels, allowing them to learn the underlying input-output relationships directly from data. During training, parameters of the NN are adjusted to minimize a loss function that quantifies the mismatch between the NN predictions and the labeled targets. This way, the model is able to learn patterns in the data and apply the learned computations to unseen data. Once trained, the NN performs inference, producing predictions efficiently by applying the learned representation to new inputs. To ensure that the model also performs well on unseen data, i.e., shows the ability to generalize, the training data needs to represent the complete input space, thus enabling the NN to learn all relevant underlying patterns (Aggarwal, 2018).

Many real-world systems exhibit complex dynamics governed by physical laws, which motivates extensions of deep learning that incorporate known physical structure. In physics-informed deep learning, observed data from a physical system is enriched with synthetically generated data obtained from analytical or numerical models of the governing physical relations, ensuring that the joint dataset reflects both the relevant sampling space and the underlying physical laws (Geng, et al., 2023). During training, the model implicitly learns these physical dependencies

through the enriched data structure (Mo, Shi, and Di, 2021). A more direct and explicit way of incorporating physical knowledge is offered by PINNs. Instead of relying solely on labeled data, PINNs embed the governing equations of the system, typically expressed as partial differential equations (PDEs), directly into the loss function (Raissi, Perdikaris, and Karniadakis, 2019). By penalizing deviations of the NN predictions from the governing equations during training, PINNs constrain the space of admissible network-representable solutions to remain consistent with the underlying physics and available observations. Therefore, PINNs generalize more reliably in settings where observational data are scarce or incomplete, while still capturing the essential dynamics dictated by the physics of the problem (Wei, Yin, and He, 2024).

An additional mechanism to incorporate physical constraints into NN predictions is the physics guard layer (PGL), introduced in (Chrosniak, Ning, and Behl, 2024). This approach ensures that the network outputs remain within a physically feasible range by constraining them through a parametrized activation function. The PGL applies a sigmoid transformation to the output of the last hidden layer, rescaling the output to lie between specified lower and upper bounds. With z denoting the output of the last hidden layer, $\sigma(\cdot)$ the sigmoid function, Φ the desired output parameter, and $\bar{\Phi}$ and $\underline{\Phi}$ the respective upper and lower bounds of this parameter, the PGL is

$$\Phi = \sigma(z)(\bar{\Phi} - \underline{\Phi}) + \underline{\Phi} = \frac{1}{1 + e^{-z}}(\bar{\Phi} - \underline{\Phi}) + \underline{\Phi}, \quad (6)$$

enforcing physically feasible predictions independent of the quality or amount of training data.

3. Methodology

The structure of the proposed PINN-based MCA is presented in Fig. 1. It consists of the simulator coordinate transformation, a PGL, and three distinct NNs jointly contributing to the goal of the PINN-based MCA, i.e., computing control signals u^S that provide a motion experience that is perceived as realistic. To this end, the loss function is designed as

$$\mathcal{L} = \|y^V - y^S\|_Q^2 + r, \quad (7)$$

expressing the realism of the simulator motion as a combination of the vehicle motion tracking $\|y^V - y^S\|_Q^2$ weighted by Q and the desire to achieve low modeled subjective ratings of motion incongruence r .

Before discussing the core concepts, implementation details, and training procedure of the PINN-based MCA in the following subsections, we summarize the PINN-based MCA framework. First, using the observed actuator position q_{t-1} and velocity $\dot{q}_{t-1}$ together with the vehicle motion y_t^V provided by the vehicle model at the current time t , the Actor NN presented in Subsec. 3.1 is trained to generate actuator accelerations $\ddot{q}_t \in \mathbb{R}^{n_{\text{DoF}}}$. Its objective is to minimize the loss in (7), that is, to produce acceleration commands that result in a high-quality motion experience. Due to the importance of producing control signals that lead to motion within the admissible actuator space, the Actor NN is followed by a PGL, presented in Subsec. 3.2. It processes the actuator accelerations $\ddot{q}_t$ provided by the Actor NN and

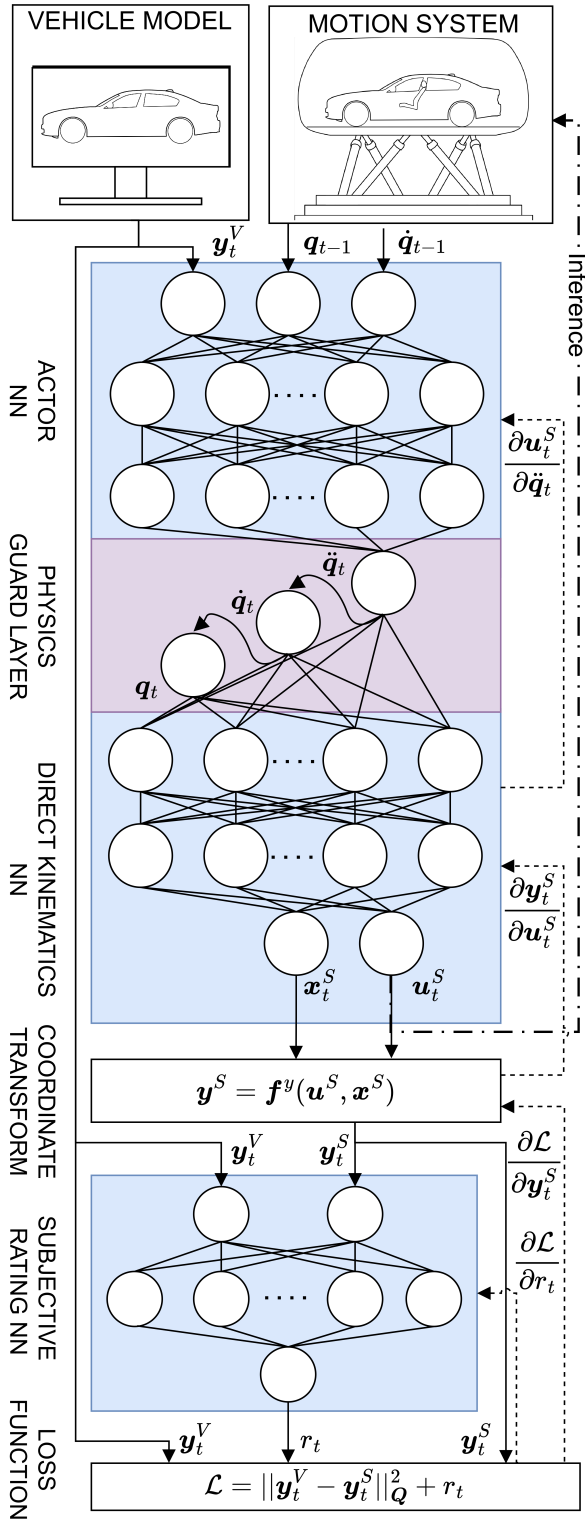


Figure 1: Structure of the PINN-based MCA.

enforces consistent and feasible actuator position q_t , velocity $\dot{q}_t$, and acceleration values $\ddot{q}_t$ at each step. For the PGL to work in the actuator space, necessary to enforce actuator constraints more effectively, the Actor NN is designed to operate directly in actuator space. This is a difference to other MCA formulations, where control signals are typically computed in the workspace, i.e., $u^S = [\ddot{\beta}^S, a^S]^\top$. To obtain the resulting simulator motion y^S , the coordinate trans-

formation (4) must be performed, requiring the simulator states x^S and controls u^S in workspace instead of actuator space. Hence, the direct kinematics are applied to q , $\dot{q}$, and $\ddot{q}$. As for typical simulator structures, computing the direct kinematics is a complex task with no closed form solution, leading to problems regarding real-time capability of the MCA (Ellensohn, et al., 2018), we pre-train a PINN to learn the direct kinematics of the simulator motion system in Subsec. 3.3. To obtain the predicted subjective rating values r for the resulting simulator motion y^S , that is needed in the loss function (7), Subsec. 3.4 presents a NN, pre-trained on a dataset of participant ratings. By backpropagating the loss to the actuator acceleration $\ddot{q}$, the Actor NN learns to generate high-quality control signals.

3.1. Actor NN

Training the Actor NN to infer actuator accelerations $\ddot{q}$ based on the current simulator position q_{t-1} , simulator velocity $\dot{q}_{t-1}$, and the reference vehicle motion y_t^V is the main objective and key challenge of the PINN-based MCA. To achieve this, the loss $\mathcal{L}$ is backpropagated all the way to the actuator controls, allowing the Actor NN to adjust its parameters so that the inferred accelerations produce motion with good rating values and low tracking errors.

Using the chain rule, the gradients of the loss w.r.t the simulator actuator controls $\ddot{q}_t$ are computed as

$$\frac{\partial \mathcal{L}}{\partial \ddot{q}_{t,j}^S} = \sum_{k=1}^m \left[\sum_{i=1}^n \left(\frac{\partial \mathcal{L}}{\partial y_{t,i}^S} + \frac{\partial \mathcal{L}}{\partial r_t} \frac{\partial r_t}{\partial y_{t,i}^S} \right) \frac{\partial y_{t,i}^S}{\partial u_{t,k}^S} \right] \frac{\partial u_{t,k}^S}{\partial \ddot{q}_{t,j}^S}. \quad (8)$$

for $i = 1, \dots, 6$, $k = 1, \dots, n_{\text{DoF}}$, and $j = 1, \dots, n_{\text{DoF}}$ during the backwards path. This equals a backpropagation of the loss through y^S and r to u^S , followed by a propagation through the direct kinematics to $\ddot{q}^S$. Using the Jacobian $J_y(u^S)$ of y^S w.r.t. u^S and $J_u(\ddot{q}^S)$ of u^S w.r.t. $\ddot{q}^S$, we rewrite the gradient of the loss w.r.t $\ddot{q}^S$ more compactly as

$$\nabla_{\ddot{q}^S} \mathcal{L} = J_u(\ddot{q}^S)^\top J_y(u^S)^\top \left[\nabla_y \mathcal{L} + \frac{\partial \mathcal{L}}{\partial r} \nabla_y r \right]. \quad (9)$$

The backpropagation is performed at each timestep t , as shown in Fig. 1. Such backpropagation is only possible because the entire pipeline, including the Rating NN, the PGL, and the Direct Kinematics NN is formulated as a fully differentiable computational graph. This ensures that gradients can be propagated through every component and that the resulting updates lead to dynamically consistent, feasible, and perceptually optimized control signals.

3.2. Physics Guard Layer

The PGL is an important component of the PINN-based MCA framework, ensuring safe operation of the MCA. We enforce that actuator positions q , velocities $\dot{q}$, and accelerations $\ddot{q}$, i.e., its PVA values, all stay within physically feasible ranges, i.e., that no actuator limitations of the simulator motion systems are violated at any timestep and under all possible inputs y^V . To this end, at each timestep, we compute a range $\ddot{Q}_{\text{PVA}} \in [\ddot{q}_{\text{PVA}}^{\min}, \ddot{q}_{\text{PVA}}^{\max}]$ such that for each value $\ddot{q} \in \ddot{Q}_{\text{PVA}}$ it is guaranteed that all actuator constraints of the simulator are respected, i.e.,

it holds that $\mathbf{q} \in [\mathbf{q}^{\min}, \mathbf{q}^{\max}]$, $\dot{\mathbf{q}} \in [\dot{\mathbf{q}}^{\min}, \dot{\mathbf{q}}^{\max}]$, and $\ddot{\mathbf{q}} \in [\ddot{\mathbf{q}}^{\min}, \ddot{\mathbf{q}}^{\max}]$. The dynamic upper and lower limits for the acceleration value of actuator i at the current time t , depending on its observed position $q_{i,t-1}$ and velocity $\dot{q}_{i,t-1}$ are

$$\ddot{q}_{PVA,i,t}^{\max} = \max\left(\min(\ddot{q}_i^{\max}, \xi_{\dot{q},t}^{\max}, \xi_{q,t}^{\max}), \ddot{q}_i^{\min}\right), \quad (10)$$

$$\ddot{q}_{PVA,i,t}^{\min} = \min\left(\max(\ddot{q}_i^{\min}, \xi_{\dot{q},t}^{\min}, \xi_{q,t}^{\min}), \ddot{q}_i^{\max}\right). \quad (11)$$

with the abbreviations

$$\xi_{\dot{q},t}^{\kappa} = \frac{\dot{q}_i^{\kappa} - \dot{q}_{i,t-1}}{T_s}, \quad \xi_{q,t}^{\kappa} = c_3 \left(-c_1 + \frac{\sqrt{c_1^2 - 4T_s^2 c_2}}{2T_s^2} \right),$$

$$c_1 = 2T_s \dot{q}_{i,t-1} - T_s^2 \ddot{q}_i^{\bar{\kappa}},$$

$$c_2 = \dot{q}_{i,t-1}^2 + 2\ddot{q}_i^{\bar{\kappa}}(q_i^{\bar{\kappa}} - q_{i,t-1} - \dot{q}_{i,t-1}T_s),$$

and $\kappa \in \{\max, \min\}$, $\bar{\kappa}$ the respective other option, and $c_3 = 1$ if $\kappa = \max$ and $c_3 = -1$ if $\kappa = \min$. Then

$$\ddot{\mathbf{q}}_t = \sigma(\ddot{\mathbf{q}}_t)(\ddot{q}_{PVA,t}^{\max} - \ddot{q}_{PVA,t}^{\min}) + \ddot{q}_{PVA,t}^{\min}, \quad (12)$$

and the new actuator velocity and position values are

$$\dot{\mathbf{q}}_t = \dot{\mathbf{q}}_{t-1} + \ddot{\mathbf{q}}_t T_s, \quad (13)$$

$$\mathbf{q}_t = \mathbf{q}_{t-1} + \dot{\mathbf{q}}_{t-1} T_s + \ddot{\mathbf{q}}_t \frac{T_s^2}{2}. \quad (14)$$

These PGL signals serve as inputs to the Direct Kinematics NN. The resulting workspace quantities are then used in subsequent steps to compute the simulator motion $\mathbf{y}^S$, the modeled subjective rating r , and the control signals $\mathbf{u}^S$ in workspace.

3.3. Direct Kinematics NN

For typical parallel simulator configurations, determining the current pose in Cartesian coordinates $\mathbf{w}$, $\dot{\mathbf{w}}$, and $\ddot{\mathbf{w}}$ from the actuator states $\mathbf{q}$, $\dot{\mathbf{q}}$, and controls $\ddot{\mathbf{q}}$ is a challenging direct kinematics problem for which no closed-form solution is generally available. To capture these nonlinear dependencies and to avoid repeatedly solving computationally expensive iterative schemes, we pre-train a PINN to approximate the direct kinematics of the simulator motion system.

To this end, training data are generated using an iterative Newton-Raphson solver, which computes the workspace pose corresponding to a given actuator state. The Newton-Raphson procedure, together with the underlying formulation of the direct kinematics problem is used to infer physical knowledge into the training process. To ensure that the provided data covers the whole actuator space, an n_{DoF} -dimensional, equidistant grid in actuator space is constructed and grid points are used as input samples. Instead of training the network solely to minimize the discrepancy between its outputs $(\cdot)^{\text{NN}}$ and the provided reference data, we incorporate physical constraints directly into the loss function, becoming

$$\mathcal{L}_{\text{DKNN}} = \mathcal{L}_{\text{pos}} + \mathcal{L}_{\text{vel}} + \mathcal{L}_{\text{acc}}, \quad (15)$$

with

$$\mathcal{L}_{\text{pos}} = \|\mathbf{f}(\mathbf{q}) - \mathbf{w}^{\text{NN}}\|^2, \quad (16)$$

$$\mathcal{L}_{\text{vel}} = \left\| \left(\frac{\partial \mathbf{f}(\mathbf{q})}{\partial \mathbf{q}} \dot{\mathbf{q}} \right) - \dot{\mathbf{w}}^{\text{NN}} \right\|^2, \quad (17)$$

$$\mathcal{L}_{\text{acc}} = \left\| \left(\frac{\partial \mathbf{f}(\mathbf{q})}{\partial \mathbf{q}} \ddot{\mathbf{q}} + \dot{\mathbf{q}}^\top \frac{\partial^2 \mathbf{f}(\mathbf{q})}{\partial \mathbf{q}^2} \dot{\mathbf{q}} \right) - \ddot{\mathbf{w}}^{\text{NN}} \right\|^2. \quad (18)$$

This physics-informed design encourages the network not only to match the training samples but also to satisfy the governing geometric and kinematic relationships of the simulator.

3.4. Subjective Rating NN

Since human motion perception is complex, the perceived realism of a reproduced vehicle motion cannot be assessed solely through a simple mismatch measure such as $\|\mathbf{y}^V - \mathbf{y}^S\|_Q^2$ (Jacumet, et al., 2025). It is therefore desirable to employ more holistic metrics that are able to evaluate the control inputs of the simulator before they are actually applied (Cleij, 2020; Kolff, et al., 2024a). A pure motion-mismatch measure fails to account for cueing error types such as false or missing cues, phase and shape errors, or discrepancies in the frequency content of the reproduced motion (Kolff, et al., 2025; 2024b). Existing MCAs do not incorporate perceptual models that extend beyond basic vestibular representations, and therefore do not capture these broader aspects of perceived motion cueing quality.

As the level of motion realism perceived by participants is the ultimate indicator of cueing quality, we employ machine-learning methods to predict a driver's subjective rating of realism of the simulated motion. A learning-based approach is able to capture patterns and dependencies in rating behavior that are not represented by the purely model-based approaches in (Cleij, 2020; Kolff, et al., 2024a). Furthermore, using a NN to predict subjective ratings of realism also enables automatic differentiation, allowing gradients to be propagated from the outputs of the Subjective Rating NN back to the Actor NN. Doing so ensures that the parameters of the Actor NN are systematically updated to improve the predicted rating value r . Similar to the models for subjective ratings, the Subjective Rating NN takes the reference motion $\mathbf{y}_t^V$ together with the specific forces and angular velocities $\mathbf{y}_t^S$ as inputs and predicts a rating r_t reflecting the perceived realism.

To pre-train the Subjective Rating NN, data from three previous participant studies is gathered. During each study, participants gave a continuous rating of perceived realism between 0 and 10. The first study featured 52 participants experience three different simulator motions in an urban region, with each motion being played three times and for four minutes. A second study featured 28 participants experiencing three motion profiles twice for seven minutes each, where a complex highway scenario was performed. Study three had 40 participants drive through a rural area for four minutes, with two simulator motion profiles being evaluated twice. All rating data is sampled at 10 ms. The test data is obtained from a fourth study conducted with 40 participants, each performing four drives in a different rural area, with two different motion profiles that both lasted three minutes.



Figure 2: The 9-DoF simulator used in the study.

A fully differentiable NN architecture is adopted to emulate the behavior of a random forest, with a small NN trained for each tree to approximate its leaf assignments and corresponding output values. Each tree-NN learns a soft, differentiable representation of the original decision boundaries, and the final prediction is obtained by combining the learned leaf probabilities with the leaf values of the pre-fitted forest. In this way, the model preserves the strong predictive performance of tree-based methods while remaining fully compatible with backpropagation.

4. Evaluation

We evaluate the performance of the proposed PINN-based MCA with regards to motion cueing quality, its ability to generalize, and the achieved runtime.

4.1. Simulator Study Evaluation

In order to assess the motion cueing performance of the proposed PINN-based MCA (MCA_{PINN}), the algorithm is compared to a benchmark filter-based MCA (MCA_{FB}) (Ellensohn, 2019). To check how the PINN-based MCA is able to work with the more complex case of a redundant motion system, BMW's 9-DoF Sapphire Space simulator (Jacumet, et al., 2026) is used. This simulator consists of a yaw table, hexapod, and X - Y rails mounted on top of each other. A picture of the simulator is shown in Fig. 2. We compare the MCAs in three typical driving simulation maneuvers, being an acceleration from 0 to 75 km/h, a braking from 75 km/h to standstill, and a double lane change at 70 km/h. All three maneuvers were set up on a four-lane endless highway shown in Fig. 3, depicting also the double lane change.

For the study, 10 participants experienced each maneuver four times, twice with each algorithm. While the sequence of the maneuvers stayed constant for all participants and the participants knew in advance which maneuver they would drive, the algorithm was

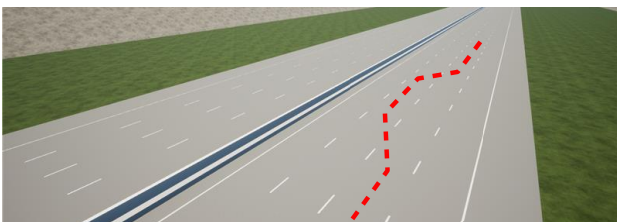


Figure 3: Endless highway with the trajectory of the double lane change.

unknown to the participants. The MCA sequence was determined by a Latin Square design to rule out learning and habituation effects. The mean age of all participants was 28 years, participants drove 5790 km annually on average, and the median number of previous simulator drives is 1. Three expert drivers participated, each having performed over 300 previous drives. At the end of each drive, the participants were asked to submit an overall rating on a scale of 0 to 10, where a value of 0 means that the motion was perfectly realistic and a value of 10 corresponds to a completely unrealistic motion experience.

First, Fig. 4 shows the training loss of the Direct Kinematics NN with two hidden layers, each with 64 neurons and ReLU activation functions for a hexapod system, representing the most common driving simulator configuration.

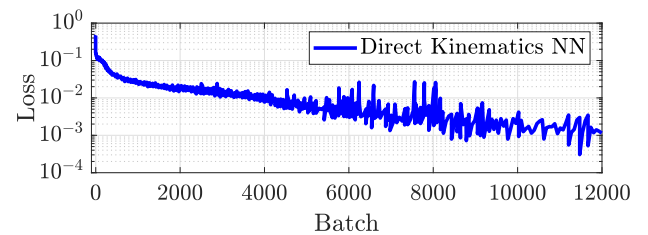


Figure 4: Training loss of the Direct Kinematics NN.

The actuator space is sampled at over 64 million points, followed by a linear interpolation, allowing the network to learn a highly accurate representation of the direct kinematics. It is observed that a fast decrease in loss occurs the initial 400 batches, followed by a slower decrease until a mean squared error of 0.0026, which can be interpreted as an error of less than 5 cm in the workspace position of all axes combined. The model therefore captures the direct kinematics with high accuracy and is suitable for the application in the loss computation for the Actor NN.

Second, Fig. 5 shows the predictions of the NN regressor compared to the ground-truth ratings. Although the network exhibits noticeable deviations, particularly around local highs and lows, it still reproduces the general rating trends for a different test drive. While the overall accuracy remains limited, the model captures perceptual tendencies that are not reflected in the reference-tracking error alone and thus serves as a useful additional component in the loss function.

Building on the pre-trained direct kinematics PINN and the pre-trained Subjective Rating NN, a separate Actor NN is trained for each maneuver, with the corresponding vehicle motion serving as training data. The

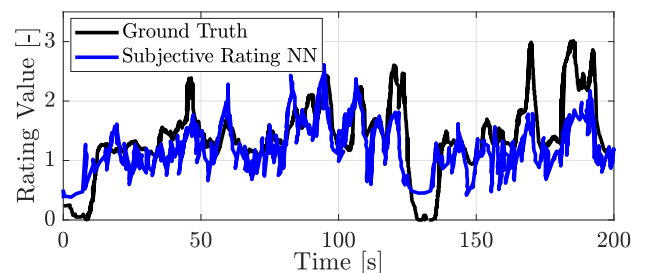


Figure 5: Subjective Rating NN test results.

forward pass was executed according to (9). Training is performed using the Adam optimizer with a learning rate of 10^{-3} for 3000 epochs. The entries of the diagonal matrix Q highlights the importance of the motion channels in the recreation of different maneuvers. The chosen values for all three maneuvers are shown in Tab. 1. The Actor NN was designed as a feed-forward NN with 2 hidden layers and 32 neurons each. A LeakyReLU was chosen as the activation function for the input and hidden layers.

The resulting motion for all three maneuvers and both MCAs is shown in Fig. 6, Fig. 7, and Fig. 8 for the respective relevant specific forces and angular velocities. It is observed that the PINN-based MCA recreates the specific forces f_x^V and f_y^V much better than the filter-based benchmark MCA. Besides an increased lateral hexapod and rail translation, more tilt-coordination is used, leading to the deviations observed from the vehicle reference motion in the ω_y motion channel. However, all rotational deviations are below 6 deg/s. The Root Mean Square Error (RMSE) for the three most important motion channels regarding human motion perception are provided in Tab. 2, where these coincide with the observations made in Figs. 6 – 8. The achieved participant ratings for all maneuvers and both MCAs are presented in Tab. 3. It is observed that participants rated the perceived motion realism nearly equal for the double lane change, whereas the filter-based

Table 1: Motion Channel Weight Parameters Q .

Weight	Acceleration	Braking	Lane Change
ω_x	50	50	50
ω_y	1	1	50
ω_z	150	150	15
f_x	40	40	50
f_y	150	150	50
f_z	20	20	20

Table 2: RMSE Comparison between the two MCAs.

PINN/CWA	Acceleration	Braking	Lane Change
f_x	0.10 / 0.43	0.07 / 0.92	0.03 / 0.04
f_y	0.02 / 0.00	0.01 / 0.00	0.04 / 0.32
ω_z	0.02 / 0.00	0.01 / 0.00	0.20 / 0.06

Table 3: Rating Comparison between the two MCAs.

	Mean		Standard Dev.	
	MCA _{PINN}	MCA _{FB}	MCA _{PINN}	MCA _{FB}
Acceleration	2.45	3.40	1.63	1.43
Braking	4.05	2.95	1.56	1.69
Lane Change	2.20	2.30	1.93	1.58

MCA achieved better (lower) subjective ratings when braking, and the PINN-based MCA achieved superior ratings during the acceleration. Participants verbally reported to negatively perceive the pitch onset produced by the PINN-based MCA, being identified as the main cause for the worse rating. The PINN-based MCA pre-tilts the simulator in anticipation of the upcoming deceleration, which participants perceived negatively.

4.2. Generalization Evaluation

We evaluate how the PINN-based MCA generalizes to different sampling times, scalings of the reference signal, as well as different reference motions.

For evaluating how well the PINN-based MCA deals with both changes in the sampling time and its reference, we train the PINN on data scaled by a factor of 0.3 and sampled at 50 ms, but provide the unscaled reference motion with a sampling of 2 ms to the PINN during deployment. The unscaled reference motion and the resulting simulator motion for the f_x channel is depicted in Fig. 9 for the acceleration maneuver. It is observed that the PINN-based MCA succeeds in this aspect of generalization, allowing a faster training of the PINN by subsampling the training data. To assess the applicability of the learned model over a broader range of motions, the reference signal is scaled by a factor of -1 relative to the training data. As shown in Fig. 10, the resulting motion during the acceleration maneuver exhibits more pronounced deviations, yet the PINN-based MCA still produces a braking response, despite having been trained exclusively on acceleration data. Last, the generalization of the PINN-based MCA is evaluated by adding random noise to the input data during inference. A noisy reference is created by superimposing a zero-mean, normally distributed signal with a standard deviation of 0.5 onto the normalized data (zero mean, unit standard deviation). As shown in Fig. 11, the PINN-based MCA still manages to reproduce the reference motion, despite these perturbations.

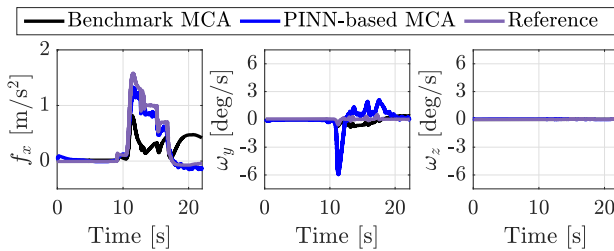


Figure 6: Comparison of the motions in the acceleration scenario achieved by both MCAs.

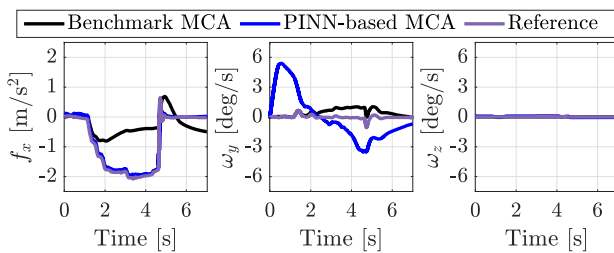


Figure 7: Comparison of the motions in the deceleration scenario achieved by both MCAs.

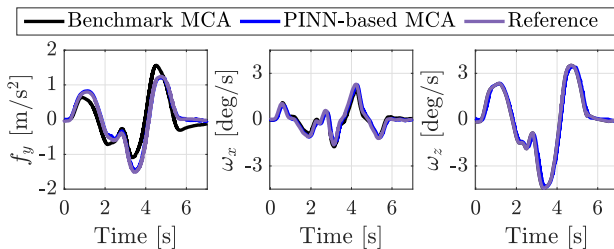


Figure 8: Comparison of the motions in the double lane change scenario achieved by both MCAs.

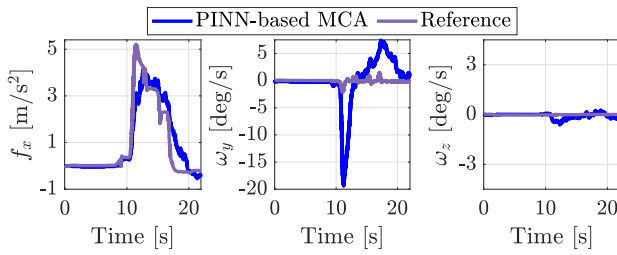


Figure 9: Performance of the PINN-based MCA in the acceleration scenario when trained with different sampling time and scaling factor.

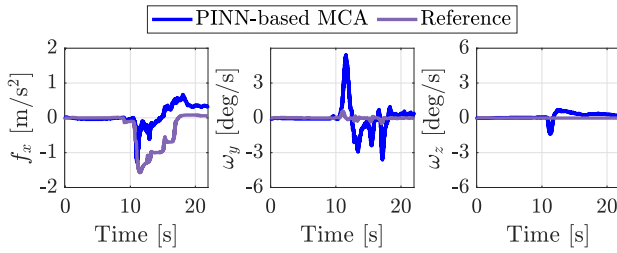


Figure 10: Performance of the PINN-based MCA in the acceleration scenario with inverted motion during inference compared to the training.

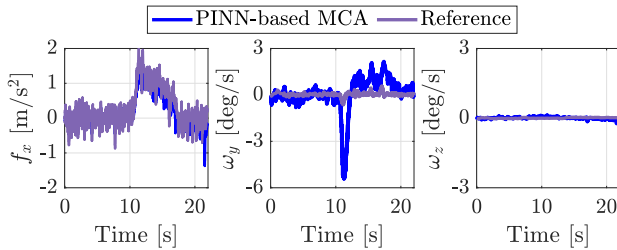


Figure 11: Performance of the PINN-based MCA in the acceleration scenario with noise during inference.

After conducting the study, the applicability of the PINN-based MCA to a whole drive of nearly 4 minutes was evaluated. The chosen drive takes part in a rural setting and includes a 90° turn, various dynamic curves, braking and acceleration, as well as a roundabout in the end. The results after training for 5000 epochs are shown in Fig. 12 for f_x , f_y , and ω_z . The results show that the PINN provides simulator controls that recreate the reference motion with high accuracy in the three most important motion channels. Hence, the PINN-based MCA demonstrates applicability to longer simulated drives that encompass a broad range of representative maneuvers. When using the PINN trained for the whole drive to perform the double lane change maneuver used in the study, i.e., the only maneuver from the participant study not included in the training data for the long drive, the motion shown in Fig. 13 is obtained. It is observed that the PINN manages to recreate the new double lane change maneuver well.

4.3. Runtime Evaluation

The NNs were trained and run on a Linux workstation using an Intel(R) Xeon(R) @ 2.00 GHz CPU and an NVIDIA RTX 5000 Ada GPU. The models used in the study were trained for 3000 Epochs, taking approximately 90 minutes. Computing the forward path for the direct kinematics NN takes on average 0.07 ms. The inference time of the complete

PINN-based MCA is 0.9 ms. These inference times are significantly shorter than the computation times of optimization-based MCAs, marking a clear advantage. The short training time enables retraining the PINN for different studies, which is crucial for its practical deployment.

5. Discussion

Overall, the evaluation results demonstrate that integrating perceptual modeling, physics-informed learning, and actuator-level constraint handling is a promising approach. In this section, we discuss limitations with regards to the participant study, the role of incorporating further available knowledge into the PINN-based MCA, the rating NN performance, and further use cases of the trained NNs.

First, the simulator study has a comparably small number of 10 participants, which prevents statistically significant conclusions and limits the generalizability of the perceived motion cueing quality. Nevertheless, the obtained participant ratings and observed motions provide a valuable first indication of its potential and highlight which aspects warrant further investigation in larger studies.

While during training, the PINN learns how to work with a series of simulator and vehicle motions, knowledge on future simulator states and future vehicle motions is not incorporated so far. Designing the PINN to incorporate predictions of the future simulator and vehicle motion would improve the motion cueing quality and generalization of the PINN. However, in its current form, especially with the incorporation of a PGL, the PINN already shows able to provide a good cueing for various maneuvers and acceptable generalization capabilities.

Initial experiments with feedforward NNs, Long Short-Term Memory, and random forest regressors showed that tree-based models captured the rating dynamics most accurately, consistently outperforming standard NN architectures. However, because random forests rely on non-differentiable decision nodes, it is not possible to integrate them into a gradient-based training pipeline. This motivated the design of the custom, fully differentiable NN architecture (see Subsec. 3.4), trained to imitate a random forest regressor. Because a random forest predicts by averaging training samples that fall into the same learned classification regions, its performance is strongly tied to the coverage of the training data. To improve the generalization of the rating NN, the training set should therefore capture the full diversity of scenarios, maneuvers, and participant responses. With the current training, the model captures perceptual tendencies that extend beyond reference tracking, making it a beneficial, however imperfect addition to the loss function. Having a more accurate rating prediction would then favor to shift focus from minimizing the tracking error $\|y^V - y^S\|_Q^2$ to minimizing the modeled rating r .

Last, both the Direct Kinematics NN and the Rating NN designed in this work are not exclusively relevant for a PINN-based MCA, but are also of interest to other types of MCAs. The very low inference times of 0.07 ms achieved by the direct kinematics NN allow to speed up MPC-based MCAs, as well as the formulation of optimization problems fully in actuator space. So far, this required approximations and simplification of the direct kinematics (Ellensohn, et al.,

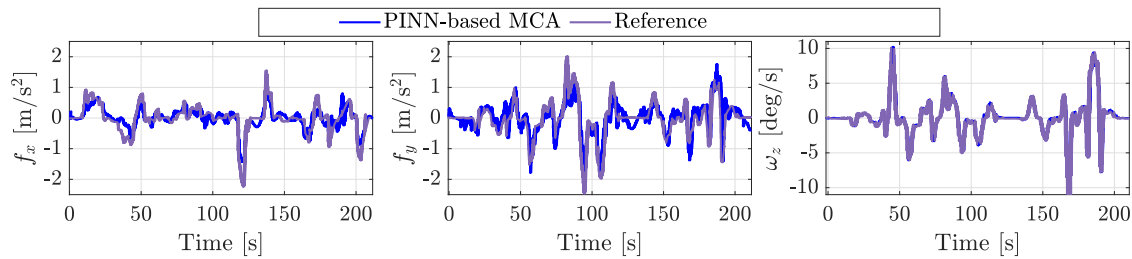


Figure 12: Performance of the PINN-based MCA for a whole vehicle drive.

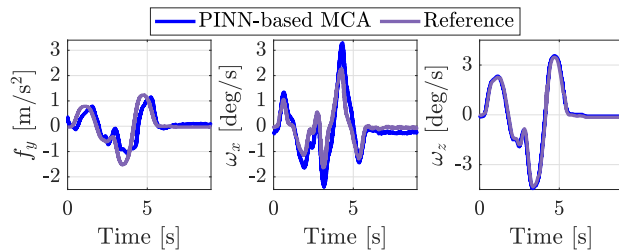


Figure 13: Performance of the PINN-based MCA in a double lane change scenario, where the PINN is trained on the whole vehicle drive which does not contain a double lane change.

2018), posing as the bottleneck of such formulations. Similarly, the Rating NN is of interest for use in other optimization-based MCAs, due to its ability to further put the human at the center of cueing algorithms. Its ability to capture the complexity of human motion perception without the need for specific models eases up applicability and helps to directly optimize for improved ratings given by participants.

6. Conclusion

This work introduced a PINN-based MCA that combines subjective rating prediction, physics-informed direct kinematics modeling, and constraint-aware control generation into a differentiable architecture suitable for real-time application. By embedding domain knowledge through the PINN and a PGL, safe operation and constraint adherence is enabled, while a Rating NN allows the Actor NN controller to consider perceptual factors beyond a pure tracking error. The evaluation on a 9-DoF driving simulator shows that the PINN-based MCA reproduces vehicle motions with high accuracy and generalizes across sampling times, large scaling variations, noisy inputs, and unseen maneuvers. A conducted study indicates that participants perceive the motion fidelity similar to that of a benchmark filter-based MCA. Hence, the framework demonstrates the feasibility and potential of embedding physics and human perception modeling into learning-based MCAs. Future work focuses on advancing the predictive capabilities of the PINN-based MCA and enlarging the training dataset of the Rating NN to further enhance practical applicability and motion cueing fidelity of the PINN-based MCA.

References

- Aggarwal, C. C., 2018. *Neural Networks and Deep Learning: A Textbook*. Cham: Springer International Publishing.
- Arango, C. G., Asadi, H., Qazani, M. R. C., and Lim, C. P., 2025. *Learning-based approximate nonlinear model predictive control motion cueing*. arXiv: 2504.00469 [cs.RO].
- Asadi, H., Lim, C. P., Mohamed, S., Nahavandi, D., and Nahavandi, S., 2019. Increasing motion fidelity in driving simulators using a fuzzy-based washout filter. *IEEE Trans. Intell. Veh.*, 4(2), pp. 298–308.
- Bruschetta, M., Cenedese, C., and Beghi, A., 2019. A real-time, MPC-based Motion Cueing Algorithm with Look-Ahead and driver characterization. *Transp. Res. F: Traffic Psychol. Behav.*, 61, pp. 38–52.
- Chrosniak, J., Ning, J., and Behl, M., June 2024. Deep Dynamics: Vehicle Dynamics Modeling With a Physics-Constrained Neural Network for Autonomous Racing. *IEEE Robotics and Automation Letters*, 9(6), pp. 5292–5297.
- Cleij, D., 2020. Measuring, Modelling and Minimizing Perceived Motion Incongruence. PhD. Delft University of Technology, Faculty of Aerospace Engineering.
- Cleij, D., Venrooij, J., Pretto, P., Pool, D. M., Mulder, M., and Buelthoff, H. H., 2018. Continuous Subjective Rating of Perceived Motion Incongruence During Driving Simulation. *IEEE Transactions on Human-Machine Systems*, 48(1), pp. 17–29.
- Ellensohn, F., 2019. Urban Motion Cueing Algorithms. PhD. Munich: Technical University of Munich.
- Ellensohn, F., Oberleitner, F., Schwenbacher, M., Venrooij, J., and Rixen, D., July 2018. Actuator-Based Optimization Motion Cueing Algorithm. In: *2018 IEEE/ASME Int. Conf. Adv. Intell. Mechatron. (AIM)*, pp. 1021–1026.
- Ellensohn, F., Venrooij, J., Schwenbacher, M., and Rixen, D., 2019. Experimental Evaluation of an Optimization-Based Motion Cueing Algorithm. *Transp. Res. F: Traffic Psychol. Behav.*, 62, pp. 115–125.
- Geng, M., Li, J., Xia, Y., and Chen, X. M., Sept. 2023. A physics-informed Transformer model for vehicle trajectory prediction on highways. *Transp. Res. Part C Emerg. Technol.*, 154.
- Huang, X., Shi, X., Wang, P., Xu, H., Tang, X., and Zhang, G., 2025. An analytic policy gradient-based deep reinforcement learning motion cueing algorithm for driving simulators. *IEEE access*, 13, pp. 81507–81523.
- Jacumet, R., Wagner, S., Schwenbacher, M., Wollherr, D., and Leibold, M., 2024. Sequential Motion Cueing Algorithms: Using The Best Algorithm in Each Situation. In: *Proc. Driv. Simul. Conf. Europe VR*. Strasbourg, France, pp. 111–117.
- Jacumet, R., Wetzel, J., Wagner, S., Schwenbacher, M., Wollherr, D., and Leibold, M., 2025. Optimization-based Motion Cueing Algorithm for Worst Case Maneuvers. en. In: *Proc. Driv. Simul. Conf. Europe VR*. Stuttgart, Germany.
- Jacumet, R., et al., 2026. A Review on Optimization-based Motion Cueing Algorithms in Driving Simulation. *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–25.
- Kolff, M., Jacumet, R., Wagner, S., Wollherr, D., and Leibold, M., 2025. A Frequency-Aware Model Predictive Control Motion Cueing Algorithm. In: *Proc. Driv. Simul. Conf. Eur. XR*. Stuttgart, Germany.
- Kolff, M., Venrooij, J., Schwenbacher, M., Pool, D. M., and Mulder, M., 2024a. Reliability and Models of Subjective Motion Incongruence Ratings in Urban Driving Simulations. *IEEE Trans. on Human-Mach. Syst.*, pp. 1–12.
- Kolff, M., et al., 2024b. Motion Cueing in BMW's Driving Simulation Center: Experiences versus Common Knowledge. In: *Proc. Driv. Simul. Conf. Europe VR*. Strasbourg, France, pp. 119–125.
- Koyuncu, A. B., Ercelik, E., Comulada-Simpson, E., Venrooij, J., Kaboli, M., and Knoll, A., Oct. 2020. A Novel Approach to Neural Network-based Motion Cueing Algorithm for a Driving Simulator. In: *2020 IEEE Intell. Veh. Symp. (IV)*. Las Vegas, NV, USA: IEEE, pp. 2118–2125.

- Mo, Z., Shi, R., and Di, X., Sept. 2021. A physics-informed deep learning paradigm for car-following models. *Transportation Research Part C: Emerging Technologies*, 130.
- Raissi, M., Perdikaris, P., and Karniadakis, G., 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.*, 378, pp. 686–707.
- Scheidel, H., Asadi, H., Bellmann, T., Seefried, A., Mohamed, S., and Nahavandi, S., 2024. A deep reinforcement learning based motion cueing algorithm for vehicle driving simulation. *IEEE Trans. Veh. Technol.*, 73(7), pp. 9696–9705.
- Wei, W., Yin, G., and He, T., Jan. 2024. Physics-Informed Data-Based LPV Modeling and Validations of Lateral Vehicle Dynamics. *IEEE Trans. on Intell. Veh.*, 9(1), pp. 2459–2468.

Enhancing Longitudinal Motion Cueing via AI-Based Driver Behavior Prediction

Z. Fang¹, F. de Oliveira¹, D.-C. Qiu², M. Sommervogel³, F. Colombet¹, S. Régnier¹

(1) Renault TCR, France, {zhou.fang, fabien.de-oliveira, florent.colombet, stephane.regnier}@renault.com

(2) UTC Compiègne France, dongcheng.qiu@etu.utc.fr

(3) MPDATA, Boulogne-Billancourt, France, matthieu.sommervogel-extern@renault.com

Abstract – Renault's high-performance driving simulator, ROADS, is an advanced tool for vehicle development that utilizes virtual proving grounds to evaluate dynamic performance. While lateral motion perception has reached high fidelity through an internal Model Predictive Control-based Motion Cueing Algorithm (MPC-MCA), longitudinal performance remains limited by the unpredictable nature of driver behavior. This paper investigates various AI architectures, specifically LSTM, Embedding LSTM, and Informer, to enhance longitudinal predictive performance within the ROADS real-time platform. By forecasting future driver actions, the MPC-MCA optimizes platform pre-positioning and workspace management. Preliminary numerical results demonstrate a 20% increase in motion restitution amplitude of longitudinal acceleration, significantly improved workspace utilization, and lower required tilt velocity.

Keywords: Motion cueing algorithm, ROADS, AI-based predictive MPC, driver behavior prediction.

Copyright: © 2026 by the authors.

Licensee Driving Simulation Association.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.82157/dsa/2026/15>

1. Introduction

As digital transformation accelerates within Renault's engineering workflows, immersive simulation has emerged as a cornerstone of vehicle development. By integrating the human element at the core of the design cycle, a methodology known as Driver-in-the-Loop (DiL), it is now possible to evaluate complex dynamic behaviors from the earliest conceptual phases. This approach allows for a precise analysis of the interactions between the driver, the chassis, powertrains (thermal or electric), and Advanced Driver Assistance Systems (ADAS), relying on high-fidelity numerical models rather than physical prototypes.

To support this ambition, Renault developed ROADS (Renault Operational Advanced Driving Simulator) in 2023. Designed as a validation laboratory for vehicle safety, handling performance, ADAS pre-tuning and driving "DNA", ROADS aims to converge objective engineering data with subjective driver perception.

The facility achieves technical excellence through a massive 15 m x 15 m workspace and a 92-ton mobile mass, providing full 360° visual immersion. Its 9-Degrees-of-Freedom (9DoF) kinematic architecture, comprising a yaw table, a hexapod, and X-Y longitudinal and lateral rails, enables high-dynamic driving experience with extreme precision.

Table 1 illustrates the ROADS performances.

Operating at a 1000 Hz numerical simulation processing frequency, the simulator with low latency (<30ms) replicates real-world proving grounds with ultra-precise 3D environments, as illustrated in the ROADS ecosystem (Fig. 1).

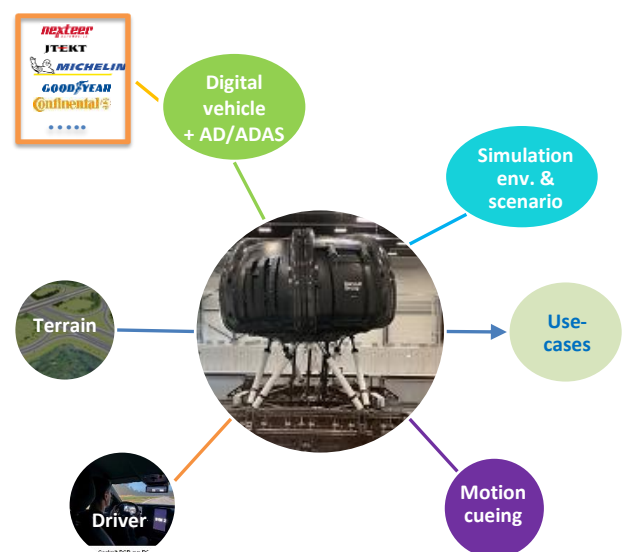


Figure 1: The ROADS ecosystem: integration of vehicle dynamics, environmental simulation, and motion cueing

Table 1: Performance of ROADS

	Rail						Hexapod				Yaw table	
	length (m)		velocity (m/s)		accel. (m/s ²)		length (m)	accel (m/s ²)	angle (°)	velocity (°/s)	angle (°)	velocity (°/s)
	X	Y	X	Y	X	Y	Z	Z	Pitch, Roll	Pitch, Roll	Yaw	Yaw
	±7.5	±7.5	±9	±9	±10	±10	±0.68	±12	±23	±65	±165	±60

Despite these advanced capabilities, a persistent challenge remains in the fidelity of longitudinal motion cueing. While lateral motion rendering is highly rated by professional drivers, largely because it is constrained by road curvature and tire grip, longitudinal rendering often fails to accurately replicate heavy braking or rapid acceleration.

The root of this discrepancy lies in predictability. Standard Model Predictive Control-based Motion Cueing Algorithms (MPC-MCA) struggle with longitudinal dynamics because they are heavily dependent on the driver's inherently unpredictable behavior. Unlike lateral forces, which can be anticipated via path geometry, longitudinal inputs are reactive and subjective.

Recent literature has explored Artificial Intelligence to enhance Motion Cueing Algorithms [1–5]. However, most existing research is limited to single-driver datasets or simplified proofs-of-concept.

This study introduces a generalized AI-based MCA trained on multi-driver datasets, specifically tailored for high-dynamic scenarios typical of proving-ground tests. By integrating an AI-based prediction of longitudinal acceleration (AccX) into the MPC framework, the algorithm can "anticipate" driver intent. This allows for optimized platform pre-positioning and more efficient workspace management. The following sections detail the end-to-end process, from multi-driver training to real-time deployment of AI based MPC-MCA on the ROADS simulator.

II. AI model architectures & training dataset

To enhance longitudinal motion cueing with MPC-MCA, the primary requirement is a model capable of forecasting driver inputs based on historical vehicle and environmental data. This study focuses on predicting longitudinal acceleration by investigating two state-of-the-art architectures for time-series forecasting: Long Short-Term Memory (LSTM) and the Informer.

II.1. Model Descriptions

A. Long Short-Term Memory (LSTM)

As a specialized Recurrent Neural Network (RNN), the LSTM (Hochreiter et al. 1997) is designed to mitigate the vanishing gradient problem. It utilizes a gating mechanism (input, forget, and output gates) to maintain a "cell state," allowing the model to

capture long-term dependencies in driving dynamics sequences. In our context, the LSTM processes a sliding window of historical sequences, including pedal positions, vehicle acceleration, velocity and position, to extrapolate the AccX signal.

B. Informer Architecture

The Informer (Zhou et al. 2021) is a Transformer-based model specifically optimized for long-sequence time-series forecasting. It introduces a ProbSparse Self-Attention mechanism, which significantly reduces computational complexity from $O(L^2)$ to $O(L \log(L))$. For the ROADS platform, this efficiency is critical for real-time integration, enabling the model to identify relevant historical "driving motifs" to predict high-dynamic events, such as heavy braking.

II.2. Input Features and Objective

The models are fed a multidimensional input vector X_t representing the vehicle's state at time t :

$$X_t = [v_x, a_x, \theta_{pedal}, I_{road}, \dots]$$

where:

- v_x is the longitudinal velocity.
- a_x is the longitudinal acceleration
- θ_{pedal} represents gas and brake pedal positions, serving as primary indicators of driver intent.
- I_{road} is the road segment number, serving as a distance marker, representing the rolling distance at 0.5 m intervals.

The objective is to predict a sequence of future accelerations $[\hat{a}_{x,t+1}, t, \hat{a}_{x,t+2}, \dots, \hat{a}_{x,t+H}]$ over a horizon H of 5 seconds, aligning with the MPC's prediction window.

II.3. Dataset and Generalization

A key contribution of this work is the generalization of the model across diverse profiles:

- **Diverse Profiles:** Data was collected from over 30 different drivers, ranging from professional test pilots to standard customers.
- **High-Dynamic Scenarios:** Maneuvers were conducted on virtual proving grounds, including high-deceleration events (up to 1g).
- **Real-time Integration:** Models were evaluated for deployment within the ROADS MPC-MCA real-time platform.

Initial offline testing indicated that the Informer outperformed the standard LSTM in capturing peak magnitudes during sudden transitions when using smaller datasets (Qiu 2025). Fig. 2 compares these preliminary Informer results against a constant predictive law via internal MPC-MCA. It has been shown that the Informer predictive law can increase 20% the reproduced amplitude of motion cueing compared to constant predictive law which is Renault's internal standard MPC-MCA, while the tilt velocity is kept in the same order in these two motion cueing algorithms.

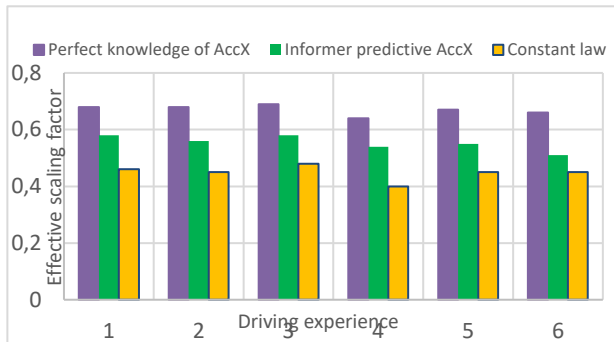


Figure 2: Motion cueing results comparison between primarily Informer predictive signal and constant predictive law

Encouraging by this study, a more extensive study was led to find the optimal configuration which will be implemented in ROADS simulator.

II.4. Investigated Configurations

1. Basic LSTM

This baseline model utilizes longitudinal acceleration AccX, vehicle speed, and road segment data as inputs. It was trained on 151 ROADS driving records using a 5-second historical look-back window.

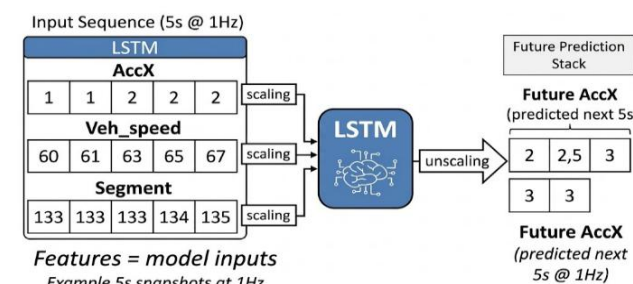


Figure 3: LSTM basic prediction model (example of 1Hz)

2. LSTM + Embedding (Fig. 4)

To better exploit the road segment signal, we implemented an embedding technique. This automatically transforms the segment sequence into multiple feature sequences, allowing the model to capture finer contextual dependencies and future road characteristics.

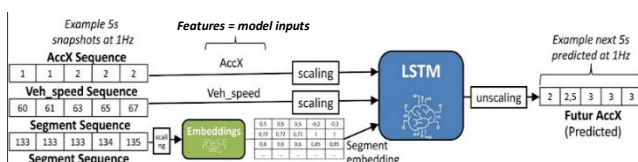


Figure 4: LSTM + Embedding model (example of 1Hz)

3. LSTM + Extensive Inputs (Fig. 5)

This configuration evolves the basic model by adding supplementary inputs such as lateral acceleration (AccY), pedal dynamics, and future road curvature (at x+30m to x+200m) to improve AccX prediction.

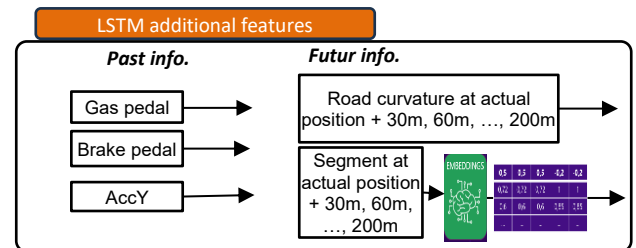


Figure 5: LSTM + extensive inputs

4. Informer (Fig. 6)

This configuration leverages the full Informer architecture, engineered for enhanced inference speed and efficiency when processing the long-sequence time-series data required for motion cueing.

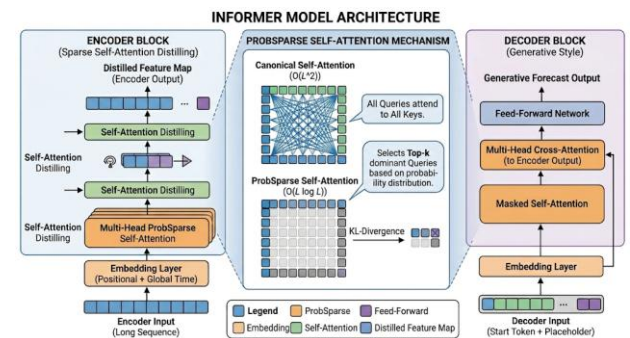


Figure 6: Informer model architecture (adapted from Zhou et al. 2021)

II.5. Data Acquisition and Validation Strategy

The experimental scenario reproduces the CTA handling track (Renault's proving ground center), which is specifically designed to validate vehicle behavior at the limit of tire grip in a controlled environment, free from traffic or signaling.

A total of 151 datasets were collected using the ROADS simulator. Fig. 7 illustrates the distribution of these driving experiences based on the mean vehicle speed and the 98th percentile of longitudinal acceleration. This percentile was selected to represent the sustained acceleration levels during the most intense maneuvers (top 2%).

The collected data covers a speed range between 50 and 115 km/h. While the distribution suggests a weak correlation between increased speed and higher longitudinal acceleration peaks, this trend is not strictly definitive across the entire sample set. To ensure the models' generalization and robustness in the target use domain, five specific datasets (highlighted in red) were reserved exclusively for validation.

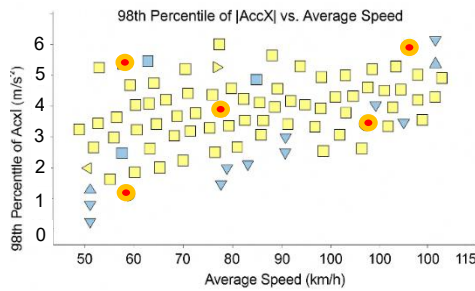


Figure 7: Dataset split distribution

III. Predictive performance comparison

To evaluate the predictive accuracy of the various deep learning methods, the Mean Absolute Error (MAE) was utilized as the primary metric. In this study, the MAE is defined as the mean absolute difference between the actual vehicle longitudinal acceleration and the predicted AccX over a 5-second horizon.

Fig. 8 compares the MAE of four distinct LSTM configurations across a prediction horizon ranging from 0.1 s to 5 s. For each time step, results were aggregated from five independent validation datasets to ensure model robustness across diverse driving samples.

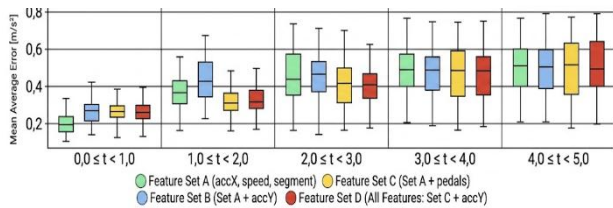


Figure 8: MAE distribution for LSTM configurations as a function of prediction time and input features

As observed in Fig. 8, for horizons where the LSTM models incorporating pedal positions or combined pedal and AccY yield superior results compared to other configurations. However, their MAE are higher than the basic LSTM model for the time predictive horizon inferior to 1 second. Given its structural simplicity and consistent performance over longer horizons, the basic LSTM remains a strong candidate for integration into the MPC-MCA strategy.

The following analysis compares the MAE between the LSTM and the Informer architectures.

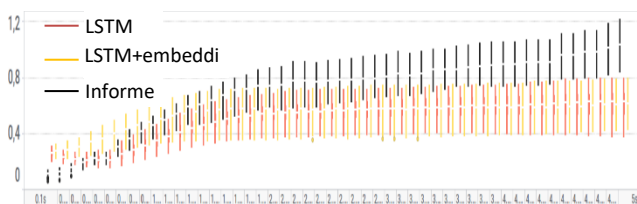


Figure 9: Predictive performance comparison of Informer, basic LSTM, and LSTM+embedding

Fig. 9 illustrates that the Informer model (black) achieves the highest accuracy in the very short term ($t < 0.6s$), recording the lowest MAE. However, its error increases rapidly, surpassing other models from the time horizon superior to 1.2s.

In contrast, both the LSTM (red) and LSTM+Embedding (orange) models exhibit a more stable, logarithmic error evolution, with the MAE plateauing around 0.7 after 2 seconds. These results suggest that when provided with sufficient driving data, the basic LSTM can intrinsically learn complex contextual dependencies similar to those captured by the LSTM+embedding approach.

While the Informer was theoretically expected to outperform the LSTM across the entire horizon, current results show its advantage is limited to short-term predictions. Nevertheless, its performance on short time horizons remains of interest. A hybrid predictive signal utilizing multiple AI models could be considered if computational resources and real-time constraints permit.

Based on the findings presented in Figures 8 and 9, it is concluded that LSTM-based models offer the most effective trade-off between precision and robustness for longitudinal acceleration prediction in the CTA scenario. Consequently, this architecture was selected for final deployment in the ROADS MPC-MCA.

To optimize computational efficiency within the real-time motion cueing environment, the predictive module was evaluated at two distinct operating frequencies: 200 Hz and 10 Hz.

Fig. 10 illustrates the Mean Absolute Error for various input window sizes and LSTM inference frequencies. In this setup, the LSTM is embedded within a Simulink model running at a base rate of 200 Hz, which matches the execution frequency of Renault's internal MPC-MCA deployed on the ROADS simulator.

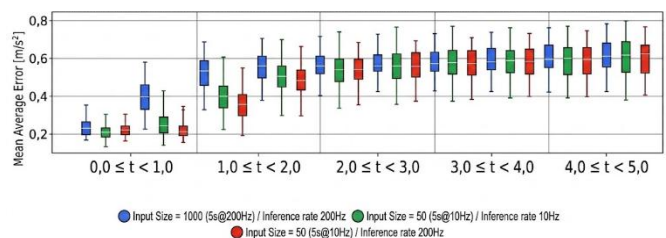


Figure 10: MAE for LSTM as a function of prediction time, input size, and inference frequency

The results indicate that the configuration utilizing a 10 Hz LSTM inference rate embedded within the 200 Hz control loop (shown in red) provides the highest accuracy. Furthermore, this approach consumes significantly fewer computational resources compared to higher frequency alternatives. Consequently, the 10 Hz predictive module is

identified as the most efficient and robust choice for integration into the real-time platform.

Once the AI model and its operating configuration were finalized, the original Python-based model was converted into a corresponding Simulink LSTM block. This conversion enables the generation of optimized C/C++ code, facilitating seamless integration into the Concurrent Real-Time system used by the ROADS driving simulator.

IV. Motion Cueing Results and Performance Analysis

To evaluate the impact of the AI-based predictive strategy on motion fidelity, we analyzed the linear correlation between the simulated vehicle's longitudinal acceleration ($AccX_{veh}$) and the specific force (fx_{MCA}) generated by the MPC-MCA.

IV.1. Longitudinal Fidelity and Signal Decomposition

Figure 11 presents a direct comparison between the vehicle's reference signal and the motion cues produced using different predictive laws within the MPC-MCA. Given the mechanical constraints of the platform, 1:1 motion restitution is unattainable in high-dynamic scenarios. To address this, a synthetic signal using dual scaling factors is fed into the MPC-MCA (Fang et al. 2024), allowing for a distinct analysis of the high-pass frequency components, which are critical for driver perception of transients.

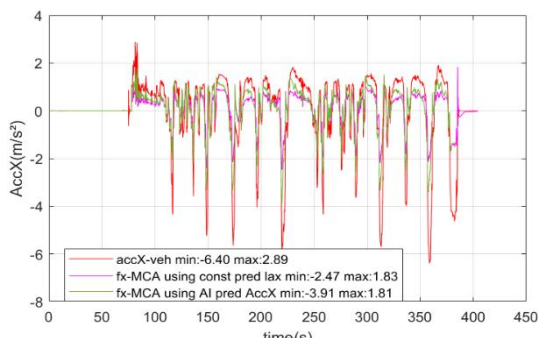


Figure 11: Comparison between the $AccX_{veh}$ and the reproduced specific forces with or without AI predictive signal

As illustrated in Fig. 12, the correlation performance of the MPC-MCA demonstrates that priority is assigned to the high-pass frequency signal. This tuning aligns with feedback from professional drivers, who prioritize accurate "jerk" and transient force perception over sustained low-frequency acceleration. Two primary metrics quantify this performance:

- **Effective Scaling Factor:** Derived from the inverse of the correlation slope.
- **Pearson Correlation Index (R):** A measure of signal linearity and timing accuracy.

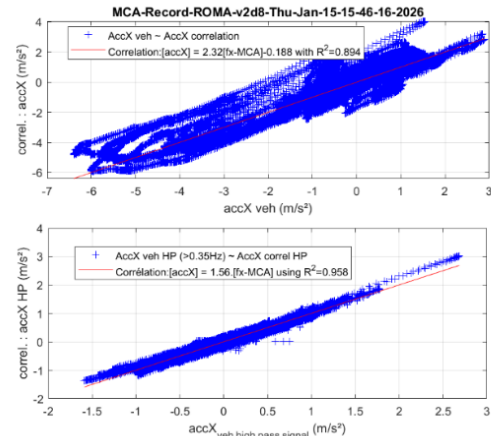


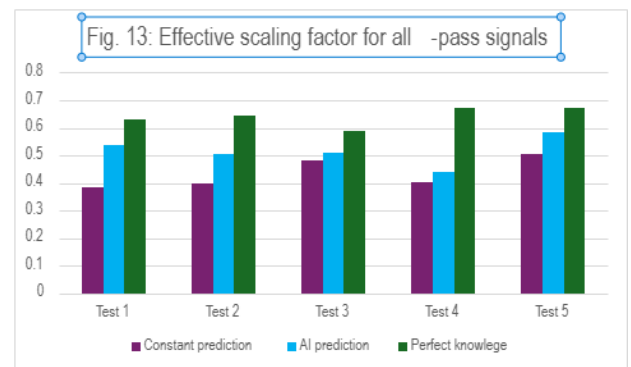
Figure 12: Linear correlation in all pass or high-pass signals for evaluating the performance of MCA

Based on empirical data from the ROADS simulator, a motion feel is considered acceptable when the R^2 for the all-pass signal exceeds 0.80–0.85. For high-frequency cues, R^2 must exceed 0.90 or even 0.95 to accurately capture rapid vehicle dynamics.

IV.2. Comparative Results: AI vs. Constant Predictive Law

Figures 13 and 14 report the effective scaling factors achieved across five validation tests (see Fig.7). Comparing the AI-based predictive model to the baseline constant predictive law reveals significant improvements:

- **Amplitude Gain:** The AI-predictive strategy achieved a 20% mean gain in the effective scaling factor for all-pass signals.
- **cues, the gain reached 25%,** significantly enhancing the driver's perception of braking and acceleration onset.



Furthermore, the integration of AI predictions allows to optimize workspace management, directly impacting the tilt coordination efficiency. Analysis of the platform's rotational dynamics shows that the mean absolute tilt velocity error was reduced by 19% compared to the constant predictive law. This reduction in "false cues" (artificial tilt rotations)

contributes to a more transparent and immersive driving experience.

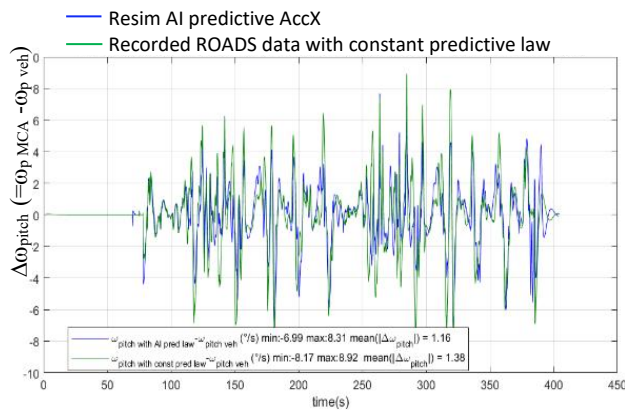
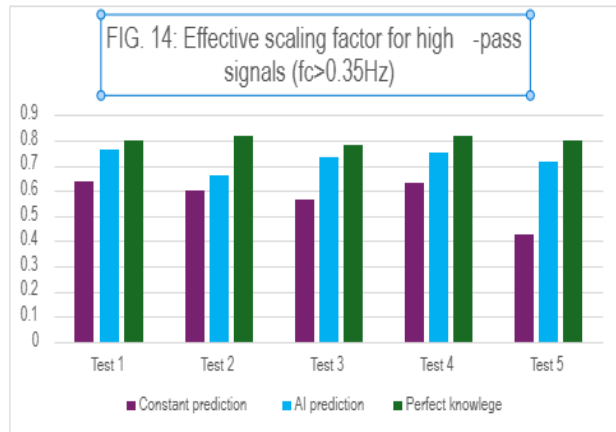


Figure 15: Delta tilt velocity comparison between AI predictive AccX and constant predictive law

V. Conclusion

This study successfully integrated an AI-based driver behavior prediction model into the MPC-MCA of the ROADS driving simulator to enhance longitudinal motion fidelity. By investigating various architectures, we identified that an LSTM-based model, operating with a 10 Hz inference rate within a 200 Hz real-time control loop, offers the optimal balance between predictive accuracy and computational efficiency.

The simulation results based on real recorded data demonstrate that forecasting driver intent allows the MPC-MCA to optimize the platform's 15m x 15m workspace more effectively. Key achievements include a 25% improvement in high-frequency motion scaling and a 19% reduction in absolute tilt velocity errors, addressing a critical limitation in current longitudinal motion cueing for high-performance handling scenarios.

Future research will focus on Informer model to reduce the MAE for long time horizon. Additionally, we aim to evaluate the model's performance in

complex urban environments by integrating external factors, such as traffic density and road signaling, into the AI input vector. Finally, subjective validation sessions with professional test drivers will be conducted to correlate these numerical gains with human-perceived immersion, utilizing the RT-ready AI-based MPC-MCA.

References

- Alsanwy S., Asadi H., Qazani M., Mohammed S., Nahvandi S., 2023. Prediction of Vehicle Motion Signals for Motion Simulators Using a Multivariate CNN-LSTM Model, DSC 2023 Europe VR.
- Arango C. G., Asadi H., Qazani M. R. C., and Lim C. P., 2025. Learning-Based Approximate Nonlinear Model Predictive Control Motion Cueing, arXiv:2504.00469v2 [cs.RO].
- Fang Z. et al., 2024. Assessment of vehicle dynamics using high-performance driving simulator, DSC 2024, Strasbourg, France.
- Hochreiter S., and Schmidhuber J., 1997. Long Short-Term Memory. *Neural Computation* 9.8, p. 1735-1780. issn : 0899-7667. doi : 10.1162/neco.1997.9.8.1735.
- Huang X., Shi X., Wang P., Xu H., Tang X., and Zhang G., 2025. An Analytic Policy Gradient-Based Deep Reinforcement Learning Motion Cueing Algorithm for Driving Simulators, *IEEE Access*, vol. 13, pp. 81507-81523, doi: 10.1109/ACCESS.2025.3564597.
- Koyuncu A. B., Erçelik E., Comulada-Simpson E., Venrooij J., Kaboli M., Knoll A., 2020. A Novel Approach to Neural Network-based Motion Cueing Algorithm for a Driving Simulator, 2020 IEEE Intelligent Vehicles Symposium (IV). Las Vegas, USA.
- Qiu D., 2025. Développement d'un modèle prédictif de comportement longitudinal de conducteur basé sur l'IA, Renault internal research report.
- Scheidel H. et al., 2025. A nonlinear real time capable motion cueing algorithm based on deep reinforcement learning, arXiv:2503.10419v3 [eess.SY]
- Zhou H. et al., 2021. Informer: Beyond efficient transformer for long sequence timeseries forecasting, *Proceedings of the AAAI conference on artificial intelligence*. T. 35. 12, p. 11106-11115 (cf. p. 29, 31).

Evaluating Data-Driven Architectures for High-Fidelity Modeling of Driving Simulator Motion

Philipp Legran, Fabian Russ, Lutz Eckstein

RWTH Aachen University – Institute for Automotive Engineering (ika)
{philipp.legran, fabian.russ, lutz.eckstein}@ika.rwth-aachen.de

Abstract - This paper investigates data-driven surrogate models for predicting the response of a highly dynamic driving simulator motion system from motion-cueing setpoints. Given the objective to replicate a specific simulator, real-world data was systematically collected using the simulator motion system and multiple data-driven modeling approaches were investigated. The paper evaluates a rollout training method for NARX models, to improve closed-loop stability. The NARX model examined is based on a physically motivated non-uniform lag window. A temporal-band ablation study analyzes the influence of individual segments of the lag window on model performance. The results demonstrate that rollout-aware training and physically motivated temporal input design are critical for accurate modeling of the driving simulator motion response.

Keywords: driving simulator, motion system, surrogate modeling, NARX, multi-step rollout training

Copyright: © 2026 by the authors.

Licensee Driving Simulation Association.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.82157/dsa/2026/16>

Introduction

Dynamic simulators have become indispensable tools for human-in-the-loop studies in the automotive, aeronautic and related industries. Realistic simulator motion is a key factor to ensure high immersion for the driver (Ronellenfisch, et al., 2019; Schöner, Häcker, and Nagel, 2026). The motion systems, including actuation hardware, control algorithms, sensors and motion cueing algorithms, are typically custom tailored to the specific simulator's use case. To account for custom hardware and requirements, motion algorithms are often tuned and tested on the final hardware, where the full control and actuation chain can be evaluated in combination. Although accurate models of the control, actuator, and sensor chain do exist, they are in most cases tightly coupled to real-time execution.

However, many dynamic simulators are unique, complex, and expensive systems, and are therefore scarce development resources that are more often reserved for operation than for extended testing and tuning (Schöner, Häcker, and Nagel, 2026). This creates a strong dependence on hardware availability when developing and evaluating control and motion-cueing algorithms. The aim of this paper is to reduce that dependence by proposing a surrogate model that reproduces the dynamic behavior of the simulator motion system. For such a surrogate to be useful, it must accurately reproduce the acceleration response, remain stable during long closed-loop rollouts, capture response timing realistically, and be efficient enough for offline development workflows.

This paper investigates a data-driven approach to modeling the entire signal chain of the simulator motion system, including both digital and physical process steps. The objective is to predict the measured simulator response directly from motion-cueing setpoints using surrogate models trained on real-hardware data. A special focus is placed on multi-step rollout training and the temporal design of a learned Nonlinear AutoRegressive with eXogenous inputs (NARX) model that predicts the motion state from past commands and past responses. In particular, the paper examines how rollout-aware training affects closed-loop stability, how a physically motivated non-uniform temporal context influences model performance, and how these design choices compare with simpler linear and nonlinear baseline models.

Research Question

This study evaluates whether data-driven modeling of a motion system can bridge the gap between computational efficiency and physical accuracy. We investigate this through the following research questions:

- **Q1:** Can data-driven surrogate models accurately reproduce the dynamic response of a driving simulator motion system from motion-cueing setpoints?
- **Q2:** Does rollout-aware training improve the closed-loop stability and predictive accuracy of NARX models compared with single-step training?
- **Q3:** How does a physically motivated non-uniform temporal input design influence NARX model performance?

Models

All models in this work predict the measured accelerations of the motion system for seven degrees of freedom: platform accelerations for Rx, Ry, Rz, Tx, Ty, Tz and rail acceleration. To reduce the model's complexity and enforce consistent derivatives, velocity and position are not predicted directly, but instead they are propagated by semi-implicit Euler integration during both rollout training and inference.

At each timestep the model receives the current and past motion-cueing setpoints together with autoregressive feedback derived from previous predictions. In closed-loop inference the model is fed only its own predicted state, so even small acceleration errors can accumulate through the integration loop and destabilize long runs. Model quality therefore depends not only on the model class, but also on temporal context and rollout training.

The paper uses a compact progression of four models:

Passthrough: The passthrough baseline forwards input commands directly, capturing neither platform dynamics nor system nonlinearities.

ARX: The Autoregressive with eXogenous inputs (ARX) model is the simplest learned baseline. It predicts acceleration from a lagged window of setpoints and past outputs using linear least-squares regression (Ljung, 1999; Verhaegen and Verdult, 2007). The ARX model tests whether a purely linear autoregressive model is sufficient to capture the platform response. It is a linear ridge-regression model with a 12-step lag window, additional static 8 ms delay and fitted analytically.

NARX: The first neural baseline is a nonlinear autoregressive model trained with a single-step objective ($k=1$) (Billings, 2013; Lin, et al., 1997). It uses the same general NARX structure as the final model but is optimized only for short-horizon prediction. The network is a two-hidden-layer MLP (1344-256-256-7, Tanh activations) trained with AdamW (10^{-3}) for 60 epochs with validation-based learning-rate decay.

NARX (rollout, $k=20$): The final model is a NARX network trained with multi-step rollout loss using a temporally structured input design. It is unrolled for $k=20$ steps during training and optimized over the resulting closed-loop trajectory. In other words, the model is trained not only to predict the next step accurately, but also to remain stable when its own predictions are fed back over many consecutive steps. The input indices follow the banded structure in Figure 2, combining short-range context for sensor dynamics, dense coverage of the dominant delay region, and sparse longer-range context for slower actuator-memory effects (Lin, et al., 1997; Peddinti, Povey, and Khudanpur, 2015). The rollout training objective is

$$\mathcal{L}_{\text{rollout}} = \frac{1}{k} \sum_{j=0}^{k-1} \|\hat{\mathbf{a}}_{t+j} - \mathbf{a}_{t+j}\|_2^2, \quad (1)$$

which reduces to ordinary single-step training for $k=1$ and directly penalizes error accumulation for $k > 1$.

The two key ablation studies directly correspond to these two design choices. First, a rollout-horizon sweep tests how the training horizon k affects closed-loop stability. Second, a leave-one-band-out study

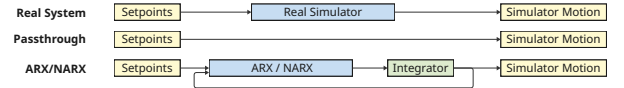


Figure 1: Schematic overview of the evaluated model classes and their integration with the Euler-based state update used during closed-loop rollout.

tests which temporal context bands are most important for accurate and stable prediction. Figure 1 summarizes the selected model classes and their role within the surrogate modeling pipeline.

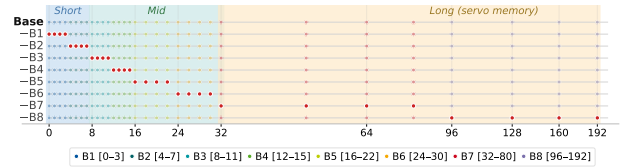


Figure 2: Temporal index design of the final NARX model and the corresponding leave-one-band-out ablations. The index set covers short-range context, the platform delay zone, and sparse long-range context.

Data Sets

The data used in this work consists exclusively of recordings captured on real hardware to reflect the actual dynamic behavior of the simulator motion system. The data distribution was chosen to balance broad coverage with realistic operating conditions. Controlled excitation data is required to systematically uncover the system dynamics, while natural driving data provides combined motions that are particularly valuable for validation.

The recordings come from ika's highly dynamic driving simulator, a 7-DoF motion system consisting of a Stewart platform and a rail axis. The dataset contains 21 reference setpoint signals and 21 measured motion signals, both sampled at 1 kHz. These signals are accelerations, velocities, and positions for the six platform degrees of freedom and the rail axis. The full measured motion vector is used during teacher-forced training, to initialize rollouts, and as reference data during training, validation, and final evaluation.

The core dataset consists of systematically designed sine-sweep experiments. Each run excites one degree of freedom with a sine sweep, while another degree of freedom is superimposed with a sinusoidal disturbance. This provides broad dynamic coverage while also introducing interactions between the axes. Data were recorded for three disturbance frequency ranges, resulting in a total of 147 runs, each lasting approximately 70 s.

The second dataset consists of step-response-like commands in which the system response is recorded up to approximately 70% of the system's capacity and the degrees of freedom are excited individually (Russ, et al., 2024). This dataset consists of 14 runs of approximately 4 to 11 minutes and is used to investigate generalization beyond the sine-sweep range and closer to the system limits.

The third dataset comprises 15 drives from real-world

simulator operation recorded by human drivers in the simulator. Most of these runs last 10 minutes and produce natural motion profiles with combined movements across all degrees of freedom.

Each dataset is split independently into training, validation, and evaluation subsets using a fixed-seed randomized 70/10/20 partition at the level of complete runs, such that no run is split internally and no global split across datasets is performed.

Results and Discussion

Models are evaluated on an eval set of runs not used during training or validation. It contains runs from all three data sources: sine sweeps, step-response-like commands, and real driving maneuvers. Inference is performed in full closed-loop rollout: predicted accelerations, velocities, and positions are fed back as context at every step with no teacher forcing, matching the conditions of real deployment.

The acceleration RMSE is the primary metric. Integrated velocity and position serve as indicators of long-run stability. Latency is the cross-correlation lag between predicted and measured signals, searched at 1 ms resolution over ± 100 ms.

Teacher-forced (TF) RMSE measures single-step prediction accuracy with ground-truth state feedback at every step, isolating model fit from closed-loop error accumulation, whereas rollout RMSE reflects true deployment conditions with no teacher forcing. Table 1 summarizes the selected main-model comparison.

Discussion

Among the selected learned models, the rollout-trained NARX achieves the best acceleration accuracy by a wide margin, reducing the mean 6-DOF acceleration RMSE from 0.042 for passthrough to 0.021. It also achieves near-zero latency (-0.4 ms mean) and keeps rail-position drift small over full-rollout inference. Notably, this remains true on real-driving runs lasting around 10 minutes. This combination makes it the most practical surrogate in the selected set.

The ARX baseline performs worse than passthrough and diverges catastrophically on the rail axis. A purely linear lagged model cannot reliably capture the nonlinear and long-horizon behavior of the motion platform in closed loop, as small persistent acceleration errors accumulate through Euler integration into growing position drift, which is then fed back as state, amplifying subsequent errors.

The single-step NARX model can be computed faster than realtime and can replicate short-horizon behavior, but it becomes unstable when evaluated in true closed-loop rollout. Its mean acceleration RMSE is an order of magnitude worse than the final rollout-trained model, and its mean acceleration latency shifts to $+27.7$ ms. This large timing error is a clear symptom of exposure bias: the model is trained on ground-truth state feedback but evaluated on its own predictions. Figure 3 illustrates this contrast on a representative closed-loop rollout from the sine-sweep dataset.

Figure 4 shows the effect of the rollout horizon k on

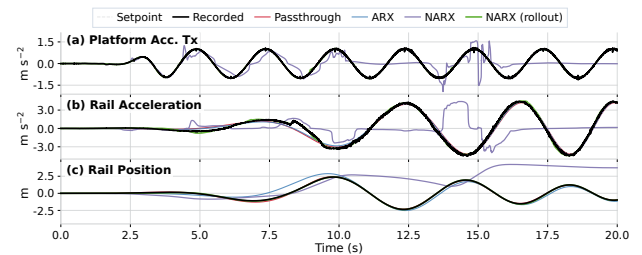


Figure 3: Representative 10-second closed-loop rollout from a held-out evaluation run. The rollout-trained NARX tracks platform and rail acceleration closely, while passthrough shows the raw system delay and weaker models show larger rollout errors.

the NARX model. Single-step training ($k=1$) yields dramatically worse full-run performance than rollout training, confirming that validation under teacher forcing is a poor proxy for deployment behavior. Rollout training also exposes the model to error accumulation over multiple steps and therefore encourages the model to learn low-frequency response behavior. Performance improves strongly from $k=1$ to $k=20$ and then saturates, which motivates the choice of $k=20$ for the final model.

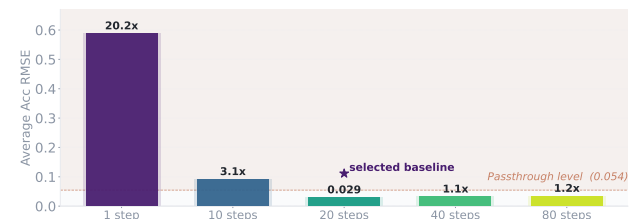


Figure 4: Effect of rollout horizon k on mean acceleration RMSE over the full evaluation set. Single-step training ($k=1$) is catastrophically worse than rollout training. Accuracy improves strongly up to $k=20$ and then largely saturates.

Figure 5 quantifies the contribution of each temporal band in the NARX model. The delay-early band (8–11 ms) and the medium band (32–80 ms) are especially important. Removing either causes a large accuracy drop, and removing the delay-early band nearly returns the model to passthrough-level performance. This supports the physically motivated design, with dense coverage in the short-lag and platform-delay zones, while sparse long-range context captures slower memory effects.

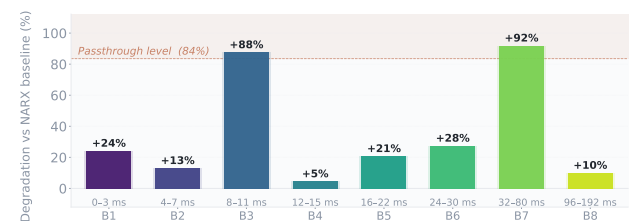


Figure 5: Acceleration RMSE increase relative to the full NARX model when each temporal band is removed. Bands in the physical delay region and long-range context region contribute most strongly to the final model's performance.

The passthrough baseline preserves the raw platform delay of roughly 12 ms, whereas the final rollout-trained NARX model brings the mean acceleration

Table 1: Evaluation results on the held-out eval set (mean over runs). Acc RMSE is averaged over Rx–Tz. Rail quantities are reported separately. Latency is the mean acceleration cross-correlation lag, and RTF (real time factor) denotes faster-than-real-time execution for values above 1.

Model	TF Acc	TF Rail	Rollout Acc	Rollout Rail	Rail Pos	Latency (ms)	RTF
Passthrough	0.072	0.165	0.042	0.127	0.009	−12.6	131
ARX	0.015	0.055	1.625	6576.910	4.2×10^6	+1.7	27.7
NARX	0.014	0.016	0.429	1.539	458.776	+27.7	5.1
NARX (rollout, $k=20$)	0.036	0.068	0.021	0.082	0.020	−0.4	5.8

latency close to zero. This indicates that the learned model does not merely reduce amplitude error, but also reproduces the timing of the physical system response much more faithfully than the baseline models.

Conclusions and Implications

This paper evaluates whether data-driven surrogate models can reproduce a simulator's motion response directly from motion setpoints, whether rollout-aware training improves closed-loop stability over single-step training, and how a physically motivated non-uniform temporal input design influences model performance. ARX, single-step NARX, and rollout-trained NARX were evaluated using real-hardware data and a passthrough reference model.

All data-driven models examined were able to predict the simulator's acceleration accurately under teacher forcing. However, both the ARX model and the single-step NARX model degraded strongly in closed-loop rollout in comparison to passthrough prediction. Despite its strong performance under teacher forcing, the single-step NARX model increased from a teacher-forced platform acceleration RMSE of approximately 0.014 to a rollout platform acceleration RMSE of approximately 0.429, confirming that single-step training is insufficient for closed-loop deployment **Q2**. With respect to **Q3**, the results indicate that the temporal design of a NARX model is a key factor influencing prediction quality. In particular, the delay-early band (8–11 ms) and the long-range context band (32–80 ms) have the strongest influence on accuracy, supporting the use of dense delay-region coverage together with sparse longer-range context.

The NARX model trained using the rollout method offered the best overall trade-off between accuracy, stability, and timing. The model, trained with a 20-step rollout, achieved the best overall accuracy for acceleration in the closed-loop system. The rollout acceleration RMSE was approximately 0.021, compared to 0.042 for the passthrough. Also noteworthy is the latency of approximately −0.4 ms, which differs significantly from the passthrough latency, which corresponds to the average system latency of approximately 12.6 ms.

Future Work

In addition to refining the temporal design of the NARX model, velocity and position prediction via Euler integration currently represents a significant limitation. The model is trained using real-world measured accelerations, velocities, and positions, which do not exhibit exact differential properties due to sensor measurement errors. In the current model formulation, these measurement inconsistencies inevitably lead to a trade-off between stability and accuracy.

In future studies, the model should be validated on additional simulators. Data beyond system limits could teach limit handling, and the model's robustness could be improved by collecting additional real-world driving data. Furthermore, the proposed NARX approach should be directly compared with recurrent sequence models such as LSTMs, which may capture longer temporal horizons.

Rollout-capable NARX surrogates with a physically motivated temporal input design are promising tools for the offline development and benchmarking of motion-cueing algorithms.

References

- Billings, S. A., 2013. *Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains*. Chichester: John Wiley & Sons, Ltd. ISBN: 978-1-119-94359-4. <https://doi.org/10.1002/9781118535561>.
- Lin, T. N., Giles, C. L., Home, B. G., and Kung, S. Y., 1997. A Delay Damage Model Selection Algorithm for NARX Neural Networks. *IEEE Transactions on Signal Processing*, 45(11), pp. 2719–2730. <https://doi.org/10.1109/78.650098>.
- Ljung, L., 1999. *System Identification: Theory for the User*. 2nd. Upper Saddle River, NJ: Prentice Hall.
- Peddinti, V., Povey, D., and Khudanpur, S., 2015. A Time Delay Neural Network Architecture for Efficient Modeling of Long Temporal Contexts. In: *Proceedings of Interspeech 2015*, pp. 3214–3218. <https://doi.org/10.21437/Interspeech.2015-647>.
- Ronellenfitch, A., Dong, S., Baumgartner, E., Reuss, H.-C., and Schramm, D., Sept. 4, 2019. Objective Criteria for Motion-Cueing Evaluation. In: *Proceedings of the Driving Simulation Conference 2019 Europe VR*. Strasbourg, France, pp. 85–91. ISBN: 978-2-85782-749-8.
- Russ, F., Legran, P., Beckmann, J., and Eckstein, L., Sept. 18, 2024. Performance and Frequency Analysis to Identify Limits of Dynamic Driving Simulators. In: *Proceedings of the Driving Simulation Conference 2024 VR*. Vol. 9. Strasbourg, France. ISBN: 978-2-9573777-5-6.
- Schöner, H.-P., Häcker, J., and Nagel, K., 2026. Dynamic Driving Simulators. In: *Handbook Assisted and Automated Driving*. Springer Vieweg, pp. 105–123. https://doi.org/10.1007/978-3-658-45276-6_8.
- Verhaegen, M. and Verdult, V., 2007. *Filtering and System Identification: A Least Squares Approach*. Cambridge University Press. <https://doi.org/10.1017/CB09780511618888>.

Generative AI tools were used for language and grammar checking. The authors reviewed and edited the content and take full responsibility for the content of the publication.

Funded by the European Union under Horizon Europe, Grant Agreement No. 101203230 (CERTAIN). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.



**Funded by
the European Union**

VIRTUAL REALITY FOR DRIVING SIMULATION

Assessing the Sensitivity of a VR Driving Simulator to Single-Day Highway Work Zones

Paolo Pretto¹, Manuela Prior¹, Mateo Jukić¹, Inge Gsellmann¹, Thorsten Muik¹, Manfred Rosenberger¹, Christina Hager², Alexander Maletzky²

(1) Virtual Vehicle Research GmbH, Inffeldgasse 21a, 8010 Graz, Austria, e-mail: paolo.pretto@v2c2.at

(2) RISC Software GmbH, Softwarepark 32a, 4232 Hagenberg, Austria

Abstract – Driving simulators provide a controlled platform for studying driver behavior in complex traffic environments, but their usefulness depends on their ability to reproduce measurable responses to variations in roadway demand. This study evaluates the sensitivity of a virtual reality-based driving simulator to experimentally induced changes in driver behavior associated with single-day highway work zones and unexpected hazards. Thirty-five licensed drivers completed three simulated highway conditions in counterbalanced order: baseline driving, a work-zone configuration with a lane closure and reduced speed limit, and a work-zone scenario including unexpected hazards requiring avoidance maneuvers. Driver responses were assessed using a multimodal set of indicators, including vehicle control behavior, gaze-based measures of visual attention, physiological signals, and subjective workload ratings. Results showed systematic changes across conditions. Drivers exhibited reduced lateral variability during navigation of the work-zone environment, indicating compensatory tightening of lane control in response to increased roadway demand. When unexpected hazards were introduced, lateral variability increased and was accompanied by greater gaze dispersion and electrodermal activity, indicating increased task demand and control effort. Together, these findings demonstrate that drivers showed both sustained adaptations to work-zone geometry and event-driven responses to unexpected hazards. The results support the use of immersive driving simulation as a tool for experimentally evaluating driver behavior in temporary work-zone scenarios and for testing alternative traffic-control strategies under controlled conditions.

Keywords: work zones; driver behavior; physiological measures; simulator sensitivity, virtual reality

Copyright: © 2026 by the authors.

Licensee Driving Simulation Association.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.82157/dsa/2026/17>

Introduction

Driver Demand in Single-Day Highway Work Zones

Single-day highway work zones introduce temporary lane closures, altered traffic flow, and constrained maneuvering space that can substantially alter the demands placed on drivers. In this paper, the term *single-day work zone* refers to maintenance operations deployed and removed within the same day, typically without long-term infrastructure changes. Unlike long-term construction zones, single-day work zones rely primarily on mobile and rapidly deployable traffic-control devices, such as trailer-mounted signs and cone-based lane delineation. These elements are often installed over shorter distances and with less structural permanence than in long-term configurations, which may reduce visibility, placement precision, and stability of lane boundaries. As a result, temporary

work zones can introduce higher levels of uncertainty for drivers, requiring quicker interpretation of traffic-control cues and more active monitoring of the roadway environment. As a consequence, these temporary configurations can influence driver state at multiple levels (Edwards et al., 2017). Drivers may need to interpret traffic-control cues more rapidly while maintaining increased visual monitoring and tighter vehicle control, which can lead to measurable changes in driver behavior, attention allocation, perceived workload, and physiological arousal (Huey et al., 2016; Ma et al., 2023; Chatterjee et al., 2018; Crundall et al., 2012). In this sense, single-day work zones constitute a naturally occurring driving context in which measurable changes in driver demand and monitoring behavior can emerge without requiring extreme or crash-imminent hazards.

Driving Simulators for Work-Zone Research

Controlled experimental investigation of work-zone driving is challenging under real-world conditions.

Field studies in active work zones are costly, difficult to standardize, and raise safety concerns, particularly when examining low-frequency hazards such as unexpected obstacles or ambiguous traffic-control cues. As a result, systematic comparison of alternative work-zone layouts or hazard scenarios is difficult to achieve in naturalistic environments (Federal Highway Administration, 2024; Savolainen et al., 2024). Driving simulators provide a promising platform for overcoming these limitations (Bella, 2008). Simulation environments allow researchers to manipulate roadway configurations, hazard scenarios, and traffic conditions while maintaining experimental control and participant safety. This capability makes simulators particularly well suited for studying roadway configurations such as single-day work zones, where layout variations and unexpected hazards can strongly influence driver behavior. However, the usefulness of driving simulators for work-zone research depends on their ability to elicit and measure consistent changes in driver responses when roadway demand varies. Demonstrating such sensitivity is an important prerequisite for using simulators to evaluate alternative work-zone layouts and temporary traffic-control strategies before real-world deployment (Kaptein et al., 1996; Wynne et al., 2019).

Driver Workload in Work Zones

Previous studies have shown that highway work zones can increase driver workload and alter vehicle-control behavior, particularly in configurations involving lane closures, reduced lane widths, or merging areas. Simulator studies have reported elevated subjective workload and increased variability in vehicle control when drivers navigate work-zone environments characterized by altered lane geometry and dense traffic-control cues (Shakouri et al., 2014; Ma et al., 2023). These effects are commonly attributed to increased information-processing demands and the need for drivers to continuously monitor roadway geometry, traffic-control devices, and surrounding traffic. Drivers often respond to these increased demands by adapting their control strategies, for example by adjusting speed or regulating lane position more tightly. Such behavioral adjustments are commonly interpreted as compensatory responses that help maintain stable vehicle control under constrained roadway conditions (Lewis-Evans and Rothengatter, 2009).

Unexpected Hazards Perception

Unexpected hazards represent an important source of driver demand, particularly when they occur in environments that already restrict maneuvering options. Sudden obstacles or displaced traffic-control devices require rapid perception, decision-making, and motor responses, often under time pressure. Research on hazard perception shows that such events can substantially increase driver

workload and visual scanning behavior, especially when escape options are limited (Underwood et al., 2005; Crundall et al., 2012). In work-zone environments, minor hazards such as fallen cones or small objects encroaching on the travel lane may therefore trigger localized changes in driver behavior and attention.

Emotional State While Driving

In addition to workload-related changes in vehicle control, driving performance can also be influenced by drivers' emotional and physiological state. Elevated arousal has been associated with changes in visual scanning behavior, risk perception, and control stability during demanding driving tasks (Matthews et al., 2012). Although work zones may not necessarily induce extreme stress, the combination of constrained roadway geometry and unexpected hazards can produce moderate increases in driver arousal and perceived task demand. Such changes may influence attentional allocation and vehicle control, making physiological indicators a useful complement to behavioral and gaze-based measures when assessing driver responses to demanding traffic situations.

Multimodal Assessment of Driver State

Assessing driver responses to demanding roadway conditions benefits from combining multiple measurement modalities. Behavioral indicators provide direct evidence of vehicle-control strategies, while gaze-based metrics capture visual attention and monitoring behavior. Physiological measures such as electrodermal activity and heart-rate variability can additionally reflect changes in driver arousal associated with increased task demand (Brookhuis and de Waard, 2010). Integrating these complementary indicators enables a more comprehensive assessment of driver state. However, these modalities differ in their sensitivity to task demands. Sustained increases in roadway complexity, such as those introduced by work-zone configurations, are typically reflected in adaptive changes in vehicle control and visual monitoring behavior (Lewis-Evans and Rothengatter, 2009; Ma et al., 2023). In contrast, electrodermal activity is particularly sensitive to transient, event-driven changes in task demand (Boucsein, 2012). Accordingly, physiological effects may be more pronounced during localized hazard encounters than during sustained driving under increased but stable demand.

Research Hypotheses

To examine both sustained adaptations to work-zone configurations and localized responses to unexpected hazards, the present study distinguishes between broader roadway-context effects and event-

driven responses within the simulated driving scenarios. The study evaluates whether a virtual reality-based driving simulator is sensitive to experimentally induced changes in roadway demand associated with single-day highway work zones and unexpected hazards. By combining behavioral, gaze-based, and physiological indicators, the study examines whether changes in roadway configuration produce coherent multimodal responses in driver behavior and attention.

H1: Driving in a work-zone configuration will produce measurable changes in vehicle control and visual attention compared to baseline highway driving.

H2: Unexpected hazards within the work zone will produce additional localized responses in vehicle control, visual attention, and physiological indicators compared to the work-zone condition without hazards.

H3: Changes in roadway demand will produce coherent responses across multiple measurement domains, including behavioral, gaze-based, and physiological indicators.

Methods

Simulator Setup

A VR-based driving simulator was developed to reproduce a three-lane highway environment with single-day work-zone configurations representative of common short maintenance operations. The simulator consisted of a fixed-base driving station equipped with a force-feedback steering wheel and pedals (Moza R5 bundle) and a fully adjustable driver seat, combined with an immersive VR headset (Meta Quest Pro) providing a wide field of view (approximately 106° horizontal and 96° vertical) and stable visual flow. The driving environment was implemented in the Unity game engine (Unity 2022). The virtual roadway reproduced an approximately 8 km section of one carriageway of an Austrian three-lane motorway (E60) with realistic lane widths, markings, signage, cone-based lane closures, mild wide-radius curves, and no visibility-limiting curvature; the work-zone configuration extended over roughly 5 km. Ambient traffic was simulated only in the leftmost lane across conditions, with vehicles generated at a constant 4-s temporal headway, visually varied in model and color, and travelling at constant speeds within each route section.

Experimental Design

A within-subject design was used, with each participant completing three conditions:

- **Baseline:** normal highway driving with a posted speed limit of 130 km/h and participants driving on the rightmost lane.

- **Work-Zone:** closure of the rightmost lane using standard cone delineation and signage, requiring participants to drive on the center lane under a reduced speed limit of 80 km/h.
- **Hazard:** identical to the Work-Zone condition, with the addition of minor unexpected hazards (e.g., displaced cones or small objects on the driving lane) requiring avoidance maneuvers.

In the Hazard condition, six hazards were introduced at predefined locations along the work-zone segment. Three hazard types were presented twice each in randomized order. To support the analysis of driver responses, the driving scenarios were segmented spatially around predefined hazard locations. Each hazard defined an *event segment* spanning 100 m upstream and 184 m downstream of the hazard location. These segments were defined to capture anticipatory and post-event responses to the hazards while ensuring that adjacent event segments remained non-overlapping. Roadway sections between event segments were defined as *inter-event segments*, representing sustained driving behavior within the work-zone environment. This segmentation enabled sustained adaptations to be distinguished from localized hazard responses in the subsequent analyses. Figure 1 illustrates the schematic layout of the three conditions, including speed-limit changes, lane-closure geometry, ambient traffic, and representative hazard types.

Participants

Thirty-five licensed drivers participated in the study (24 male, 11 female). Participants had a mean age of 30.4 years (SD = 9.7) and had held a driving license for an average of 12.2 years (SD = 9.6). Fifty-seven percent of participants reported driving at least once per week, with a median annual driving distance of approximately 6,000 km. On average, participants estimated that 47% of their driving occurred on highways, and reported no prior issues with simulator sickness. Sensation seeking scores (M = 2.87, SD = 0.74 on a five-point Likert scale) indicated moderate tendencies toward risk-taking behavior. All participants reported normal or corrected-to-normal vision and provided informed consent prior to participation.

Measures

Subjective mental workload was assessed using the Driver Activity Load Index (DALI) (Pauzié, 2008), and **emotional state** was evaluated using the Self-Assessment Manikin (SAM) (Bradley and Lang, 1994) following each experimental ride. **Driving behavior** was characterized using longitudinal and lateral vehicle-control metrics. Longitudinal control included mean speed and speed variability relative to the applicable speed limit. Lateral control was assessed using lane-position-based measures, including the standard deviation of lateral position

(SDLP), to quantify steering stability and lane-keeping performance. **Gaze behavior** was analyzed within predefined road segments using region-of-interest-based metrics. Primary indicators included gaze switches between predefined regions of interest (ROIs), grouped into road infrastructure, construction-site elements, event items, traffic, scenery, and ego-vehicle elements; lane markings were included in the road-infrastructure ROI. Horizontal gaze variability was computed as the standard deviation of horizontal gaze position. **Physiological indicators** of driver arousal and workload included heart-rate variability (HRV) and electrodermal activity (EDA) (Mehler et al., 2012). Signals were inspected for quality, and segments with unreliable recordings were excluded from further analysis. HRV indices and EDA summary metrics (mean, range and standard deviation) were extracted.

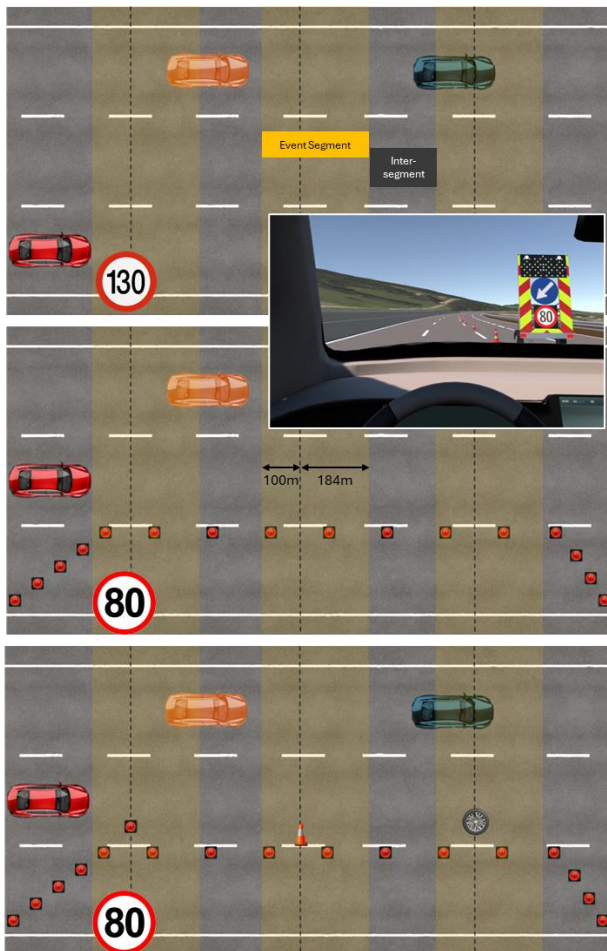


Figure 1: Schematic top-view of the three experimental conditions, with inset showing a first-person VR view of the work-zone environment. Participants drove under a Baseline condition (top; 130 km/h, no roadworks), a Work-Zone condition (middle; cone-based closure of the rightmost lane with reduced speed limit of 80 km/h), and a Hazard condition (bottom; identical work zone with additional hazards on the driving lane). Shaded regions indicate event and inter-event segments used for the analyses; ambient traffic was present in the leftmost lane across all conditions.

Procedure and Data Analysis

Participants first completed a familiarization drive to adapt to the simulator. They then completed the three experimental conditions in counterbalanced order. After each drive, subjective questionnaires were administered. Behavioral, gaze, and physiological data were synchronized and analyzed using within-subject comparisons across conditions. Parametric analyses were conducted using repeated-measures ANOVA, with equivalent nonparametric tests (Friedman and Wilcoxon signed-rank tests) applied when assumptions were violated. The analysis aimed to evaluate the sensitivity of the simulated conditions to experimentally induced changes in roadway demand.

Results

The results are reported for both *event segments*, capturing localized driver responses to hazards, and *inter-event segments*, representing roadway sections between hazards where sustained adaptations to the work-zone configuration could be observed. This segmentation allows behavioral, gaze-based, and physiological indicators to be examined both in proximity to unexpected hazards and during sustained driving within the work-zone environment. Results are presented first for driving behavior, followed by visual attention and physiological responses, and finally subjective workload and emotional ratings, which provide contextual interpretation of the objective measures.

Behavioral Responses

Lateral Vehicle Control

Lateral vehicle control was assessed using the standard deviation of lateral position (SDLP) (Figure 2). Analyses were conducted separately for event segments, capturing driver responses in proximity to hazards, and inter-event segments, representing driving between hazard locations. Within *event segments*, SDLP differed significantly across experimental conditions (Friedman $\chi^2(2) = 66.23, p < .001$). The Hazard condition showed the highest lateral variability, reflecting avoidance maneuvers required to circumvent displaced cones or objects encroaching on the lane. In contrast, both Baseline and Work-Zone conditions showed lower variability in these segments. Outside hazard locations, analyses of *inter-event segments* also revealed significant condition effects (Friedman $\chi^2(2) = 25.89, p < .001$). SDLP was lowest in the Work-Zone condition compared with both the Baseline and Hazard conditions. In the Hazard condition, SDLP increased again and differed significantly from the Work-Zone condition, indicating greater lateral variability even during roadway sections without

immediate hazard encounters. Taken together, these findings indicate a graded pattern of lateral control across conditions. Baseline driving was characterized by relatively relaxed lane-keeping behavior. In the Work-Zone condition, drivers exhibited reduced lateral variability, consistent with compensatory tightening of vehicle control in response to increased task demand within the work-zone environment. In the Hazard condition, however, lateral variability increased again, suggesting that the additional task demands associated with the presence of unexpected hazards led to more variable control despite the same underlying roadway geometry.

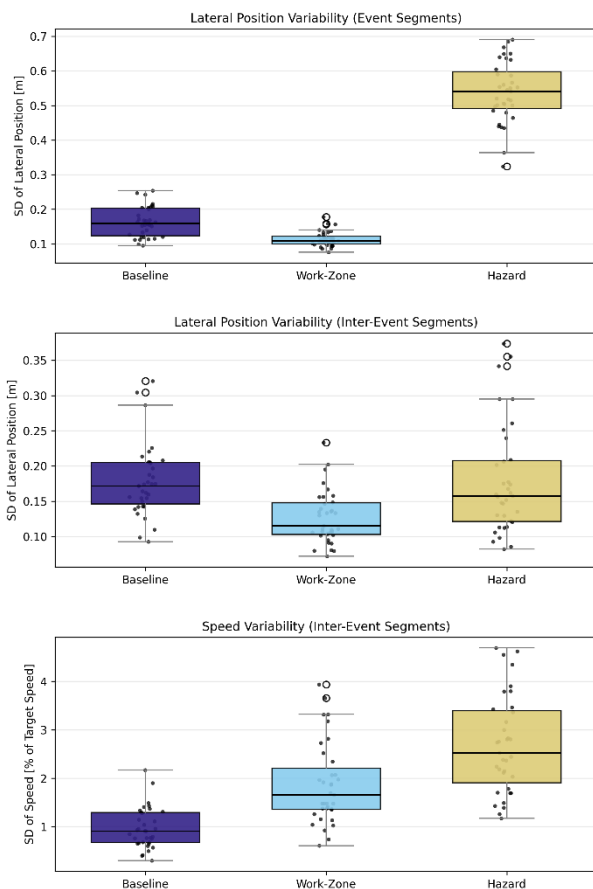


Figure 2. Behavioral responses across driving conditions. (top) SD of lateral position within event segments. (middle) SD of lateral position within inter-event segments. (bottom) SD of speed relative to the applicable speed limit.

Longitudinal Speed Control

Mean speeds closely followed the posted limits, with drivers travelling at approximately 129 km/h in the Baseline condition and approximately 81 km/h in both the Work-Zone and Hazard conditions. Speed variability relative to the applicable limit differed significantly across conditions ($F(2,68) = 12.58$, $p < .001$), being lowest in Baseline and higher in both Work-Zone and Hazard, which did not differ significantly from each other. Thus, hazards did not substantially alter mean speed or speed variability beyond the work-zone effect.

Collisions

Collisions with obstacles or traffic-control devices occurred only in the Hazard condition ($F(2,68) = 16.42$, $p < .001$), reflecting the additional demands associated with navigating unexpected hazards.

Visual Attention

Visual attention was analyzed using gaze-based indicators reflecting drivers' monitoring behavior during the task (Figure 3). Analyses focused on two metrics: gaze switching between predefined ROIs and horizontal gaze variability, computed within predefined road segments.

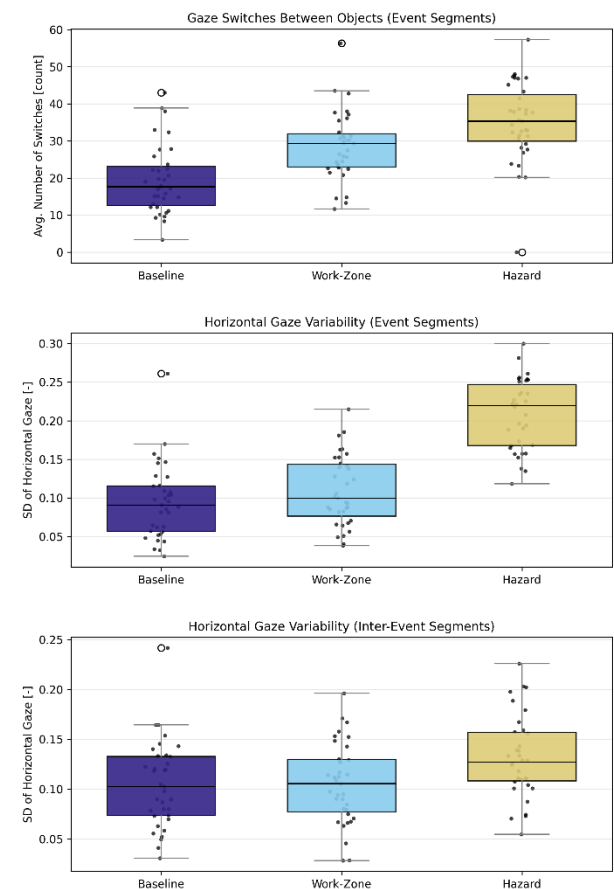


Figure 3. Visual attention measures across driving conditions. (top) Gaze switching within event segments. (middle) Horizontal gaze variability within event segments. (bottom) Horizontal gaze variability within inter-event segments.

Gaze Switching

The number of gaze switches between ROIs differed across experimental conditions ($F(2,68) = 7.24$, $p < .01$). Post hoc comparisons indicated that gaze switching was significantly higher in the Work-Zone condition than in the Baseline condition ($p < .01$) and increased further in the Hazard condition compared with the Work-Zone condition ($p < .05$). This progressive increase in gaze switching across conditions suggests greater visual monitoring activity as driving demands increased from normal highway

driving to work-zone navigation and finally to scenarios involving unexpected hazards.

Horizontal Gaze Variability

Horizontal gaze variability also showed significant condition effects. Within *event segments*, variability differed significantly across conditions (Friedman $\chi^2(2) = 33.52$, $p < .001$), with the Hazard condition showing the highest variability. This pattern reflects increased visual scanning in the vicinity of hazard locations. In *inter-event segments*, horizontal gaze variability remained elevated in the Hazard condition compared with both the Baseline and Work-Zone conditions (Friedman $\chi^2(2) = 16.24$, $p < .001$). This indicates that increased visual monitoring was not limited to immediate hazard encounters but persisted during subsequent driving sections. Together, these results indicate that gaze behavior changed systematically across conditions, with both increased gaze switching and greater horizontal gaze dispersion associated with the presence of work-zone configurations and unexpected hazards.

Physiological Responses

Electrodermal activity (EDA) and heart-rate variability (HRV) were analyzed as indicators of driver arousal (Figure 4). Given the short event windows and raw electrodermal sensor output, EDA SD was retained as an exploratory within-study summary metric; conventional SCR peak metrics were examined but did not yield robust, interpretable condition effects. Due to signal quality constraints, EDA analyses were conducted on 19 participants. EDA SD differed significantly across conditions (Friedman $\chi^2(2) = 24.947$, $p < .001$), with higher values in the Hazard condition than in both Baseline and Work-Zone, indicating increased arousal during unexpected hazard encounters. RR-interval variability did not differ significantly, indicating lower sensitivity of HRV than EDA in the present scenario.

Hazard Effects and Habituation

Hazard-specific analyses examined physiological responses across successive hazard encounters. The standard deviation of the electrodermal activity (EDA SD) attenuated across repeated events ($\beta = -0.078$, $p < 0.05$), suggesting a gradual reduction in physiological reactivity over time (Figure 4). This pattern is consistent with habituation effects, where repeated exposure to similar stimuli produces reduced physiological responses. No significant differences were observed across the three hazard types, indicating that physiological responses were primarily associated with the presence of unexpected events rather than with specific hazards.

Workload and Emotional Responses

Subjective ratings of workload and emotional state were collected after each experimental condition using the Driver Activity Load Index (DALI) and the

Self-Assessment Manikin (SAM). Detailed statistical analyses of these self-report measures are reported in Pretto et al. (2026), where the focus was on the human-factors implications of work-zone driving. In summary, subjective workload ratings indicated increasing perceived task demand across the experimental conditions ($F(2, 68) = 14.03$, $p < .001$). Mental and temporal workload were lowest during Baseline driving, increased in the Work-Zone condition, and reached the highest levels in the Hazard condition. Similarly, self-reported stress ($F(2, 68) = 26.44$, $p < .001$) and emotional arousal ($F(2, 68) = 12.95$, $p < .001$) were highest when hazards were present, while valence shifted toward a more negative emotional state in the Hazard condition ($F(2, 68) = 4.40$, $p < .05$). These results confirm that the work-zone configuration and the presence of hazards increased perceived driving demand, providing subjective support for the behavioral, gaze-based, and physiological differences observed across conditions.

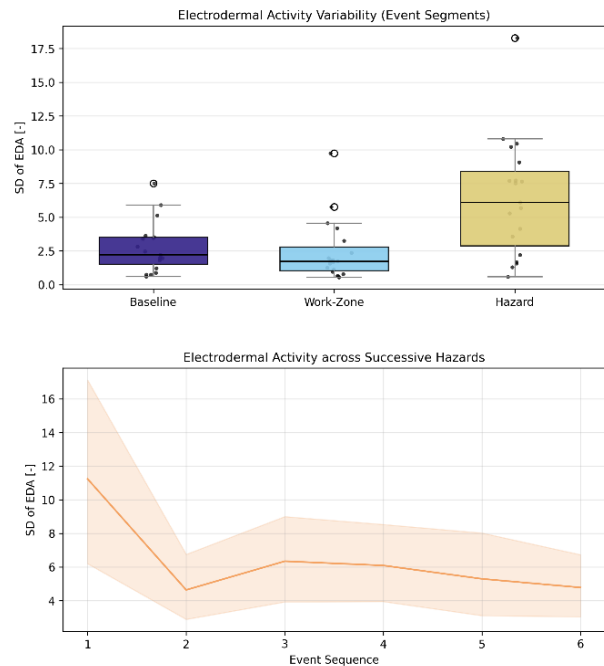


Figure 4. Physiological responses. (top) Electrodermal activity variability (EDA SD) across driving conditions. (bottom) EDA SD across successive hazard encounters.

Discussion

Demand-Based Adaptive Control

The results indicate that drivers responded differently to sustained changes in roadway configuration and to localized hazard events within the work-zone environment. These two types of demand produced distinct patterns of driver behavior and attention. The behavioral results suggest that drivers adopted different control strategies as task demands increased across conditions. During baseline driving, lane keeping appeared relatively

relaxed, reflecting lower demands of unobstructed highway driving. In the Work-Zone condition, lateral variability decreased relative to baseline, indicating compensatory tightening of vehicle control under constrained but predictable lane geometry, consistent with previous work on demand- or risk-related control adaptation (Lewis-Evans and Rothengatter, 2009; Ma et al., 2023). In the Hazard condition, however, SDLP increased both around hazards and in inter-event sections, suggesting that unexpected obstacles shifted drivers from precision lane keeping to reactive avoidance and anticipatory corrective steering under uncertainty. Thus, the Work-Zone and Hazard conditions appear to reflect different control modes rather than contradictory effects: predictable narrowing promoted tighter lane keeping, whereas unexpected obstacles required avoidance-oriented corrections.

Reactive Responses to Hazards

In contrast to the sustained adaptations observed during routine work-zone driving, the presence of unexpected hazards triggered localized reactive responses. Within event segments surrounding hazards, SDLP increased significantly, reflecting lateral maneuvers required to avoid obstacles encroaching on the driving lane. These behavioral changes were accompanied by increased horizontal gaze variability and higher electrodermal activity, indicating heightened visual scanning and physiological arousal during hazard encounters. The increase in collisions observed exclusively in the Hazard condition further supports the interpretation that unexpected hazards introduced additional task demands beyond those imposed by the work-zone configuration alone. Together, these results suggest that drivers transitioned from a stabilized control regime during sustained work-zone navigation to a reactive avoidance mode when confronted with sudden obstacles.

Multimodal Evidence of Driver State

An important contribution of the present study lies in the integration of behavioral, gaze-based, and physiological indicators to assess driver responses. Behavioral measures revealed changes in vehicle control, gaze metrics indicated increased visual monitoring, and electrodermal activity reflected elevated physiological arousal during hazard encounters. Together, these measures suggest that the introduction of hazards increased the overall difficulty of the driving task, leading drivers to exert greater control effort while maintaining heightened vigilance. Subjective workload and emotional ratings further support this interpretation, with participants reporting increasing mental and temporal demand across the experimental conditions. By examining sustained work-zone configurations and localized hazard events within a single experimental design, the present study addresses a gap in previous work,

which has typically investigated these sources of driver demand in isolation.

Implications for Simulator-Based Work-Zone Research

From a methodological perspective, the results demonstrate that the VR-based simulator elicited measurable changes in driver behavior and state in response to variations in roadway demand. The work-zone configuration alone produced systematic adaptations in vehicle control and visual monitoring behavior, while the addition of hazards elicited distinct reactive responses involving avoidance maneuvers and increased physiological arousal. These findings indicate that the simulator is sensitive to both sustained and event-driven changes in task demand within a controlled experimental setting. While the present study does not directly assess behavioral equivalence between simulated and real-world driving, the findings indicate that the simulator reliably captures systematic changes in driver behavior and state as roadway demand varies. In this sense, the study shows that such sensitivity extends to both sustained roadway configurations and localized hazard events within a single experimental framework. The results further suggest that different measurement domains capture complementary aspects of driver response. Behavioral and gaze-based measures reflected both sustained adaptations to the work-zone environment and localized responses to hazards, whereas physiological responses were primarily sensitive to transient, event-driven demand. This differentiation highlights the value of combining multiple indicators when assessing driver state in simulation environments. Taken together, these findings support the use of immersive driving simulation as a tool for experimentally evaluating driver behavior in temporary work-zone scenarios and for assessing the impact of both roadway configurations and localized hazards under controlled conditions.

Limitations

Several limitations should be considered when interpreting the results. First, the experimental design manipulated multiple aspects of the driving environment simultaneously, including lane closures, speed-limit reductions, and the introduction of hazards. While this approach increases ecological realism, it makes it difficult to isolate the contribution of individual factors to the observed behavioral changes. Second, although the simulator provided an immersive virtual environment, it lacked motion feedback. While previous studies have shown that visual and task-based cues can strongly influence driver behavior in simulation, the absence of vestibular and motion cues may affect the magnitude of certain responses. The present results demonstrate sensitivity to systematic changes in driver behavior as roadway demand varies, rather

than providing a direct validation of real-world behavioral equivalence. Finally, the traffic model and hazard exposure were intentionally limited: ambient vehicles were confined to the leftmost lane, and only six hazards were presented, increasing experimental control but reducing interaction complexity and possibly contributing to EDA habituation. Future studies should examine a wider range of traffic interactions, hazard types, and hazard frequencies.

Acknowledgments

This work was supported by SafeRoadWorks project, funded by the program “Digitale Technologien” (grant n. 910803) of the Austrian Federal Ministry for Innovation, Mobility and Infrastructure (BMIMI). The Austrian Research Promotion Agency (FFG) was authorized for the program management. The authors would like to thank Stelios Damalas, Nikolaus Kasper, and Marion Seidenberger.

References

- Bella, F., 2008. Driving simulator for speed research on two-lane rural roads. *Accident Analysis & Prevention*, 40(3), pp. 1078–1087.
- Bradley, M.M. and Lang, P.J., 1994. Measuring emotion: The Self-Assessment Manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), pp. 49–59.
- Boucsein, W., 2012. *Electrodermal activity*. 2nd ed. New York: Springer.
- Brookhuis, K.A. and de Waard, D., 2010. Monitoring drivers' mental workload in driving simulators using physiological measures. *Accident Analysis & Prevention*, 42(3), pp. 898–903.
- Chatterjee, K., Hounsell, N., Firmin, P. and Bonsall, P., 2018. Driver response to variable message sign information in work zones. *Transportation Research Part F: Traffic Psychology and Behaviour*, 56, pp. 241–255.
- Crundall, D., Chapman, P., Trawley, S., Collins, L., van Loon, E., Andrews, B. and Underwood, G., 2012. Some hazards are more attractive than others: Drivers' hazard perception and response. *Accident Analysis & Prevention*, 45, pp. 600–609.
- Edwards, J., Davey, J. and Armstrong, K., 2017. Temporary roadworks, driver behaviour and work zone safety: A human factors perspective. *Safety Science*, 98, pp. 21–30.
- Federal Highway Administration (FHWA), 2024. *Work zone road safety audits: Guidelines and prompt lists*. U.S. Department of Transportation.
- Huey, R., McDonald, A. and Porter, R., 2016. Driver workload and behavior in highway work zones. *Transportation Research Record*, 2555(1), pp. 1–9.
- Kaptein, N.A., Theeuwes, J. and Van der Horst, A.R.A., 1996. Driving simulator validity: Some considerations. *Transportation Research Record*, 1550, pp. 30–36.
- Lewis-Evans, B. and Rothengatter, T., 2009. Task difficulty, risk, effort and comfort in a simulated driving task – Implications for risk allostasis theory. *Accident Analysis & Prevention*, 41(5), pp. 1053–1063.
- Ma, Y., Li, Z. and Wang, Y., 2023. Driving workload and performance in highway work zone transition areas: A driving simulator study. *Applied Sciences*, 13(21), 11669.
- Mehler, B., Reimer, B. and Coughlin, J.F., 2012. Sensitivity of physiological measures for detecting systematic variations in cognitive demand from a working memory task: Implications for driver state monitoring. *Human Factors*, 54(3), pp. 396–412.
- Paazié, A., 2008. A method to assess the driver mental workload: The driving activity load index (DALI). *IET Intelligent Transport Systems*, 2(4), pp. 315–322.
- Pretto, P., Jukic, M., Prior, M., Gsellmann, I., Muik, T., & Rosenberger, M. (2026). Driver cognitive, emotional, and behavioral responses to single-day highway work zones with unexpected lane hazards. In *Proceedings of the Applied Human Factors and Ergonomics International Conference (AHFE 2026)*, Vol. XX.
- Savolainen, P.T., Lord, D. and McGee, H., 2024. Improving the effectiveness of merge control strategies in work zones. *Transportation Research Record*, 2678(3), pp. 456–468.
- Underwood, G., Chapman, P., Bowden, K. and Crundall, D., 2005. Visual search while driving: Skill and awareness during inspection of the scene. *Transportation Research Part F: Traffic Psychology and Behaviour*, 8(2), pp. 87–97.
- Wynne, R., Beanland, V., Salmon, P.M. and Lenne, M.G., 2019. Systematic review of driving simulator validation studies. *Transportation Research Part F: Traffic Psychology and Behaviour*, 65, pp. 138–159.

Expectation-Driven Differences in Pedestrian Gap Acceptance toward Automated Vehicles

Melina Bergen¹, Gerald Temme¹, Wolfgang Einhäuser², Martin Fischer¹

(1) German Aerospace Center (DLR), Institute of Transportation Systems, Lilienthalplatz 7, 38108 Braunschweig, Germany, e-mail: {melina.bergen, gerald.temme, ma.fischer}@dlr.de

(2) Chemnitz University of Technology, Reichenhainer Str. 70, 09126 Chemnitz, Germany, e-mail: {wolfgang.einhaeuser-treyer}@physik.tu-chemnitz.de

Abstract -

As automated vehicles (AVs) become increasingly integrated into urban traffic, understanding their interaction with pedestrians is essential. This study investigates pedestrian gap acceptance in interactions with AVs and human-driven vehicles (HVs) using a virtual reality experiment with continuous locomotion on an omnidirectional treadmill. Vehicle behavior was intentionally held identical to isolate expectation-based effects of vehicle type. The analysis focused on unregulated crossings with systematically varied time-to-arrival (TTA). Gap acceptance was primarily driven by TTA. When directly comparing AV and HV crossing opportunities within mixed trials, participants showed higher acceptance rates for gaps followed by HVs. This preference became more pronounced after repeated exposure. Subjective measures were not associated with observed behavioral differences. These findings suggest that under specific conditions, pedestrian decisions may be influenced not only by objective vehicle behavior, but also by expectations associated with vehicle type.

Copyright: © 2026 by the authors.

Licensee Driving Simulation Association.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.82157/dsa/2026/18>

Introduction

As automated vehicles (AVs) gradually become integrated into urban traffic, understanding their interactions with vulnerable road users (VRUs) such as pedestrians becomes essential. While pedestrians have developed expectations about conventional human-driven vehicles (HVs) through experience, interactions with AVs may introduce uncertainty, as AVs differ in consistency, rule adherence, and communication, potentially requiring pedestrians to adapt their decision-making strategies.

Previous work on AV interaction has primarily focused on trust and cooperation from the driver perspective, while far less attention has been paid to pedestrians and other VRUs, despite the fact that they must also interpret and respond to the AVs' behavior (Jayaraman, et al., 2019).

Pedestrians typically infer vehicle intent from three sources: explicit communication (e.g., driver gestures or external human-machine-interface eHMI; Dey, et al., 2021; Šucha, 2014), contextual cues such as traffic regulations (Havard and Willis, 2012), and implicit communication through vehicle kinematics (Rasouli and Tsotsos, 2019). While many studies have focused on explicit communication systems for AVs (Carmona, et al., 2021; Jayaraman, et al., 2019; Song, et al., 2023), no standardized communication system currently exists (Saleh, Hossny, and Naha-vandi, 2017), making it important to study interac-

tions under conditions where pedestrians must rely on other information sources.

At unregulated crossings, pedestrians rely on vehicle kinematics, trust, and prior expectations. As AVs are often associated with increased uncertainty, these expectations may directly shape crossing decisions (Jayaraman, et al., 2019; Rad, Almeida Correia, and Hagenzieker, 2020). Recent findings suggest that people respond differently to AVs and HVs even when vehicle behavior is identical. For example, Sun, Ge, and Qu (2024) showed in a simulator study, that drivers interacted differently with AVs compared to HVs despite identical kinematics and the only difference being the visual appearance of the vehicles. This indicates that expectations alone can influence behavior, which raises the question of whether pedestrians similarly adjust their crossing decisions based on perceived vehicle identity rather than observable motion cues. To isolate such expectation-based effects, the present study examines pedestrian interactions with AVs and HVs while keeping vehicle kinematics identical across conditions. The study focuses on unregulated crossings without explicit communication, where pedestrians must rely primarily on implicit expectations when deciding whether to cross.

Previous studies have often placed pedestrians directly at the curb using button-press tasks, simplified virtual reality (VR) setups, or projection-based environments. While these approaches provide strong



Figure 1: Omnideck, MoSAIC Laboratory at the Institute of Transportation Systems, Braunschweig

experimental control, they constrain natural locomotion and remove the continuous approach phase that characterizes real-world street crossings. Since crossing decisions often begin already during approach rather than only at the curb (Gorrini, et al., 2018; Rasouli, Kotseruba, and Tsotsos, 2017), preserving continuous locomotion is important for capturing realistic pedestrian behavior.

To address this, in the present study an omnidirectional treadmill with a high-fidelity virtual environment is used. A realistic time-pressure task was additionally implemented to promote ecologically valid decision-making and reduce overly cautious simulator behavior. The present paper reports a focused analysis of a subset of conditions from a larger ongoing experimental study. Specifically, we analyze behavioral differences between AV and HV interactions in unregulated crossings, as well as potential adaptation effects over repeated exposure. Based on these considerations, the following research questions are addressed:

- Does pedestrian gap acceptance differ between interactions with AVs and HVs when vehicle behavior is held constant?
- Do potential differences emerge only after repeated exposure, indicating learning or adaptation effects over time?
- Do subjective impressions and expectations toward AVs relate to observed crossing decisions?

Method

Experimental Setup

The study was conducted using the Omnideck, an omnidirectional treadmill by *W5 Omnifinity AB* (see Fig. 1). The system consists of a circular platform (diameter 3.7 m) with motorized roll segments that continuously move users back toward the center, enabling unrestricted movement in all directions within a virtual environment (Guilfoyle and Thor, 2023; Theisen, et al., 2025). The virtual environment was presented via a *Varjo XR-3* head-mounted display. Participants walked through a virtual city toward a destination that required crossing three sequential roads. The first two roads consisted of a single lane, while the third road additionally included a bus lane (see Fig. 2). However, all roads were designed to have the same overall width, ensuring comparable crossing distances across conditions. Vehicles approached from the left and were presented as either AVs or HVs. No explicit communication cues

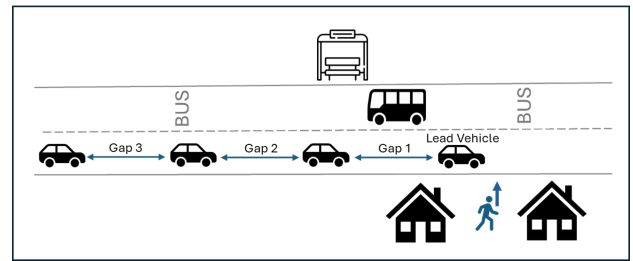


Figure 2: Example of the third street, where the bus stop was located.

(e.g., driver gestures or eHMI) were present. Vehicle kinematics, including speed and braking behavior, were intentionally held identical across conditions, such that participants could only rely on implicit expectations associated with vehicle type when making crossing decisions.

On the third road, a bus approached from the right as part of a broader time-pressure task. However, the bus always passed the crossing area before the participant reached the road, ensuring that it did not influence crossing decisions.

Behavioral measures focused on pedestrians' gap acceptance decisions. Gap acceptance was defined as the crossing within a temporal gap between two consecutive vehicles within a traffic stream. Based on the position of the pedestrian relative to the sequence of vehicles, gaps were classified as accepted or rejected.

In addition, subjective measures were collected before and after the experiment, including familiarity with AVs, perceived likelihood of being detected by the vehicle, appropriateness of vehicle reactions, and safety concerns. These measures were explored in relation to observed behavioral differences.

Participants

A total of 25 participants took part in the study (30% female, 70% male; Age: $M = 35$ years, $SD = 10.57$). The ethics commission of the DLR determined that no formal ethical review was required. One participant was excluded from the behavioral analyses due to an incorrect simulator configuration during data collection. Thus, the final sample for the gap acceptance analyses consisted of 24 participants. Due to a technical issue with the questionnaire tool, subjective questionnaire data were only available for a subset of participants ($n = 19$).

Design and Procedure

A within-subjects design was used with vehicle type (AV vs. HV) as the main factor. Participants completed multiple crossing trials in which vehicles approached the pedestrian with a speed of 30 km/h and with systematically varied time-to-arrival (TTA), representing the temporal gaps between consecutive vehicles (5 s, 6 s, & 7 s). This paper reports a focused analysis of a subset of conditions from a larger experimental study. While the full design included additional crossing contexts, the present analysis focuses exclusively on unregulated crossings.

Across the experiment, each participant completed

10 unregulated crossing scenarios, including four mixed (AV and HV), three HV-only, and three AV-only conditions (N = 240 street trials overall). Each street trial included three consecutive gap opportunities.

Mixed trials were presented at the beginning and end of the experiment, in order to assess whether potential behavioral differences between interactions with AVs and HVs change after repeated exposure. The AV-only and HV-only conditions were counterbalanced across participants. Vehicle kinematics were identical for both vehicle types, ensuring that any observed differences in behavior could be attributed to vehicle type rather than kinematic differences.

To ensure that participants could reliably distinguish both vehicle types, written instructions were provided prior to the experiment, including an image of an AV. Vehicles differed clearly in visual appearance. All AVs looked identical and passengers were seated facing backwards, providing a salient visual cue that distinguished them from HVs. In contrast, HVs resembled conventional vehicles familiar to participants.

To encourage naturalistic decision-making, participants were instructed to reach a bus stop within a limited time window. Missing the bus resulted in a delay before the next trial. Remaining time and navigation instructions were continuously displayed on a virtual smartwatch worn on the right wrist.

Prior to the main experiment, participants completed training trials to familiarize themselves with the simulator and task.

Results

To investigate the influence of temporal gap size and vehicle type on pedestrian gap acceptance, generalized estimating equation (GEE) models were employed in combination with follow-up analyses such as Chi-square tests. GEE models were chosen to account for the non-independence of observations within participants.

Gap Acceptance

Across all unregulated trials, gap acceptance rates increased with increasing TTA, indicating that participants were more likely to cross when presented with larger temporal gaps. This was confirmed by a GEE model, which revealed a significant main effect of TTA ($\beta = 1.09, SE = 0.40, p = .007$).

In addition, a significant main effect of gap position was observed ($\beta = -2.43, SE = 1.20, p = .042$), indicating that participants were less likely to accept later gaps compared to earlier ones. No significant interaction between TTA and gap position was found. Fig. 3 shows gap acceptance as a function of TTA and gap position. While acceptance rates increased with TTA across all gaps, differences between gap positions are visible, with lower acceptance rates for later gaps. This likely reflects the sequential task structure, as later gaps were only encountered after earlier gaps had already been rejected. As a result, participants with a generally higher willingness to cross were more likely to accept earlier gaps, meaning later gaps were primarily encountered by more cautious participants. In addition, the first gap was evaluated during continuous approach, whereas later gaps were observed from a stationary position at the curb, potentially affecting movement dynamics and gap perception.

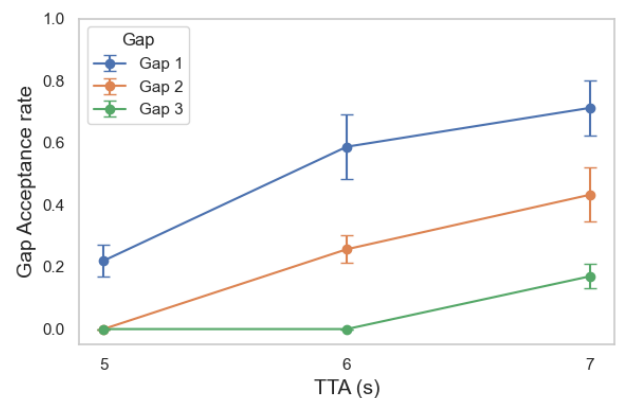


Figure 3: Gap acceptance by TTA and gap position (unregulated trials, excluding training). Values represent participant means; error bars indicate standard errors.

Vehicle-Type Effects on Gap Acceptance

The single-vehicle-type blocks showed no consistent differences between AVs and HVs. Instead, acceptance rates primarily reflected block order, with higher acceptance rates in the second block irrespective of vehicle type, suggesting a general learning or habituation effect.

Mixed trials were specifically designed to allow direct comparisons between interactions with both vehicle types within the same scenario. The following analyses therefore focus on these trials. In these trials, participants were repeatedly faced with immediate choices between gaps followed by either AVs or HVs, allowing a more direct comparison than in single-vehicle-type trials. Gaps were categorized according to the surrounding vehicles as either AV–HV or HV–AV. Across mixed trials, acceptance rates were consistently higher for gaps followed by HVs (37%) than for gaps followed by AVs (22%). This pattern was visible for both small gaps (29% vs. 12%) and intermediate gaps, where the strongest difference emerged (48% vs. 20%).

Mixed trials were presented at both the beginning and end of the experiment, in order to compare whether this preference changed after repeated exposure. Early trials showed no consistent pattern of vehicle-type differences, with only an isolated effect for the largest gap (7 s: $\chi^2 = 3.99, p = .046$). In contrast, late trials revealed significant differences for both small and intermediate gaps (5 s: $\chi^2 = 4.94, p = .026$; 6 s: $\chi^2 = 9.60, p = .002$), with higher acceptance rates for gaps followed by HVs. A GEE analysis confirmed this pattern, showing a significant vehicle-type effect for intermediate gaps in late mixed trials ($\beta = -2.87, SE = 1.013, p = .005$).

Subjective Measures

Although vehicle kinematics were identical across conditions, 31% of participants reported perceiving behavioral differences between AVs and HVs, while 47% reported perceiving no differences and 22% were undecided.

No significant differences between vehicle types

were found regarding perceived likelihood of being detected by the vehicle, perceived appropriateness of vehicle behavior, or safety concerns. Participants reported moderate familiarity with AVs ($M = 2.85$, $SD = 0.75$). Higher familiarity was associated with more positive evaluations of automated vehicle behavior, shown by a significant correlation with perceived appropriateness of vehicle reactions ($r(17) = .51$, $p = .027$). However, neither familiarity with AVs nor any subjective rating measures significantly predicted gap acceptance behavior.

Discussion

The present study investigated whether pedestrians respond differently to AVs and HVs when objective vehicle behavior is held constant. Across all trials, gap acceptance was primarily determined by the temporal and spatial characteristics of approaching vehicles, consistent with previous research (Oxley, et al., 2005; Rasouli and Tsotsos, 2019). However, analyses of mixed comparison trials, specifically designed to allow direct comparisons between both vehicle types within the same scenario, revealed systematic differences in crossing behavior. Participants more frequently accepted gaps followed by HVs, with effects becoming stronger after repeated exposure and most pronounced for intermediate gap conditions. While these differences were apparent in mixed trials, where participants had the choice of crossing either before an AV or an HV, single-vehicle-type trials were less sensitive to such effects, as participants had to select among gaps involving only one vehicle type. Since vehicle kinematics were intentionally identical across conditions, these differences cannot be explained by objective behavioral differences, but from expectation-based responses toward different vehicle types. This interpretation is supported by the observation that nearly one third of participants reported perceiving behavioral differences between AVs and HVs despite objectively identical vehicle behavior. Vehicle-type differences emerged primarily after repeated exposure and were strongest for intermediate temporal gaps, suggesting that expectation-based effects may become more relevant under conditions of greater decision uncertainty. In contrast, when gaps were clearly safe or unsafe, objective traffic cues such as TTA likely dominated decision-making regardless of vehicle type.

Several limitations should be considered. Although vehicle behavior was intentionally controlled to isolate expectation-based effects, this may have reduced ecological validity, as real-world differences between AVs and HVs often involve behavioral variation. In addition, repeated exposure may have introduced learning effects independent of vehicle type, and the sequential gap structure created differences between early and later crossing opportunities, as only the first gap was evaluated during continuous approach. Finally, the relatively small sample size limits generalizability.

Overall, the findings suggest that while objective traffic cues dominate most pedestrian decisions, expectation-based effects may influence behavior when pedestrians directly compare different vehicle types under conditions of uncertainty. Future studies should examine whether similar effects emerge in more naturalistic interactions involving more distinct behavioral differences between AVs and HVs.

Acknowledgment

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.



Funded by
the European Union

References

- Carmona, J., Guindel, C., Garcia, F., and De La Escalera, A., 2021. eHMI: Review and guidelines for deployment on autonomous vehicles. *Sensors*, 21(9), p. 2912.
- Dey, D., Matvienko, A., Berger, M., Pfleging, B., Martens, M., and Terken, J., 2021. Communicating the intention of an automated vehicle to pedestrians: The contributions of eHMI and vehicle behavior. *it – Information Technology*, 63(2), pp. 123–141.
- Gorini, A., Crociani, L., Vizzari, G., and Bandini, S., Nov. 2018. Observation results on pedestrian-vehicle interactions at non-signalized intersections towards simulation. *Transportation Research Part F: Traffic Psychology and Behaviour*, 59, pp. 269–285. <https://doi.org/10.1016/j.trf.2018.09.016>.
- Guilfoyle, R. and Thor, P., 2023. *Omnideck – Product Sheet*. Available at: <http://bit.ly/2VpakmLi>.
- Havard, C. and Willis, A., 2012. Effects of installing a marked crosswalk on road crossing behaviour and perceptions of the environment. *Transportation Research Part F: Traffic Psychology and Behaviour*, 15(3), pp. 249–260.
- Jayaraman, S. K., Creech, C., Tilbury, D. M., Yang, X. J., Pradhan, A. K., Tsui, K. M., and Robert, L. P. J., 2019. Pedestrian trust in automated vehicles: Role of traffic signal and AV driving behavior. *Frontiers in Robotics and AI*, 6, p. 117.
- Oxley, J. A., Ihlen, E., Fildes, B. N., Charlton, J. L., and Day, R. H., 2005. Crossing roads safely: an experimental study of age differences in gap selection by pedestrians. *Accident Analysis & Prevention*, 37(5), pp. 962–971.
- Rad, R. S., Almeida Correia, G. Homem de, and Hagenzieker, M., 2020. Pedestrians' road crossing behaviour in front of automated vehicles: Results from a pedestrian simulation experiment using agent-based modelling. *Transportation Research Part F: Traffic Psychology and Behaviour*, 69, pp. 101–119.
- Rasouli, A., Kotseruba, I., and Tsotsos, J. K., 2017. Are they going to cross? a benchmark dataset and baseline for pedestrian crosswalk behavior. In: *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 206–213.
- Rasouli, A. and Tsotsos, J. K., 2019. Autonomous vehicles that interact with pedestrians: A survey of theory and practice. *IEEE transactions on intelligent transportation systems*, 21(3), pp. 900–918.
- Saleh, K., Hossny, M., and Nahavandi, S., 2017. Towards trusted autonomous vehicles from vulnerable road users' perspective. In: *Proceedings of the 2017 Annual IEEE International Systems Conference (SysCon)*. IEEE, pp. 1–7.
- Song, Y., Jiang, Q., Chen, W., Zhuang, X., and Ma, G., 2023. Pedestrians' road-crossing behavior towards eHMI-equipped autonomous vehicles driving in segregated and mixed traffic conditions. *Accident Analysis & Prevention*, 188, p. 107115.
- Šucha, M., 2014. Road users' strategies and communication: Driver-pedestrian interaction. In: *Proceedings of the Transport Research Arena (TRA)*. Vol. 1.
- Sun, H., Ge, Y., and Qu, W., 2024. Greater prosociality toward other human drivers than autonomous vehicles: Human drivers' discriminatory behavior in mixed traffic. *Accident Analysis & Prevention*, 203, p. 107623.
- Theisen, M., Bergen, M., Einhäuser, W., and Schießl, C., 2025. Head movements predict pedestrian crossing decisions earlier than walking speed. *Transportation Research Part F: Traffic Psychology and Behaviour*, 113, pp. 586–608.

Using Mixed Reality for Safety-Critical Testing in Railway Systems

Raphael Kling, Manuel Wild, Marvin Wieschollek, Martin Cichon

Karlsruhe Institute of Technology - Institute of Vehicle System Technology, 76131 Karlsruhe Rintheimer Querallee 2, e-mail: {raphael.kling, manuel.wild, marvin.wieschollek, martin.cichon}@kit.edu

Abstract – Rail transport is essential for meeting climate goals, but challenges such as congestion, high costs, staff shortages, and low predictability limit efficiency. Automated Train Operation can improve availability and reduce costs, yet the development requires extensive safety validation. Field testing is limited by infrastructure constraints, operational disruptions, and the rarity of hazardous scenarios. This paper proposes a Mixed-Reality-based validation approach that injects virtual sensor data into real-world sensor streams, enabling safe and realistic testing of rare and safety-critical situations. The method enhances field-test value, improves validation completeness, and helps bridge the simulation-to-reality gap for automated rail systems.

Keywords: Vehicle In the Loop, Augmentation, Scaled Testing, Automated Train Operation

Copyright: © 2026 by the authors.

Licensee Driving Simulation Association.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.82157/dsa/2026/19>

Introduction

Rail transportation is one of the safest modes of transport due to strict regulatory requirements and comprehensive validation of safety-critical systems. However, railway operators face increasing challenges, including growing traffic demand, limited network capacity, and a demographic shift that is leading to a shortage of qualified operational staff. Automated Train Operation (ATO) is therefore considered a key technology for improving operational efficiency, availability, and capacity utilization. While ATO is already established in metro systems and isolated industrial railways (Rossberg 2008; Rio Tinto 2017), its deployment in open railway networks remains challenging.

On-sight operations including shunting are particularly demanding because trains often operate without signal protection. Continuous observation of the environment is required, including the detection of obstacles, assessment of switch positions and track conditions, and the corresponding adaptation of train movements. For automation, these perception- and decision-making responsibilities must be transferred to the automated system. For higher Grades of Automation (GoA), the system must therefore be capable of perceiving, interpreting, and responding to dynamic environmental conditions while ensuring safe operation (IEC 2014).

The validation of such perception-driven systems requires extensive testing under a wide range of operational conditions. Although field testing provides

realistic results, it is costly, difficult to scale, and often unacceptable for hazardous scenarios. Simulation offers a safe, repeatable, and cost-effective alternative, but purely virtual environments are limited by the simulation-to-reality gap (Manivasagam et al. 2023). To address these challenges, this paper presents a Mixed Reality (MR) approach for railway applications that combines simulation and field testing by augmenting real sensor streams with virtual sensor data. By synchronizing real and virtual environments, the approach enables safe and realistic testing of safety-critical railway scenarios while reducing the limitations of both purely simulated and purely real-world testing. To enable faster development of this approach without relying on field testing, scaled testing is used for development (Hofmeier and Cichon 2025a).

State of the Art and Research Gap

The validation of highly automated vehicles is increasingly shifting from mileage-based testing toward scenario-based methodologies, as rare safety-critical events cannot be efficiently covered through real-world operation alone (Bagschick et al. 2018). To achieve systematic and traceable validation, the Operational Design Domain (ODD) defines the conditions under which the System under Test (SuT) is expected to operate safely, while scenario-based methods derive representative test cases from these conditions (SAE International 2021).

To evaluate such scenarios, Software-in-the-Loop, Hardware-in-the-Loop, and Vehicle-in-the-Loop progressively integrate real-world components into simulation environments (Park, Chung, and Lee 2020). These approaches enable efficient, repeatable, and increasingly realistic testing of scenario-based test cases. However, they remain limited in accurately capturing the interaction between perception systems and real operating conditions. At the same time, purely field-based validation is often impractical, as hazardous and safety-critical scenarios occur only rarely and cannot or are unacceptable to be reproduced on demand. While the automotive domain benefits from reusable and sensor-realistic ADAS test dummies, comparable solutions are difficult to implement in railway environments because trains are constrained to fixed tracks. Consequently, a gap remains between scalable simulation-based testing and realistic field validation, highlighting the need for alternative validation approaches.

To address this gap, this paper presents a MR validation framework for railway environments that integrates virtual sensor data into real sensor data. The implementation is based on Vir2Rail (Hofmeier 2026), the approach combines the scalability and repeatability of simulation with the realism of field testing, enabling the validation of safety-critical railway scenarios while reducing the simulation-to-reality gap.

Methodology

This chapter introduces the general approach of this method. The objective is to create synthetic sensor measurements that are indistinguishable from those produced by physical objects, enabling safe testing of safety-critical scenarios while preserving the realism of the surrounding environment. By combining real and virtual elements, the approach reduces the simulation-to-reality gap and supports realistic validation under controlled conditions. Although demonstrated using LiDAR and RGB camera, the concept is transferable to other sensor modalities, including radar and ultrasonic sensors. Successful implementation requires compatible virtual object representations and continuous spatiotemporal synchronization between real and virtual worlds.

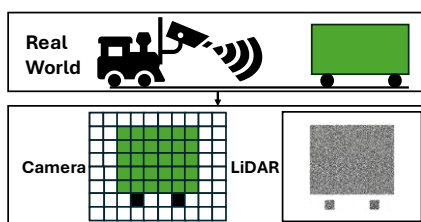


Figure 1 Real-World Sensor Data Generation Process

Figure 1 illustrates the sensor data generation process in a railway scenario when approaching a wagon. An RGB camera records image data as pixel

arrays influenced by optical and sensor characteristics, while LiDAR generates point clouds representing distance measurements derived from reflected laser signals.

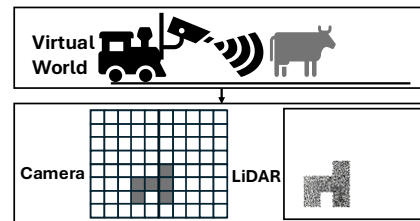


Figure 2 Virtual Sensor Data Generation Process

Figure 2 shows the process for virtual sensor data generation. The virtual train with similar sensor setup as in Figure 1 is approaching the virtual obstacle to be augmented, in this case a cow.

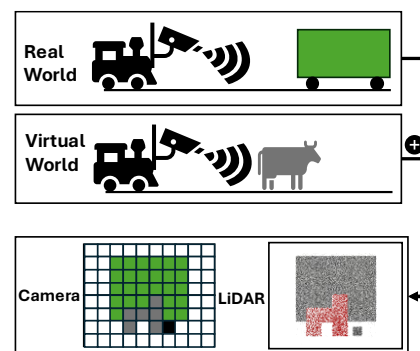


Figure 3 Fused Sensor Data

Figure 3 illustrates the fusion of real sensor data with virtual objects. In this example, a virtual cow is inserted into the sensor stream of a train approaching a wagon. For the LiDAR, the wagon is shown in black, while the virtual cow is represented by red points. Pixels belonging to the virtual cow replace the corresponding wagon pixels while virtual points from the LiDAR overwrite real points, enabled by the availability of pixel- and point-wise ground-truth annotations in virtual sensor data. By aligning real and virtual train trajectories and using realistic sensor models, virtual obstacles can be accurately positioned within real data streams.

The central hypothesis is that perception systems respond to virtual obstacles in the same way as real ones. To validate this, datasets with real and augmented objects within the same scenery will be created and evaluated. This enables testing the reaction of already existing object detectors.

Implementation

To reduce cost and increase availability while supporting fast adaptation and transferability of the results to full-scale applications at first a 1:32 scaled environment is used which is combined with a digital twin, the setup is similar to (Hofmeier and Cichon 2025b). After successful implementation in 1:32 scale the transformation to full scale is shown.

Creating Virtual Sensor Data

To ensure realism, both the geometric perspective between sensor and environment and the underlying sensing principles must be accurately represented and match between the virtual and the real-world.

Correct perspective is achieved by leveraging the fact that railway vehicles operate on predefined trajectories. Track data is imported in GeoJSON format into Unreal Engine (UE), where it is used to construct a virtual track. A virtual locomotive model matching the real train in geometry, axles, and bogies is constrained to this track. Virtual sensors are then positioned to mirror the placement of their real counterparts on the locomotive. Together, the real track layout, realistic vehicle model, and accurate sensor placement ensure a faithful geometric representation.

In addition, sensor-specific measurement principles are modelled. A LiDAR module from Vir2Rail is adapted, and a camera module is newly developed. Both can be parametrized to replicate the behaviour of their real-world counterparts and tag which pixels/points belong to specific virtual objects.

Scaled Implementation

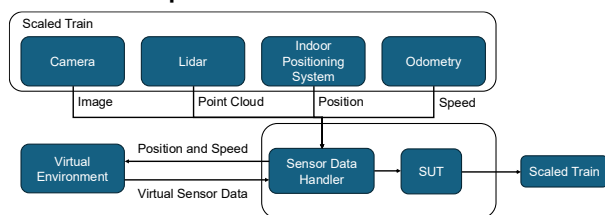


Figure 4 Concept: MR within the Scaled Environment

Figure 4 illustrates the implementation of MR within the scaled environment. The scaled train is equipped with an RGB camera, LiDAR, an odometry sensor and an indoor positioning system serving as a GNSS surrogate (Hofmeier and Cichon 2025a). A Jetson Nano is used for data processing and train control. For each test scenario, the selected track for the real train and its virtual replica is defined in advance, together with the identifiers and spatial coordinates of the augmented objects.

The functionality for augmenting virtual obstacles, introduced in Figure 3, is implemented in the sensor data handler. This component is coupled with the virtual environment in UE, which includes virtual sensor models aligned with the physical sensor setup of the SuT similar to (Hofmeier and Cichon 2025b). All sensors provide their information to the sensor data handler running on the Jetson Nano. The handler forwards the train position and speed to the virtual environment. Based on this information, the virtual train is aligned with the track in the virtual world. Once the virtual train is positioned, virtual sensor data is generated and sent back to the sensor data handler. The handler then combines the real and virtual sensor data and forwards the combined output to the SuT.

The SuT processes this enriched data stream without knowledge of the augmentation and controls the scaled train accordingly.

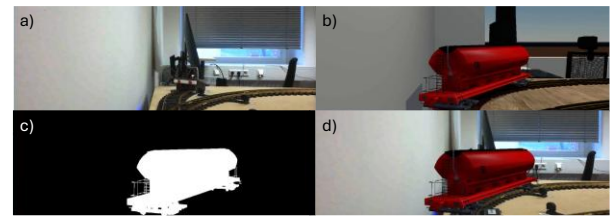


Figure 5 Results Scaled MR Camera

Figure 5 illustrates the augmentation process by combining a real image (a) with a virtual image (b) using a virtual segmentation mask (c), resulting in the final MR image (d). The augmented image is generated by fusing real and virtual images based on the mask. Pixels corresponding to white regions in the mask are taken from the virtual image, while pixels corresponding to black regions are retained from the real image. Consequently, the virtual wagon is seamlessly integrated into the real scene, producing an image that contains both real and virtual elements.

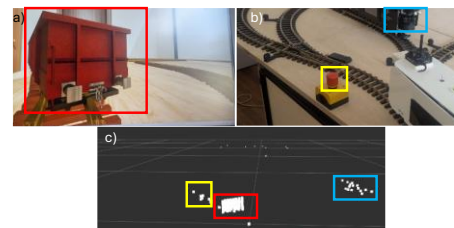


Figure 6 Idealized Results Scaled MR LiDAR

Figure 6 illustrates LiDAR augmentation in the scaled environment, showing the original scene (a), the virtual scene (b), and the resulting combined point cloud (c). The virtual wagon can be seen augmented into the real point cloud; for better visualization, all ground points were removed. The virtual point cloud is generated using ground truth point to object mapping and merged with the real data. To model realistic occlusions, a spherical occupancy grid based on LiDAR azimuth and elevation bins is created, allowing real points behind virtual objects to be removed (Manivasagam et al. 2023).

Transfer of the Method to Full Scale

After developing the methods within the scaled environment, the method was transferred to full scale. To enable scalability, recorded but unprocessed data is used. The data was recorded within the Munich Nord Cargo Yard using a class BR 294 locomotive with an in-house sensor setup. The locomotive and the associated track infrastructure were already available within the UE environment from previous work, which served as the basis for the development of Vir2Rail (Hofmeier 2026). Finally, virtual sensors had been placed at their real counterparts' positions while also adapting the sensing principles to follow their real counterparts.

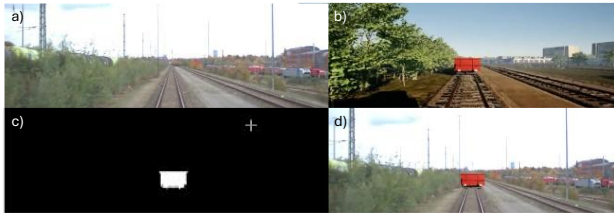


Figure 7 Results Full Size MR Camera

Figure 7 illustrates the fusion of virtual and real camera data on the shunting yard recordings augmenting a virtual wagon. The arrangement of the images within Figure 7 is similar to Figure 5.

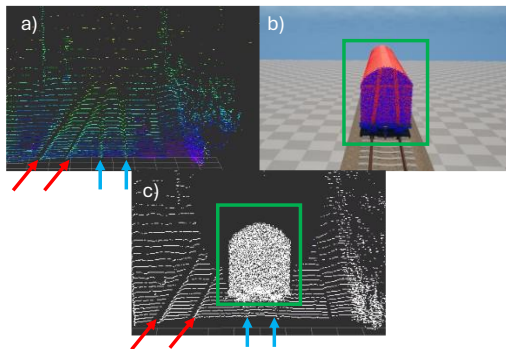


Figure 8 Results Full Size MR LiDAR

Figure 8 shows the full-scale MR Lidar scenario with a) the real scene, b) the virtual scene and c) the combined scene. The methodology is similar to the scaled approach. Both camera and LiDAR augmentations show that the concept in general is applicable.

Outlook

This paper demonstrates the feasibility of the proposed MR concept for enhancing railway field testing through the rapid generation of safety-critical scenarios. However, several challenges remain. First, the method requires more comprehensive validation by comparing the SuT response to virtual and real obstacles, as well as through pixel- and pointwise analyses of virtual and real scenes. Second, the visual quality of inserted objects, particularly for camera-based perception, is affected by inconsistent lighting. This will be addressed by AI-based lighting adaptation using environmental light sensors, which will also enable realistic shadow generation. Furthermore, the current approach depends on accurate track-related data to represent train movement within the virtual environment. Future work will investigate the use of rail detectors to automatically map objects to the railway infrastructure, reducing this dependency. Additionally, the framework will be extended by dynamic virtual objects that can react to real obstacles. Finally, the method shall also be applied in real time at full scale.

Acknowledgements

This work was partly accomplished within the project VAL, FKZ 53T200008KT, EBA Ref. 8fd/003-1255#008-VAL, funded by the German Federal Ministry of Transport. We also thank DB Cargo for its co-operation.

References

- Hofmeier, Tobias. 2026. "Vir2Rail – Modular LiDAR Simulation Framework for ATO Research in Railway Environments". <<https://doi.org/10.5281/zenodo.18769047>> [accessed 16 March 2026].
- Hofmeier, Tobias and Martin Cichon. 2025a. "Abschlussbericht EduRail - Educational Railway Demonstrator for Future Mobility". Projektbericht. <<https://doi.org/10.5445/IR/1000180934>> [accessed 17 September 2025].
- Hofmeier, Tobias and Martin Cichon. 2025b. "Scaled Model Development and Testing for Automatic Train Operation: Creating a Digital Twin". *SN Computer Science* 6. <<https://doi.org/10.1007/s42979-025-04529-6>> [accessed 4 June 2026].
- IEC. 2014. "IEC 62290-1:2014 Railway Applications – Urban Guided Transport Management and Command/Control Systems – Part 1: System Principles and Fundamental Concepts". IEC.
- "J3016_202104 : Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles - SAE International". 2021. 2021 <https://www.sae.org/standards/j3016_202104-taxonomy-definitions-terms-related-driving-automation-systems-road-motor-vehicles> [accessed 9 July 2026].
- Manivasagam, Sivabalan, Ioan Andrei Bărsan, Jingkang Wang, Ze Yang and Raquel Urtasun. 2023. "Towards Zero Domain Gap: A Comprehensive Study of Realistic LiDAR Simulation for Autonomy Testing". In: . *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. Paris, France: IEEE. 8238–8248. <<https://doi.org/10.1109/ICCV51070.2023.00760>> [accessed 27 February 2026].
- Park, Changwoo, Seunghwan Chung and Hyeongcheol Lee. 2020. "Vehicle-in-the-Loop in Global Coordinates for Advanced Driver Assistance System". *Applied Sciences* 10: 2645. <<https://doi.org/10.3390/app10082645>>.
- Rio Tinto. 2017. "Rio Tinto Completes First Fully Autonomous Rail Journey in Western Australia". October 2, 2017 <<https://www.riotinto.com/en/news/releases/2019/autohaul-successfully-deployed>> [accessed 6 April 2023].
- Rossberg, Ralf Roman. 2008. "In Nürnberg Fahren U-Bahnen Der Linie 3 Nunmehr Ohne Führer". *Ingenieur.de - Jobbörse Und Nachrichtenportal Für Ingenieure*, June 13, 2008. <<https://www.ingenieur.de/technik/fachbereiche/verkehr/in-nuernberg-fahren-u-bahnen-linie-3-nunmehr-fuehrer/>>.

SCENARIO MODELING, CONTROL AND AUTHORING TOOLS

A Unified Framework for Scenario Coverage Estimation and Representative Selection in Simulation-Based Testing of Automated Vehicles

Anasse Essalih^{1,2}, Abdel-Hakim Arab^{1,3}, Laurent Durville¹, Mohamed-Cherif Rahal¹,
Mustapha Lebbah², Walid Kherji¹, Fouad Hadj Selem¹

(1) VEDECOM Institut, firstname.lastname@vedecom.fr

(2) DAVID Lab, UVSQ, Paris-Saclay University, Versailles, France, firstname.lastname@uvsq.fr

(3) ESTACA'Lab—Paris-Saclay, ESTACA, firstname.lastname@estaca.fr

Abstract - Scenario-based testing is now central to automated vehicle validation, yet a key challenge remains: how to select a limited set of scenarios that faithfully represents a given Operational Design Domain (ODD). The combinatorial explosion of scenario types, continuous parameters, and categorical factors makes exhaustive testing infeasible and current approaches fragmented across scenario classes. This paper proposes a unified smart sampling framework for multi-scenario selection: Driving situations are encoded as time series in a scenario-agnostic representation and projected into a common latent space, enabling the selection of a minimal, statistically representative subset that preserves both inter-scenario proportions and intra-scenario distributions, including categorical conditions. Applied to more than 280,000 real-world scenarios, the method reduces the test set by up to 97% while maintaining statistical fidelity. It provides quantifiable guarantees on representativeness, directly linking sample size to uncertainty, and identifies missing conditions for targeted data completion. This approach enables manufacturers to drastically reduce testing cost while preserving coverage and reliability.

Keywords: Scenario-based testing, representative selection, coverage estimation, generative modeling

Copyright: © 2026 by the authors.

Licensee Driving Simulation Association.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.82157/dsa/2026/20>

Introduction

The rapid deployment of automated driving systems requires safety validation methods that are both rigorous and scalable. Scenario-based testing has emerged as the industrial standard, combining large-scale data collection, scenario extraction, test selection, and residual risk estimation within a given Operational Design Domain (ODD) Gelder, et al., 2025. However, a central challenge remains: ensuring that the selected test scenarios are both representative of real-world conditions and sufficient to support reliable safety conclusions, Riedmaier, et al., 2020. In practice, the diversity of driving situations leads to a combinatorial explosion of scenarios, driven by heterogeneous behaviors, continuous parameters, and contextual categorical factors such as environment, infrastructure, or location. Existing approaches typically address completeness and representativeness within individual scenario classes Gelder, et al., 2025; Riedmaier, et al., 2020. While effective in restricted settings, this per-class strategy becomes impractical at the scale of a full ODD, where multiple scenario types must be considered jointly. As a result, it prevents a global assessment of coverage and lim-

its the ability to identify missing or underrepresented conditions.

A key limitation explaining this fragmentation lies in the way scenarios are traditionally parameterized Menzel et al. 2018. Each scenario type is described using variables that are specific to its own dynamics, often defined at particular key moments. For instance, a braking scenario may be characterized by speed or distance at the time of deceleration, while a lane-change or cut-in scenario relies on variables defined at the moment of lateral interaction. These definitions are not shared across scenario classes, making it impossible to construct a unified representation where all scenarios can be jointly analyzed. Consequently, previous methods could not project heterogeneous scenarios into a common space and apply a single selection or coverage estimation strategy. In our previous work Hadj Selem, et al., 2024, we introduced a latent-space-based framework to estimate completeness and select representative scenarios within a single scenario class. While this approach demonstrated strong potential for reducing dataset size and quantifying coverage, it inherently relied on homogeneous parameterizations and could not be extended to multi-scenario datasets for the

reasons described above. A second limitation concerns the integration of categorical variables. In realistic ODDs, factors such as weather, lighting conditions, infrastructure type, or country are essential dimensions of variability Gelder, et al., 2025. However, existing approaches either ignore these variables or handle them through dataset fragmentation (e.g., separate analyses per condition). When directly integrated into generative models, they often lead to poorly structured latent spaces, preventing reliable partitioning and statistically controlled sampling Hadj Selem, et al., 2024.

By contribution. This paper addresses these limitations by proposing a unified framework for multi-scenario coverage estimation and representative selection. The key idea is to represent all driving situations in a common scenario-agnostic form based on time series, enabling their projection into a shared latent space independent of scenario classes. In addition, we introduce a deterministic normalization step that restores a well-structured latent space even in the presence of mixed continuous and categorical variables, which is essential for enabling statistically controlled sampling. From an industrial perspective, the proposed approach provides a practical solution to drastically reduce test scenarios while preserving statistical fidelity to real-world data. It enables direct control of the cost-uncertainty trade-off, identifies missing ODD conditions, and strengthens safety validation within frameworks such as the SAF Gelder, et al., 2025.

Related Work

Scenario-based testing is now the reference framework for the safety validation of automated driving systems. Within this paradigm, two complementary challenges must be distinguished: *scenario generation*, which creates new scenarios, and *scenario selection*, which extracts representative subsets from real-world data.

Scenario generation approaches Scenario generation relies on the functional-logical-concrete hierarchy introduced by Menzel et al. 2018, where logical scenarios are parameterized and sampled to produce concrete instances. Extensions include structured logical scenario definition Weber, et al., 2019 and data-driven parameter estimation with representativeness metrics Gelder, et al., 2022.

Adversarial and importance-sampling approaches further improve efficiency by focusing on safety-critical situations, as in Feng, et al., 2021. In parallel, several frameworks translate abstract or recorded scenarios into executable simulations (e.g., CARLA). While effective for validation, these methods do not address the selection of representative subsets from real-world datasets.

Scenario selection and coverage from real data

Despite its importance, scenario selection from large naturalistic datasets remains less explored Riedmaier, et al., 2020; Zhong, et al., 2021. Coverage-oriented approaches Gelder, et al., 2019; 2025; Wang, Liu, and Zhao, 2017 define metrics such as completeness and diversity but do not provide operational selection procedures. Data-driven methods extract or prioritize relevant scenarios from real data, e.g., challenging scenario identification Ponn, et al., 2020 or realistic parameter estimation Nakamura, et al., 2022. More recently, Chen, et al., 2026 formulates scenario selection as a sampling optimization

problem balancing representativeness and safety-critical coverage. However, these approaches remain feature-based and lack explicit control over statistical uncertainty in heterogeneous datasets, and therefore cannot be directly extended to multi-scenario selection.

Critical scenario identification and falsification A complementary research direction focuses on identifying failure-inducing scenarios rather than preserving data distributions. Methods such as Song, et al., 2023 and the survey in Zhang, et al., 2023 highlight the importance of search-based and optimization-driven testing. Extensions including parameter coverage Laurent, et al., 2023 and causality-driven testing Giamattei, et al., 2024 refine scenario relevance for validation. Frameworks such as Sorokin, et al., 2024 enable integrated search-based testing pipelines. While highly effective for stress testing, these approaches target criticality rather than statistical representativeness.

Heterogeneous scenario representation and limitations

Handling heterogeneous variables remains a key challenge. Real-world ODDs involve mixed continuous and categorical variables (e.g., weather, infrastructure), which complicate latent modeling and partitioning Shukla and Marlin, 2022. Existing approaches typically ignore these variables or fragment datasets, preventing unified reasoning Riedmaier, et al., 2020. Our previous works Hadj Selem, et al., 2024; Kheriji, et al., 2024 showed that latent-space representations enable completeness estimation and representative selection in constrained settings, but remain limited to homogeneous or single-class scenarios.

Positioning of this work Existing approaches either rely on scenario-specific parameterizations. Chen, et al., 2026, focus on coverage metrics without selection procedures, Gelder, et al., 2025, or target critical scenario discovery rather than distribution-preserving selection Song, et al., 2023. As a consequence, they are typically limited to a single scenario class or to a fixed categorical condition, preventing their direct application to unified multi-scenario selection under heterogeneous representations. In contrast, we propose a unified framework that selects representative scenarios directly from heterogeneous real-world datasets. By projecting multi-scenario time series into a shared latent space and restoring its Gaussian structure despite mixed variable types, our method enables statistically controlled sampling with explicit guarantees on representativeness and uncertainty at the ODD level. Moreover, the framework naturally supports selection driven by both empirical frequency and scenario criticality, enabling the identification of rare and safety-critical edge cases within a unified probabilistic setting.

Methodology

Data Transformation

The proposed methodology is validated on ego-centric scenarios extracted from naturalistic driving data. Each occurrence is described by multivariate time series capturing both the ego-vehicle dynamics and the relative dynamics of surrounding obstacles. However, these raw signals cannot be directly exploited by learning algorithms for several reasons.

First, the number of surrounding obstacles varies across occurrences, which leads to inconsistent input dimensionality. Moreover, their ordering in the raw data is not stable, and their spatial configuration relative to the ego vehicle changes from one occurrence to another. As a result, the obstacle driving the interaction may appear at different positions in the raw representation, even across similar scenarios. In addition, keeping all detected obstacles would lead to an excessive and highly variable input dimensionality, making the representation both inefficient and poorly

suitable for learning. In addition, scenario durations are not fixed, resulting in variable-length sequences. Furthermore, many signals are only partially observed, yielding pervasive missing values due to unequal visibility, late appearance, or early disappearance of obstacles over time.

Addressing these issues requires a structured pipeline as illustrated in Fig. 1 that successively filters, organizes, and compresses the raw data into a fixed-size, semantically consistent representation suitable for downstream analysis.

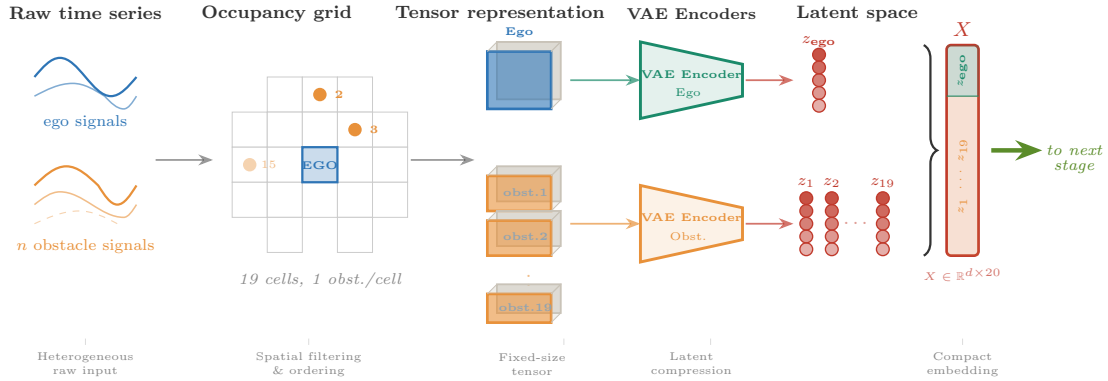


Figure 1: Data transformation pipeline.

Occupancy grid. To handle the variable number of obstacles and their unstable arrangement, we adopt the ego-centered occupancy grid introduced in Kheriji, et al., 2024. The space around the ego vehicle is partitioned into a fixed number of cells, each roughly the size of a vehicle. For each occurrence, detected obstacles are assigned to cells according to their relative position with respect to the ego vehicle at their first appearance in the scene. When multiple obstacles fall into the same cell, only the one that appears first is retained. This ensures that each cell contains at most one obstacle, yielding a fixed-size spatial structure regardless of the actual number of surrounding road users. This representation serves two purposes. First, it bounds the input dimensionality by limiting the number of tracked obstacles to the number of grid cells. Second, it anchors each obstacle to a spatially meaningful and reproducible position relative to the ego vehicle. As shown in Kheriji, et al., 2024, the obstacle driving the interaction consistently appears within the same grid region across occurrences of a given scenario class, confirming the stability of this spatial encoding for downstream learning.

Tensor representation. Once obstacles are assigned to occupancy-grid cells, the next step is to encode the resulting multivariate time series into a fixed-size tensor representation, despite variable scenario durations and the heterogeneous nature of the recorded signals. To this end, we adopt the tensorization approach introduced in Kheriji, et al., 2024, which transforms temporal signals into structured image-like channels. More precisely, for the ego vehicle and for each obstacle assigned to a grid cell, the selected temporal variables are organized into pairs of signals (x, y) . Each pair is then projected onto a two-dimensional surface, where time and x define the coordinates, and y is encoded as the pixel intensity.

In this way, each variable pair produces one channel, and the concatenation of all associated with the ego vehicle and the surrounding obstacles yields a single three-dimensional tensor of fixed dimensions.

This representation ensures a common temporal horizon across occurrences, avoids artificial zero-padding for shorter scenarios, preserves the semantics of absent obstacles through empty channels, and maintains spatial consistency thanks to the occupancy-grid structure. As a result, the learning model can identify stable scenario-specific patterns regardless of the actual duration or number of surrounding obstacles.

Latent representation learning. The tensor representation described above produces a high-dimensional but fixed-size input suitable for deep learning. However, its dimensionality remains too large for direct statistical analysis or sampling. To compress this representation into a compact and informative feature space, a set of Variational Autoencoders (VAEs) is trained in a fully unsupervised manner on the corpus of unlabeled occurrences. The key architectural choice is to train *separate* encoders for structurally distinct components of the tensor: one VAE dedicated to the ego vehicle dynamics, and one shared-weight VAE applied independently to each of the 19 obstacle slots. Each encoder compresses its input into a low-dimensional latent vector by minimizing a joint objective combining reconstruction fidelity and KL-divergence regularization, the latter enforcing a smooth and continuous latent manifold. Once all encoders are trained, each occurrence is represented by the concatenation of its latent vectors: the ego embedding followed by the 19 obstacle embeddings, *concatenated in the canonical order of the occupancy grid cells*. This fixed ordering ensures that each position in the resulting vector X consistently corresponds to the same spatial region around the

ego vehicle across all occurrences, preserving the spatial semantics of the scene in the latent representation. Thus the vector X captures both the ego behavior and the spatial configuration of interactions with surrounding obstacles, while reducing the data volume by several orders of magnitude compared to the raw tensor. This latent representation X constitutes the input for all subsequent stages of the framework.

Second Stage Transformation Projection into a Scenario-Agnostic Latent Space via β -VAE

The first stage of the pipeline produces, for each occurrence, a structured latent tensor $X \in \mathbb{R}^{d \times 20}$, composed of the ego-vehicle embedding and the 19 obstacle embeddings arranged according to the occupancy grid. While this representation is compact relative to the raw signals, it remains scenario-specific in the sense that the identity of the obstacle driving the interaction—and hence its position within the grid—varies across scenario classes. To obtain a scenario-agnostic representation, the tensor X is first flattened into a vector of dimension $d \times 20$. A discrete index is then appended, indicating the grid position of the obstacle of interest for the given occurrence (with index 0 assigned to the ego vehicle). This index encodes the spatial role of the interacting agent without relying on any scenario-specific parameterization, thereby producing a mixed-type feature vector combining continuous latent variables with a categorical identifier. The resulting dataset $\{X_i\}_{i=1}^n$, now comprising heterogeneous observations across all scenario classes, is projected into a lower-dimensional latent space using a β -Variational Autoencoder (β -VAE) Burgess, et al., 2018. The training objective takes the form:

$$\mathcal{L}(\theta, \phi) = \mathbb{E}_{q_\phi(Z|X)}[-\log p_\theta(X|Z)] + \beta D_{\text{KL}}(q_\phi(Z|X) \| p(Z)) \quad (1)$$

To account for the mixed nature of the input, the reconstruction term decomposes into two components corresponding to the continuous and categorical parts of the input, respectively:

$$-\log p_\theta(X | Z) = \text{MSE}(x_c, \hat{x}_c) + \text{CE}(x_d, \hat{x}_d) \quad (2)$$

where $\text{MSE}(x_c, \hat{x}_c) := \frac{1}{N_c} \sum_{i=1}^{N_c} (x_{c,i} - \hat{x}_{c,i})^2$ denotes the mean squared error over the N_c continuous dimensions, and $\text{CE}(x_d, \hat{x}_d) := -\sum_{j=1}^{N_d} x_{d,j} \log(\hat{x}_{d,j})$ denotes the cross-entropy loss over the N_d categorical dimensions. The total loss is therefore:

$$\mathcal{L} = \text{MSE}(x_c, \hat{x}_c) + \text{CE}(x_d, \hat{x}_d) + \beta D_{\text{KL}}(q_\phi(Z|X) \| p(Z)) \quad (3)$$

The hyperparameter β controls the trade-off between reconstruction fidelity and latent regularization. Higher values of β encourage a smoother and more disentangled latent geometry, at the potential cost of reconstruction accuracy. In our setting, β is tuned to ensure that the encoder produces a continuous latent manifold—a necessary prerequisite for the deterministic Gaussianization step described below—while preserving sufficient reconstruction quality for both the continuous and discrete components. This second-stage encoding serves a dual purpose. First, it reduces the dimensionality of the concatenated first-stage representations, which would otherwise be too high for reliable density estimation and stratified sampling. Second, it absorbs the categorical obstacle index into a unified continuous latent space, enabling the joint treatment of all

scenario classes within a single probabilistic framework.

Deterministic Gaussianization of the Latent Space

Although the β -VAE encourages the aggregate posterior to approximate a standard Gaussian, the presence of categorical variables in the input prevents the learned latent distribution from achieving this structure in practice. Categorical features are inherently non-continuous, and their encoded distributions remain far from Gaussian regardless of training quality. The modified reconstruction loss improves the handling of mixed data types, but the resulting latent distribution is typically irregular and dataset-dependent. Nevertheless, the encoder provides a key property: a *continuous* latent representation of the data. This continuity, inherited from the VAE framework, enables the application of deterministic distribution-matching transformations that do not require the latent space to carry generative meaning. Training an additional VAE on this intermediate representation to enforce Gaussianity is not a viable strategy. The intermediate latent distribution has no generative interpretation, and a cascaded VAE would learn an arbitrary transformation that neither preserves the geometric structure of the space nor improves its statistical properties in a controlled manner. Instead, we apply a sequence of two deterministic mappings that reshape the empirical latent distribution into a standard multivariate Gaussian. This pipeline, illustrated in Fig. 2, consists of the Rosenblatt transformation followed by an inverse Gaussian CDF mapping.

Rosenblatt transformation. The Rosenblatt transformation Rosenblatt, 1952 converts a random vector $\mathbf{x} = (x_1, \dots, x_n)$ with a continuous joint density $f_{\mathbf{x}}$ into a vector of independent uniform variables $\mathbf{u} = (u_1, \dots, u_n)$ via the recursive mapping

$$u_i = F_i(x_i | x_1, \dots, x_{i-1}), \quad i = 1, \dots, n, \quad (4)$$

Inverse Gaussian CDF mapping. After the Rosenblatt step, all $u_i \sim \mathcal{U}(0, 1)$ are mutually independent. Applying the inverse standard Gaussian CDF component-wise,

$$z_i = \Phi^{-1}(u_i), \quad i = 1, \dots, n, \quad \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \quad (5)$$

Together, these two deterministic steps transform the irregular empirical latent distribution into a standard multivariate normal distribution. The resulting Gaussian space can then be partitioned into equiprobable cells along each dimension using Gaussian quantiles, enabling both representative sampling and coverage estimation within a unified and statistically well-characterized framework.

Latent Space Exploitation

Given the standard Gaussian latent space obtained above, a representative subset is selected by stratifying the space using Gaussian quantiles independently along each axis, following the procedure introduced in our previous work Hadj Selem, et al., 2024. This stratification partitions the latent space into equiprobable cells, providing direct coverage guarantees: if the space is discretized into K strata per dimension and all strata are populated, one can certify that at least $(1 - \frac{1}{K}) \times 100\%$ of the distributional support has been covered.

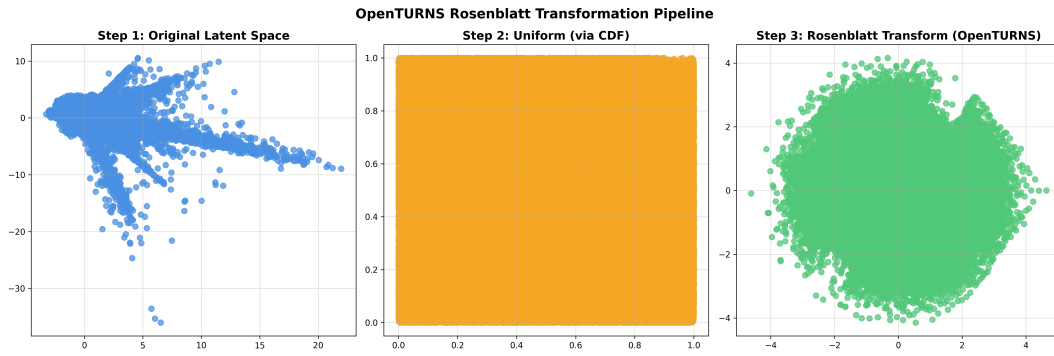


Figure 2: Full second stage data transformation pipeline: VAE maps to a continuous latent space; Rosenblatt yields uniform marginals; inverse Gaussian CDF gives a standard normal distribution.

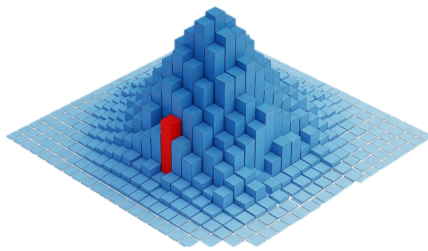


Figure 3: Latent space discretization into equiprobable strata; empty cells (highlighted) indicate gaps.

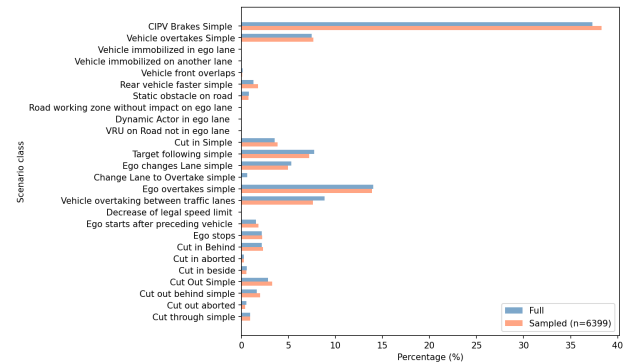


Figure 4: Distribution of scenario classes in the full dataset and the selected subset ($n = 6,399$). The relative frequencies are closely preserved across all classes, including both dominant and rare scenarios.

Sample Validation Using Classifier-Based 2 sample test

To objectively assess whether the selected subset preserves the statistical properties of the full dataset, we adopt a classifier-based two-sample test Chakravarti, et al., 2022. The principle is to train a binary classifier—in our case, XGBoost—to discriminate between observations drawn from the original population and those from the selected sample. If the classifier cannot distinguish between the two sets, this constitutes strong evidence that they share the same underlying distribution. The test statistic is the Misclassification Error (MCE), defined as:

$$\text{MCE} = \frac{1}{2} \left[\frac{1}{m} \sum \mathbb{I}(h(X_i) > \pi) + \frac{1}{n} \sum \mathbb{I}(h(W_j) < \pi) \right], \quad (6)$$

where $h(\cdot)$ is the classifier score, m and n are the sizes of the original and selected sets respectively, π is the decision threshold, and $\mathbb{I}(\cdot)$ is the indicator function. An MCE close to 0.5 indicates that the classifier performs no better than random guessing, meaning the two samples are statistically indistinguishable. Statistical significance is then established through a permutation test. The class labels are randomly shuffled B times, and the MCE is recomputed for each permutation, yielding a null distribution under the hypothesis that both samples originate from the same distribution. The empirical p-value is then computed as $\hat{p} = \frac{1}{B} \sum_{b=1}^B \mathbb{I}(\text{MCE}_b \leq \text{MCE}_{\text{obs}})$. A high p-value indicates that the observed MCE is consistent with the null distribution, confirming representativeness. This metric provides a unified and rigorous criterion, jointly accounting for continuous and categorical variables without variable-by-variable testing.

Results

The proposed framework is evaluated on a dataset of more than 280,000 real-world driving scenarios spanning 25 scenario classes. The selected subset contains approximately 6,400 occurrences — corresponding to a reduction of roughly 97% of the original dataset — though the framework naturally accommodates any target sample size. The quality of this selection is assessed along two complementary axes: (i) statistical fidelity, evaluated both visually through distributional comparisons and quantitatively via the classifier-based two-sample test described above, and (ii) comparative performance against baseline selection methods. To enable a controlled and meaningful comparison, all three methods — the proposed approach, K-Prototypes clustering, and uniform random sampling — are evaluated on a common subset of 400 samples, and the resulting distributional histograms are reported below.

Inter-Scenario Fidelity

A first requirement is that the selected subset preserves the relative proportions of scenario classes observed in the full dataset. Figure 4 compares the class distributions of the full population and the selected sample across all 25 scenario types. The proportions are closely matched for both dominant classes such as CIPV Brakes and rare classes such as Cut Out Aborted, confirming that the stratified selection in the shared latent space does not introduce

systematic bias toward over- or under-represented scenario types. This result is notable given that the scenario class labels were never used during latent space construction: the preservation of inter-scenario proportions emerges naturally from the equiprobable partitioning of the unified Gaussian space.

Intra-Scenario Fidelity

Beyond class proportions, the selected subset must also preserve the internal distributional structure within each scenario class, for both continuous and categorical variables. We illustrate this on two representative scenarios: CIPV Brakes, the most frequent class in the dataset, and Ego Overtakes Simple, a less frequent but structurally distinct class. Figure 5 shows the distributions of ego speed at the initial time T_0 and the number of lanes on the principal road for the CIPV Brakes scenario. In both cases, the selected sample closely reproduces the shape and support of the original distribution, including the tail behavior of the continuous variable and the relative frequencies of each lane category. Figures 8 and 9 presents analogous results for the Ego Overtakes Simple scenario, comparing the ego speed at the key 'overtake' time T_1 and the distribution of countries in which the scenario was recorded. Again, the selected subset faithfully reflects the original data, preserving both the continuous speed profile and the multinomial country distribution despite the drastic reduction in sample size. These results confirm that the unified latent space not only maintains inter-scenario balance but also preserves the fine-grained intra-scenario structure across heterogeneous variable types. Importantly, these variables are never used during selection, as the model operates directly on complete raw time series without access to predefined features or key moments; their accurate preservation therefore demonstrates that it captures the global semantic structure of scenarios and consistently retains critical patterns across classes.

Classifier-Based Validation

To move beyond visual inspection, we apply the classifier-based two-sample test described in the previous section. Table 1 reports the MCE statistic and the permutation p-value for three selection strategies: K-Prototypes clustering Huang, 1998, uniform random sampling, and the proposed method. K-Prototypes achieves an MCE of 46.4%, significantly below the 50% baseline, indicating that the classifier can reliably distinguish the selected subset from the original population. The associated p-value of 0.2% confirms that this difference is statistically significant: the K-Prototypes selection does not faithfully reproduce the joint distribution of the data. In contrast, both random sampling and the proposed method yield MCE values close to 50% (50.08% and 50.16%, respectively), with high p-values (52.45% and 55.74%), indicating that the classifier cannot distinguish the selected subset from the full dataset in either case. While random sampling achieves comparable distributional fidelity overall — a result that can be attributed to the natural correlations between features such as position and speed — it performs relatively worse than the proposed method on uncorrelated features such as country of recording and number of lanes, and yields slightly less accurate mean estimates, as illustrated in Fig. 6. More critically, random sampling provides no structural guarantees: it

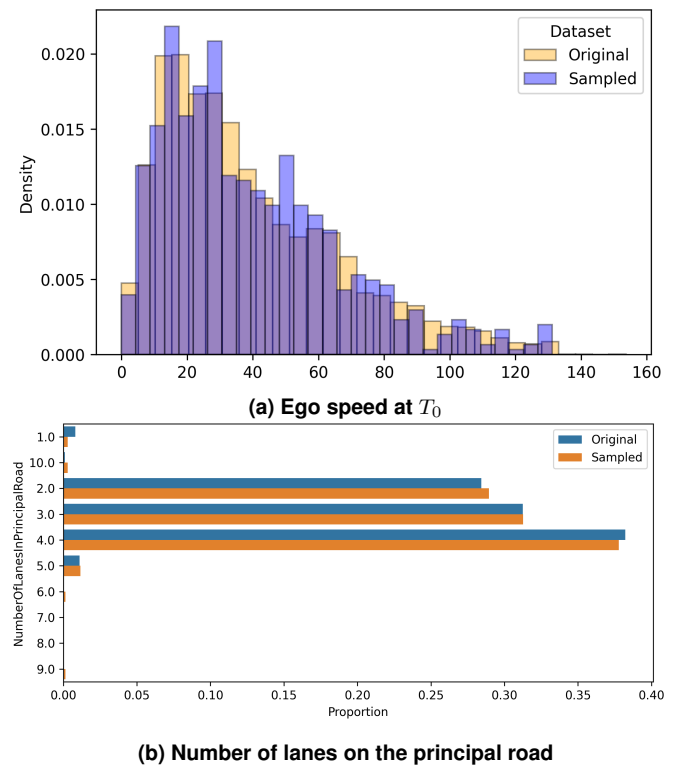


Figure 5: CIPV Brakes — intra-scenario fidelity

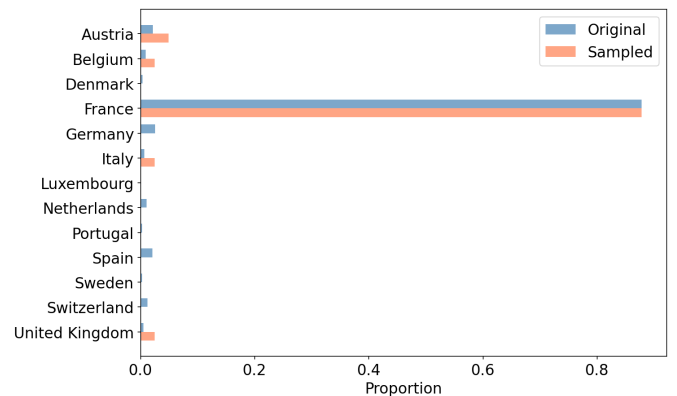


Figure 6: Country distribution — Proposed method

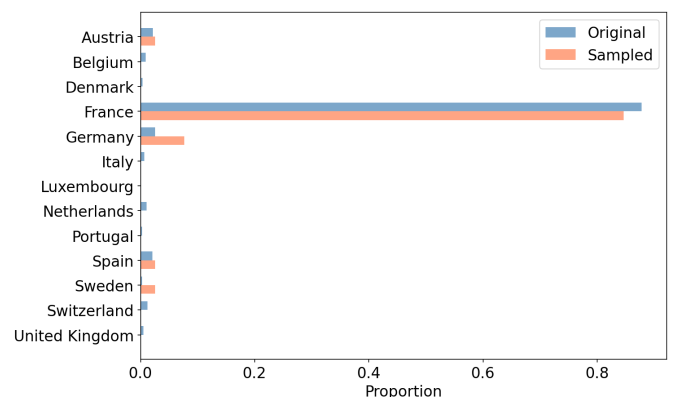


Figure 7: Country distribution — Random sampling

does not partition the latent space, cannot quantify coverage, and offers no mechanism for identifying missing or underrepresented regions. The proposed method, by contrast, matches the statistical fidelity of random sampling while additionally providing equiprobable stratification, explicit coverage guarantees through the fill rate of latent cells, and the ability to detect data gaps for targeted completion — capabilities that are essential for rigorous safety validation within structured frameworks such as the SAF Gelder, et al., 2025.

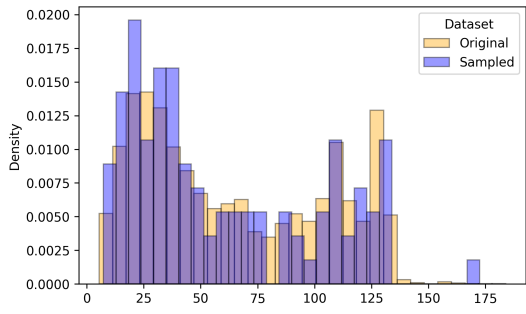


Figure 8: Ego speed at keytime T_1

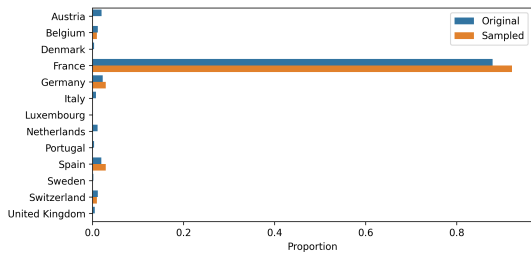


Figure 9: Country distribution

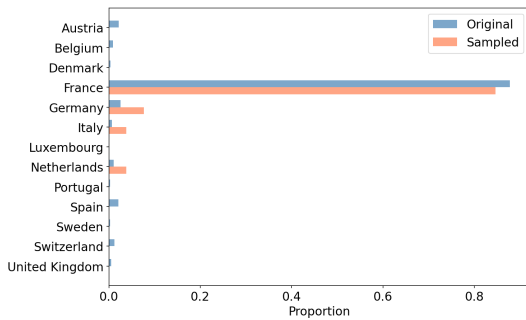


Figure 10: Country distribution — K-Prototypes

Table 1: Classifier-based two-sample test: MCE (%) and permutation p-value (%) for three methods; MCE $\approx$ 50% with high p-value indicates the subset is statistically indistinguishable from the full dataset.

Method	MCE (%)	P-Value (%)
K-Prototypes	46.40	0.20
Random sampling	50.08	52.45
Proposed method	50.16	55.74

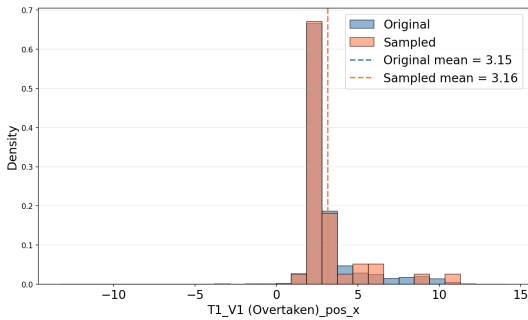


Figure 11: Position of the Overtaken Vehicle at T_1 — Proposed method

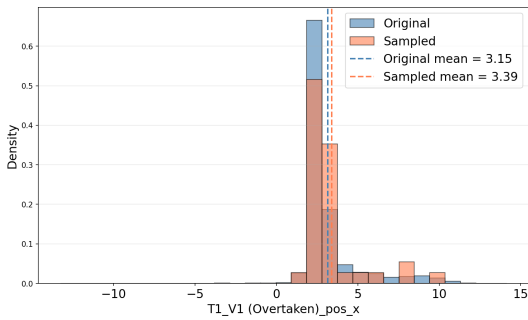


Figure 12: Position of the Overtaken Vehicle at T_1 — Random sampling

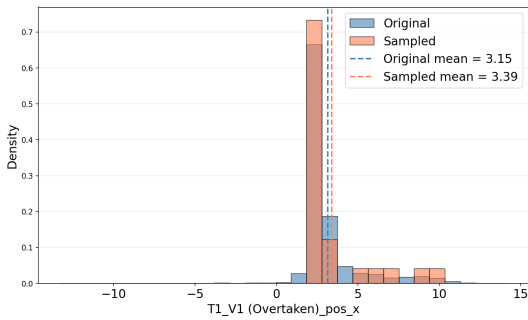


Figure 13: Position of the Overtaken Vehicle at T_1 — K-Prototypes

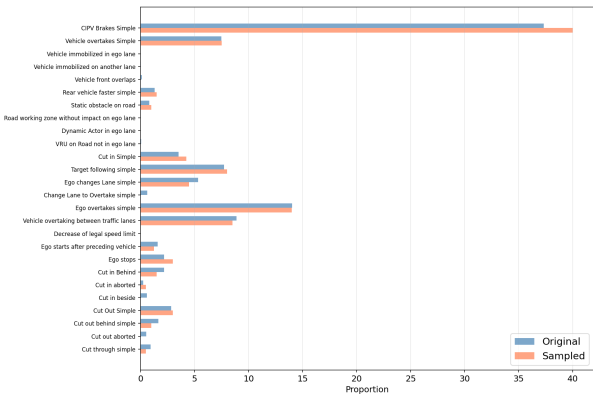


Figure 14: Scenario distribution — Proposed method

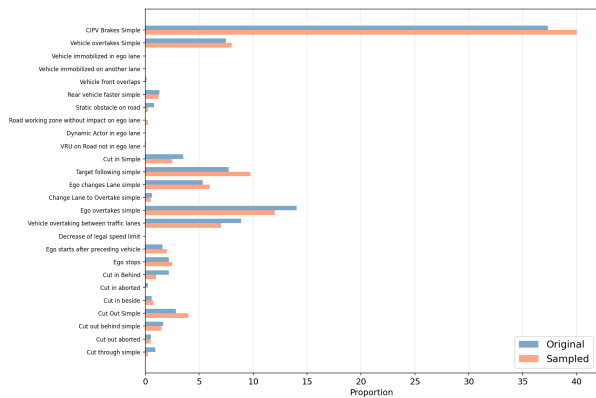


Figure 15: Scenario distribution — Random sampling

Conclusions

This paper addresses a critical bottleneck in autonomous vehicle validation: how to select a minimal yet statistically faithful set of test scenarios across a full Operational Design Domain. By projecting heterogeneous driving situations into a shared Gaussian latent space, the method achieves a 97 % reduction in test volume while preserving distributional fidelity across all scenario classes and categorical conditions — with explicit, quantifiable coverage guarantees. Beyond efficiency, it provides a principled way to detect missing conditions and quantify residual uncertainty, directly supporting safety argumentation for regulatory certification. This makes it a concrete and industrially deployable foundation for scalable, trustworthy validation of automated driving systems.

References

- Burgess, C. P., Higgins, I., Pal, A., Matthey, L., Watters, N., Desjardins, G., and Lerchner, A., 2018. *Understanding disentangling in β -VAE*. arXiv: 1804.03599 [stat.ML]. Available at: <<https://arxiv.org/abs/1804.03599>>.
- Chakravarti, P., Kuusela, M., Lei, J., and Wasserman, L., 2022. *Model-Independent Detection of New Physics Signals Using Interpretable Semi-Supervised Classifier Tests*. arXiv: 2102.07679 [stat.AP]. Available at: <<https://arxiv.org/abs/2102.07679>>.
- Chen, R., et al., 2026. Test Case Sampling Optimization for Safety Validation of Automated Driving Systems. *Nature Communications*. DOI à vérifier. <https://doi.org/10.1038/s41467-026-69675-8>.
- Feng, S., Yan, X., Sun, H., Feng, Y., and Liu, H. X., 2021. Intelligent Driving Intelligence Test for Autonomous Vehicles with Naturalistic and Adversarial Environment. *Nature Communications*, 12, p. 748. <https://doi.org/10.1038/s41467-021-21007-8>.
- Gelder, E. de, Hof, J., Cator, E., Paardekooper, J.-P., Op den Camp, O., Ploeg, J., and De Schutter, B., 2022. Scenario Parameter Generation Method and Scenario Representativeness Metric for Scenario-Based Assessment of Automated Vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 23(10), pp. 18794–18807. <https://doi.org/10.1109/TITS.2022.3154774>.
- Gelder, E. de, Paardekooper, J.-P., Op den Camp, O., and De Schutter, B., 2019. Safety Assessment of Automated Vehicles: How to Determine Whether We Have Collected Enough Field Data? *Traffic Injury Prevention*, 20(sup1), S162–S170. <https://doi.org/10.1080/15389588.2019.1602727>.
- Gelder, E. de, Singh, T., Hadj-Selem, F., Vidal Bazan, S., and Op den Camp, O., 2025. Scenario Metrics for the Safety Assurance Framework of Automated Vehicles: A Review of Its Application. *Vehicles*, 7(3), p. 100. <https://doi.org/10.3390/vehicles7030100>.
- Giamattei, L., Guerriero, A., Pietrantuono, R., and Russo, S., 2024. Causality-driven Testing of Autonomous Driving Systems. *ACM Transactions on Software Engineering and Methodology*, 33(3). <https://doi.org/10.1145/3635709>.
- Hadj Selem, F., Ben Nejma, G., Kheriji, W., Durville, L., Rahal, M. C., Geronimi, S., and Arnoux, E., 2024. Developing a Methodology to Assess Data Completeness of Driving Scenarios for Testing Autonomous Vehicles: A Focus on ODD-Specific Objectives. In: *DSC*. Strasbourg, France.
- Huang, Z., 1998. Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values. *Data Mining and Knowledge Discovery*, 2(3), pp. 283–304. ISSN: 1573-756X. <https://doi.org/10.1023/A:1009769707641>. Available at: <<https://doi.org/10.1023/A:1009769707641>>.
- Kheriji, W., Hadj-Selem, F., Ben Nejma, G., Bontemps, T., Durville, L., and Rahal, M.-C., 2024. Extracting Scenarios for Automated Driving Systems: A Statistical Based Occupation-Grid Approach with Deep Semi-Supervised Learning. In: *ITSC*, pp. 2659–2664. <https://doi.org/10.1109/ITSC58415.2024.10919875>.
- Laurent, T., Klikovits, S., Arcaini, P., Ishikawa, F., and Ventresque, A., 2023. Parameter Coverage for Testing of Autonomous Driving Systems under Uncertainty. *ACM Transactions on Software Engineering and Methodology*, 32(3). <https://doi.org/10.1145/3550270>.
- Nakamura, H., Muslim, H., Kato, R., Préfontaine-Watanabe, S., Kaneko, H., Imanaga, H., Antona-Makoshi, J., Kitajima, S., Uchida, N., Kitahara, E., Ozawa, K., and Taniguchi, S., 2022. Defining Reasonably Foreseeable Parameter Ranges Using Real-World Traffic Data for Scenario-Based Safety Assessment of Automated Vehicles. *IEEE Access*, 10, pp. 37743–37760.
- Ponn, T., Breituß, M., Yu, X., and Diermeyer, F., 2020. Identification of Challenging Highway-Scenarios for the Safety Validation of Automated Vehicles Based on Real Driving Data. In: *EVER*, pp. 1–10. <https://doi.org/10.1109/EVER48776.2020.9242539>.
- Riedmaier, S., Ponn, T., Ludwig, D., Schick, B., and Diermeyer, F., 2020. Survey on Scenario-Based Safety Assessment of Automated Vehicles. *IEEE Access*, 8, pp. 87456–87477. <https://doi.org/10.1109/ACCESS.2020.2993730>.
- Rosenblatt, M., 1952. Remarks on a Multivariate Transformation. *Annals of Mathematical Statistics*, 23, pp. 470–472. Available at: <<https://api.semanticscholar.org/CorpusID:119994384>>.
- Shukla, S. N. and Marlin, B. M., 2022. Heteroscedastic Temporal Variational Autoencoder for Irregular Time Series. In: *ICLR*. Available at: <<https://openreview.net/forum?id=Az7opqbQE-3>>.
- Song, Q., Tan, K., Runeson, P., and Persson, S., 2023. Critical scenario identification for realistic testing of autonomous driving systems. *Software Quality Journal*, 31, pp. 441–469. <https://doi.org/10.1007/s11219-022-09604-2>.
- Sorokin, L., Munaro, T., Safin, D., Liao, B. H.-C., and Molin, A., 2024. OpenSBT: A Modular Framework for Search-based Testing of Automated Driving Systems. In: *Proceedings of the IEEE/ACM 46th International Conference on Software Engineering: Companion Proceedings*. <https://doi.org/10.1145/3639478.3640027>.
- Wang, W., Liu, C., and Zhao, D., 2017. How Much Data Are Enough? A Statistical Approach with Case Study on Longitudinal Driving Behavior. *IEEE Transactions on Intelligent Vehicles*, 2(2), pp. 85–98. <https://doi.org/10.1109/TIV.2017.2736910>.
- Weber, H., Bock, J., Klimke, J., Roesener, C., Hiller, J., Krajewski, R., Zlocki, A., and Eckstein, L., 2019. A Framework for Definition of Logical Scenarios for Safety Assurance of Automated Driving. *Traffic Injury Prevention*, 20(sup1), S65–S70. <https://doi.org/10.1080/15389588.2019.1630827>.
- Zhang, X., Tao, J., Tan, K., Törnngren, M., Sánchez, J. M. G., Ramli, M. R., Tao, X., Gyllenhammar, M., Wotawa, F., Mohan, N., Nica, M., and Felbinger, H., 2023. Finding Critical Scenarios for Automated Driving Systems: A Systematic Mapping Study. *IEEE Transactions on Software Engineering*, 49(3), pp. 991–1026. <https://doi.org/10.1109/TSE.2022.3170122>.
- Zhong, Z., Tang, Y., Zhou, Y., Oliveira Neves, V. de, Liu, Y., and Ray, B., 2021. *A Survey on Scenario-Based Testing for Automated Driving Systems in High-Fidelity Simulation*. arXiv: 2112.00964 [cs.SE].

A Framework for Describing and Controlling Surrounding Traffic in Driving Simulators

Sven Tarlowski, Philipp Legran, Lutz Eckstein

RWTH Aachen University – Institute for Automotive Engineering (ika)
{sven.tarlowski, philipp.leggran, lutz.eckstein}@ika.rwth-aachen.de

Abstract - Surrounding traffic in driving simulator studies is often specified only qualitatively, which weakens reporting, reproducibility, and cross-study comparability. This paper presents an ego-centred framework for describing and controlling surrounding traffic in a moving reference frame around the ego vehicle. First, requirements for a suitable traffic-state description are derived and used to assess candidate traffic metrics. Based on this review, the framework structures traffic-state variables into environment, interaction, and risk components and, for the present highway use case, uses density, speed distribution, and time-headway distribution as core descriptors for specification and reporting. Second, the paper instantiates this framework in a rule-based highway traffic controller in VTD that acts on selected environment variables through vehicle spawning, vehicle removal, and one-time desired-speed assignment at spawn. Third, repeated highway simulations evaluate whether this controller can establish and maintain the selected target traffic descriptors reproducibly. The results support the technical feasibility of maintaining selected ego-centred traffic descriptors for the tested highway scenarios. Future work should apply the broader framework in concrete driving-simulator studies and extend the evaluation toward human-driven ego vehicles and perceptual realism.

Keywords: Driving simulator – Ego-centred traffic state – Runtime traffic control – Study design – Reproducibility

Copyright: © 2026 by the authors.

Licensee Driving Simulation Association.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.82157/dsa/2026/21>

Introduction

In recent years, research on Advanced Driver Assistance Systems (ADAS), automated driving, and human-machine interaction has expanded rapidly, increasing the need for efficient methods to test and validate increasingly complex systems (Yurtsever, et al., 2019). Driving simulators are attractive in this context because they enable experiments at an early development stage, provide a safe and reproducible environment, and reduce the dependence on costly physical prototypes. Driving simulators are therefore widely used in studies on driver distraction, cabin interaction, and automated driving functions under controlled highway traffic conditions (Wang and Qin, 2025; Wang, et al., 2026).

However, for simulator-based findings to be transferable, the simulated environment must provide sufficient fidelity. Prior work shows that simulator fidelity is multi-dimensional and that unrealistic or missing cues can bias behaviour (Fischer, et al., 2015; Himmels, et al., 2024). One particularly important cue is the surrounding traffic, because local traffic density, speed patterns, and interactions with other vehicles affect workload, stress, attention allocation, and tactical driving decisions (Hock, et al., 2018). Despite this relevance, surrounding traffic is often specified only qualitatively in driving simulator studies, for ex-

ample as “low” or “high” traffic, while omitting quantitative descriptors such as local density, speed distributions, or time-headway statistics (Wang and Qin, 2025; Wang, et al., 2026). This weakens study reporting and limits reproducibility and comparability across experiments.

Existing traffic modelling approaches do not fully resolve this problem. Although current traffic tools can simulate surrounding vehicles, they do not provide a concise way to specify and maintain a prescribed local traffic state around a moving vehicle in a driving simulator. The missing link is a quantitative, ego-centred description that supports both study reporting and runtime control.

This paper addresses this issue by providing a structured, conceptual basis and a technical instantiation specific to highways. First, it derives requirements for an ego-centred traffic-state description and proposes a framework that structures traffic variables into environment, interaction, and risk components. For the present highway use case, density, speed distribution, and time-headway distribution form the core descriptors for specification and reporting. Second, it instantiates this framework in a rule-based runtime controller in Virtual Test Drive (VTD) that maintains selected environment descriptors through vehicle spawning, vehicle removal, and one-time desired-

speed assignment at spawn. Third, it evaluates this highway controller instantiation in repeated closed-loop highway simulations to test whether the selected target traffic descriptors can be built up and maintained reproducibly.

State of the Art

Previous work on driving simulators has shown that the concept of fidelity depends on task-relevant cues rather than on a single, overarching definition (Fischer, et al., 2015; Himmels, et al., 2024). Surrounding traffic is one such cue because it influences workload, tactical driving behaviour, and the overall driving experience (Hock, et al., 2018). In many simulator studies, explicitly scripted situations are embedded within longer driving segments, so surrounding traffic also acts as a background condition between focal events. For the present problem, this implies that traffic should be specified in a reproducible way for study design and reporting rather than only through qualitative labels such as “low” or “high” traffic.

Traffic-flow research offers a broad range of variables for characterising traffic, but their usefulness depends on scale and application. Macroscopic models describe aggregate behaviour over larger road sections, whereas microscopic approaches describe individual vehicles and their local relations (Ferrara, Sacone, and Siri, 2018). For a driving simulator, the relevant viewpoint is the local traffic experienced by a single ego vehicle (Olstam, 2005). Established quantities such as density, vehicle speed, and time headway are therefore natural candidate descriptors because they capture complementary aspects of traffic amount, motion, and spacing structure (Treiber, 2013). At the same time, these quantities are not interchangeable: density alone cannot distinguish traffic states that differ in speed or spacing, and no single scalar is sufficient to characterise a local traffic state comprehensively.

The driving-simulator context introduces additional constraints beyond those of conventional traffic analysis. Surrounding traffic must be handled in a moving ego-centred reference frame and within a limited spatial window around the ego vehicle (Olstam, 2005; 2009). Vehicles have to be inserted and removed without producing visible discontinuities, and their subsequent behaviour should remain governed by the underlying driver model rather than by continuous direct intervention. As a result, both the traffic description and the traffic-generation method must be tailored to the ego-centred simulator use case rather than transferred directly from infrastructure-oriented traffic simulation.

Existing microscopic traffic simulators provide important building blocks, but they do not by themselves resolve this problem. SUMO is a well-established microscopic traffic simulator, and previous work has shown that such tools can be coupled with driving simulators (Lopez, et al., 2018; Salles, et al., 2022). However, these contributions mainly demonstrate traffic simulation capability and software integration, not how a prescribed local traffic state around the ego vehicle can be defined for study design and reporting and then maintained online during runtime. Scenario scripting standards such as OpenSCENARIO (ASAM e.V., 2022) address a different

problem again: they are well suited to defining discrete situations and event logic, but not to sustaining a quantitative background traffic state around a moving ego vehicle over longer driving segments.

Research Question

This shows that there is a need for a systematic approach to further describe and control surrounding traffic within human-in-the-loop driving simulator studies. This motivates two coupled research questions:

1. **Traffic description:** How can the traffic state around an ego vehicle be described quantitatively in a moving, ego-centred reference frame such that it is (i) microscopic, (ii) sufficiently discriminative, and (iii) interpretable for study design and reporting?
2. **Traffic generation & control:** How can selected ego-centred environment descriptors be generated and maintained during runtime in a highway driving simulator through controller actions that remain compatible with the underlying traffic model?

The paper addresses these questions in two corresponding steps. RQ1 is addressed analytically through requirement derivation and a structured review of candidate traffic metrics. RQ2 is addressed through the implementation and technical evaluation of a rule-based highway traffic controller that actuates selected environment variables in closed-loop simulation.

Methodology

The methodology addresses the two research questions (RQ) in sequence. First, an ego-centred traffic-state description is derived to identify metrics suitable for study design, reporting, and controller specification. Second, a runtime controller is implemented and evaluated on that basis to generate and maintain the desired local traffic state while keeping vehicle behaviour compatible with the underlying driver model.

Traffic State Description

The goal of the traffic state description is not to reproduce every detail of highway traffic, but to characterise the aspects of surrounding traffic that are relevant to the ego driver during background driving phases, i.e., between explicitly scripted events. For this use case, the description must support both technical control and study reporting. As most candidate metrics originate from traffic engineering and simulation, they must be evaluated in the context of the current use case. This subsection is organised as follows: First, the requirements for suitable metrics are derived. Second, the candidate metrics are reviewed against these requirements. Third, the resulting traffic-state representation is defined.

Requirement Derivation

The requirements follow from RQ1 and from the intended use in simulator study phases between scripted scenarios. Because conflict measures such as time-to-collision or pairwise time headway are most informative for short-horizon interactions,

whereas density and speed statistics describe more persistent conditions, the description should primarily capture the sustained local traffic state around the ego vehicle rather than a single instantaneous interaction.

Accordingly, the metric set must be microscopic, sufficiently discriminative, interpretable, and applicable in a moving ego-centred reference frame. Operationally, metrics must therefore be applicable in a moving ego-centred reference frame (R1) (Olstam, 2005), characterise the local surrounding traffic on a microscopic level (R2), enable a sufficiently discriminative traffic-state description either individually or as part of a set (R3), and remain interpretable for study design and reporting (R4). These requirements already suggest that a single traffic metric will rarely be sufficient.

Table 1: Requirements for ego-centred traffic metrics

ID	Requirement
R1	Applicable within a moving, ego-centred reference frame.
R2	Applicable on a microscopic level around the ego vehicle.
R3	Enable a sufficiently discriminative traffic state description.
R4	Interpretable for study design and reporting.

Table 1 summarises the resulting requirements. Candidate metrics were then reviewed with respect to these four criteria.

Candidate Metric Review

The candidate metrics were selected as an application-oriented subset of commonly used traffic measures rather than as an exhaustive taxonomy. For the present framework, they are grouped by temporal role: environment descriptors characterise more persistent traffic conditions over roughly 10 s to 60 s and longer, interaction descriptors capture mesoscopic relations over about 3 s to 10 s, and risk descriptors represent short-horizon states up to about 3 s. This temporal layering clarifies why the framework separates variables by function rather than by formula alone. Environment variables describe the sustained background traffic that can be specified for longer study phases, whereas interaction variables describe how this state is experienced locally by the ego vehicle over shorter horizons. Risk indicators describe short-horizon conflict situations that are important for evaluation but too transient and ego-dependent to define the surrounding traffic state on their own. The same physical quantity can therefore change role with the aggregation level and temporal scope: pair-specific time headway is an interaction or risk measure, while the distribution of time headways across the local traffic window serves as an environment descriptor.

Table 2 summarises the authors' analytical assessment of these metrics against R1 to R4. No single metric satisfies R3 on its own, so the traffic state must be described by a complementary metric set rather than by a single scalar quantity.

Density and related normalised variants satisfy R1 and R2 when computed in a moving window around the ego vehicle and remain intuitive for reporting, but

they are weak on discrimination because they cannot distinguish, for example, sparse high-speed traffic from denser slow traffic. Speed distributions provide the complementary information that density lacks by capturing overall pace and heterogeneity in an interpretable form.

Time headway and similar interaction-oriented measures are more sensitive to individual constellations and are therefore not sufficient as sole state descriptors. However, once sampled across the ego-centred window and represented as a distribution, time headway characterises the spacing structure of the surrounding traffic and can therefore serve as an environment descriptor rather than only a pair-specific interaction measure. Lane-change rate and similar metrics may enrich the description further, but they are more scenario-dependent and harder to interpret consistently across studies.

Selected Traffic State Representation

To enable a consistent description and control of surrounding traffic, the traffic state is formulated as a vector-valued representation in an ego-centred reference frame. The traffic state at time t is defined as:

$$x(t) = \begin{bmatrix} x_E(t) \\ x_I(t) \\ x_R(t) \end{bmatrix} \quad (1)$$

where x_E , x_I , and x_R denote the subvectors of environment variables, interaction variables, and risk or conflict indicators, respectively. Because these metrics generally evolve over time, the traffic state is defined as time-dependent. Accordingly, an initial state can be specified as $x(0)$, and a desired target state can be specified as x_{des} .

Within the framework, each subvector is populated with metrics appropriate to the use case. Table 2 lists candidate metrics, and any metric that satisfies the application requirements can, in principle, be incorporated into the traffic-state vector. For the present paper, the primary operational focus lies on the *environment variables*, because these can be specified, actuated, and evaluated most directly in the highway controller.

Density ρ reflects the overall vehicle presence, while the velocity distribution (μ_v, σ_v) characterises how traffic evolves dynamically. The time-headway distribution $(\mu_{THW}, \sigma_{THW})$ describes the spacing structure of the traffic flow, which strongly influences perceived traffic conditions. Importantly, representing velocity and time headway as distributions rather than scalar values allows traffic heterogeneity to be captured (Treiber, 2013). Together, these measures provide a compact and interpretable description that is more informative than a single density value alone. They were selected for the present use case because they can be measured in an ego-centred microscopic window and, in combination, provide sufficient discrimination between relevant local traffic states without requiring a complete specification of the full microscopic traffic configuration. For the current highway use case, the environment subvector is defined as follows:

$$x_E = [\rho, \mu_v, \sigma_v, \mu_{THW}, \sigma_{THW}]. \quad (2)$$

Table 2: Evaluation of traffic metrics against the requirements. Symbols indicate whether a metric satisfies a requirement well (+), partially (o), or poorly (–).

Metric	Type	R1	R2	R3	R4
Traffic Density	Environment	+	+	–	+
Lane Normalized Traffic Density	Environment	+	+	o	o
Normalized Traffic Density	Environment	+	+	o	o
Velocity Distribution	Environment	+	+	o	+
THW Distribution	Environment	+	+	o	o
Lane-Change Rate	Interaction	+	o	–	–
Relative Speed	Interaction	+	+	–	o
Nearby Vehicle	Interaction	+	+	o	o
TTC	Risk	–	–	–	o

Interaction variables, such as relative speed or lane-change rate, and risk indicators, such as time headway or time-to-collision, remain part of the full framework but are not independent control targets in the present controller. Their values should emerge from the generated surrounding traffic and the ego vehicle's behaviour rather than be prescribed directly. This creates a practical distinction between configuration variables and dependent evaluation variables without reducing the generality of the framework. The exact composition of the traffic-state vector may vary with study context, but the proposed structure provides a consistent basis for selecting, documenting, and interpreting application-specific traffic descriptors.

Traffic State Controller

The preceding subsection defined the full traffic-state vector $x(t) = [x_E(t), x_I(t), x_R(t)]^T$, with the descriptive environment subvector $x_E = [\rho, \mu_v, \sigma_v, \mu_{THW}, \sigma_{THW}]$. For the present highway controller, only ρ , μ_v , and σ_v are actuated directly. Traffic density, speed distribution, and time-headway distribution are tightly coupled in highway traffic, because a given density and speed distribution jointly constrain the resulting spacing structure (Treiber, 2013). Time headway therefore remains part of the descriptive framework, but is treated as a dependent outcome rather than as an independent control target. Interaction and risk metrics are likewise excluded from the control objective because they depend strongly on ego behaviour, and enforcing quantities such as time-to-collision or relative speed would constrain the ego vehicle indirectly and undermine the experimental validity of the simulator study. This restriction is deliberate. Directly commanding accelerations, lane changes, or ego-relative gaps would increase control authority over interaction-level quantities, but it would also override the behaviour generated by the underlying traffic model and increase the risk of controller-induced artefacts. The controller therefore targets the surrounding traffic only at the environment level, so coupled quantities such as time headway and relative speed remain dependent outcomes of the generated traffic state rather than prescribed signals.

In the present implementation, the controller is implemented in VTD and evaluated in a highly dynamic driving simulator for a highway use case in which reproducible background or transition traffic must be maintained around the ego vehicle. Its task is to continuously maintain a specified target density ρ^* and target speed distribution (μ^*, σ^*) in the

local traffic surrounding the ego vehicle. The controller follows a rule-based architecture. Compared with model-predictive or learning-based alternatives, this design offers full interpretability, requires no offline training data, and introduces only negligible runtime cost, which is advantageous for real-time simulator integration. The controller actuates the traffic state exclusively through vehicle spawning, vehicle removal, and a one-time desired-speed assignment at spawn. It does not intervene in vehicle behaviour after spawn. This separation between traffic management and driver model ensures that subsequent vehicle dynamics, including car-following and lane-changing, remain governed by the underlying driver model rather than by direct controller commands.

Spatial Zones

The controller operates over two concentric longitudinal zones centred on the ego vehicle, illustrated in Figure 1.

The *metric zone* extends R_m ahead of and behind the ego vehicle, yielding a total measurement length of $L_m = 2 \cdot R_m$. For the present use case, this zone is intended to cover the visible surroundings of the ego vehicle that are relevant to the subject, so the traffic metrics are evaluated only within this region. All metrics are computed over the vehicles located within this region. The surrounding *buffer zone* extends from R_m to R_b on each side and serves as a reservoir in which newly spawned vehicles can evolve before entering the metric zone. Vehicles that move beyond R_b and continue to separate from the ego vehicle are removed. Spawning and removal are thus confined to the outer controller region rather than the measurement region itself, reducing the likelihood of visible traffic discontinuities near the ego vehicle.

Metric Measurement and Smoothing

At each control step, density and speed statistics are computed over all vehicles and lanes in the metric zone:

$$\rho = \frac{N}{L_m}, \quad \mu_v = \frac{1}{N} \sum_{i=1}^N v_i, \quad \sigma_v = \sqrt{\frac{1}{N} \sum_{i=1}^N (v_i - \mu_v)^2}. \quad (3)$$

Raw measurements are smoothed with an exponential moving average (EMA)

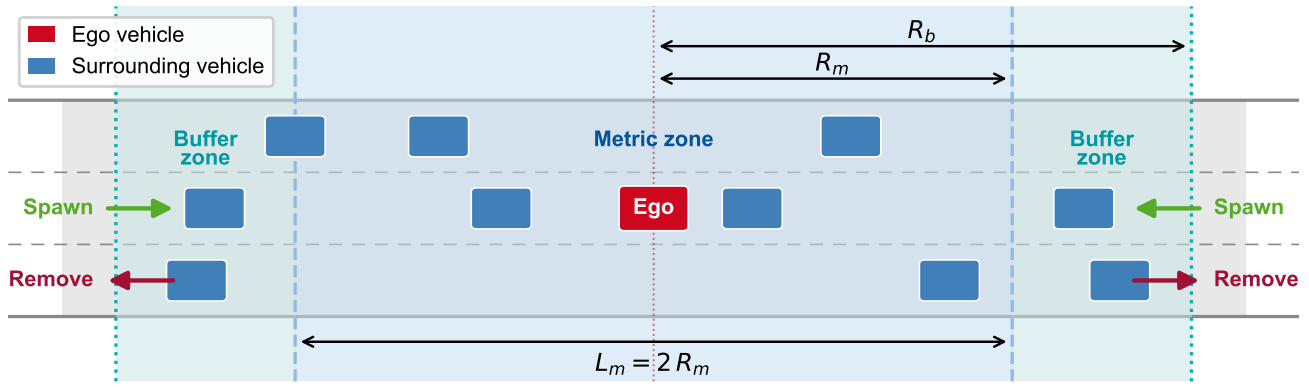


Figure 1: Metric zone ($\pm R_m$) and buffer zone ($\pm R_b$) centred on the ego vehicle. Density and speed statistics are measured inside the metric zone. Vehicles are spawned at the inner boundary of the buffer zone.

$$\bar{x}_k = \alpha x_k + (1 - \alpha) \bar{x}_{k-1}, \quad \alpha = 0.2, \quad (4)$$

which suppresses short-term fluctuations from individual vehicle crossings without introducing excessive lag. All subsequent control decisions use the smoothed values $\bar{\rho}$, $\bar{\mu}_v$, $\bar{\sigma}_v$.

Control Law

The controller manages density through a *spawn-credit accumulator* S that integrates a feedforward and a feedback term:

$$S += (q_{\text{net}} + K_\rho (\rho^* - \bar{\rho})) \cdot \Delta t. \quad (5)$$

The feedforward term q_{net} denotes the net ego-relative vehicle outflow across the front and rear boundaries of the metric zone, obtained by subtracting the buffer-to-metric inflow from the metric-to-buffer outflow. Positive values therefore indicate an impending density loss in the metric zone.

$$q_{\text{net}} = q_{\text{out,front}} + q_{\text{out,back}} - q_{\text{in,front}} - q_{\text{in,back}} \quad (6)$$

It anticipates an upcoming density drop before the EMA-smoothed density measurement fully reflects it, thereby improving responsiveness. The feedback term $K_\rho (\rho^* - \bar{\rho})$ corrects residual density error that the feedforward term does not capture. When $S \geq 1$, one vehicle is spawned and S is decremented by 1. Anti-windup clamping limits $S \in [-2, 5]$ and a minimum spawn interval of $T_{\text{min}} = 1$ s prevents rapid-fire spawning.

Each spawned vehicle is assigned a desired speed drawn from the target distribution:

$$v_s \sim \mathcal{N}(\mu^*, \sigma^{*2}), \quad v_s \in [v_{\text{min}}, v_{\text{max}}]. \quad (7)$$

The desired speed is set once at spawn via the driver model interface and is never modified thereafter. Spawn position is determined by the sampled speed relative to ego speed: a faster vehicle is placed behind ego so that it naturally catches up into the metric zone, a slower vehicle is placed ahead so that

the ego catches up to it. A gap check ensures no existing vehicle is within a defined range of the spawn point.

Table 3 summarises the controller parameters used in the current evaluation. The values were selected empirically for the driving simulator setup.

Table 3: Traffic state controller parameters.

Parameter	Symbol	Unit	Value
Metric zone half-length	R_m	m	400
Buffer zone half-length	R_b	m	600
EMA smoothing factor	α	—	0.2
Density feedback gain	K_ρ	(veh/s)/(veh/km)	0.02
Min. spawn interval	T_{min}	s	1.0
Min. spawn gap	—	m	20
Min. Speed	v_{min}	km/h	90
Max. Speed	v_{max}	km/h	130
Credit lower clamp	S_{min}	veh	-2
Credit upper clamp	S_{max}	veh	5
Control interval	Δt	s	1.0

Evaluation

The evaluation combines an argument-based assessment of the traffic-state description with a technical controller evaluation in repeated closed-loop simulations within the highly dynamic driving simulator at ika (Wagener, et al., 2023). With respect to RQ1, the traffic-state description satisfies the requirements in Table 1: the selected variables are computed in a moving ego-centred window, remain microscopic and interpretable, and become sufficiently discriminative only in combination, because density, speed distribution, and time-headway distribution capture complementary aspects of traffic amount, motion, and spacing. However, the influence of this on subjective perceptions of the traffic state cannot be determined here and is therefore subject to future research.

With respect to RQ2, the controller evaluation is conducted on a 12 km three-lane highway test road within the highly dynamic driving simulator at ika. The controller is integrated into the environmental control tool and uses the parameter set summarised in Table 3. Each run starts from an empty surrounding traffic state. At $t = 0$ s, the controller is activated while the ego vehicle accelerates from standstill toward $v_{\text{ego,des}} = 110$ km/h. To reduce behavioural variance

and isolate controller performance, the ego vehicle is driven in all runs by an automated driving system comprising adaptive cruise control, lane keeping, and automated lane change functions.

Density ρ is the primary evaluation variable, because establishing and maintaining the target amount of surrounding traffic is the controller's central task and therefore provides the clearest convergence criterion from the empty initial state. Mean speed μ_v and speed standard deviation σ_v within the ego-centred metric zone are evaluated as complementary variables that indicate whether the generated traffic also matches the target motion distribution once the desired density has been established. Time headway is reported only as a dependent supporting variable, because in highway traffic the spacing structure is jointly shaped by density and speed distribution and is therefore not treated here as an independent tracking target. The analysis therefore addresses three questions: whether the controller converges from the empty initial condition, whether the generated traffic state can be maintained over time, and whether this behaviour is reproducible across repeated runs.

Five target configurations are evaluated. The baseline scenario (1) uses $\rho^* = 15$ veh/km, $\mu^* = 110$ km/h, and $\sigma^* = 14.4$ km/h ($n = 7$). Four additional scenarios vary density and target mean speed: a low-density scenario (2) with $\rho^* = 10$ veh/km and $\mu^* = 110$ km/h, a low-density low-speed scenario (3) with $\mu^* = 100$ km/h, a low-density high-speed scenario (4) with $\mu^* = 120$ km/h, and a high-density scenario (5) with $\rho^* = 20$ veh/km and $\mu^* = 110$ km/h, each repeated seven times. In all five scenarios, $\sigma^* = 14.4$ km/h. Steady-state errors are computed over the final 60 s of each run. Accordingly, build-up time is defined as the first time the EMA-smoothed density enters and remains within $\pm 5\%$ of the density target for at least 10 s.

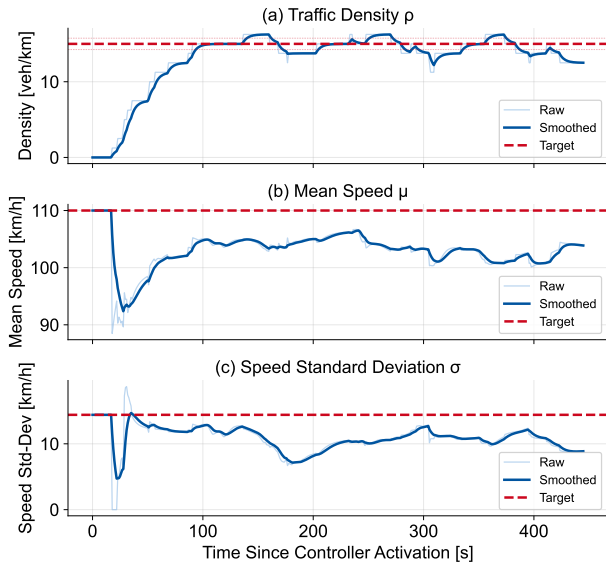


Figure 2: Representative baseline run. Raw and smoothed measurements for density, mean speed, and speed standard deviation are shown together with their respective target values.

Figure 2 first illustrates the closed-loop behaviour in one representative baseline run. Starting from an

empty road, the controller builds up traffic in the ego-centred metric zone until density approaches the target value of $\rho^* = 15$ veh/km. After the transient phase, density remains close to the target with only moderate residual oscillation. Mean speed shows the same qualitative pattern, with an initial build-up phase followed by a stable operating region near the target. The speed standard deviation also converges, but remains below the target value. This offset is consistent with the controller design, because desired speeds are sampled from a bounded distribution at spawn and subsequently modified by car-following interactions, both of which reduce the realised variance.

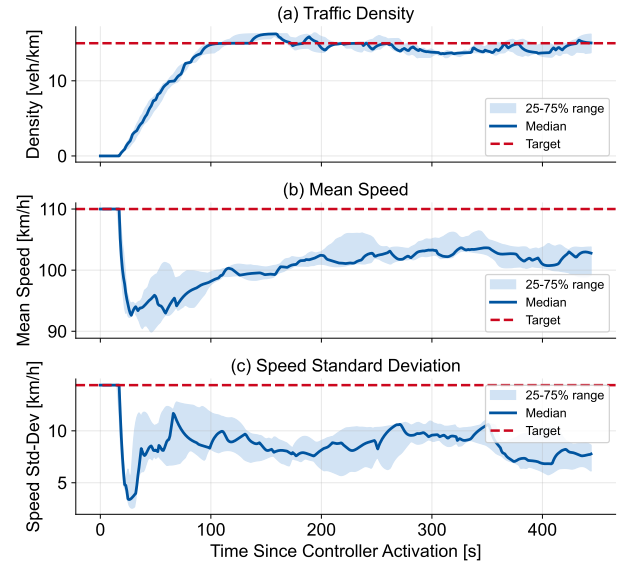


Figure 3: Reproducibility across seven repeated baseline runs. The solid line shows the median trajectory and the shaded area shows the interquartile range for density, mean speed, and speed standard deviation.

Figure 3 addresses within-scenario reproducibility for the baseline by aggregating the results of seven repeated runs. Instead of plotting all runs individually, the figure reports the median value together with the interquartile range for each controlled variable. This makes the run-to-run dispersion directly visible while preserving the overall convergence behaviour. The spread decreases once the traffic state has been established, indicating that the controller not only converges in a single run but also produces a comparable stabilised traffic state across baseline repetitions. Compared with density tracking, the residual speed error remains larger. However, without a human-subject evaluation, its practical relevance for perceived traffic conditions cannot be determined here.

Table 4 extends this analysis to all five target configurations. It reports the target values, the steady-state mean absolute errors over the final 60 s, and the build-up time for each scenario.

The cross-scenario comparison shows a clear trend. The three low-density scenarios yield the smallest density errors, each with a median e_ρ of 0.7 veh/km. At $\rho^* = 10$ veh/km, lowering the target mean speed to 100 km/h produces the smallest mean-speed error (2.9 km/h), whereas raising it to 120 km/h produces

Table 4: Summary of repeated-run reproducibility across evaluation scenarios. Steady-state errors are reported as the median per-run mean absolute error over the last 60 s; values in brackets denote the interquartile range. Build-up time is defined as the first time the EMA density enters and remains within $\pm 5\%$ of the density target for 10 s.

Scenario	n	Targets			Steady-state MAE			
		ρ^* [veh/km]	μ^* [km/h]	σ^* [km/h]	e_ρ [veh/km]	e_μ [km/h]	e_σ [km/h]	t_{build} [s]
Baseline (1)	7	15.0	110.0	14.4	1.1 [0.9, 1.6]	7.7 [6.5, 10.2]	6.7 [4.9, 8.4]	94.0 [90.0, 108.0]
(2)	7	10.0	110.0	14.4	0.7 [0.5, 1.5]	6.5 [4.5, 9.3]	6.0 [5.8, 7.3]	118.0 [111.0, 122.5]
(3)	7	10.0	100.0	14.4	0.7 [0.5, 0.9]	2.9 [2.1, 5.3]	4.6 [4.0, 6.3]	106.0 [97.0, 108.5]
(4)	7	10.0	120.0	14.4	0.7 [0.0, 0.8]	12.6 [10.7, 14.6]	7.4 [6.6, 9.3]	107.0 [94.5, 199.0]
(5)	7	20.0	110.0	14.4	2.2 [1.4, 2.4]	6.8 [6.0, 10.6]	7.2 [3.7, 8.0]	198.0 [105.5, 217.0]

the largest (12.6 km/h). The high-density scenario remains controllable, but it exhibits the largest density error (2.2 veh/km) and the longest build-up time (198 s). This behaviour may reflect both controller limitations and constraints imposed by the road infrastructure. These results indicate that the controller remains functional across all tested target configurations, while tracking performance degrades for more demanding density and speed targets relative to the baseline. However, the present evaluation does not permit a judgement as to which residual errors remain acceptable for simulator studies, because their practical relevance depends on subjective traffic perception and task context and would therefore require human-subject validation.

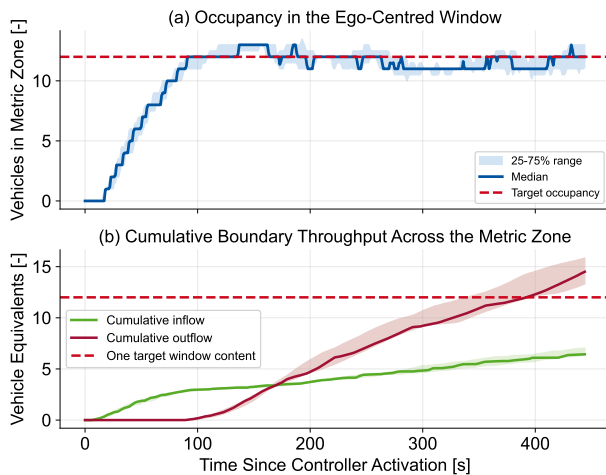


Figure 4: Supporting evidence of traffic exchange in the baseline scenario. The upper panel shows the vehicle occupancy in the metric zone, while the lower panel reports cumulative inflow and outflow in flow-equivalent units across the zone boundaries.

Figure 4 addresses a possible alternative interpretation of the baseline density result, namely that the target density might be maintained by retaining nearly the same vehicles around the ego vehicle. Because the available logs do not contain vehicle identities, exact set membership cannot be reconstructed. For the baseline target density of 15 veh/km and a metric-zone length of 0.8 km, the nominal target occupancy is approximately 12 vehicles. The cumulative boundary throughput exceeds one nominal metric-zone content and is therefore consistent with continuous refresh of the local traffic window. This evidence should be interpreted conservatively as an indirect indicator of turnover rather than as proof of exact vehicle replacement.

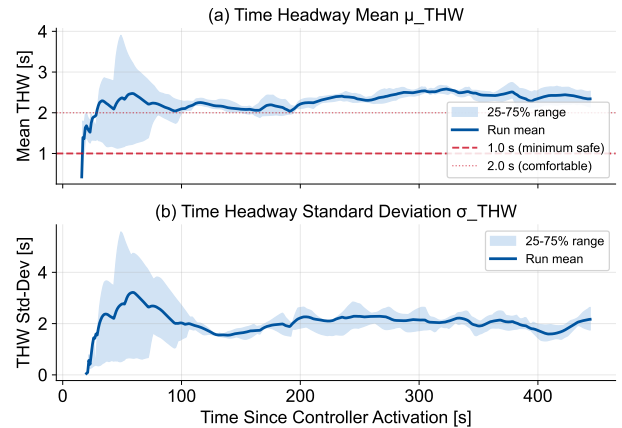


Figure 5: Time headway as a dependent outcome across seven repeated baseline runs. The run-mean trajectory and interquartile range are reported to assess whether the controlled density and speed state yields plausible spacing behaviour.

Figure 5 provides supporting evidence that the controlled density and speed state yields a plausible spacing structure in the baseline scenario. During the initial build-up phase, both the mean time headway and its standard deviation exhibit a wider interquartile range, which is expected while the traffic state is still forming with larger gaps between vehicles. Once the steady phase is reached, the runs converge toward a mean time headway of approximately 2.5 s and a standard deviation of approximately 2.0 s. This indicates that similar spacing behaviour emerges across repeated baseline runs even though time headway is not controlled directly.

The present evaluation is limited to technical controller performance under an ADS-driven ego vehicle. No formal artefact metric, vehicle-identity-based turnover analysis, or human-subject realism assessment was performed. The findings should therefore be interpreted as evidence that the controller can reproducibly maintain the targeted density and speed distribution within the tested highway scenarios for the evaluated use case, but not as a general claim of perceptual realism.

Conclusions and Future Work

This paper presented an ego-centred framework for describing surrounding traffic in driving simulators and, for the present highway use case, a rule-based runtime controller for generating and maintaining selected target traffic descriptors around the ego ve-

hicle. The metric review showed that density alone is not sufficient to specify surrounding traffic for the present use case. Instead, density, speed distribution, and time-headway distribution provide a more robust and interpretable basis for traffic-state specification, reporting, and analysis. On this basis, the proposed highway controller maintains selected environment-level traffic descriptors through vehicle spawning, vehicle removal, and one-time desired-speed assignment at spawn without continuous direct intervention in vehicle behaviour.

The evaluation supports the claim that selected ego-centred environment descriptors can be generated and maintained reproducibly in the investigated simulator setup. Across the tested target configurations, the controller achieved stable traffic build-up and stable steady-state behaviour, with the strongest reproducibility in the baseline scenario and predictable performance changes as density and target mean speed were varied. The supported claim is therefore limited to technical feasibility for the evaluated highway-specific controller instantiation, target range, and simulator integration. The results do not yet establish perceptual realism, generality across broader traffic scenarios, or validation of the full framework under human-driven ego behaviour.

Future work should address these open points directly. First, the proposed traffic-state description framework should be applied in a concrete driving-simulator study in order to assess its practical value for scenario specification, study reporting, and reproducibility. Second, the controller should be evaluated with human-driven ego vehicles in order to quantify the effect of behavioural variability on the generated traffic state. Third, the perceptual realism of the resulting traffic should be assessed explicitly, for example through artefact metrics and user-centred validation. Fourth, the controller should be examined beyond the present parameterisation through sensitivity analyses, broader scenario classes, and comparisons with alternative control architectures such as predictive or learning-based approaches. These steps would extend the present contribution from technical feasibility toward validated methodology for reproducible traffic design in driving simulator studies.

References

- ASAM e.V., 2022. *ASAM OpenSCENARIO V1.2.0*. Standard. Zugriff am: 17.04.2026. ASAM e.V. Available at: <https://www.asam.net/standards/detail/openscenario/>.
- Ferrara, A., Sacone, S., and Siri, S., 2018. Microscopic and Mesoscopic Traffic Models. In: A. Ferrara, S. Sacone, and S. Siri eds. *Freeway Traffic Modelling and Control*. Advances in Industrial Control. Cham: Springer International Publishing, pp. 113–143. ISBN: 978-3-319-75959-3. https://doi.org/10.1007/978-3-319-75959-3_5.
- Fischer, M., Labusch, A., Bellmann, T., and Seehof, C., 2015. A Task-oriented Catalogue of Criteria for Driving Simulator Evaluation. In: H. Bülthoff, A. Kemeny, and P. Pretto eds. *Proceedings of the Driving Simulation Conference 2015*, pp. 139–150. Available at: <https://elib.dlr.de/99561/>.
- Himmels, C., Venrooij, J., Parduzi, A., Peller, M., and Riener, A., 2024. The bigger the better? Investigating the effects of driving simulator fidelity on driving behavior and perception. *Transportation Research Part F: Traffic Psychology and Behaviour*, 101, pp. 250–266. ISSN: 13698478. <https://doi.org/10.1016/j.trf.2024.01.007>. Available at: <https://www.sciencedirect.com/science/article/pii/S1369847824000135>.

- Hock, P., Kraus, J., Babel, F., Walch, M., Rukzio, E., and Baumann, M., 2018. How to Design Valid Simulator Studies for Investigating User Experience in Automated Driving. In: *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM Conferences. New York, NY: ACM, pp. 105–117. ISBN: 9781450359467. <https://doi.org/10.1145/3239060.3239066>.
- Lopez, P. A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y.-P., Hilbrich, R., Lücken, L., Rummel, J., Wagner, P., and Wießner, E., 2018. Microscopic Traffic Simulation using SUMO. In: *The 21st IEEE International Conference on Intelligent Transportation Systems*. IEEE. Available at: <https://elib.dlr.de/124092/>.
- Olstam, J., 2005. Simulation of rural road traffic for driving simulators. In: *Proceedings of the 84th Annual meeting of the Transportation Research Board, Washington D.C., USA, 2005*. Available at: <https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A687680&dsid=7763>.
- 2009. *Simulation of surrounding vehicles in driving simulators*. Vol. 1248. Linköping studies in science and technology / Dissertations. Norrköping: Department of Science and Technology, Linköping University. ISBN: 9789173936606.
- Salles, D., Lang, L., Kehr, M., and Reuss, H.-C., 2022. A Modular Co-Simulation Framework with Open Source Software and Automotive Standards. In: M. Bargende, H.-C. Reuss, and A. Wagner eds. *22. Internationales Stuttgarter Symposium*. Proceedings. Wiesbaden: Springer Fachmedien Wiesbaden, pp. 207–223. ISBN: 978-3-658-37010-7. https://doi.org/10.1007/978-3-658-37011-4_18.
- Treiber, M., Jan. 2013. *Traffic Flow Dynamics*. ISBN: 978-3-642-32459-8. <https://doi.org/10.1007/978-3-642-32460-4>.
- Wagener, N., Ruß, F. S., Beckmann, J. N. B., Legran, P., and Eckstein, L., 2023. ika's highly dynamic driving simulator. In: *22. Driving Simulation Conference, Antibes (France)*, 6 Sep 2023 - 8 Sep 2023. Available at: <https://publications.rwth-aachen.de/record/980509>.
- Wang, X. and Qin, H., 2025. Impact of Combined Cognitive and Visual Distraction Level on Driving Behavior in Different Traffic Environments. *International Journal of Automotive Technology*, 26(7). Cited by: 0, pp. 1691–1699. <https://doi.org/10.1007/s12239-025-00220-0>. Available at: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85217215786&doi=10.1007%2fs12239-025-00220-0&partnerID=40&md5=a290c9273302634564d3593684e3f012>.
- Wang, Z., Sheng, H., Gu, F., Zhao, L., Wang, Z., and He, D., 2026. The influence of cabin environment on takeover performance in conditional automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, 117, p. 103478. ISSN: 1369-8478. <https://doi.org/https://doi.org/10.1016/j.trf.2025.103478>. Available at: <https://www.sciencedirect.com/science/article/pii/S1369847825004334>.
- Yurtsever, E., Lambert, J., Carballo, A., and Takeda, K., 2019. A Survey of Autonomous Driving: Common Practices and Emerging Technologies. *CoRR*, abs/1906.05113. arXiv: 1906.05113. Available at: <http://arxiv.org/abs/1906.05113>.

During the preparation of this document, generative AI tools were used for language and grammar checking. The authors reviewed and edited the content as needed and take full responsibility for the publication's content.

Acknowledgement

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.



**Funded by
the European Union**

Applying a Representativity-Based Model Selection Framework to an Industrial ADAS Use Case

Yara Hallak^{1,2}, Layale Saab¹, Dominique Blouin², Baptiste Laborie¹, Laurent Pautet²,

(1) Renault, 1 avenue du Golf, 78084 GUYANCOURT, e-mail: yara.hallak@renault.com, layale.saab@renault.com, baptiste.laborie@renault.com

(2) Telecom Paris, 19 Pl. Marguerite Perey, 91120 Palaiseau, e-mail: yara.hallak@telecom-paris.fr, dominique.blouin@telecom-paris.fr, laurent.pautet@telecom-paris.fr

Abstract - Simulation is increasingly used to replace costly and time consuming physical tests in ADAS development. However, selecting a simulation model that is representative for a specific simulation need remains a major challenge. Current industrial practices often rely on trial error methods and experts experiences, which may lead to extended selection timelines and a process that is difficult to formalize and trace. Although existing work addresses concepts such as fidelity, validation, and credibility, these notions are treated independently and do not provide a unified basis for systematic model selection. To address this limitation, this work introduces the concept of representativity, defined as the combination of fidelity and validity while also accounting for the constraints imposed by the simulation means. Although widely used in industry when selecting simulation models to build simulation platforms, representativity remains absent from scientific literature. This paper presents an automated representativity based evaluation approach designed to support objective, efficient, and repeatable model selection. The method is applied to a Lane Support System non regression case study involving four Vehicle Dynamics model variants. The proposed approach reduces effort, enables non-experts to perform model selection, avoids over or under dimensioned simulation platforms, and ensures traceability throughout the evaluation chain.

Keywords: Modeling and Simulation, Model Representativity, Advanced Driver Assistance Systems, Lane Support System, Simulation Application

Copyright: © 2026 by the authors.

Licensee Driving Simulation Association.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.82157/dsa/2026/22>

Introduction

Modeling and Simulation (M&S) is very important in the development and validation of complex cyber-physical systems (CPS), particularly in the automotive industry where it supports the design and assessment of new functionalities such as the Advanced Driver Assistance Systems (ADAS). Simulation plays a central role in the development and validation of ADAS, enabling engineers to explore a wide range of operating conditions long before physical testing is possible. As simulation usage expands, industrial teams increasingly rely on multiple variants of models, each differing in abstraction level, intended purpose, and compatibility with specific simulation environments. Selecting the most appropriate model for a given simulation need has therefore become a critical step in the ADAS development workflow.

In practice, this selection process is often guided by engineering experience and informal notions of how "representative" a model is for the defined simulation need. Although widely used in industry, the concept of representativity remains loosely defined and is not

well supported by structured evidence. This creates challenges for transparency, repeatability, and communication across teams, especially as simulation becomes a key element in safety-related decision making. Current model selection relies on trial and error methods and on experts experiences, which can sometimes lead to time and resources waste.

This work proposes an evaluation of our requirement-driven approach that formalizes representativity to support systematic and automatic model selection. The work is motivated by the need for clearer criteria, stronger traceability, and more consistent decision-making when comparing simulation model variants. We evaluate it on an industrial Lane Support System (LSS) use case, which is an ADAS functionality to automatically maintain a vehicle in the correct lane while driving.

Research Question

The central research question guiding the study is:
How can the concept of representativity be trans-

formed into a structured, evidence-based framework that supports systematic, transparent, and reproducible selection of simulation model variants for specific simulation need?

Methodology

We define **model representativity** as a model's ability to accurately capture essential characteristics and relationships of the real system, for a specific simulation need, while taking into consideration the constraints of the employed simulation means. This tool application paper builds on our previous work (Hallak, et al., 2026), and focuses on demonstrating the practical application of our approach. More detailed explanations of the method, along with additional examples, can be found in this article [†] (Hallak, et al., 2026).

Representativity Assessment Scale

The proposed approach is based on our Representativity Assessment Scale (RAS) which proposes multiple factors, each of which capturing a specific dimension of a model's ability to support a given simulation need. The considered factors include: **Functions Coverage**, the extent to which a model implements the functions required for the simulation task. **Interfaces Existence**, the availability and correctness of a model's input/output interfaces required for integration into a simulation platform with the other components. **Model Verification**, the evidence that a model has successfully passed internal verification activities e.g., technical checks. **Maturity**, the completeness, stability, and documentation level of a model. **Validity**, the alignment of a model's behavior with reference data, physical principles, or expected system characteristics e.g. compliance with validation tests scenarios. **Simulator Compatibility**, the compatibility and ability of a model to execute correctly within a target simulation platform. Each factor is assigned a discrete level from 0 (insufficient evidence) to 3 (full evidence), allowing a qualitative and quantitative assessment of model suitability.

Automation of Model Selection Process

To automate the model selection process, we use the Architecture Analysis and Design Language (AADL) (SAE International, 2017) with the Architecture-Led Incremental System Assurance (ALISA) (Blouin, et al., 2023) assurance cases workbench on OSATE environment. This approach achieves representativity evaluation through a structured and requirement-oriented assessment framework. Simulation requirements are explicitly formulated and categorized according to these RAS factors to capture both system-level and simulation-specific concerns. Assurance activities are then specified to verify each requirement, producing structured evidence that directly supports scale level assignment. The execution of the ALISA verification activities generates assurance lists, which are automatically processed by the tool

to compute requirement satisfaction ratios per factor. These ratios are mapped onto the levels of the representativity scale. The resulting scores allow a systematic comparison between models and the selection of the one. At the end of the assessment routine, an automatically generated assurance-case report consolidates all evidences, requirement satisfaction results, factor scores, and scale levels, providing decision makers with a traceable justification for model selection that can be reviewed, reused, and audited.

Motivations and Benefits of Approach

The proposed representativity-based evaluation framework addresses several limitations observed in current industrial practices for simulation model selection which often rely on trial-and-error procedures or experts expertise. Such processes can be time-consuming and resource-intensive. In contrast, our approach enables automated model selection based on objective criteria and explicit requirements, reducing time and effort. Therefore, it empowers non-experts to participate in the model selection process. By evaluates all dimensions of representativity, through the RAS factors, the approach avoids selecting over or under dimensioned models, ensuring alignment between the chosen model and the simulation need while avoiding excessive computational cost and insufficient behavioral fidelity. Finally, the approach provides full traceability; each decision is backed by explicit requirements, verification evidence, and quantitative scores, enabling consistent and repeatable model selection practices within industrial workflows.

Use Case

In this paper, we apply our approach to an industry-driven case study focused on a Lane Support System (LSS). The LSS is an ADAS function responsible for monitoring the vehicle's lateral position and providing corrective steering assistance to help the driver stay within the lane, prevent unintended lane departures, and enhance overall lane-keeping stability. For this case study, the simulation is required to select the most representative vehicle dynamic model for evaluating the performance of the LSS in a non-regression context. As the ADAS system evolves, these changes may influence the behavior of the LSS. To ensure that such evolutions do not degrade LSS performance, it is essential to choose a vehicle dynamics model that accurately represents the conditions under which the LSS must be validated.

The selection process considers four Vehicle Dynamics (VD) model variants. The **first variant** is a vehicle compatible model originally developed for general vehicle dynamics analysis but not intended for ADAS evaluation. The **second variant** is a compatible standalone model designed for both the vehicle and ADAS domains. The **third variant** is incompatible with the vehicle configuration expected. The **fourth variant** is interfacing with the Electric Power Steering (EPS) virtual calculator to exchange steering related signals. In this case study, we will consider the VD and the EPS as one block component.

Although the four model variants are described here with clear distinctions for illustration purposes, such

[†] <https://zenodo.org/records/19597191?>

Evidence	Pass	Fail	Error
~  Case DynamicVehicleLSS1	175	54	38
~  Plan DynamicVehicleV1_API	175	54	38

Figure 1: VD Model Variant 1 Assurance List

Evidence	Pass	Fail	Error
~  Case DynamicVehicleLSS4	267		
~  Plan DynamicVehicleV4_API	267		

Figure 2: VD Model Variant 4 Assurance List

information is rarely available upfront in industrial contexts. In practice, details on model intent, domain compatibility, and limitations are often distributed across documentation, tools, or individual expertise, making model selection far from trivial. The proposed framework is therefore not intended to confirm an obvious choice, but to reconstruct this decision in a systematic, evidence-based way. By explicitly linking requirements, verification results, and representativity factors, the framework exposes the underlying rationale that leads to the selection of the most appropriate model.

Results and discussion

Results Overview

Each requirement was associated with one of the RAS factors. For this case study, we defined 157 requirements, verified through 267 verification activities. ALISA assurance lists (see Fig 1, 2) contains the verification outcomes i.e. pass, fail, error for each requirement. A pass indicates that the requirement is satisfied, while fail and error indicate that the requirement is not satisfied. Using these results, we computed satisfaction ratio for each factor. These ratios, expressed as percentages, are mapped to corresponding levels on the RAS. The resulting ratios provide a quantitative view of the representativity profile of each model variant with respect to the defined simulation need. Table 1 summarizes the results obtained for the four VD variants. Table 2 shows the obtained RAS for the first variant after mapping the percentages to their corresponding scale levels.

The results reveal clear differences across the variants, even in cases where some models share similar levels of functional detail or behavioral complexity. This demonstrates that representativity does not depend on any single factor in isolation, but emerges from the combined contribution of multiple factors, each addressing a different axis of the simulation need.

Interpretation of Results

The representativity assessment highlights clear differences across the four DV model variants, despite some of them sharing similar functional or structural characteristics.

The first variant, although vehicle-compatible, shows a heterogeneous profile. It achieves full Maturity, however, it achieves partial Functions Coverage.

It achieves limited Interfaces Existence and low Validity, revealing significant gaps in the interfaces and contextual alignment needed for LSS evaluation. Its moderate Model Verification and Simulator Compatibility indicate partial but incomplete suitability for the ADAS simulation environment. **The second variant**, presents a more consistent profile with full Functions Coverage, full Maturity, higher Model Verification rate, and full Simulator Compatibility indicating that it is better aligned with the LSS simulation environment. But its Validity score of 63% reflects its inability to pass several validation tests. This variant generates a steering torque that is not physically correlated and fails to apply it correctly. **The third variant** achieves full Functions Coverage and full Maturity. Its Model Verification score drops due to incompatibility with the target vehicle configuration. Despite good Validity and Simulator Compatibility, its incompatibility with the target vehicle affected the representativity of the variant for the intended LSS use case. **The fourth variant**, achieves 100% across all factors. Its integration with the EPS virtual calculator ensures realistic steering behavior, making it the most representative model for the defined simulation need.

Overall, the results demonstrate that representativity is multi-dimensional. This confirms that representativity cannot be inferred from a single characteristic; it emerges from the combined contribution of multiple factors, each addressing a different axis of the simulation need.

Related Work

Several approaches exist for describing and evaluating simulation models, but they focus on aspects other than representativity. Credibility-oriented frameworks, such as the NASA's Credibility Assessment Scale (CAS) (Babula, et al., 2009) and the AFS project, developed by IRT SystemX, focus on evaluating the degree of confidence that a stakeholder places in the model based on evidence management and structured argumentation. Fidelity focused approaches, such as the Fidelity Evaluation Framework (FEF) (Schricker, Franceschini, and Johnson, 2001), evaluates how closely a model reproduces the behavior of a reference system.

Our work differs in that it focuses specifically on **model representativity**, i.e. the suitability of a model for a specific simulation need, combining fidelity and validity while accounting for simulation means constraints. Inspired by existing approaches but pursuing a different objective, our framework automates the evaluation of whether a model is representative for a given need, thereby supporting systematic model selection.

As part of this work and in collaboration with the AFS project at IRT SystemX, we are developing an ontology to formalize the concepts used to describe simulation models and their relationships. Its goal is to clarify terms such fidelity, maturity, precision, accuracy, and to make their distinctions explicit for consistent communication between parties such as OEM and suppliers. This ontology will provide a foundation for future extensions of the representativity framework and support alignment with broader simulation credibility initiatives.

Table 1: Factors success percentage values for vehicle dynamic model variants

Variant-Factor	Functions Coverage	Interfaces Existence	Model Verification	Maturity	Validity	Simulator Compatibility
1st Variant	71%	65%	80%	100%	38%	71%
2nd Variant	100%	74%	87%	100%	63%	100%
3rd Variant	100%	74%	67%	100%	75%	100%
4th Variant	100%	100%	100%	100%	100%	100%

Table 2: Representativity Assessment Scale with achieved levels highlighted for Variant 1

Level - Factor	Function coverage	Model Verification	Interface existence	Maturity	Validity	Simulation Mean Compatibility
Level 3	Full Coverage , Realistic Behavior	No or minor numerical error	All interfaces, correctly typed	Operational	Valid in required context	Fully compatible
Level 2	Enough Coverage for Ideal Behavior	Formal error estimation	Missing interfaces, correctly typed	Pre-deployment	Valid in similar context	Compatible with minor adaptations
Level 1	Partial Coverage	Passes some tests	Missing interfaces, Incorrectly typed	Prototype	Validation in progress	Partially Compatible
Level 0	Insufficient Evidence	Insufficient Evidence	Insufficient Evidence	Experimental	Insufficient Evidence	Not Compatible

Discussion and Future Work

The current prototype demonstrates that representativity assessment can be effectively implemented using AADL and ALISA within OSATE, providing a functional and operational workflow that formalizes the definition of requirements and automatically verifies them through associated verification activities and claims.

However, to fully address industrial needs, the next step is to integrate this workflow with the Model Identity Card (MIC) (IRT SystemX, n.d.), the standard mechanism for specifying and characterizing simulation models and building simulation platforms architectures Hallak, et al., 2024. While the MIC captures many model attributes, some parameters required for representativity assessment are not yet included or formally defined. At present, the variants in OSATE are characterised through manually entered AADL properties filled with their gathered values from various sources.

The long term ambition of this work is to extend the MIC with the necessary fields such as compatibility indicators, domain integrated constraints, validation tests results, and connect it directly to the assessment workflow. This would allow the tool to retrieve all required information automatically without manual data entry.

In such a workflow, model characteristics are maintained in the MIC, the assessment tool queries them directly, and ALISA performs automated comparison and requirements verification. The prototype therefore provides a solid foundation, and future work aims to industrialize the process by linking it to the MIC for full consistency, automation, and scalability.

Conclusions/implications

This paper showed how the industrial notion of representativity can be formalized and operationalized to support systematic selection of simulation model variants. By structuring simulation needs as explicit requirements, linking them to verification activities,

and aggregating evidence across six representativity factors, the framework provides a transparent and reproducible basis for comparing simulation models. Applied to an LSS case study involving four VD model variants, the framework generated representativity profiles that differentiated the strengths and limitations of each variant. From an industrial perspective, the approach reduces reliance on expert intuition, improves communication between modeling and simulation teams, and provides traceable justification for model selection decisions. The structured evidence generated by the framework can be reused across projects, contributing to long term model governance and supporting multi organization collaboration.

References

- Babula, M., Bertch, W., Green, L., Hale, J., Mosier, G., Steele, M., and Woods, J., Jan. 2009. NASA Standard for Models and Simulations: Credibility Assessment Scale. en. In: Orlando, Florida. ISBN: 978-1-60086-973-0. <https://doi.org/10.2514/6.2009-1011>.
- Blouin, D., Crisafulli, P., Maxim, C., and Caron, F., Oct. 2023. An Introduction to ALISA and Its Usage for an Industrial Railway System Case Study. en. *ACM SIGAda Ada Letters*, 43(1), pp. 69–72. ISSN: 1094-3641. <https://doi.org/10.1145/3631483.3631493>.
- Hallak, Y., Blouin, D., Pautet, L., Saab, L., and Laborie, B., 2026. A Structured Approach for Automated Representativity Evaluation of Simulation Models. <https://doi.org/https://doi.org/10.5281/zenodo.19597191>. Available at: <https://zenodo.org/records/19597191?i>.
- Hallak, Y., Blouin, D., Pautet, L., Saab, L., Laborie, B., and Mittal, R., Sept. 2024. Model Management at Renault Virtual Simulation Team: State of Practice, Challenges and Research Directions. en. In: Linz Austria: ACM, pp. 1005–1014. ISBN: 979-8-4007-0622-6. <https://doi.org/10.1145/3652620.3688223>.
- IRT SystemX, n.d. *Model Identity Card (MIC)*. <https://mic.irt-systemx.fr/mic>.
- SAE International, 2017. *SAE AS5506: Architecture Analysis & Design Language (AADL)*. <https://www.sae.org/standards/as5506-architecture-analysis-design-language-aadl>.
- Schricker, B., Franceschini, R., and Johnson, T., 2001. Fidelity evaluation framework. en. In: Seattle, WA, USA: IEEE Comput. Soc, pp. 109–116. ISBN: 978-0-7695-1092-7. <https://doi.org/10.1109/SIMSYM.2001.922122>.

SIMULATION DESIGN

CycloCARLA: Design, Implementation, and Evaluation of a High-Fidelity Bicycle Simulator

Patrick Rebling, Yannick Rauch, Nico Sander, Philipp Nenninger, Reiner Kriesten

Institute of Energy Efficient Mobility, Karlsruhe University of Applied Sciences, Karlsruhe, Germany, e-mail: {patrick.rebling, yannick.rauch, nico.sander, philipp.nenninger, reiner.kriesten}@h-ka.de

Abstract - Achieving physical fidelity in bicycle simulators is challenged by longitudinal resistance mismatches and a lack of active steering torque. This paper introduces CycloCARLA, an accessible, high-fidelity research simulator framework built on CARLA using commercial off-the-shelf hardware. Longitudinal fidelity is validated via CAN bus data from an e-bike, showing near-identical energy consumption (4.99 Wh real vs. 4.98 Wh simulated). A user study ($N = 30$) confirms that our active force feedback pipeline significantly improves perceived realism without inducing simulator sickness. The framework's source code is publicly available at <https://git.ieem-ka.de/simulator-environments/carlaverse/cyclocarla>.

Keywords: Bicycle Simulator, Digital Twin, E-Bike, CARLA

Copyright: © 2026 by the authors.

Licensee Driving Simulation Association.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.82157/dsa/2026/23>

Introduction

The transition toward sustainable urban mobility has established the electric bicycle (e-bike) as a core component of modern transportation ecosystems (Younes, et al., 2023). However, testing the interactions between vulnerable road users (VRUs) and connected automated vehicles (CAVs) in real-world environments entails critical safety risks and high operational costs (Grigoropoulos, 2023). While simulation tools offer a reproducible alternative (Busch, et al., 2020), achieving physical fidelity remains a major hurdle. Automotive simulators benefit from the vehicle's dominant mass, whereas bicycle simulation demands highly sensitive haptic interfaces to map the rider's continuous stabilizing inputs (Dialynas, et al., 2020).

Current low-cost research setups with static mockups face two primary challenges: a mismatch in longitudinal dynamics that distorts acceleration and braking perception due to inertial issues (Powell, Stroh, and Thomas, 2018), and the absence of active steering torque required for stabilization (Martinez Garcia, et al., 2023). Custom robotic testbeds resolve these issues but remain prohibitively expensive and difficult to replicate (Dialynas, Happee, and Schwab, 2019). Conversely, commercial off-the-shelf (COTS) fitness devices lack the low-latency control and multi-physics coupling needed for valid behavioral research.

To bridge this gap, this paper presents CycloCARLA, a modular, open-source human-in-the-loop framework built on the Car Learning to Act (CARLA) platform (Dosovitskiy, et al., 2017). To verify the plat-

form's validity, this research systematically addresses two core questions:

RQ1: *To what extent does the simulation's pedal resistance model accurately replicate the real-world torque and power dynamics of an instrumented e-bike?*

RQ2: *How can CARLA's Application Programming Interface (API) be used to convert virtual micro-terrain irregularities and impulse events into stable, realistic steering torque feedback?*

System Architecture

The software framework of CycloCARLA operates as a closed-loop system dividing tasks between a Python-based client and a custom C++ microcontroller firmware (see Figure 1). To ensure high accessibility, all active hardware components are mounted entirely to an external framework base rather than the bicycle itself. This non-invasive design allows researchers to detach and exchange unmodified bicycle frames within minutes to test varying geometries without sensor recalibration.

The Python client serves as the central orchestrator, mapping hardware outputs to compute target kinematics ($\vec{v}_{tgt}$) and tracking environmental physics within CARLA. Hardware interaction is managed via an ESP32 microcontroller using its dual-core architecture to prevent serial communication bottlenecks. To ensure strict timing constraints and reliable data flow, the firmware design is based on the principles of a time-controlled Hardware Abstraction Layer (HAL)

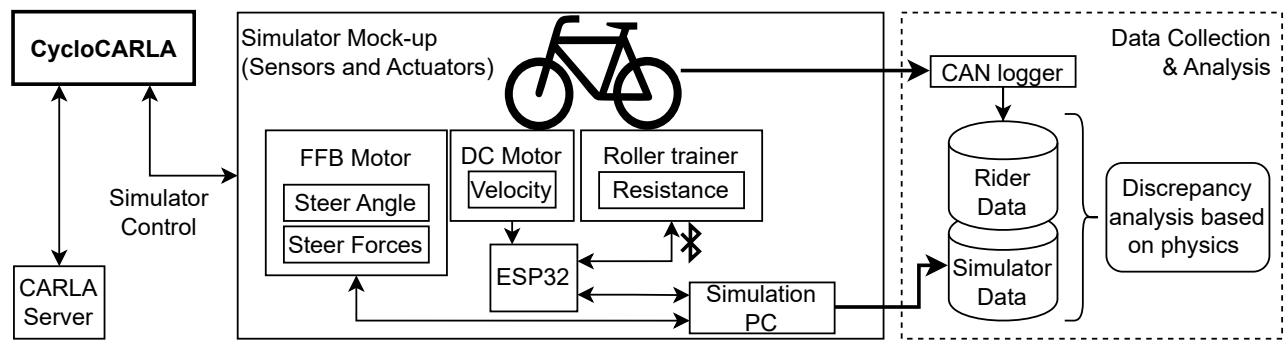


Figure 1: System architecture and validation methodology of CycloCARLA. The loop connects the software framework with the physical mock-up, using a CAN logger for ground-truth discrepancy analysis.

(Simmann, Veeranna, and Kriesten, 2024). In terms of program execution flow, a continuous three-step closed-loop sequence is executed: (1) Core 0 executes a dedicated sensor task, sampling an analog Direct Current (DC) flywheel motor voltage every 10 ms to calculate wheel speed, and broadcasting the trainer's revolutions per minute (RPM) (n_{trainer}) every 100 ms via serial to the Python client. (2) The Python client processes this speed to update the virtual bicycle's kinematics ($\vec{v}_{\text{igt}}$) in CARLA, tracks environmental interactions, and computes the relative resistance target ($F_{\text{res,rel}}$). (3) Concurrently, Core 1 polls incoming values of $F_{\text{res,rel}}$, converts them into proprietary byte arrays, and transmits them to a Wahoo KICKR Snap roller trainer via Bluetooth Low Energy (BLE). Simultaneously, the custom force feedback (FFB) server reads physical handlebar angles, queries environmental micro-terrain and collision events from CARLA, and commands the Moza R12 direct-drive wheelbase via DirectInput to apply active steering forces (combining road surface vibrations, obstacles, and velocity-dependent auto-centering) as shown in Figure 3.

Longitudinal Calibration and Fidelity (RQ1)

To evaluate RQ1 and eliminate non-linear hardware-induced discrepancies, a digital twin approach was implemented. Ground-truth telemetry was captured from a physical e-bike equipped with a custom CAN logger recording rider cadence, speed, and propulsive motor power. In this context, the e-bike's *propulsive force* represents the forward force actively exerted by the rider and electric motor, whereas the simulator's *resistance force* corresponds to the target braking load commanded by the simulation to mimic external physical factors (aerodynamic drag, rolling resistance, and gravity). The calibration protocol involved systematic stationary test rides across an operational range from 0% to 100% resistance in 10% increments. To maintain a constant cadence and velocity against high resistance loads, the e-bike's internal electrical support (50–300% intensity) was actively used. Based on this empirical dataset, a two-dimensional Look-Up Table (LUT) was generated to dynamically interpolate the exact braking force percentage required from the smart trainer based on instantaneous wheel velocity and the requested resistance force calculated by the simulation (Rauch, et al., 2023).

To validate the compensated physics engine, a final test ride was executed over a replicated virtual track featuring alternating flat, uphill, and downhill segments (Figure 2). Telemetry comparisons revealed minor speed deviations due to hardware sampling limits: the real e-bike captures wheel rotation via a single magnetic sensor pair on a 27.5-inch tire, resulting in slight signal lag during dynamic transitions. In contrast, the simulator continuously samples the DC flywheel motor, which rotates approximately 14 times faster than the rear wheel, providing higher data granularity.

Force analysis highlights distinct physical characteristics across different riding phases. During acceleration phases, the propulsive force measured at the e-bike exceeds the simulation's requested load (by approximately 15–20 N) because the roller trainer's physical flywheel mimics structural vehicle inertia. Conversely, during deceleration and coasting, the real propulsive force drops to zero while the simulator's resistance force remains active to brake the wheel. This highlights a physical limitation of COTS trainers, which maintain a mechanical baseline minimum internal resistance of 21 N and cannot generate positive propulsive forces to push the wheel and simulate uninhibited downhill coasting. Consequently, the close agreement in total integrated energy expenditure (4.99 Wh of propulsive energy from the physical e-bike versus 4.98 Wh from the simulator, showing a minor 0.2% deviation) verifies steady-state metabolic and resistance calibration, but it does not fully prove high fidelity across all transient dynamic transitions. Under steady-state conditions (uphill climbs and flat sections), the forces align closely, confirming the model's validity for typical training and commuting speeds.

Subjective Haptic Evaluation (RQ2)

To address RQ2 and assess the steering pipeline's realism, a user study was conducted with a sample of $N = 30$ valid participants (22 male, 8 female) aged 21 to 51 years ($M = 29.0$, $SD = 7.0$). The experiment used a within-subjects, counterbalanced repeated-measures design. Under two distinct conditions (a baseline setting featuring a static handlebar with deactivated FFB, and an active condition using the high-torque FFB pipeline illustrated in Figure 3), participants were allowed to explore the virtual urban environment freely. Rather than following a rigid

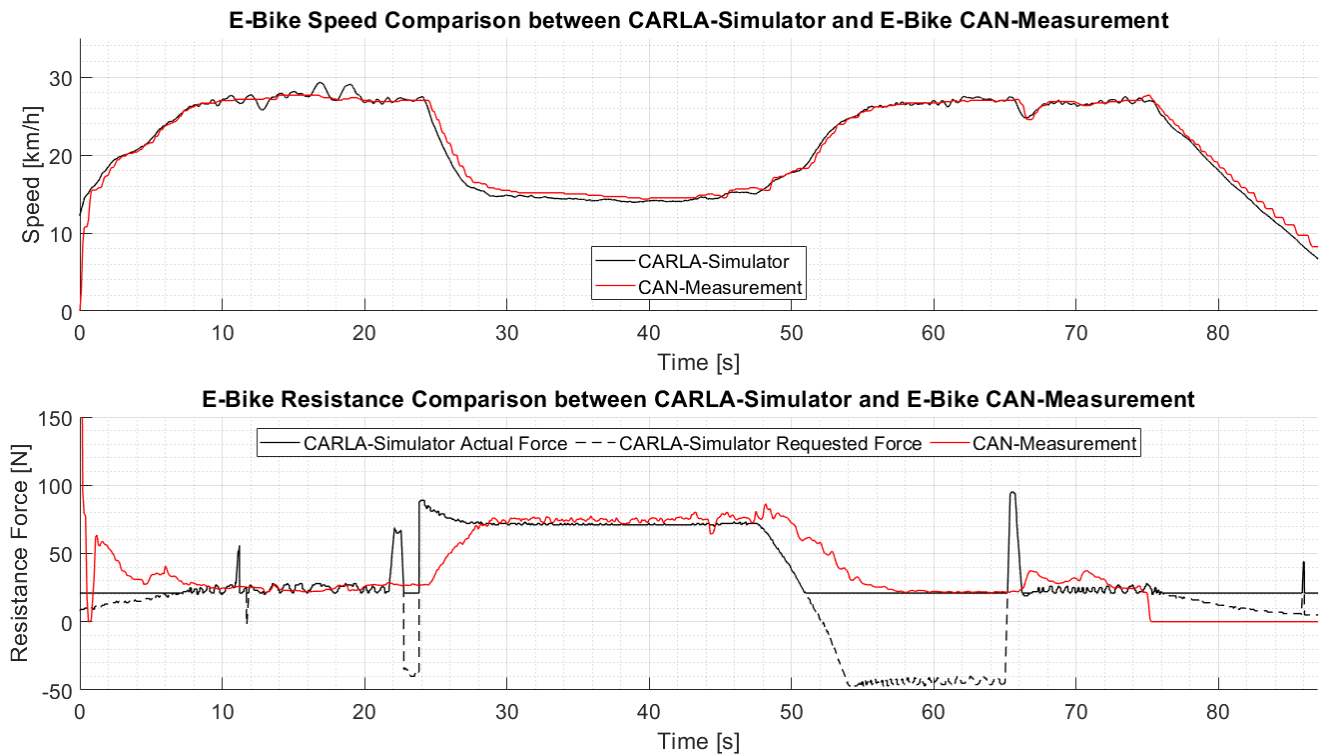


Figure 2: Evaluation of e-bike longitudinal dynamics: Comparison of speed and resistance forces between CAN-bus measurements and the CARLA simulator during up- and downhill segments.

path, this free-roaming approach enabled users to intentionally repeat physical events (such as crossing obstacles, turning, and transitioning between varied micro-terrain textures like cobblestones, grass, and asphalt) to focus intensively on the haptic sensation and response of the handlebars. Subjective metrics were gathered immediately after each ride using standardized 4-point Likert scales (1 = strongly disagree, 4 = strongly agree).

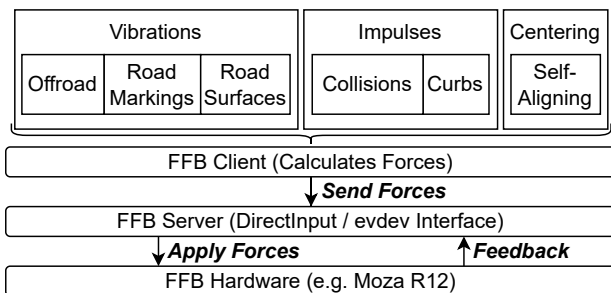


Figure 3: The force feedback pipeline of CycloCARLA. Different force types are calculated based on virtual environment and semantic information and sent to the FFB server that applies the forces to the FFB hardware.

Quantitative Results

The evaluation focused on three major functional categories, analyzed using Wilcoxon signed-rank tests to evaluate differences between the deactivated baseline and active FFB conditions ($N = 30$):

- **Control and Stability:** Self-reported tracking stability was high across both environments ($Mdn =$

3.50, $IQR = 0.50$ for baseline; $Mdn = 3.38$, $IQR = 1.00$ for active FFB). A Wilcoxon signed-rank test confirmed no statistically significant difference in stability ($W = 126.5$, $p = 0.210$). The slightly higher variance in the active FFB condition reflects the physical effort required to compensate for high-frequency terrain forces. However, forced-choice interviews revealed that 63.3% of participants preferred the active FFB condition for superior lane-keeping support.

- **Realism and Feedback:** The haptic pipeline exerted a substantial impact on environmental plausibility. The baseline setup received the lowest possible rating regarding road textures and obstacle traversals ($Mdn = 1.00$, $IQR = 0.00$), whereas active FFB led to a highly significant increase in perceived realism ($Mdn = 3.25$, $IQR = 0.25$; Wilcoxon $W = 0.0$, $p < 0.001$), validating the framework's capability to render micro-terrain features effectively.
- **Simulator Sickness:** Physical comfort remained optimal across all trials, yielding identical scores ($Mdn = 4.00$, $IQR = 0.00$) for both conditions. A Wilcoxon signed-rank test indicated no significant difference in sickness or discomfort ($W = 2.0$, $p = 0.564$), proving that the active FFB pipeline does not induce additional visuo-vestibular sensory conflicts or dizziness, matching strict requirements for impedance causality in haptic setups (Schwab and Recuero, 2013).

Ultimately, 96.7% of all participants confirmed that active FFB established a significantly stronger connection between the visual scene and physical steering responses. Exploratory Spearman correlations indicated no strong demographic biases ($r_s = -0.19$, $p = 0.319$ for frequency of cycling vs. active realism),

demonstrating consistent immersion across varying real-world cycling backgrounds.

Qualitative Feedback and Limitations

Semi-structured post-study interviews supported the quantitative findings. When navigating the baseline configuration, users reported a distinct lack of presence, frequently describing the bicycle as *floating* or completely disconnected from the virtual asphalt. In the active FFB condition, participants commented positively on the haptic texture but noted minor technical imperfections, such as excessive steering resistance during sharp maneuvers.

A primary structural limitation of the current mock-up is the lack of physical lateral tilt, a constraint often observed in low-cost virtual reality configurations (Khundam, Thamrongrat, and Panwisawas, 2025). This rigid vertical setup limits research on high-speed cornering dynamics and rapid evasive maneuvers where body leaning is crucial. However, the evaluation demonstrates that it remains valid for low-speed urban scenarios, intersection negotiations, and cyclist-pedestrian interactions where handlebar steering inputs dominate balancing behavior.

Conversely, the use of affordable, modular COTS components represents a significant advantage. Rather than relying on expensive custom robotic setups, this low-cost architecture makes it economically and technically feasible to deploy multiple synchronized simulator units. This enables future large-scale, multi-user studies to explore cooperative cycling dynamics, cyclist-to-cyclist interactions, and cyclist-to-vehicle cooperative behavior within a shared virtual environment.

Conclusion and Future Work

CycloCARLA establishes that a high-fidelity, reproducible research simulator can be successfully engineered using affordable COTS components and open-source software. The digital twin validation confirms exceptional longitudinal accuracy with just 0.2% energy deviation. Meanwhile, the user study proves that the active haptic pipeline significantly enhances perceived realism without inducing simulator sickness.

Future work will focus on optimizing the steering algorithms to eliminate excessive counter-torque during sharp maneuvers and mapping complex lateral balancing equations more smoothly. Furthermore, leveraging CARLA's network architecture, we aim to integrate CycloCARLA into a distributed, multi-simulator framework (Rebling, et al., 2025). This will enable the cross-institutional study of large-scale, synchronized interactions between human cyclists, e-bike assistance systems (Rauch and Kriesten, 2024), automated vehicles, and pedestrians within a shared virtual environment.

Acknowledgments

This work was funded by the projects KIIWI (16DHBK1060, German Federal Ministry of Research, Technology and Space (BMFTR)), move.mORE (Innovation University (IHS)) and Real2Sim (BW7_2199/2, Baden-Wuerttemberg Ministry of Economic Affairs, Labor and Tourism).

References

- Busch, F., Kath, H., Keler, A., Hosseini, S. A., Grigoropoulos, G., and Kath, J., Sept. 2020. *Fahrradsimulator: Anwendungsorientierter Erfahrungsbericht zu Aufbau und Nutzung*. Tech. rep. Technical University of Munich, p. 13. <https://doi.org/10.14459/2019MD1575160>. [Accessed Feb. 26, 2025].
- Dialynas, G., De Haan, J. W., Schouten, A. C., Happee, R., and Schwab, A. L., June 2020. The Dynamic Response of the Bicycle Rider's Body to Vertical, Fore-and-Aft, and Lateral Perturbations. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, 234(7), pp. 1944–1957. ISSN: 0954-4070, 2041-2991. <https://doi.org/10.1177/0954407019891289>. [Accessed Mar. 30, 2026].
- Dialynas, G., Happee, R., and Schwab, A. L., Jan. 2019. Design and Hardware Selection for a Bicycle Simulator. *Mechanical Sciences*, 10(1), pp. 1–10. ISSN: 2191-916X. <https://doi.org/10.5194/ms-10-1-2019>. [Accessed Mar. 30, 2026].
- Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., and Koltun, V., 2017. CARLA: An Open Urban Driving Simulator. In: S. Levine, V. Vanhoucke, and K. Goldberg eds. *Proceedings of the 1st Annual Conference on Robot Learning*. Vol. 78. Proceedings of Machine Learning Research. PMLR, pp. 1–16. [Accessed Aug. 4, 2025].
- Grigoropoulos, G., 2023. Resolving Interactions among Fully Automated Vehicles and Bicyclists in Complex Traffic Situations. PhD. Technische Universität München. [Accessed Mar. 20, 2026].
- Khundam, C., Thamrongrat, P., and Panwisawas, C., 2025. A Novel Virtual Reality Cycling System Using Controllers for Pedal Motion-Based Tracking: Movement Techniques, Motion Sickness, and User Experience. *IEEE Access*, 13, pp. 152346–152364. ISSN: 2169-3536. <https://doi.org/10.1109/ACCESS.2025.3604557>. [Accessed Mar. 20, 2026].
- Martinez Garcia, D., Gröne, K., Fischer, M., Zhao, M., and Bergen, M., Oct. 2023. Validation of a Bicycle Simulator Based on Objective Criteria. In: *The Evolving Scholar - BMD 2023, 5th Edition*. The Evolving Scholar - BMD 2023, 5th Edition. <https://doi.org/10.59490/650d765b1dbdfef75c1c3a17>. [Accessed Dec. 23, 2025].
- Powell, J., Stroh, O., and Thomas, G. W., Jan. 2018. Hardware Design for an Electro-Mechanical Bicycle Simulator in an Immersive Virtual Reality Environment. *International Journal of Virtual Reality*, 18(2), pp. 1–18. ISSN: 1081-1451. <https://doi.org/10.20870/IJVR.2018.18.2.2903>. [Accessed Mar. 30, 2026].
- Rauch, Y. and Kriesten, R., July 2024. Approach for an Assistance System for E-Bikes to Implement Rider-Adaptive Support. *SAE International Journal of Advances and Current Practices in Mobility*, 07(3), pp. 1308–1317. ISSN: 2641-9645. <https://doi.org/10.4271/2024-01-2979>. [Accessed Dec. 23, 2025].
- Rauch, Y., Rall, J., Ruhe, M., and Kriesten, R., Nov. 2023. Approach to a Holistic Modelling of Cycling Dynamics. In: *SIMUL2023 - The Fifteenth International Conference on Advances in System*.
- Rebling, P., Beeh, L., Nenninger, P., and Kriesten, R., Sept. 2025. Highly-Modular and Immersive Human-in-the-Loop Driving Simulators Using the CARLA Simulation Environment. In: *SIMUL 2025, The Seventeenth International Conference on Advances in System Modeling and Simulation*, pp. 9–14. ISBN: 978-1-68558-300-2. [Accessed Dec. 23, 2025].
- Schwab, A. L. and Recuero, A. M., 2013. Design and Experimental Validation of a Haptic Steering Interface for the Control Input of a Bicycle Simulator. In: *Proceedings of the ECCOMAS Thematic Conference on Multibody Dynamics 2013*. Proceedings of the ECCOMAS Thematic Conference on Multibody Dynamics 2013, pp. 103–110. ISBN: 978-953-7738-22-8.
- Simmann, G., Veeranna, V., and Kriesten, R., July 2024. Design of an Alternative Hardware Abstraction Layer for Embedded Systems with Time-Controlled Hardware Access. In: *2024 Stuttgart International Symposium*. SAE International. <https://doi.org/10.4271/2024-01-2989>.
- Younes, H., Noland, R. B., Von Hagen, L. A., and Sinclair, J., Nov. 2023. Cycling during and after COVID: Has There Been a Boom in Activity? *Transportation Research Part F: Traffic Psychology and Behaviour*, 99, pp. 71–82. ISSN: 1369-8478. <https://doi.org/10.1016/j.trf.2023.09.017>. [Accessed Feb. 26, 2025].

AI & MACHINE LEARNING FOR SCENARIO GENERATION

Open-Source Scenes and Scenarios for various Intelligent Transportation Systems Applications

Kevin Malena¹, Abdelmawla Rizk², Niklas Fittkau², Sandra Gausemeier², Ansgar Trächtler^{2,3}

(1) Private University of Applied Sciences for Business and Technology (PHWT), Rombergstrasse 40, 49377 Vechta, Germany,

(2) Chair of Control Engineering and Mechatronics, Heinz Nixdorf Institute, Paderborn University, Fürstenallee 11, 33102 Paderborn, Germany,

(3) Fraunhofer Institute for Mechatronic Systems Design IEM, Zunkunftsmeile 1, 33102 Paderborn, Germany.

Abstract - This paper presents a standardized framework for scalable, open and reproducible virtual traffic scenes and scenarios for Intelligent Transportation Systems (ITS). Based on public resources such as OpenStreetMap, traffic statistics and sensor-derived data, the modular pipeline generates exemplary open-source scenarios. These accessible environments shall provide common references for cross-domain studies in traffic control, V2X communication, driving simulation and lighting or sensor evaluation. By harmonizing semantics, geometry and metadata, they improve consistency, interoperability, reproducibility and comparability across simulation platforms. Future work shall expand the collection to foster transparency, collaboration and knowledge exchange in academic and industrial ITS research.

Keywords: Traffic Scenarios, Driving Simulation, Traffic Simulation & Control, ADAS.

Copyright: © 2026 by the authors.

Licensee Driving Simulation Association.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.82157/dsa/2026/24>

Introduction

Research on Intelligent Transportation Systems (ITS) increasingly depends on realistic and reproducible simulation environments for evaluating automation, traffic management and cooperative mobility under safe and repeatable conditions. Tools such as CARLA, SUMO and the ASAM-standardized OpenDRIVE and OpenSCENARIO provide open and interoperable infrastructures, but the scenarios built upon them are often project-specific, tailored to individual datasets or optimized for narrow research questions. While some domains, such as pedestrian intention estimation, have established benchmark datasets Girase, et al., 2021; Rasouli, et al., 2019, other ITS applications still lack comparable resources; likewise, dedicated simulation architectures such as mixed-autonomy traffic control Link, et al., 2023 remain poorly standardized. As a result, comparable and reusable reference scenarios are still missing in many ITS domains, unlike standardized manoeuvres in vehicle dynamics such as the double lane-change test (ISO 3888-2).

To address this limitation, we introduce open-source virtual scenes and traffic scenarios^A generated from openly available data sources such as OpenStreetMap and extended through additional datasets. The contribution is a structured pipeline and ex-

emplary reference environments that support cross-domain studies in traffic control, V2X communication, driving simulation, ADAS and lighting or sensor evaluation. By combining open data, standardized formats and interoperable simulation software, the approach supports transparency, reproducibility and comparability across diverse ITS applications.

Motivation and Research Focus

The main hypothesis of this contribution is that standardized, open-source simulation scenarios can markedly improve reproducibility and comparability across ITS research domains. In many prior studies, scenes and scenarios were created ad hoc using similar data inputs but differing modelling conventions, coordinate systems and dynamic behaviours, even when applying the OpenDRIVE and OpenSCENARIO. Consequently, results from related research areas such as adaptive traffic light control, advanced driver assistance systems (ADAS), headlight control, or communication in mixed traffic could not be validated across independent setups.

This work addresses that fragmentation by establishing a shared reference framework for scene and scenario generation as well as description. Following the philosophy of open infrastructures such as I-24 MOTION (see Gloudemans, et al., 2023), provid-

^A See Git: [EEBEO Open-Source Data](#).

ing code and data repositories for the testbed^B and the Simulation of Urban MObility (SUMO) project^C (see Lopez, et al., 2018), the proposed approach extends openness from raw traffic data to comprehensive digital test environments directly integrable into (co-)simulations.

While frameworks such as ScenarioNet of Li, et al., 2023 concentrate on aggregating and unifying real-world driving data from existing large-scale datasets for machine learning applications and CommonRoad of Maierhofer, Klischat, and Althoff, 2021 focus on providing tools for map conversion and editing, our work offers both a modular pipeline for generating such scenarios and openly accessible, fully developed reference environments, thereby establishing a shared and reproducible benchmark for ITS research.

Beyond individual applications such as the traffic light optimization or traffic estimation in Malena, et al., 2024; 2021, the setup enables these and similar domain-specific studies to be conducted within a unified environment. It thereby supports cross-domain analyses and lays the foundation for future co-simulations. Promising examples are coupling headlight models (of the headlight control in Fittkau, et al., 2026) with LiDAR sensor models of Bußemas, et al., 2024 for combined optical and perception testing or integrating traffic control algorithms with sensor data simulation for situation-aware decision-making. By promoting such joint evaluation capabilities, the proposed open-source scenarios and their generation pipeline establish a practical foundation for reproducible, interoperable and cross-domain ITS experiments.

From Scenario Generation to (Co-)Simulation

The standardized scenarios are generated through a modular, data-driven workflow that links open data sources with automated processing tools and simulation platforms. Figure 1 illustrates the overall pipeline,

^B See [code and data repositories](#) for the I-24 MOTION testbed.

^C See [general documentation](#) and [exemplary traffic scenarios](#).

which transforms publicly available geodata into validated virtual traffic environments suitable for multiple ITS applications. The workflow combines several mandatory and optional stages and ensures that the resulting scenarios remain consistent across heterogeneous domain-specific simulation frameworks (e.g., MATLAB/Simulink, SUMO, CARLA, Unity).

Scene and Scenario Generation Workflow

The workflow starts with extracting OSM data within a user-defined bounding box. OSM entities such as nodes, ways and relations are classified into semantic categories including roads, buildings, vegetation and surrounding infrastructure. Complementary datasets, for example satellite imagery, elevation tiles, terrain models, weather records and traffic counts, enrich the static map representation with spatial, temporal and environmental context. Elevation information is incorporated using kriging-based interpolation (see Stein, 1999), transforming sparse terrain samples into continuous road height profiles.

Semantic segmentation is applied to extract road-level and environment-specific structures from aerial and satellite imagery. An EfficientNet-B0 encoder like Tan and Le, 2019 (pretrained on ImageNet^D) combined with a U-Net-style decoder (see Ronneberger, Fischer, and Brox, 2015) performs dense per-pixel classification of roads, sidewalks, buildings, vegetation and traffic-specific infrastructure. The resulting segmentation, illustrated in Figure 2, provides contextual attributes for subsequent geometric reconstruction and planned generative road modelling.

The extracted road topology is converted into an OpenDRIVE-compatible representation. Roads are modelled as ordered point sequences from which centerlines are reconstructed using straight and circular primitives. This preserves geometric continuity while avoiding computationally expensive spline interpolation.

^D Free available [Database of ImageNet](#).

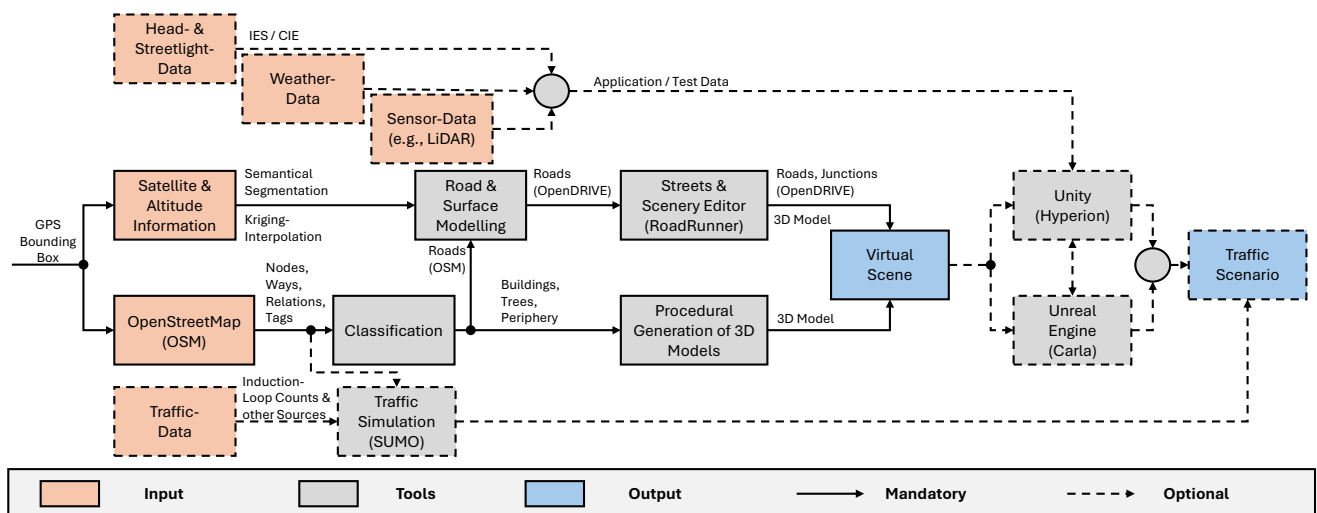


Figure 1: General pipeline for generating virtual scenes and traffic scenarios

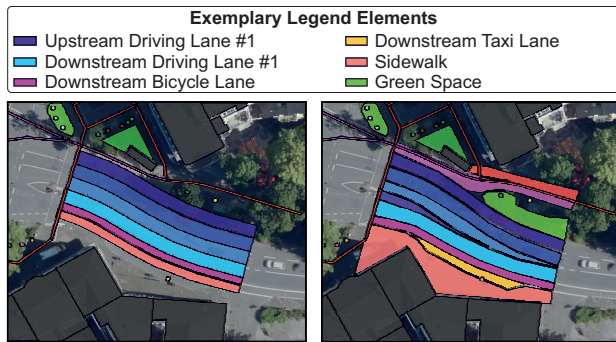


Figure 2: Semantic segmentation results for urban scenes; *left*: segmented OSM-based satellite image; *right*: refined human-annotated segmentation

Towards Generative Reconstruction of Lane-Level Structure

Building upon the segmented imagery, generative reconstruction is investigated to infer missing lane-level details and complex junction structures where OSM data lacks sufficient resolution. The concept uses a conditional WGAN-GP like Arjovsky, Chintala, and Bottou, 2017; Gulrajani, et al., 2017 combined with image-to-image translation (see Isola, et al., 2016). An OSM-derived representation serves as input, while a semantically enriched layout with lane markings, sidewalks, bicycle areas, pedestrian regions and public transport infrastructure acts as target output.

The generator follows a U-Net-like encoder-decoder architecture with an EfficientNet-B0 encoder, while a conditional PatchGAN critic evaluates the local realism of generated structures. Instead of producing RGB imagery directly, the model predicts semantic class logits $\hat{z}$ that are transformed into a differentiable color representation $\hat{y}$ for adversarial training. The objective combines adversarial realism with semantic supervision:

$$\mathcal{L}_G = -\mathbb{E}_{\hat{y}}[D_x(\hat{y})] + \alpha \mathcal{L}_{CE}(y, \hat{z}),$$

where $D_x(\cdot)$ denotes the conditional critic and α weights the sparse categorical cross-entropy term. This loss preserves semantic consistency, whereas the adversarial component encourages spatial coherence and realistic lane-level topology.

The generative model is expected to reconstruct detailed lane markings, connection topologies and intersection layouts beyond the current limitations of OSM-based extraction. As illustrated in Figure 2, such augmented results can enrich the OSM-based input representation with additional lane-level annotations, including dedicated areas for bicycles, pedestrians and public transportation. The resulting augmented representations are intended to provide transferable, simulation-ready digital road models.

Domain Example: Headlight Simulation

Following the standardized workflow in Figure 1, the headlight evaluation use case shows how the generic scenario generation process can be extended to a domain-specific testing application. Starting from OSM data and OpenDRIVE geometry, the virtual

road network is processed with tools such as RoadRunner and exported to rendering engines for lighting and perception analysis. A representative Hyperion environment (see Rüdtenklau, et al., 2019), the *Greitelerweg* in Paderborn, is shown in Figure 3.

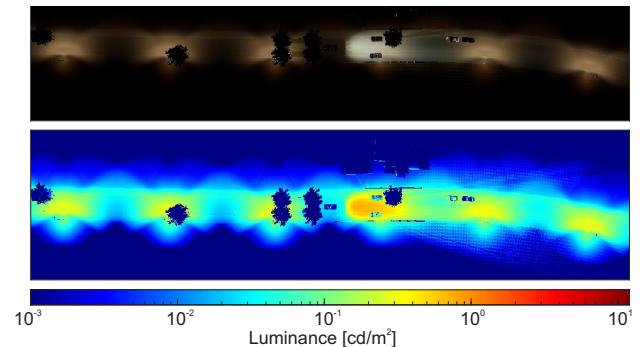


Figure 3: Representative small-scale scene (*Greitelerweg*, Paderborn) in top view; *top*: photorealistic Unity 3D rendering; *bottom*: corresponding luminance heatmap

The figure combines a Unity 3D top view with low beam headlights and batwing-shaped streetlight distributions with a luminance heatmap for evaluating lighting uniformity and headlight ECU performance.

In the final pipeline stage, the scene is imported into Unity or CARLA for Software-in-the-Loop (SiL) and Hardware-in-the-Loop (HiL) validation. For headlight testing, the following domain-specific steps are applied:

1. **Material assignment:** Road and surface objects are enhanced with physically based textures, e.g., asphalt materials from ambientCG^E.
2. **Streetlight integration:** IES photometric files^F, e.g., from TRILUX^G, are used to model luminaire distributions according to DIN EN 13201.
3. **Position validation:** Lamp spacing and alignment are checked against map-based measurements and DIN EN 13201 criteria.
4. **Luminance calibration:** Scene illumination is optimized using the TRILUX Planning Aid to ensure realistic brightness and uniformity.

The resulting scene can be coupled with Matlab-based headlight-control algorithms and executed with the headlight ECU in virtual or real environments. Virtual sensors, such as camera or LiDAR models behind the windshield, generate synthetic perception data under realistic illumination conditions. This enables consistent X-in-the-Loop (XiL) testing and demonstrates the modular, cross-domain co-simulation concept of the open-source scenario framework.

Results and Discussion

The harmonized open-source scenarios have been deployed in pilot studies covering traffic control, V2X communication and lighting and sensor evaluation. Figure 4 shows an exemplary urban ADAS test environment extracted from the standardized scenario database.

^E Find 3D Assets from ambientCG for various materials.

^F ANSI/IES TM-33-23 Standard, see Illuminating Engineering Society (IES), 2023.

^G TRILUX homepage providing Photometric data (IES), Lighting Classes Assistant and Planning Aid.

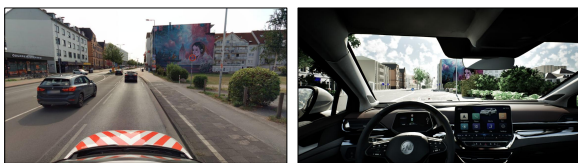


Figure 4: Real and virtual ADAS test environment at Bahnhofstraße, Paderborn; left: real scene; right: Unreal rendering (see Fig. 1)

The generated outputs include realistic road geometries, signalized intersections, surrounding infrastructure and optional dynamic traffic participants via co-simulation with SUMO. They support interactive driving simulator experiments on the Gforcefactory platform, where real-time rendering, vehicle motion cues and driver interaction enable studies on driver behaviour, automated driving functions and cooperative approaches such as GLOSA (see Malena, et al., 2026). The standardized scenario definition reduces inconsistencies found in earlier project-specific setups, particularly regarding coordinate frames, object semantics and behaviour modelling.

For headlight simulations, result quality depends on realistic materials, light-source distributions and suitable sun, night and sky models. The proposed approach therefore aggregates openly available assets such as photogrammetry-based asphalt materials, IES streetlight files and sky models into a shared library. Future replication of previous traffic-flow and headlight studies with the standardized datasets will help quantify reproducibility, statistical robustness and cross-platform comparability.

Conclusion

This work presents open and standardized ITS scenes and scenarios for reproducible cross-domain experimentation. By sharing both the generation pipeline and resulting scenarios, it supports transparent evaluation and benchmarking across traffic control, autonomous driving, headlight simulation and sensor modelling.

Future work will focus on refining the generative reconstruction, expanding the scenario library and integrating additional simulation toolchains to further improve reproducibility and comparability.

References

- Arjovsky, M., Chintala, S., and Bottou, L., 2017. *Wasserstein GAN*. Available at: <https://arxiv.org/pdf/1701.07875>.
- Bußemas, L., Fittkau, N., Gausemeier, S., Trächtler, A., and Rüddenklau, N., 2024. LiDAR-Sensormodell basierend auf zeitabhängigem Photon Mapping. *VDI Mechatroniktagung Dresden 2024*. Available at: <https://ris.uni-paderborn.de/record/53106>.
- Fittkau, N., Bußemas, L., Malena, K., Gausemeier, S., and Trächtler, A., 2026. Intelligent headlights boost energy efficiency and drive decarbonization: (accepted full paper). In: *IEEE IV 2026 The IEEE Intelligent Vehicles Symposium (IV)*. IEEE.
- Girase, H., Gang, H., Malla, S., Li, J., Kanehara, A., Mangalam, K., and Choi, C., 2021. *LOKI: Long Term and Key Intentions for Trajectory Prediction*. Available at: <https://arxiv.org/pdf/2108.08236>.
- Gloudemans, D., Wang, Y., Ji, J., Zachár, G., Barbour, W., Hall, E., Cebelak, M., Smith, L., and Work, D. B., 2023. I-24 MOTION: An instrument for freeway traffic science. *Transportation Research Part C: Emerging Technologies*, 155, p. 104311. ISSN: 0968-090X. <https://doi.org/10.1016/j.trc.2023.104311>.

- Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A., 2017. *Improved Training of Wasserstein GANs*. Available at: <https://arxiv.org/pdf/1704.00028>.
- Illuminating Engineering Society (IES), 2023. *ANSI/IES TM-33-23: Standard Format for the Electronic Transfer of Luminaire Optical Data*. Ed. by Illuminating Engineering Society (IES). New York, NY.
- Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A., 2016. *Image-to-Image Translation with Conditional Adversarial Networks*. Available at: <https://arxiv.org/pdf/1611.07004>.
- Li, Y., Cheng, Z., Yao, X., Kong, Z., Wang, Z., and Liu, M., 2023. Multi-Objective Optimal Deployment of Road Traffic Monitoring Cameras: A Case Study in Wujiang, China. *Sustainability*, 15(15), p. 12011. <https://doi.org/10.3390/su151512011>.
- Link, C., Malena, K., Gausemeier, S., and Trächtler, A., 2023. Simulation Environment for Traffic Control Systems Targeting Mixed Autonomy Traffic Scenarios. In: A. Vinel ed. *9th International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS)*. Setúbal: SCITEPRESS - Science and Technology Publications Lda, pp. 367–377. ISBN: 978-989-758-652-1. <https://doi.org/10.5220/0011987600003479>.
- Lopez, P. A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y.-P., Hilbrich, R., Lücken, L., Rummel, J., Wagner, P., and Wießner, E., 2018. Microscopic Traffic Simulation using SUMO. In: *The 21st IEEE International Conference on Intelligent Transportation Systems*. IEEE. Available at: <https://elib.dlr.de/124092/>.
- Maierhofer, S., Klischat, M., and Althoff, M., 2021. CommonRoad Scenario Designer: An Open-Source Toolbox for Map Conversion and Scenario Creation for Autonomous Vehicles. In: *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*. IEEE, pp. 3176–3182. ISBN: 978-1-7281-9142-3. <https://doi.org/10.1109/ITSC48978.2021.9564885>.
- Malena, K., Link, C., Gausemeier, S., and Trächtler, A., 2024. Implementation and Results of a Multi-Stage Model Predictive Traffic Light Control System. In: *2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, pp. 331–337. ISBN: 979-8-3315-0592-9. <https://doi.org/10.1109/ITSC58415.2024.10919569>. Available at: <https://ieeexplore.ieee.org/document/10919569>.
- 2026. ML-based Prediction Framework for varying Traffic Signal Control Strategies and its GLOSA-application: (presented full paper; not yet published). In: *2025 IEEE 28th International Conference on Intelligent Transportation Systems (ITSC)*. [Piscataway, NJ]: IEEE.
- Malena, K., Link, C., Mertin, S., Gausemeier, S., and Trächtler, A., 2021. Online State Estimation for Microscopic Traffic Simulations using Multiple Data Sources. In: K. Berns ed. *7th International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS)*. Setúbal: SCITEPRESS - Science and Technology Publications Lda, pp. 386–395. ISBN: 978-989-758-513-5. <https://doi.org/10.5220/0010414903860395>.
- Rasouli, A., Kotseruba, I., Kunic, T., and Tsotsos, J., 2019. PIE: A Large-Scale Dataset and Models for Pedestrian Intention Estimation and Trajectory Prediction. In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, pp. 6261–6270. ISBN: 978-1-7281-4803-8. <https://doi.org/10.1109/ICCV.2019.00636>.
- Ronneberger, O., Fischer, P., and Brox, T., 2015. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Available at: <https://arxiv.org/pdf/1505.04597>.
- Rüddenklau, N., Biemelt, P., Mertin, S., Gausemeier, S., and Trächtler, A., 2019. Real-Time Lighting of High-Definition Headlamps for Night Driving Simulation. *IARIA SysMea*, 12. Available at: https://personales.upv.es/thinkmind/dl/journals/sysmea/sysmea_v12_n12_2019/sysmea_v12_n12_2019_8.pdf.
- Stein, M. L., 1999. *Interpolation of spatial data: Some theory for kriging*. Springer series in statistics. New York, Berlin, and Heidelberg: Springer. ISBN: 978-1-4612-7166-6. <https://doi.org/10.1007/978-1-4612-1494-6>.
- Tan, M. and Le, Q. V., 2019. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *International Conference on Machine Learning*. Available at: <https://arxiv.org/pdf/1905.11946>.

A Comparative Evaluation of LLM Prompting Strategies for OpenSCENARIO Generation from ADScene Platform

Sofiane Sadat^{1,2}, Hakim Mohellebi¹, Chouki Sentouh², Eric Vaillant¹, Emmanuel Arnoux¹, Stéphane Régnier¹, Jean-Christophe Popieul²

(1) Renault, Technocentre, 1, Avenue du Golf, F-78084 Guyancourt, e-mail : {sofiane.sadat, hakim.mohellebi, eric.vaillant, emmanuel.arnoux, stephane.regnier}@Renault.com

(2) LAMIH - UMR 8201, Université Polytechnique Hauts-de-France and INSA Hauts-de-France, Campus Mont Houy 59313 Valenciennes Cedex 9, e-mail: {chouki.sentouh, jean-christophe.popieul}@uphf.fr

Abstract – The rapid progress of Generative Artificial Intelligence (GenAI) has created new opportunities for automating engineering workflows (Gozalo-Brizuela and Merchan, 2024), including simulation scenario generation for the verification and validation of Autonomous Driving Systems (ADS). In this context, this study investigates the automatic generation of Python OpenSCENARIO code (ASAM, 2026) from structured ADScene (Guyonvarch et al., 2023) descriptions using Large Language Models (LLMs). Building on the limitations identified in earlier multi-agent approaches (Mohellebi, 2025), this work explores several alternative generation strategies, including direct single-step generation, retrieval-augmented generation, structured constrained generation, reasoning-based class selection, and full-documentation prompting. To assess these methods, we introduce a hierarchical evaluation framework based on executability, behavioral fidelity, and trigger fidelity, complemented by the number of retries required to obtain a correct generation. The results show that simpler generation pipelines consistently improve executability compared with the multi-agent baseline, while documentation grounding plays a central role in improving behavioral correctness and trigger conformity. Among the evaluated methods, the full-documentation variant provides the strongest overall results while remaining straightforward to implement. Overall, the findings suggest that, for this task, simpler generation strategies are more effective than complex multi-agent pipelines.

Keywords: AI scenario generation, ADScene platform, Generative Artificial Intelligence, Large Language Models (LLMs), ASAM OpenSCENARIO

Copyright: © 2026 by the authors.

Licensee Driving Simulation Association.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.82157/dsa/2026/25>

1. Introduction

The deployment of Autonomous Driving Systems is expected to improve road safety, traffic efficiency, and mobility accessibility (Koopman and Wagner, 2017). However, as these systems move from controlled environments to public roads, ensuring their reliability and safety remains a major challenge. Statistical evidence suggests that demonstrating safety improvements over human drivers would require extremely large amounts of real-world driving data, making physical testing costly, time-consuming, and potentially unsafe, especially for rare or critical situations (Kalra and Paddock, 2016).

To address these limitations, the field has increasingly adopted simulation-based verification and validation (V&V). Simulation enables safe, repeatable, and controlled testing of ADS in a wide

range of traffic situations. This transition has been supported by initiatives such as the PEGASUS project (Winner et al., 2019) and by the adoption of ASAM OpenSCENARIO as a standard language for describing simulation scenarios across tools and platforms.

Despite this progress, authoring OpenSCENARIO scenarios remains a significant bottleneck. The process requires both domain expertise and detailed knowledge of the scenario description language, and becomes increasingly difficult as scenario complexity grows. In this context, LLMs offer a promising way to automate the generation of simulation code from structured scenario descriptions.

Previous work introduced a multi-agent LLM-based pipeline (Mohellebi, 2025) for translating ADScene descriptions into Python OpenSCENARIO code. Although this initial approach showed encouraging

results, it also revealed important limitations, including frequent manual corrections, reduced robustness, and a tendency to generate inappropriate trigger logic, particularly through the overuse of temporal triggers.

Building on these observations, the present study investigates several alternative generation strategies for automatic OpenSCENARIO synthesis, including direct generation, retrieval-augmented generation, constrained structured generation, reasoning-based class selection, and full-documentation prompting. To compare these methods, an evaluation framework is introduced based on executability, behavioral fidelity, and trigger fidelity, together with the number of retries required to obtain a correct generation. The objective is not only to improve code validity but also to assess whether generated scenarios faithfully reproduce the intended behavior and triggering logic. The remainder of this paper is organized as follows. Section 2 formalizes the ADScene-to-OpenSCENARIO conversion problem. Section 3 describes the baseline multi-agent architecture used as a reference throughout this study. Section 4 presents the experimental setup, while Section 5 introduces the proposed generation strategies. Section 6 reports the results and discusses the main findings. Finally, Section 7 concludes the paper and outlines directions for future work.

2. Problem Formulation

This research addresses the problem of converting logical scenarios described in the ADScene platform into executable ASAM OpenSCENARIO 1.x specifications.

ADScene is an industrial scenario database and description framework designed to support the design, validation, and homologation of Automated Driving Systems (ADS) and Advanced Driver Assistance Systems (ADAS). It provides a structured and harmonized repository of driving scenarios originating from multiple sources, including accidentology data, expert knowledge, regulatory requirements, and real-world driving data. Scenarios in ADScene are expressed using a dedicated data model and grammar that facilitate a shared understanding of complex driving situations among stakeholders.

In alignment with the PEGASUS scenario classification, ADScene organizes scenarios at different abstraction levels, namely functional scenarios, logical scenarios, and concrete scenarios. This study specifically focuses on logical scenarios.

An ADScene logical scenario describes a driving situation at a parameterized and semantically structured level, without fixing numerical values. It is typically composed of a scenario title, a global textual

description, and a storyboard representation. The storyboard decomposes the scenario into a sequence of stages, each corresponding to a semantic step in the evolution of the driving situation.

Each stage may include a snapshot image representing the beginning of the stage, a textual description of the ongoing situation, a list of actors involved in the scene, parameterized actor states such as longitudinal position, lateral position, and speed, as well as behaviours and maneuvers such as braking or lane changing. Actor states are described through intervals or qualitative descriptors, reflecting the logical nature of the scenario. In addition, some parameters may be annotated with semantic roles such as input parameters or trigger parameters. Input parameters are intended to remain variable to enable the systematic derivation of multiple concrete scenarios, whereas trigger parameters are used to activate specific events or maneuvers when given conditions are satisfied.

One of the objectives is to preserve the semantics embedded in the logical scenario description. In particular, the generated OpenSCENARIO instance must remain consistent with the storyboard structure, the actors involved, the parameterized vehicle states, and the temporal and causal relations implied by the maneuvers and their triggers, while also preserving the possibility of varying relevant parameters to cover the associated space of concrete scenarios.

Ensuring the correct interpretation of these parameters is essential to preserve the intended meaning of the scenario. Each maneuver must be associated with trigger conditions that accurately reflect the causal and temporal structure specified in ADScene. Although multiple OpenSCENARIO encodings may appear syntactically or visually acceptable, only a subset faithfully represents the logical structure and trigger semantics defined in the original ADScene scenario.

The objective of this research is to design a method for converting ADScene logical scenarios into executable OpenSCENARIO representations that:

- remain consistent with the ADScene description,
- preserve parameter variability,
- respect trigger semantics,
- remain interpretable and adjustable by scenario users.

The target specification is OpenSCENARIO 1.x. To ensure flexibility and maintain compatibility across OpenSCENARIO 1.x versions, the proposed approach relies on the generation of Python code using the ScenarioGeneration (pyoscx, 2026) library. This library provides programmatic abstractions to construct OpenSCENARIO descriptions and automatically generate the corresponding XML files.

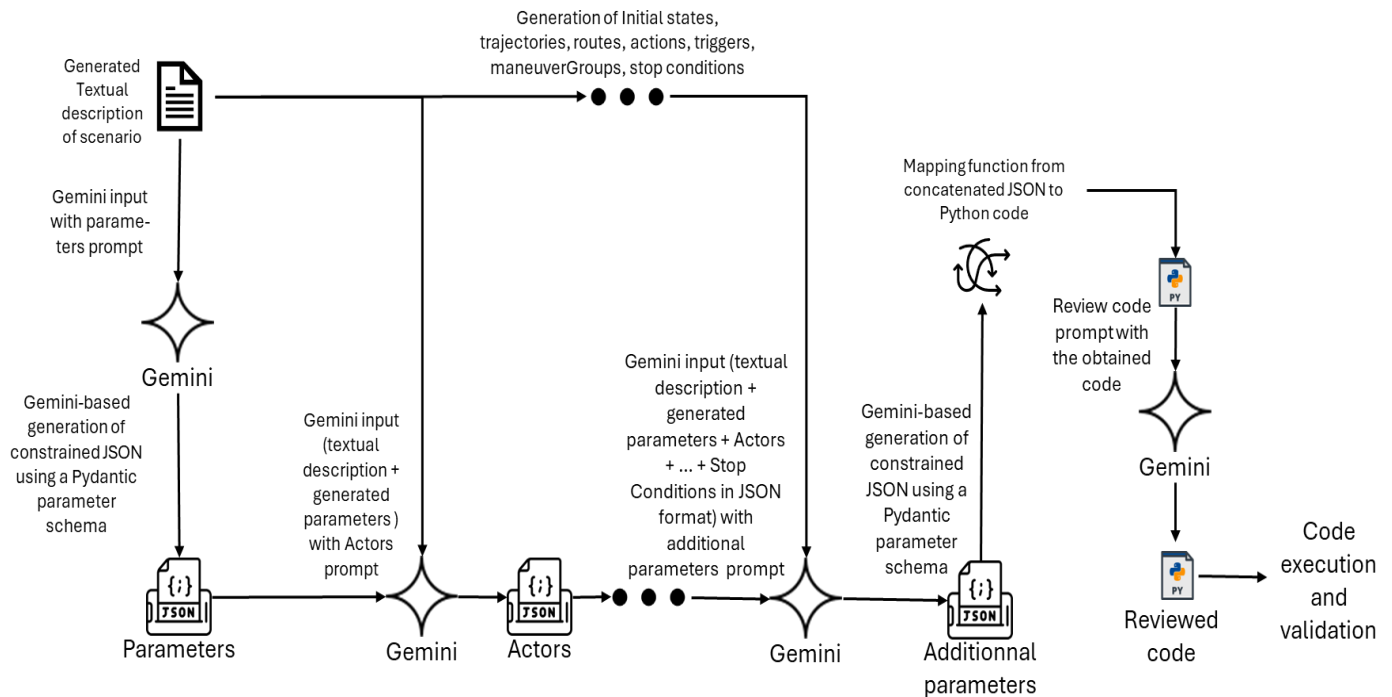


Figure 1: Baseline Method Pipeline from Textual Scenario Description to Code Execution

By using Python as an intermediate representation, scenarios can be generated for different OpenSCENARIO 1.x versions, such as 1.1, 1.2, or 1.3, with the library taking care of the necessary adaptation.

3. Baseline Method: Multi-Agent Architecture for ADScene-to-OpenSCENARIO Conversion

A first relevant work addressing the conversion of ADScene scenarios into ASAM OpenSCENARIO is presented in (Mohellebi, 2025). In this approach, the authors propose a multi-agent pipeline in which several specialized agents operate in sequence to generate an OpenSCENARIO specification from an ADScene logical scenario.

The process begins with a data formatting stage in which the ADScene logical scenario, originally provided as a JSON structure is enriched and normalized. During this stage, identifier fields representing actors (e.g., numeric IDs such as 4322) are replaced with meaningful actor names (e.g., *ego*, *vehicle_1*, etc.). Other structural transformations are also applied to ensure that the logical scenario data becomes more interpretable for downstream agents.

Once the reformatted data is available, a dedicated agent extracts the stop conditions from the natural-language descriptions. Another agent

processes the infrastructure description and its associated image to select the most appropriate OpenDRIVE file from a pre-established database. After selecting the most consistent road network, an additional agent generates a structured description of the chosen OpenDRIVE file, extracting key elements required to construct the OpenSCENARIO representation.

The core of the workflow consists of a stage in which the previously transformed logical scenario data together with the stop conditions, the road description, and the scenario images are provided to a multimodal agent. This agent produces a detailed step-by-step description of the scenario, identifying the essential semantic elements of each stage of the storyboard.

These textual descriptions are then passed to a chain-of-agents architecture (see Figure 1) that uses Pydantic schemas to constrain the outputs of the LLMs. By enforcing strict Python class structures, the system prevents hallucinated parameters or unexpected fields. Each agent is responsible for generating a specific section of the scenario:

- The **scenario parameters**, with explicit instructions to only use parameters present in the logical description
- The **actors**, including their names, roles, and vehicle models
- The **initial states**, such as position and speed

- The **routes** and **trajectories**, if necessary to describe the scenario
- The **actions**, **triggers**, and **maneuver groups**
- the **stop conditions**
- An **additional-parameters** agent, which reviews the generated structure to detect parameters used in the code but missing from the parameter list, adding them when necessary.

Once the full JSON representation has been generated and concatenated, a mapping function converts the JSON to a single Python file built using the *ScenarioGeneration* library. This Python file, when executed, generates the corresponding OpenSCENARIO XML.

A dedicated **code review** agent analyzes the produced Python file in the context of the *ScenarioGeneration* documentation, correcting inconsistencies or API-related errors. Finally, the generated code is executed, and logs are captured. If execution errors occur either Python errors or esmini simulation errors the incorrect code and error log are sent back to a **correction** agent. Here, esmini is used as a lightweight, open-source OpenSCENARIO player. It serves as an efficient tool for rapid scenario validation and visual playback while strictly complying with ASAM OpenSCENARIO standards. This correction loop is applied up to two iterations before the system terminates.

Overall, this first approach proposes a multi-agent environment using 12 different agents.

The initial implementation of this approach relied on Gemini-2.0-Flash (Google, 2026a), a model primarily designed for fast and cost-efficient everyday tasks. However, all experiments and results reported in the present article are conducted using Gemini-2.5-Flash (Google, 2026b). This choice was made in order to compare the proposed generation strategies under the same underlying model, so that the observed differences can be more directly attributed to the pipeline design rather than to the base model itself.

This update is further justified by the Gemini 2.5 technical report (Comanici et al., 2025), which positions Gemini 2.5 Flash as a hybrid reasoning model for more complex tasks, whereas Gemini 2.0 Flash is described as a non-thinking model optimized for speed. The objective of this study is therefore not to benchmark different foundation models, but rather to evaluate alternative prompting and generation strategies under a common model configuration.

4. Experiment setup

4.1 Dataset construction and scenario selection

To evaluate the different models, we constructed a dataset based on highway scenario container, which correspond to relatively simple driving situations. These scenarios provide an appropriate benchmark for comparing different generation approaches on elementary but representative cases. Our underlying hypothesis is that if a model can correctly generate these selected scenarios, its behavior can then be generalized to the broader set of the highway scenario container.

To cover a diverse range of maneuvers, trigger types, and road configurations, we selected a set of 12 scenarios. These scenarios were chosen for the specific challenges they introduce in scenario coding. The selected scenarios are the following:

L001.001.01.CIPV_brakes_simple: a simple scenario in which the vehicle in front of the ego vehicle brakes, triggered by the relative distance between the two vehicles.

L001.002.01.L.CIPV_brakes_lane_change_to_left_busy_lane: after braking, vehicle *v1* performs a lane change to the left. The maneuver is triggered by the distance between ego vehicle and *v1*.

L001.003.01.L.CIPV_brakes_to_exit_to_left: a cut-out to the left scenario in which the braking maneuver ends once the vehicle reaches the adjacent lane.

L001.003.03.L.CIPV_brakes_then_exit_to_left: similar to the previous scenario, except that the braking phase must be completed before the lane change begins. This case was specifically selected to assess whether the model can capture such subtle ordering constraints.

L002.001.01.L.Cut_in_from_left_simple: a simple cut-in from the left.

L002.002.01.L.Cut_in_from_left_while_CIPV_brakes: vehicle *v1* cuts into the lane while *v2* is braking, and *v1* is also braking during the maneuver.

L002.003.02.L.Cut_in_from_left_Y_Junction: this scenario was included to test whether the model can handle more complex road topologies, such as a Y-junction.

L004.001.01.R.Cut_in_from_right_aborted: this case was added because it involves an aborted maneuver in which the vehicle changes its lateral offset and then returns, making it structurally different from standard lane changes.

L006.002.01.L.Cut_out_to_left_Brake_hidden_by_lead: a cut-out to the left occurring while the CIPV is braking, with partial occlusion by the lead vehicle.

L011.001.01.R.Vehicle_immobilized_in_ego_lane_and_lane_shift_right: this scenario was selected because it involves an action performed by the ego vehicle itself.

L014.001.02.Rear_vehicle_faster_Rear_End_collision: this case was added because collisions are usually modeled as end conditions. Here, however,

the collision is part of the scenario and should not terminate it.

L024.002.03.R.Vehicle_overtaking_between_traffic_lanes_right: this scenario was included because it requires specific lateral offset behaviours and a particular overtaking configuration, making it interesting from a coding perspective.

A limitation of this study concerns reproducibility. Since the generation process relies on a non-zero temperature, identical prompts may yield different outputs across runs. Therefore, the reported results should be interpreted as comparative observations under a common experimental setting rather than strictly deterministic outcomes. Nevertheless, all evaluated strategies were executed using the same model version, prompts, and generation parameters, allowing a fair comparison between approaches.

4.1 Evaluation methodology

Evaluating generated scenarios is challenging. One simple evaluation metric that we can consider is **Executability**, i.e., whether the generated scenario can be executed successfully in esmini without runtime errors. This metric is particularly useful because it enables large-scale batch evaluation as it does not require human evaluation and can be automated.

A second criterion is **behavioral fidelity**. This metric aims to evaluate whether the executed generated scenario respects the described scenario, regardless of whether the triggers are correct or not. To assess this, the simulation must be executed, and its outcome must be examined through human evaluation, to determine whether the intended scenario is reproduced or not.

A third criterion we consider is **trigger fidelity**, which evaluates whether the generated scenario employs semantically appropriate triggers. This aspect is particularly important because LLMs tend to overuse temporal triggers derived from surface-level textual descriptions (e.g., “the maneuver starts after 3 seconds”), while ignoring the semantic labels that actually define the intended triggering logic. To evaluate this criterion, the generated code must also be examined manually to determine whether actions are initiated using the correct trigger conditions.

As expected, the evaluation criteria follow a hierarchical structure: **Executability > Behavioral Fidelity > Trigger Fidelity**, since each criterion is assessed only if the previous one has been successfully validated. The choice of road infrastructure is not considered a challenging aspect in this setup, since the scenarios in this dataset rely on simple road layouts.

Finally, to evaluate the models, we also consider the **number of attempts** required to obtain a correct generation, noting that the correction pipeline is identical for all models. This pipeline consists of a simple agent that receives the generated code

together with the error it produced and uses a tool to query the documentation and retrieve relevant classes as many times as needed. The agent can be called up to two times; if it fails twice, the correction process stops.

5. Proposed Generation Strategies

To improve the performance of the models, we enrich the prompt with more representative examples that use correct and meaningful triggers. In addition, we also explore different generation strategies. Here are the five methods that we have tested:

5.1 Single-generator method

This model adopts a simplified generation workflow in which the formatted ADScene description is directly translated into Python OpenSCENARIO code. By replacing the previous multi-agent decomposition with a single, strongly instructed generation step, the approach reduces pipeline complexity. The prompt also includes examples of Python OpenSCENARIO scripts that were manually corrected, providing the model with concrete references that help guide and stabilize the generation process.

This approach became practically viable with the transition to Gemini 2.5 Flash, which provides substantially stronger reasoning capabilities than the earlier Gemini 2.0 Flash model. This difference helps explain why the previous work relied on a multi-agent decomposition: at the time, such decomposition was necessary to structure the task and compensate for the limitations of direct generation. In the current setting, the stronger reasoning abilities of Gemini 2.5 Flash make this additional architectural complexity less necessary.

5.2 RAG-augmented generator

In this approach, each class in the documentation is associated with a set of keywords that are semantically close to its intended use. For example, the class *AbsoluteSpeedAction* may be linked to terms such as *accelerating*, *decelerating*, or *braking*. Each class, along with its associated keywords, is then embedded into a vector space.

In parallel, the ADScene description is also embedded. A cosine-similarity comparison between the scenario embedding and all class embeddings is then performed. Based on these similarity scores, the method selects the top-K classes that appear most relevant to the scenario. The documentation corresponding to these selected classes is then inserted into the prompt.

The underlying hypothesis is that providing the model with documentation that is potentially most

relevant to the scenario could help guide the generation process and improve the correctness of the produced code.

5.3 Constrained structured-output generator

In this configuration, free-form code generation is replaced by a constrained approach in which the model produces JSON conforming to a predefined schema. This structured representation is then deterministically translated into Python code and validated or repaired if necessary. Such an approach may reduce syntax errors, increases determinism and facilitates debugging through the inspection of intermediate structured outputs.

This method is inspired by the baseline, which also relied on Pydantic-based constrained outputs. However, whereas the baseline applied these constraints separately to different parts of the scenario, the present method uses a single schema to represent the entire scenario. This design simplifies the pipeline, but it also reduces representational granularity. Modeling all class-specific parameter constraints within a single Pydantic schema is not practically feasible, because the number of possible combinations becomes too large to handle. For this reason, we use a generic action structure that includes many possible parameters and rely on the model to choose the ones that match the selected class. This makes the approach workable, but it also means that the schema controls the global structure of the output more than the exact semantic validity of each class definition.

5.4 LLM class-selector generator

This approach is similar to the RAG-based method, since it also selects a subset of the documentation before generation. The main difference is that, instead of relying on embeddings and cosine similarity, it relies on the reasoning ability of the LLM to choose the most relevant classes for the scenario. The documentation of these selected classes is then added to the prompt to guide code generation. The idea is that this may provide more relevant context while keeping the prompt more focused.

5.5 Full-documentation variant

This variant adopts a maximal-context strategy by injecting the entire documentation of ScenarioGeneration library into the prompt without any filtering or retrieval. The documentation is generated from the library source code. It includes class descriptions, method signatures, parameter documentation, and enumeration definitions, resulting in a comprehensive textual representation of the API comprising 21,423 words. The objective is to provide the model with complete access to the available knowledge during generation, in contrast to approaches that rely on selective context, such as

RAG-augmented generator or LLM class selector generator.

6. Results and Discussion

6.1 Baseline Results

The baseline results highlight the limitations of the initial generation approach (see Table 1). Out of the 12 evaluated scenarios, only 6 were executable, corresponding to an executability rate of 50%. Among these executable scenarios, only 3 achieved behavioral fidelity. In addition, trigger fidelity was not achieved in any of the validated cases, which confirms that the baseline tends to rely on incorrect triggering logic, especially temporal triggers, rather than semantically appropriate ones.

The number of retries also reflects the limited robustness of the baseline model. These limitations motivate the introduction of new models designed to improve performance.

Table 1: Baseline evaluation results per scenario

Scenario	Exec.	Beha. Fid.	Trig. Fid.	Retries
L001.001.01	Yes	Yes	No	0
L001.002.01.L	Yes	No	No	1
L001.003.01.L	No	No	No	2
L001.003.03.L	No	No	No	2
L002.001.01.L	Yes	Yes	No	0
L002.002.01.L	No	No	No	2
L002.003.02.L	No	No	No	2
L004.001.01.R	Yes	Yes	No	1
L006.002.01.L	Yes	No	No	2
L011.001.01.R	No	No	No	2
L014.001.02	No	No	No	2
L024.002.03.R	Yes	No	No	0

6.1 Results of the proposed new generation strategies

The experimental results (see Table 2) show that all proposed methods outperform the legacy multi-agent baseline in terms of executability, indicating that simpler generation pipelines are generally more robust. The strongest performance is obtained with the full-documentation variant, which achieves the highest executability (100.00%), the highest trigger fidelity (41.67%), the joint-best behavioral fidelity (58.33%), and the lowest correction effort (0.42 mean retries). The RAG-augmented generator also performs strongly, reaching the same behavioral fidelity as the full-documentation variant while maintaining high executability (91.67%) and a low number of retries. In contrast, the single-generator method substantially improves executability over the baseline but remains limited in trigger fidelity, suggesting that direct generation alone is insufficient to ensure semantically appropriate trigger selection. The LLM class-selector generator provides

intermediate performance, indicating that reasoning-based documentation filtering may be beneficial, although it is less effective than direct retrieval or full-context conditioning in the present setup.

Table 2: Comparison of generation methods tested over the twelve scenarios

Model	Exec. (%)	Behav. Fid. (%)	Trig. Fid. (%)	Mean Retries
Single generator	91.67	41.67	8.33	1.08
RAG-augmented	91.67	58.33	25.00	0.50
Constrained struct.	83.33	25.00	16.67	1.58
LLM class selector	91.67	33.33	16.67	0.75
Full documentation	100.00	58.33	41.67	0.42
Multi-agent baseline	50.00	25.00	0.00	1.33

Finally, the constrained structured-output generator improves structural regularity but shows lower behavioral fidelity and higher repair costs, which may reflect the difficulty of representing rich class-specific semantics within a single global schema.

Table 3: Token consumption and average duration per model

Model	Avg Tokens	Total Tokens	Avg Duration (s)
Single generator	26329	315956	85
RAG-augmented	26905	322871	66
Constrained struct.	28017	336211	69
LLM class selector	51236	614832	91
Full documentation	60789	729470	63
Multi-agent baseline	70727	848724	219

Overall, the results indicate that the full-documentation variant is a particularly relevant option in this experimental setting. It combines the strongest generation quality with the lowest average execution time (63s), while remaining straightforward to implement, since it relies on a direct prompt without additional retrieval or filtering stages. Although this approach is also associated with a higher token consumption (see Table 3) than more selective strategies, this does not appear to be a limiting factor in the present setting. This makes it an attractive solution both from a performance perspective and from a methodological standpoint.

7. Conclusion and future work

This study investigated the automatic generation of Python OpenSCENARIO code from structured

ADScene descriptions using large language models. The proposed evaluation framework, based on executability, behavioral fidelity, and trigger fidelity, enabled the assessment of generation quality beyond simple syntactic validity. The results show that the proposed methods consistently outperform the legacy multi-agent baseline. This improvement may be partly attributed to the transition from Gemini 2.0 Flash to Gemini 2.5 Flash, whose stronger reasoning capabilities appear to make direct single-prompt generation more effective. The results further underline the importance of documentation grounding for improving both behavioral correctness and trigger conformity. Among the evaluated approaches, the full-documentation variant appears particularly effective while remaining straightforward to implement.

Although the experiments were conducted using a non-zero temperature setting, making individual generations inherently non-deterministic, all evaluated approaches were compared under identical experimental conditions, allowing meaningful relative comparisons between generation strategies. Furthermore, while the selected scenarios were chosen to cover representative highway driving situations, additional experiments on larger and more diverse scenario sets will be required to fully assess the generalization capability of the proposed approach.

Future work will aim to improve the evaluation scores by optimizing prompting strategies, investigating multi-agent architectures, and employing models with stronger reasoning capabilities. Furthermore, the proposed approach will be evaluated on more complex scenarios involving challenging trajectories, complex road topologies, and actions requiring synchronization between actors.

Reference

- ASAM. 2026. *ASAM OpenSCENARIO: User Guide*. [online] Available at: <https://www.asam.net/fileadmin/Standards/OpenSCENARIO/ASAM_OpenSCENARIO_BS-1-2_User-Guide_V1-2-0.html>
- Comanici, G. et al. 2025. Gemini 2.5: Pushing the Frontier with Advanced Reasoning, Multimodality, Long Context, and Next Generation Agentic Capabilities. arXiv [Preprint]. Available at: <https://arxiv.org/abs/2507.06261>
- Google, 2026a. *Gemini 2.0 Flash | Generative AI on Vertex AI*. [online] Google Cloud Documentation. Available at: <<https://docs.cloud.google.com/vertex-ai/generative-ai/docs/models/gemini/2-0-flash?hl=fr>>
- Google, 2026b. *Gemini 2.5 Flash | Generative AI on Vertex AI*. [online] Google Cloud Documentation. Available at: <<https://docs.cloud.google.com/vertex-ai/generative-ai/docs/models/gemini/2-5-flash?hl=fr>>
- Gozalo-Brizuela, R. and Merchan, E.E.G., 2024. A Survey of Generative AI Applications. *Journal of Computer Science*, 20(8), pp.801–818. <https://doi.org/10.3844/jcssp.2024.801.818>.

- Guyonvarch, L., Hermitte, T., Kroger, R., Chauvel, C., Arnoux, E. and Geronimi, S., 2023. ADSCENE scenarios data base: Focus on accident data support for validation of Automated Driving functions. *Transportation Research Procedia*, 72, pp.9–16. <https://doi.org/10.1016/j.trpro.2023.11.316>.
- Kalra, N. and Paddock, S.M., 2016. Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability? *Transportation Research Part A: Policy and Practice*, 94, pp.182–193. <https://doi.org/10.1016/j.tra.2016.09.010>.
- Koopman, P. and Wagner, M., 2017. Autonomous Vehicle Safety: An Interdisciplinary Challenge. *IEEE Intelligent Transportation Systems Magazine*, 9(1), pp.90–96. <https://doi.org/10.1109/MITS.2016.2583491>.
- Mohellebi, H., 2025. AI-based simulation scenario generation from the ADScene scenario platform. Available at: <<https://dsc2025.org/program-dsc-2025/>>
- pyoscx, 2026. *pyoscx/scenariogeneration*. Available at: <<https://github.com/pyoscx/scenariogeneration>>
- Winner, H., Lemmer, K., Form, T. and Mazzega, J., 2019. PEGASUS—First Steps for the Safe Introduction of Automated Driving. pp.185–195. https://doi.org/10.1007/978-3-319-94896-6_16.

SHORT NOTES

Absolute Accuracy Validation of Unity3D Perpendicular Distance Measurements: Impact for Road Safety Studies

Machiel van der Stelt¹, Thomas Bräunl¹, Paul Roberts¹

(1) The University of Western Australia, 35 Stirling Hwy, Crawley WA 6009, Australia, e-mail: 22888094@student.uwa.edu.au, {thomas.braunl, paul.l.roberts}@uwa.edu.au

Abstract – This study evaluates the absolute technical validity of virtual reality distance measurements against ground truth across straight and curved road sections in Unity3D. Programmatic raycasting measurements were collected across five simulated runs on both straight and curved road profiles. Statistical analysis revealed that measurements on the straight road were highly consistent across runs, exhibiting a minor systematic negative deviation of 4 cm relative to a 1.40 m ground truth. In contrast, curved sections displayed significantly higher variance and a 10 cm mean deviation, primarily caused by reference wall faceting from mesh deformations. Overall the proposed measurement method provides a reliable technical baseline for human behavioural road safety research, though for future work multi-ratio calibration is recommended for high-precision simulation applications.

Keywords: Distance Measurement, Absolute Validation, Road Safety, Simulation, Virtual Reality

Copyright: © 2026 by the authors.
Licensee Driving Simulation Association.

Introduction

This study is an absolute accuracy validation study, following up on an earlier methodological study by Van der Stelt, et al. (2025) of a perpendicular measuring application. This absolute accuracy validation study enables researchers to increase the confidence in the accuracy of measured lateral perpendicular distances in future Unity3D Road Safety studies.

While the research study by O'Hern, et al. (2016) was a validation study, it focused on the behavioural validity, instead of absolute accuracy validation. The study by Zhao, et al. (2021) investigated situations where the lateral positioning to other moving objects was used, but no distance measurement, or absolute accuracy validation was determined in these studies.

Methods

Calibration

Calibration of this measuring method was covered to some degree in the study Van der Stelt, et al. (2025). The road section, which was modelled on the Hackett Road in Crawley, had 4 ground truth distance points, The Riverside Drive Road section in North Fremantle, had 5 ground truth values, as indicated as orange lines in Fig. 1. To quantify the technical validity of the

simulator, the Absolute Error (E_a) was calculated (Eq. 1). This represented the numerical difference between the Unity3D perpendicular distance measured (D_{VR}) and the 9 ground truth (D_{GT}) values.

$$E_a = |D_{VR} - D_{GT}| \quad (1)$$

With D_{VR} being the virtual reality measurement and D_{GT} the constant value of each ground truth.

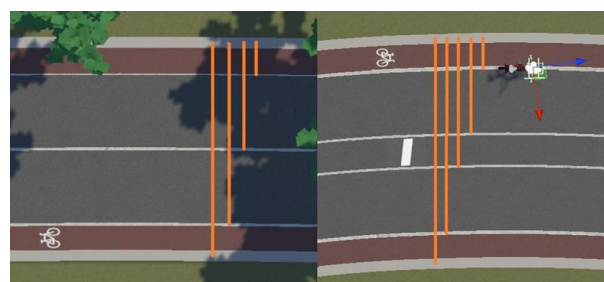


Figure 1: Ground truth distance values indicated with the orange lines, of respectively Hackett Road on the left side and Riverside Drive on the right side.

Absolute Accuracy Validation

Approximately 800 distance measurement samples were automatically recorded across five manual longitudinal runs at the bicycle lane outer-line on both the straight (Hackett Road) and curved (Riverside Drive) road sections Van der Stelt, et al. (2023). A two-way ANOVA and Tukey's HSD post-hoc test (Eq. 2)

evaluated pairwise differences between individual runs:

$$q = (\bar{y}_i - \bar{y}_j) / SE, (SE = \sqrt{MSE/n}) \quad (2)$$

Here, $\bar{y}_i - \bar{y}_j$ represents the mean difference between two test runs, SE is the standard error, MSE is the Mean Square Error derived from the ANOVA residuals, and n is the sample size per run. Main effects for Road $F = 2840.97$, Pr is $< 2 \times 10^{-16}$ and Road:Run ($F = 5.27$), Pr value of 1.87×10^{-6} were statistically significant at $\alpha = 0.05$. High residual degrees of freedom ($df = 786$) provided strong statistical power to detect small positional variations across runs.

Pairwise Tukey comparisons for Hackett Road Van der Stelt et al. (2026d) yielded p-values near or equal to 1, failing to reject the null hypothesis (H_0) and confirming high repeatability. Conversely, 8 of 10 comparisons on Riverside Drive yielded $p < 0.05$, rejecting H_0 due to increased run-to-run variance, as illustrated in the nested comparison plot Van der Stelt et al. (2026b). As shown in repository boxplot Van der Stelt et al. (2026a) and dataset Van der Stelt et al. (2026c), Hackett Road exhibited a tightly confined distribution with a mean negative deviation of 4 cm relative to the 1.40 m ground truth, whereas Riverside Drive showed a 10 cm mean deviation. This increased error stems from reference wall mesh faceting created when applying ProBuilderize along curves.

Table 1: Anova analysis between roads, and road runs.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Road	1	0.8655	0.8655	2840.966	$< 2 \times 10^{-16}$ ***
Road:Run	8	0.0129	0.0016	5.273	1.87×10^{-6} ***
Residuals	786	0.2394	0.0003		

Significant codes: 0 '***'; 0.001 '**'; 0.01 '*'; 0.05 '.'; 0.1 ' '; 1

Note: both the Pr values have a significant code of ***, which means they are very close to 0.

Conclusions

This absolute accuracy validation study demonstrated that the straight Hackett Road and curved River Side Drive had a negative deviation mean of 4 centimetres and 10 centimetres, respectively. The deviation between mean values of the five runs nested in the Hackett Road exhibited lower variance than the deviation between the mean values of the five runs nested in the Riverside Drive. This is likely due to the reference wall in the Riverside Drive being bent with the ProBuilderize function and the faceted nature of the curved reference wall. Overall, the ANOVA test showed that the measured values were considered statistically significant and fell within the error, ($\alpha = 0.05$). Overall, this study provides the technical baseline required for researchers to dependably measure perpendicular lateral distances in Unity3D-based virtual reality environments for road safety studies. Future work can focus on applying variable

calibration factors at different road sections to increase accuracy.

Acknowledgements

This research was supported by an Australian Government Research Training Program (RTP) Scholarship. Thanks to advisor Thomas Bräunl and co-supervisor Paul Roberts for their support. Thanks to David Harries for proofreading and general guidance.

References

- O'Hern, S., Oxley, J. and Stevenson, M., 2016, Validation of a bicycle simulator for road safety research, Accident Analysis and Prevention
- Van der Stelt, M., Bräunl, T., Roberts, P., 2025, Perpendicular Distance Measurement & Localization on Arbitrary Road Curves via VR Raycasting in Unity.
- Zhao, Y., Miyahara, T., Mizuno, K., Ito, D., Han, Y., 2021, Analysis of car driver responses to avoid car-to-cyclist perpendicular collisions based on drive recorder data and driving simulator experiments, Accident Analysis and Prevention, Elsevier.
- Van der Stelt, M., Bräunl, T., & Roberts, P. (2026a). Boxplot data variance of Hackett Road and Riverside Drive runs [Plot]. The University of Western Australia. <https://doi.org/10.26182/pn34-j063>
- Van der Stelt, M., Bräunl, T., & Roberts, P. (2026b). Graphical representation of the test run comparisons nested within each road [Plot]. The University of Western Australia. <https://doi.org/10.26182/z336-en50>
- Van der Stelt, M., Bräunl, T., & Roberts, P. (2026c). Mean distances nested in Hackett Road and nested in Riverside Drive [Data set]. The University of Western Australia. <https://doi.org/10.26182/n9gw-k677>
- Van der Stelt, M., Bräunl, T., & Roberts, P. (2026d). Nested test run comparisons [Data set]. The University of Western Australia. <https://doi.org/10.26182/fh10-xn96>
- Van der Stelt, M., Bräunl, T., & Roberts, P. (2023). After system calibration raw distance data for 5 runs Hackett Road and 5 runs Riverside Drive [Data set]. Mendeley Data. <https://data.mendeley.com/datasets/d6m4sy3r8t/1>

Acoustic measures to increase immersion in static driving simulators

Arne Düselder¹, Marius Dute², Mirko Djukic³

(1) fka GmbH, Steinbachstr. 7, 52074 Aachen, e-mail: arne.dueselder@fka.de

(2) HEAD acoustics, Ebertstr. 30a, 52134 Herzogenrath, e-mail: marius.dute@head-acoustics.com

(3) HEAD acoustics, Ebertstr. 30a, 52134 Herzogenrath, e-mail: mirko.djukic@head-acoustics.com

Abstract – *Static driving simulators are essential tools for conducting driving research in controlled environments. However, the lack of motion feedback can limit the realism of the simulated experience and, consequently, the validity of research findings. This study investigates the potential of acoustic measures to enhance driver immersion in static simulators. By integrating haptic and auditory cues, such as engine vibrations and realistic sounds, the research aims to improve drivers' perception of speed, driving status, and overall environmental awareness. A modal exciter and vibration transducer were installed to simulate vibrations through the seat and steering wheel, while a high-fidelity 5.1 surround sound system was implemented to provide a comprehensive auditory experience. Preliminary results indicate that the integration of these acoustic enhancements significantly improves driver immersion and perception compared to standard simulator setups. Drivers reported feeling more connected to the simulated environment, suggesting that acoustic feedback can play a crucial role in enhancing the realism and effectiveness of static driving simulators for research and training purposes. Further research is planned to fully assess the impact of these acoustic improvements on specific driver behaviors and simulator performance.*

Keywords: *Driving Simulation, Acoustic Enhancement, Immersion, Haptic Feedback, Static Simulator*

Copyright: © 2026 by the authors.
Licensee Driving Simulation Association.

Introduction

Driving simulators are invaluable tools for conducting driving research under controlled and safe conditions, applicable across various fields such as traffic safety, driver assistance systems, vehicle design, and driver training (Carsten & Jamson, 2011). The validity of simulator experiments depends heavily on the realism of the simulated driving experience (Zhang, et al., 2025), making driver immersion a critical factor for producing results that are comparable to real-world scenarios.

Besides visual perception as the key element for real world driving as well as simulated driving tasks (Andersen, 2011), the vestibular and auditory systems also play crucial roles in perceiving environmental conditions during a drive. These sensory systems work together to provide drivers with comprehensive feedback about their surroundings and vehicle dynamics. Static or fixed-base simulators, however, cannot replicate the forces experienced by drivers that are essential for determining vehicle orientation, speed, and acceleration. To address this limitation, acoustics, encompassing both sound and vibration, are employed to enhance feedback to drivers, thereby improving immersion. This presentation

demonstrates the integration of various acoustic measures into a static simulator's vehicle setup, providing additional haptic cues and creating a more realistic sound profile aimed at enhancing speed perception and overall driving realism.

Research Question

This study investigates whether incorporating haptic and auditory cues—such as engine vibrations and sounds—can help drivers better perceive their speed, driving status, and other critical factors in a static simulator. The primary aim is to enhance immersion by using acoustics to compensate for the lack of motion feedback typical in fixed-base simulators. Additionally, this approach could allow the simulator to serve as an effective tool for studying noise, vibration, and harshness (NVH) related topics by generating detailed and realistic audio and vibrational responses.

Methodology

To achieve the desired acoustic enhancements, a modal exciter (B&K Type 4808), capable of exerting a nominal force of 100 N, was installed under the car. This device was securely attached to the frame of the

driver's seat, serving as a source for engine and road-induced vibrations. To provide additional feedback, another mechanical vibration transducer (IBEAM) was mounted on the steering wheel column.

Recognizing that the existing audio setup in the BMW i3 was insufficient for reproducing the complex sound spectrum required, a high-fidelity 5.1 surround sound system was integrated into the simulator. This upgrade ensured a more comprehensive auditory experience by accurately capturing and reproducing the nuanced sounds of different vehicle types and road conditions.

All acoustic systems were precisely calibrated using a binaural head and vibration sensors to ensure they aligned perfectly with the driver's position, optimizing the realism and effectiveness of the feedback. The PreSense software, widely used for NVH studies, was adapted to process live data from the simulation environment. This adaptation allowed the system to generate appropriate vibrational and auditory responses tailored to various vehicle types and road conditions, enhancing the overall driving experience.

Results and discussion

The mechanical integration into the car and the adaptation of all related software parts proved to be successful. The integration of mechanical components within the car and the subsequent adaptation of associated software were successful in generating a detailed sound profile and vibrations transmitted through both the seat and steering column. These enhancements made a noticeable difference in the driver's sensory experience compared to previous setups that lacked such feedback mechanisms. Although time constraints limited our ability to conduct a comprehensive analysis of their impact on drivers' speed perception and overall immersion, preliminary observations were promising.

Direct comparisons with standard simulator setups revealed significant improvements when acoustic measures were employed. Drivers reported feeling more connected to the simulated environment, indicating an increase in perceived realism. These findings suggest that incorporating sound and vibration feedback can play a crucial role in enhancing driver immersion within static simulators.

Conclusions/implications

Further research is essential to thoroughly assess how these generated sound and vibration profiles affect drivers' speed perception and overall immersion levels. The initial setup, while promising, requires full integration into the mock-up car for concealment and will need a full recalibration. Once integrated seamlessly, additional studies could explore specific parameters affecting driver behaviour and simulator effectiveness.

If subsequent analyses validate our preliminary findings, static simulators could significantly benefit from increased driver immersion and feedback. This advancement would not only improve their utility in driving research but also enhance training programs by providing a more realistic experience that closely mimics real-world conditions.

References

- Andersen, G. J., 2011. Sensory and perceptual factors in the design of driving simulators. In: D. L. Fisher, M. Rizzo, J. K. Caird & J. D. Lee, eds. *Handbook of Driving Simulation for Engineering, Medicine, and Psychology*. Boca Raton: CRC Press.
- Carsten, O. & Jamson, A. H., 2011. Driving simulators as research tools in traffic psychology. In: B. E. Porter, ed. *Handbook of Traffic Psychology*. Oxford, UK: Elsevier, pp. 87-96.
- Zhang, S., Zhao, C., Zhang, Z. & Lv, Y., 2025. Driving simulator validation studies: A systematic review. In: *Simulation Modelling Practice and Theory*. s.l.:Elsevier.
- Dibs, V. and Rall, A., 2008a. Behavioral Mechanisms Underlying Speed Perception. *Movement Verlag*, p. 108.
- Dibs, V. and Rall, A., 2008b. Perceptual Mechanisms Underlying Speed Behavior. *Movement Verlag*, p. 118.
- Smile, P. and Why, P., 2006. Theory of different driving modalities. *Journal of Driving*, Volume 9(2), pp. 63-72.

Designing a Balanced Order Driving Simulator Study with OpenSCENARIO and Esmini

Anders Andersson¹

(1) Swedish National Road and Transport Research Institute (VTI), Olaus Magnus väg 35, 583 30, Linköping, Sweden, e-mail: {anders.andersson}@vti.se

Abstract – Designing driving simulator studies with human drivers in a balanced manner is important for reducing biased results. Simulator scenarios should also preferably be easy to share between simulators and an enabler for this is standardized scenario formats. In this work the scenario standard OpenSCENARIO was tested in an XR-based driving simulator study including a balanced order. The scenarios were run within the simulator software using the open-source OpenSCENARIO player Esmini. The used road network was designed by combining multiple blocks of 8 base road building blocks into two roads with approximately 30 blocks each, where a scenario only occurs within one base block. Using this block structure, a total of 48 scenarios split on two roads with 24 scenarios each was performed. The study using this block concept was successfully completed with 44 participants.

Keywords: OpenSCENARIO, Esmini, simulator study design, simulator scenarios, balanced order

Copyright: © 2026 by the authors.
Licensee Driving Simulation Association.

Introduction

Driving simulators with realistic road environments are widely used to study human behaviour and test new vehicle technologies. To reduce development time and support collaboration, it is important to enable scenarios to be run across different simulator platforms. This makes standardized approaches for generating road networks and describing scenarios essential.

OpenSCENARIO, developed by ASAM, is an open standard for describing dynamic driving scenarios in a structured, machine-readable XML format. It defines elements such as actors, manoeuvres, triggers, and environmental conditions, supports modularity through catalogues, and integrates with OpenDRIVE for road network generation (ASAM, 2022). The standard is widely used for large-scale scenario generation in virtual testing of ADAS and automated vehicles (Li et al., 2023; Baker et al., 2024; Kim et al., 2025), and methods for generating scenarios from traffic databases have been proposed (Karunakaran et al., 2023). However, most applications focus on short scenarios for offline simulations with limited experience on its use in driving simulator studies involving multiple participants, where scenario reordering is important for balanced designs. Our approach in this work is to

combine several smaller scenarios in a balanced order for a longer driving simulator study.

Study design

The simulator study focusing on driver interactions with cyclists and pedestrians was conducted in a fixed-base driving simulator equipped with a vehicle mock-up and an XR headset providing 360° field of view, see Figure 1. This setup allowed participants to view the physical vehicle controls, their own body and hands, and to freely look around the virtual environment implemented using Unreal Engine 5.



Figure 1. Fixed-base driving simulator used in the study, equipped with a vehicle mock-up and an XR headset.

The study included 16 versions of a cyclist-related scenario, 12 variations of a pedestrian scenario, and 4 versions of a car-following scenario, all requiring randomization to reduce order effects. Additionally, failures in individual scenarios should not affect

subsequent ones. OpenSCENARIO was used to implement scenarios where each scenario was linked to an OpenDRIVE road with the tag RoadNetwork. Thus, each scenario is tightly connected to a road network. To enable reordering of the 48 scenarios without creating separate road networks for each participant, the road layout was constructed using five specific blocks implementing the targeted events, see Figure 2. Block 1 consists of two four-way intersections used for the car-following events; block 2 is a pedestrian crossing for pedestrian events; block 3 is a T-junction used to clear traffic; block 4 is a four-way intersection used for cyclist events; and block 5 is an additional four-way intersection used to clear traffic. These blocks were combined into a road segment repeated 20 times to create two road networks. Because cyclist events were more frequent, some segments omitted scenarios in blocks 1, 3, and 5. In such cases, one, two, or all three of these blocks were removed to shorten the overall route and reduce participant fatigue and simulator sickness.

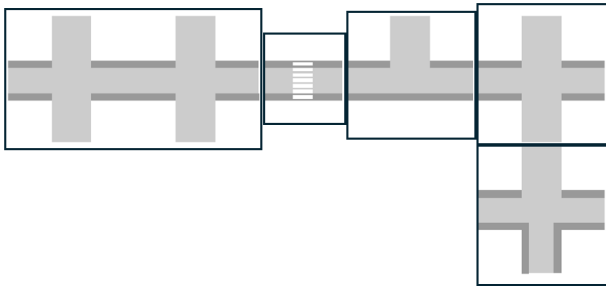


Figure 2. The study's repeated road segment, composed of five event-specific blocks.

Esmini was integrated into the VTI simulator using esminiLib, which provides a C++ API. Esmini is a lightweight, open-source simulator for running OpenSCENARIO and OpenDRIVE files (GitHub 2025). The workflow involved loading a scenario at the start of a road block, running it, and ensuring it finished before the end. Otherwise, it was forcefully stopped. As a result, Esmini was restarted multiple times during a run for different scenario variations.

When loading a scenario, which required about one second, parameters (via the OpenSCENARIO ParameterDeclarations tag) were used to select variants and adjust road IDs, enabling scenarios to be relocated and varied. Only the ego vehicle was synchronized with Esmini, meaning no interaction occurred between scenario vehicles and surrounding traffic—this was handled through careful scenario and traffic design.

Results and discussion

A driving simulator study with immersive XR visualization was conducted using OpenSCENARIO scenarios. Event ordering was balanced across participants and executed smoothly. Scenarios

started and stopped around the road block borders, resulting in a tight design with often only seconds between scenarios. By repeatedly stopping Esmini, failures in one scenario could not affect subsequent ones.

However, limitations in Esmini constrained the study design. At the time, it lacked support for traffic lights, and the surrounding traffic didn't use indicators and brake lights. Also, since Esmini runs as a separate process, any data to be logged must be explicitly sent to the main system, meaning internal data not exposed by Esmini, such as vehicle acceleration states, could not be recorded.

Conclusions/implications

This paper presents a structured approach to scenario and road design using OpenSCENARIO and Esmini to address challenges associated a balanced study design. The method works well when all traffic participants can enter and exit within a single block without affecting surrounding traffic. The proposed method enabled the successful execution of a study with 44 participants, incorporating a randomized sequence of 48 scenarios. Limitations in Esmini are expected to diminish in future versions.

Acknowledgements

This work was carried out within the Integrated 4D driver modelling under uncertainty (i4Driving) project which is funded by the European Union's Horizon Europe research and innovation programme (Grant Agreement ID 101076165).

References

- ASAM (2022) ASAM OpenSCENARIO® 2.0.0 :: OpenSCENARIO
- Baker, P., Mitchell, J., Chodowicz, E., Zhang, X., Khastgir, S., & Jennings, P. (2024). Exploring Realism in Virtual Testing: Towards a Scalable Platform Using Open-Source Solutions for Automated Driving Systems. *In 2024 IEEE Intelligent Vehicles Symposium (IV)* (pp. 2517-2522).
- GitHub (2025) <https://github.com/esmini/esmini>.
- Karunakaran, D., Berrio, J. S., Worrall, S., & Nebot, E. (2023) Automatic lane change scenario extraction and generation of scenarios in OpenX format from real-world data. *arXiv preprint arXiv:2203.07521*.
- Kim, J., Kim, S., Jeong, G., Lee, E., & Lee, J. (2025). Rule-Based Translation From Natural Language to OpenSCENARIO by Injecting Various Disturbance for Driving Simulator. *IEEE Access*, 13, 207200-207210.
- Li, Q., Peng, Z. M., Feng, L., Liu, Z., Duan, C., Mo, W., & Zhou, B. (2023). Scenarionet: Open-source platform for large-scale traffic scenario simulation and modeling. *Advances in neural information processing systems*, 36, 3894-3920.

Improving Motion Cueing with High Shelf Filters

Craig Brown¹, Edward Smith², Nikhil Garrett², Chris Williams²

(1) ProCue Ltd, London, UK, e-mail: craig.brown@procue.xyz

(2) Dynisma Ltd, Bristol, UK, e-mail: {edward.smith, nikhil.garrett, chris.williams}@dynisma.com

Abstract - In audio engineering the “high shelf” can provide a sense of “clarity” or “air” to an audio signal. By increasing high-frequency gain, it is hypothesised that such clarity could also be brought to motion. To test this hypothesis, a second order high shelf filter is combined with the classical Motion Cueing Algorithm (MCA) on a state-of-the-art, high bandwidth, ultra-low latency Dynisma DMG-1 motion simulator. Subjective feedback is gathered from six highly-skilled drivers. Preliminary results hint towards the high shelf’s ability to increase perceived power of turn-in and entry braking cues, fidelity of mid-exit balance cues as well as road surface detail. These improvements are attainable with reduced workspace usage than a ‘full-spectrum’ increase in gain or scaling of the same magnitude. Therefore, the technique is especially relevant to ‘compact footprint’, ‘high bandwidth’ motion simulators.

Keywords: ride, handling, drivability, ultra-low latency, high-bandwidth

Copyright: © 2026 by the authors.
Licensee Driving Simulation Association.

Introduction

In audio engineering the “high shelf” can provide a sense of “clarity” or “air” to an audio signal (Winer, 2017). It is hypothesised that such clarity could also be brought to motion.

A generalised low shelf filter of order M is given in Holters and Zölzer (2006). To derive the equivalent high shelving filters, the following steps are required. Firstly, application of the conformal mapping $s \leftarrow s^{-1}$ which interchanges high and low frequencies (Holters and Zölzer, 2006; Smith, 2007). Finally, denormalisation by substituting $s = s\omega^{-1}$.

These substitutions produce Equation 1, which can be used to construct high shelf filters of positive integer order M , where g is the gain and ω is the transition frequency. By construction, these filters have minimum-phase behaviour, that is to say, all poles

and zeros are located on the left side of the complex plane. This is important for motion simulator applications since phase and group delay are related to latency, which is key for (performance) driving (Melzi and Prevati, 2025).

$$H(s) = \prod_{m=1}^M \frac{\omega + \sqrt{g} (e^{j\alpha_m}) s}{\omega + (e^{j\alpha_m}) s} \quad (1)$$

By setting $g = 1 + BK^{-1}$ simulator operators may conveniently treat the shelf as an offset or boost B to the overall vehicle acceleration scaling K .

Importantly, when placed in series with a classical washout filter, washout occurs as expected.

Magnitude and phase response for first and second order (dashed) high shelving filters, with $\omega = 10 \times 2\pi$, $g = 1.05, 1.15, 1.25$ are shown in Figure 2.

In summary, the shelf filter offers the ability to *scale high frequency content separately to overall motion*.

Methodology & Results

Possible applications of high shelving includes entry braking, corresponding to the longitudinal axis, handling, corresponding to yaw and lateral axes and ride, corresponding to vertical, roll and pitch axes.

An LMP2 vehicle model was used in AVL VSM. The model is well-correlated to a contemporary, competitive car. All drivers are highly skilled: three of the six were professional and hold motorsport licences.



Figure 1: The “Air Band” is a high shelving filter known for its pleasing sonic qualities. With permission of Maag Audio.

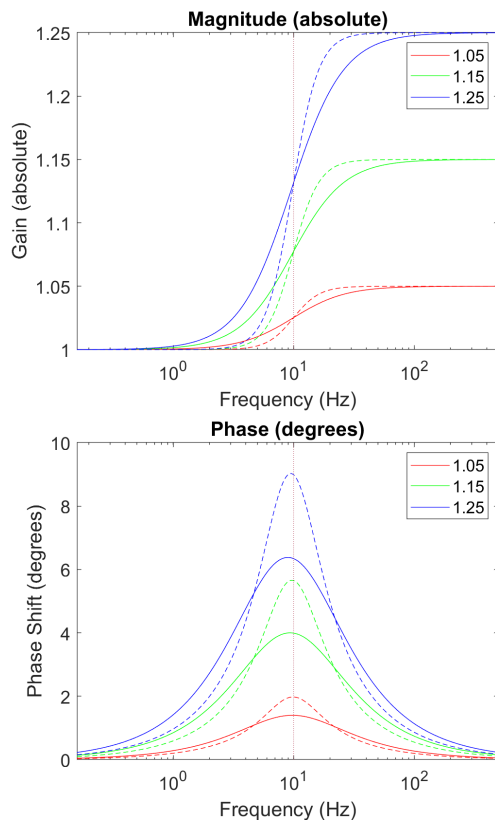


Figure 2: Comparison of high shelf filters for different values of g . Solid lines: 1st order. Dashed: 2nd order.



Figure 3: Dynisma DMG-1 test at Spa-Francorchamps.

Participants were instructed to drive full-pace, consistent laps of the Circuit of Spa-Francorchamps, and were blind to parameter changes.

Feedback was sought via an intercom system, asking the driver to rate changes on a numeric scale. Though informal, this is how technologies are typically assessed in a motorsport setting due to time pressures involved.

Results are summarised in Table 1 and are a collated summary of all feedback received.

Discussion & Conclusion

A new method of increasing realism in motion simulators was presented. It is simple, easy to implement, intuitive and may be placed in series with any MCA. It is especially powerful for high bandwidth simulators, such as the Dynisma DMG-1 used in this study (Table 2), as it increases motion fidelity without a significant

Table 1: Table summarising driver feedback.

Axes, ω , B	Comments (collated)
Lateral 2Hz, Yaw 1Hz $B \leq 0.3$	Perceived increase in: 1) bite at corner entry, increased sense of rear stability and turn in power; 2) understeer mid corner—likely induced by contrast with entry feeling; 3) oversteer on corner exit with kerb strike.
Longitudinal [1Hz 2Hz], $B \leq 0.3$	Perceived increase in: 1) initial braking efficiency; 2) pitching (despite pitch filter being unchanged); 3) rear traction when applying throttle after braking. Heavy braking zones perceived as more realistic. Other braking zones perceived as less realistic (likely by contrast).
Vertical, Roll, Pitch, 2.5Hz, $B = 0.1$	Perceived increase in road surface detail and realism. Improved sense of connection to road.

Table 2: Dynisma DMG-1 specifications

Specification	Value
Nominal payload	250kg – 400kg
Motion bandwidth	Up to 100Hz
Motion latency	<5ms
Max. excursion (long, lat, heave)	1300mm, 1300mm, 150mm
Max. excursion (roll, pitch, yaw)	$\pm 8^\circ$, $\pm 7.5^\circ$, $\pm 45^\circ$
Combined lat-yaw	500mm @ $\pm 30^\circ$
Combined lat-long excursion rad.	650mm

increase in workspace usage. Whilst feedback gave a strong indicator of improvement, more testing is required to be conclusive.

The range of potential applications include, but are not limited to, improving perception of corner entry, rear stability, oversteer, understeer, braking efficiency, rear traction, road surface detail and Noise, Vibration & Harshness (NVH).

A combined classical/high shelving MCA has 6+ parameters per axis giving 36+ in total, which becomes difficult to tune manually. Automatic parametrisation (Asadi, et al., 2015), using a tool such as ProCue (Brown, 2024), is recommended for extracting the most performance from this new approach.

References

- Asadi, H., Mohamed, S., Nelson, K., Nahavandi, S., and Zadeh, D. R., 2015. Human Perception-Based Washout Filtering Using Genetic Algorithm. In: S. Arik, T. Huang, W. K. Lai, and Q. Liu eds. *Neural Information Processing*. Cham, Germany: Springer International Publishing, pp. 401–411.
- Brown, C., 2024. ProCue: State-of-the-art Motion Cueing For All. In: A. Kemeny, J.-R. Chardonnet, F. Colombet, S. Espié, and F. Merienne eds. *Proceedings of the Driving Simulation Conference 2024 Europe VR*. Driving Simulation Association. Strasbourg, France.
- Holters, M. and Zölzer, U., 2006. Parametric higher-order shelving filters. In: *2006 14th European Signal Processing Conference*, pp. 1–4.
- Melzi, S. and Prevati, G., 2025. Driving simulator: Analysing the impact of mechanical latency on the perception of lateral dynamics. *Journal of Vibration and Control*, 31(7-8), pp. 1510–1523.
- Smith, J. O., 2007. *Introduction to Digital Filters with Audio Applications*. 1st. W3K Publishing.
- Winer, E., 2017. *The Audio Expert*. 2nd. New York: Routledge.

Fatigue-Inducing Scenario Design for Truck-Driving Simulation: From Evidence to Implementation

Tiago Fonseca¹, Telmo Costa Botelho¹, Sérgio Pedro Duarte¹, António Lobo², Sara Ferreira¹

(1) CITTA, Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal, e-mail: {tffonseca, tcbotelho, spduarte, sara}@fe.up.pt

(2) CITTA, Departamento de Engenharia Civil, Universidade de Coimbra, Rua Luís Reis Santos, 3030-788 Coimbra, Portugal, e-mail: lobo@uc.pt

Abstract – Road freight transport exposes truck drivers to conditions that increase fatigue risk, yet fatigue-induction protocols used in truck-driving simulators remain difficult to reproduce and translate into controlled operational testing environments. This study aimed to implement an evidence-informed truck-driving scenario derived from a prior systematic review and to assess the technical reproducibility and preliminary validity of the resulting simulator protocol. A two-stage design was adopted. First, review-derived scenario parameters were translated into explicit simulation elements in SILAB and implemented in a high-fidelity truck-driving simulator. Second, the resulting protocol was evaluated through technical verification and a pilot trial involving five licensed non-professional drivers. The implemented scenario comprised a continuous 110 min drive under low-variability, highly monotonous conditions, including low-light late-evening visual settings, limited traffic complexity, and sparse scripted events. Technical verification across three dry runs confirmed stable execution, correct event activation, consistent traffic behavior, and complete synchronized data logging. In the pilot trial, subjective sleepiness, assessed using the Karolinska Sleepiness Scale (KSS), increased by 2.8 points over the drive. Overall, the findings support the technical feasibility and reproducibility of the implemented scenario and provide preliminary evidence that the resulting protocol can elicit progressively increasing subjective sleepiness.

Keywords: driving simulator; truck driver; fatigue; simulation design; road safety.

Copyright: © 2026 by the authors.
Licensee Driving Simulation Association.

Introduction

Truck drivers frequently work under demanding conditions, including prolonged driving, irregular schedules, insufficient restorative sleep, and sustained exposure to monotonous environments (Fonseca & Ferreira, 2025). Together, these factors increase vulnerability to fatigue, a multidimensional psychophysiological state associated with reduced alertness, slower reactions, impaired vigilance, and a diminished capacity to respond to traffic demands (May & Baldwin, 2009; Williamson et al., 2011).

Studying driver fatigue in real traffic is both methodologically challenging and ethically constrained. Meaningful fatigue induction may require prolonged exposure, night-time driving, or post-shift testing, all of which can increase risks to participants and other road users. Driving simulators therefore provide a safer and more controllable alternative, allowing fatigue-related conditions to be reproduced while supporting the synchronized

collection of subjective, behavioral, vehicle-based, and physiological measures.

Despite these advantages, simulator-based studies of truck-driver fatigue remain difficult to compare and replicate. Prior studies vary substantially in scenario characteristics, including road geometry, traffic density, event structure, and monotony, as well as in protocol design, particularly driving duration and time of day. These elements are also frequently underreported, limiting reproducibility and hindering the translation of published findings into operational simulator implementations.

To address this gap, the authors conducted a separate systematic review examining how fatigue has been induced in truck-driving simulators and which scenario characteristics are most consistently associated with measurable fatigue-related changes (Fonseca & Ferreira, 2026). The review found that such effects were reported more consistently when prolonged time on task was combined with low-variability driving demands, particularly in monotonous environments with limited traffic complexity.

However, identifying these patterns is only the first step. Their practical value depends on whether they can be translated into a clearly specified driving scenario and implemented within a reproducible experimental protocol.

Accordingly, this study aimed to translate the design parameters identified in the review into an evidence-informed truck-driving scenario and to assess the technical reproducibility and preliminary validity of the resulting simulator protocol. By providing an implementable fatigue-induction scenario embedded within a controlled experimental protocol, the study has the potential to support future simulator-based research and the early-stage evaluation of fatigue-monitoring technologies, in-vehicle warning systems, and related countermeasures.

Methods

This study adopted a two-stage design: (i) implementation of a fatigue-induction scenario and technical verification of its execution, and (ii) pilot validation of the resulting simulator protocol in a driver-in-the-loop trial. The first stage examined whether the scenario could be implemented and executed consistently across dry runs. The second stage examined whether the protocol—defined as the experimental use of the scenario together with the exposure duration, measurements, and participant procedures—showed a preliminary ability to elicit progressive fatigue.

Simulator Setup

The experiment was conducted in the truck-driving simulator at the University of Porto, configured as a static platform incorporating a full Volvo FM 380 cabin. The visual system comprised three projection screens providing a horizontal field of view of 175° and a vertical field of view of 35°, together with virtual mirrors. Driver interfaces included a direct-drive force-feedback steering system, accelerator and brake pedals, automatic transmission, an adjustable air-suspended seat, and an audio environment.

The scenario was implemented in SILAB version 7.3 (Krueger et al., 2005). Scripted events were triggered at fixed distances, and the same event logic was maintained across runs. Simulator outputs, event markers, and subjective assessment checkpoints were time-synchronized in SILAB using the simulation start time as the temporal reference.

Scenario Parameters

The implemented scenario consisted of a continuous motorway driving task on a dual carriageway with two lanes in each direction. Lane width was 3.75 m, with shoulders measuring 3.75 m on the right and 1.0 m on the left. The horizontal alignment comprised straight and curved segments distributed throughout the route, totaling 99 km of straight sections and 101

km of curved sections, with a mean curve radius of 1,350 m. The vertical alignment was approximately centered at 0 m, with variations of up to ± 3 m. No intersections, ramps, or distinctive roadway features were included, and signage density was low.

Traffic volume was low and stable, comprising six vehicles travelling in the opposite direction and one vehicle travelling in the adjacent left lane. No lead vehicle, lane-changing traffic, heavy vehicles, or car-following demands were included. Two sparse event types were implemented: a swerve to avoid an animal on the road and an overtaking maneuver involving a vehicle in the right lane. Swerve events occurred at approximately 10 and 100 min, whereas overtaking events occurred at approximately 40 and 80 min. Participants were instructed to maintain a speed of 90 km/h and remain in the right lane unless an event required a lane change.

The simulated environment represented clear-weather, late-evening conditions. Testing sessions were conducted between 10:00 and 19:00. The protocol comprised 110 min of continuous exposure to the implemented scenario. No breaks were provided. Monotony was induced through repetitive scenery, low traffic volume, prolonged straight segments, large-radius curves, sparse events, limited task switching, and low signage variability.

Pilot Protocol

The pilot validation stage included licensed drivers recruited from the academic community of the University of Porto. Inclusion criteria were possession of a driving license, age between 18 and 65 years, absence of self-reported sleep disorders, and no exclusive night-shift work. Corrected-to-normal vision was permitted. Five non-professional drivers participated: three men and two women, with a mean age of 29.8 years (SD = 8.8) and a mean license-holding duration of 10.6 years (SD = 8.4).

Participants attended a single laboratory session. Baseline sleep–wake information was collected by self-report, including prior sleep duration, prior wake duration, and recent duty schedule. No formal sleep-restriction protocol was imposed, but participants were instructed not to nap before testing. Consumption of caffeine and energy drinks was prohibited within 6h of testing, nicotine within 2h, and alcohol within 24h. Medication use was screened, and compliance with the pre-test restrictions was verified by self-report. After briefing and confirmation of eligibility, participants completed a 5-min familiarization drive in an equivalent motorway scenario. Simulation sickness was assessed by self-report, and the session was interrupted or terminated if relevant symptoms occurred.

Outcome Measures

Technical reproducibility was assessed through three complete dry runs conducted before participant

testing. Verification focused on stable scenario execution, correct event activation, consistent traffic behavior, and complete time-stamped data logging.

During participant testing, subjective sleepiness was assessed every 10 min using the Karolinska Sleepiness Scale (KSS; Åkerstedt & Gillberg, 1990), ranging from 1 (“extremely alert”) to 9 (“very sleepy, fighting sleep”). Secondary outcomes were speed, lateral position, and line crossings. Speed and lateral position were recorded continuously at 60 Hz. Lateral position was defined as the distance from the vehicle center of gravity to the lane center, with negative values indicating leftward displacement and positive values rightward displacement. Line crossings were defined as events in which half of the vehicle extended outside the lane boundaries. To avoid distortion from scripted maneuvers, data recorded within 1.5 km before and 1.5 km after each event were excluded from the analyses.

Results

The implemented scenario executed correctly across all three dry runs. Scripted events were triggered as planned, traffic behavior followed the predefined logic, and all relevant outputs were exported with complete time-stamped records. No technical failures or missing data were observed.

All five participants completed the pilot drive without exclusions or protocol deviations. Figure 1 shows the temporal evolution of KSS, speed, lateral position, and line crossings over the 110-min drive. To align the vehicle-based measures with the 10-min KSS assessments, all outcomes were summarized at 11 checkpoints using data from the preceding 10-min window. Values represent participant means, with error bars indicating the standard error of the mean.

KSS showed the clearest temporal pattern across the session. Mean KSS increased from 1.8 at 10 min to a peak of 4.8 at 90 min and remained elevated thereafter. This pattern indicates a progressive increase in subjective sleepiness over the course of the drive.

Mean speed increased from 77 km/h at the first checkpoint to 86 km/h at 20 min. It then remained within a relatively narrow range for most of the session. Mean lateral position ranged from 0.13 to 0.31 m. Positive values were observed throughout the session, indicating a consistent rightward position relative to the lane center. Line crossings were generally infrequent, with no crossings recorded at several checkpoints and the highest mean value observed at 100 min.

Discussion

Among the outcomes examined, KSS showed the clearest temporal evolution, increasing by 2.8 points

from the first to the final checkpoint. This magnitude of change is consistent with previous evidence indicating that KSS is sensitive to time on task in driving contexts. Åkerstedt et al. (2014) reported increases of approximately 1.5–2.5 points during 40–80 min of night driving and approximately 1.5 points during daytime driving, supporting the interpretation that the present protocol elicited a measurable increase in subjective sleepiness.

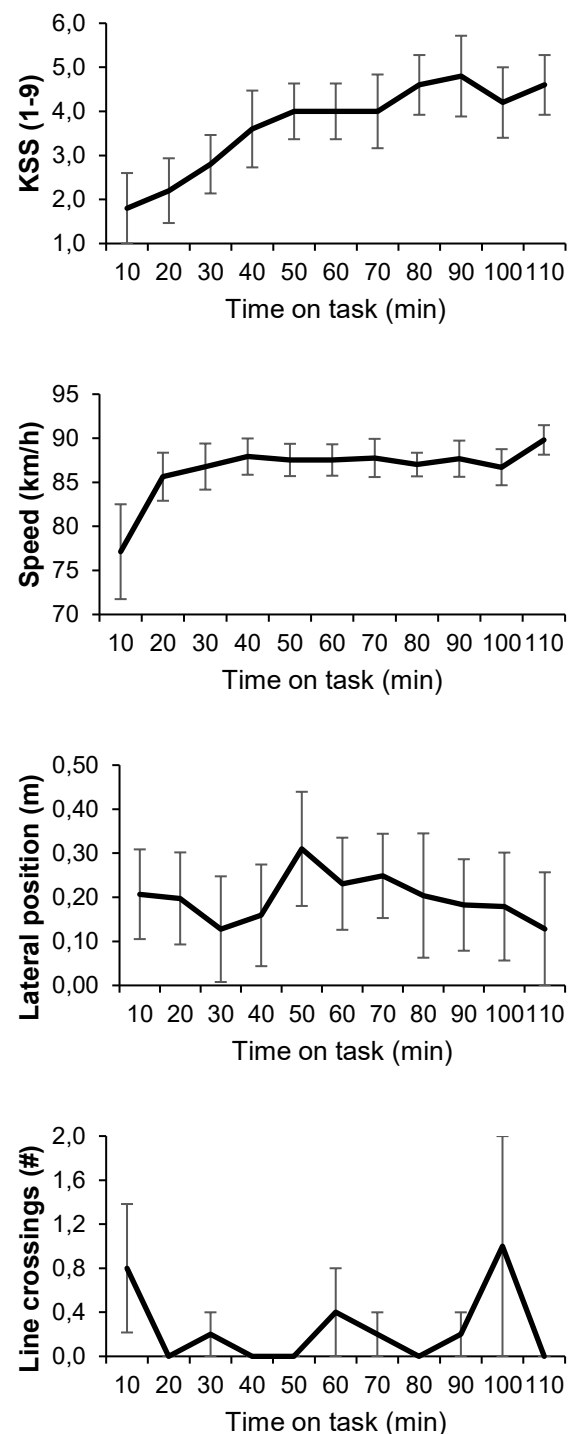


Figure 1. Mean values of KSS, speed, lateral position, and line crossings across time-on-task segments.

The vehicle-based outcomes showed less consistent temporal variation. Speed increased during the early part of the session and subsequently stabilized near the instructed target of 90 km/h. This pattern differs from the progressive time-on-task-related increase reported by Ahlström et al. (2018), but is consistent with drivers' tendency to maintain their preferred speed under the predominantly free-flow conditions of the present scenario (Lobo et al., 2014).

Lateral position remained positive throughout the drive, indicating a persistent rightward position relative to the lane center. This pattern may partly reflect the visual and geometric characteristics of the simulated road environment. Antonson et al. (2009), for example, found that landscape context can influence lateral position, with open landscapes associated with more rightward vehicle placement. The finding is also consistent with Melo et al. (2012), who observed that drivers tended to travel closer to the edge line when shoulder width and lateral clearance were greater, as in the present scenario, which included a 3.75-m right shoulder. By contrast, line crossings remained infrequent and occurred intermittently rather than increasing progressively, differing from previous evidence that line crossings may increase with sleepiness-related impairment (Ahlström et al., 2018).

Overall, the findings support the technical feasibility and reproducibility of the implemented scenario and provide preliminary evidence that the resulting protocol can elicit progressively increasing subjective sleepiness. The scenario may therefore serve as a useful foundation for future simulator-based studies of truck-driver fatigue.

Limitations

This study should be interpreted in light of respective limitations. The pilot stage included only five drivers and no reference condition, limiting generalizability, statistical inference, and causal interpretation. Validation relied primarily on subjective sleepiness rather than a broader multimodal fatigue assessment. Simulator adaptation effects cannot be excluded because no stabilization assessment was conducted after the familiarization drive. Finally, the manual and platform-dependent implementation of the scenario limits its broader reusability.

Funding: This work was financially supported by UID/04427 of the CITTA—Research Center for Territory, Transports and Environment, funded by Fundação para a Ciência e a Tecnologia (FCT), I.P./MECI through the national funds. This work was also financed by the FEDER funds through COMPETE2030 and with financial support of FCT, I.P., within the framework of the project Tec4Safe - Tailored technology and innovation for driving safely, with No. 15365, and the operation code at the Funds

Platform COMPETE2030-FEDER-00904100 and FCT code 2023.16094.ICDT. In addition, this work was supported by FCT, I.P. through the individual research grants 2025.02730.BDANA and 2025.01027.BD.

References

- Ahlström, C., Anund, A., Fors, C., & Åkerstedt, T. (2018). The effect of daylight versus darkness on driver sleepiness: a driving simulator study. *Journal of Sleep Research*, 27(3). <https://doi.org/10.1111/jsr.12642>
- Åkerstedt, T., Anund, A., Axelsson, J., & Kecklund, G. (2014). Subjective sleepiness is a sensitive indicator of insufficient sleep and impaired waking function. *Journal of Sleep Research*, 23(3), 242–254. <https://doi.org/10.1111/jsr.12158>
- 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>
- Antonson, H., Mårdh, S., Wiklund, M., & Blomqvist, G. (2009). Effect of surrounding landscape on driving behaviour: A driving simulator study. *Journal of Environmental Psychology*, 29(4), 493–502. <https://doi.org/10.1016/j.jenvp.2009.03.005>
- Fonseca, T., & Ferreira, S. (2025). Truck Driver Safety: Factors Influencing Risky Behaviors on the Road—A Systematic Review. *Applied Sciences (Switzerland)*, 15(17). <https://doi.org/10.3390/app15179662>
- Fonseca, T., & Ferreira, S. (2026). Scenario Parameters for Fatigue Induction in Truck-Driving Simulators: A Systematic Review of Experimental Designs. *Applied Sciences*, 16(6), 3057. <https://doi.org/10.3390/app16063057>
- Krueger, H.-P., Grein, M., Kaussner, A., & Mark, C. (2005). SILAB - A task-oriented driving simulation. *Driving Simulation Conference*, 323–331.
- Lobo, A., Rodrigues, C., & Couto, A. (2014). Estimating percentile speeds from maximum operating speed frontier. *Transportation Research Record*, (2404), 1–8. <https://doi.org/10.3141/2404-01>
- May, J. F., & Baldwin, C. L. (2009). Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies. *Transportation Research Part F: Traffic Psychology and Behaviour*, 12(3), 218–224. <https://doi.org/10.1016/j.trf.2008.11.005>
- Melo, P., Lobo, A., Couto, A., & Rodrigues, C. (2012). Road cross-section width and free-flow speed on two-lane rural highways. *Transportation Research Record*, (2301), 28–35. <https://doi.org/10.3141/2301-04>
- Williamson, A., Lombardi, D. A., Folkard, S., Stutts, J., Courtney, T. K., & Connor, J. L. (2011). The link between fatigue and safety. *Accident Analysis and Prevention*, 43(2), 498–515. <https://doi.org/10.1016/j.aap.2009.11.011>

Assessing Communication Behaviors Between Excavator Operators, Human Workers and Assistance Systems on a Virtual Construction Site

René Reinhard, Franziska Hekele, Jonathan Schneider, Tim Nicolai, Sebastian Emmerich, Michael Burger

Fraunhofer Institute for Industrial Mathematics ITWM, Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany, rene.reinhard@itwm.fraunhofer.de

Abstract - Construction sites are among the most dangerous workplaces, in part due to the required execution of cooperative tasks between vulnerable construction workers and heavy machinery. An application of assistance systems in such complex scenarios remains challenging: While for on-road applications, proximity warning systems and brake assistants are getting increasingly reliable in the detection of safety-critical situations, their adaptation to heavy machinery requires algorithms to distinguish between planned and unplanned individuals to allow for safe interactions on the construction site. The development, testing, and validation of such assistance systems in the real world may entail high risks for engineers and testing personnel. The current study employs a linked-lab simulator setup, consisting of an interactive excavator simulator and a VR-based worker simulator, to study how proximity warning system use and spatial task overlap affect communication between the excavator operator and the ground-level worker. Subjective and objective measures are used to gain insights into possible reasons for behavioral changes in typical worker-machine interactions on construction sites, and implications for the development of assistance systems for heavy machinery are explored.

Keywords: Cooperative Virtual Interaction, Interactive Driving Simulation, Virtual Reality

Copyright: © 2026 by the authors.
Licensee Driving Simulation Association.

Introduction

The relevance of construction site safety and its implications for workers' well-being are evident in accident statistics. In Germany, insurance providers reported 103,949 construction site injuries in 2024, of which 15.2% resulted from worker collisions with movable objects or heavy machinery (DGUV, 2025), accounting for 23.8% of fatal construction site accidents.

Simulators on and off the road

Given construction sites' temporary nature and associated safety risks (Zhang, et al., 2017), interactive simulation has been widely used to identify hazardous situations (Özel and Pekerçli, 2025; Perlman, Sacks, and Barak, 2014) and train workers (Sun, et al., 2022). It offers controllable, reproducible, and safe conditions with good real-world transferability. However, it often presents single user non-interactive scenarios (Perlman, Sacks, and Barak, 2014), thus often falling short of the crowded, communicative, and collaborative reality of actual sites (Sun, et al., 2026). Even beyond construction, where simulators are widely used in research and development (Bruck, Haycock, and Emadi, 2021), most are placing only one human in the loop while simulating or replaying surrounding actors (e.g., Mahadevan, et al., 2019).

Linked-Lab Simulators

This approach may suffice for most on-road use cases, where clear, strict rules govern road user interactions, e.g., a pedestrian crossing in front of a car can be directly classified as potentially dangerous. Advanced studies on construction sites, however, require complex inter-agent communication with multiple humans (Feldstein, et al., 2018).

In this vein, bicycle or pedestrian simulators have been used to study interactions among road users for safety or transport system optimization (Zambare and Liu, 2025). On-road linked-lab experiments have examined vehicle-pedestrian interactions at crosswalks (Lehsing, Benz, and Bengler, 2016), the influence of automated driving on pedestrian behavior (Bazilinskyy, et al., 2020) or of external communications in mixed traffic (Hübner, et al., 2022).

On construction sites, communication is even more complex, as multiple teams perform distinct tasks. Some involve cooperation between human workers and heavy machinery; others take place independently but within the same workspace. This work therefore proposes interactive, cooperative simulation as a promising method for studying the complex scenarios encountered daily on construction sites. Placing both a ground-level worker and a heavy-machinery operator in the same virtual environment enables the study of communication channels be-



Figure 1: Communication setup and interfaces for the linked-lab experiment. The VR participant (a) controls the worker avatar (b) placed in the same virtual environment as the virtual construction machine (c), controlled from the interactive driving simulator (d).

tween worker, operator, and, where applicable, machine assistance systems.

Accordingly, the present study explores task performance, communication strategies and assistance systems in a linked-lab setup, with VR-based workers and simulator-based excavator operators sharing a virtual construction site. This short paper focuses on the VR-based worker.

Methods

Sample description

Twelve participants were initially recruited, requiring normal or corrected-to-normal visual acuity, intact stereopsis, and absence of relevant medical contraindications. One participant was excluded due to marked simulator sickness during training.

The final sample of eleven participants completed all three sessions of the experiment ($M_{\text{age}} = 27.3$ years; [23; 35]; $SD_{\text{age}} = 3.7$ years; 91% male). 27% of the sample reported previous experience with an excavator (between less than a year and above 10 years). The research complied with the APA Ethics Code and the Declaration of Helsinki with latest amendments.

Apparatus & Materials

The linked-lab setup consisted of two real-time-coupled laboratories (Figure 1): the Fraunhofer ITWM Robot based Driving and Operation Simulator (RODOS[®]) and a XR laboratory worker simulation.

Driving simulator. The RODOS[®] provides realistic kinesthetic feedback via a motion system based on the serial 6-DOF kinematics of an industrial robot. Here, an adapted Volvo EW160 excavator cabin with original control elements was used. The visual scene was rendered using eight 2k dual-stereo projectors (120 Hz refresh rate) onto a 10 m diameter dome ($300^\circ \times 110^\circ$ FOV). Eye openness, gaze direction, and expressions were tracked via in-cabin camera.

XR Laboratory. The virtual construction site (Figure 2a) was displayed using a Meta Quest Pro head-mounted display (1800×1920 pixels per eye; 106° horizontal FOV; 72 Hz refresh rate) within a $7\text{ m} \times 6\text{ m}$ tracked area. Participant tracking encompassed: (1) bodily movements and posture tracking through a 15-IMU (Rebocap) setup; and using the Meta Quest Pro's built-in IR-camera systems (2) hand-and-finger movements via optical tracking; (3) facial expressions; (4) eye openness and gaze direction.

Assistance system (AS). A prototypical proximity warning system provided the excavator operator with visual information about the presence, distance and

direction of workers within the machine's configuration space, displayed on an HMI in a bird's eye view centered on the cabin (Figure 2c). During inactive conditions, it displayed a "sensor error" notification.

Measures

Individual Reactions to VR and Simulator. The participant's condition was assessed with the Karolinska Sleepiness Scale (KSS) and Simulator Sickness Questionnaire (SSQ). Presence was assessed with the Igroup Presence Questionnaire (IPQ).

System Usability Scale (SUS). The AS's usability was evaluated using the SUS score between 0 and 100, interpreted using established adjective ratings (Bangor, Kortum, and Miller, 2008).

Single Item Online Measures. Subjective safety and the attention paid to the interaction partner were assessed using a 7-point response format.

Visual Attention Allocation. Based on eye-tracking, the proportion of total trial duration with gaze fixations within an AoI around the excavator was calculated.

Procedure

Each participant completed three sessions in counterbalanced order: (1) excavator training; (2) as excavator operator; (3) as VR construction worker. A trained confederate filled the complementary role. Sessions occurred within a week of each other.

Pre- and Post-Session. Before training, participants gave informed consent and completed a sociodemographic questionnaire. Before each session, they filled in the KSS and SSQ; after, the SSQ and IPQ.

Excavator side. Excavator operators brought the bucket into contact with successive visually highlighted objects. Points were awarded according to response time. Recorded task-irrelevant workers traversed the periphery in 90-second intervals. Counterbalanced, in half of the trials, the AS was used.

VR side. In VR, participants embodied an avatar visible to the confederate, who performed the excavator task following a standardized interaction protocol. The operator's expressions, gaze, and head position were rendered on an avatar visible in VR. Participants completed two blocks of 15 trials (5 min break between; 30-40 min total VR time). Each trial, they started at the periphery of one randomly selected sector of the excavator operating area. Their task was to find five sequentially presented items and grasp them (Figure 2). Object positions were randomly one of six predefined arrangements per sector, each maintaining consistent inter-object distances. Afterwards, they returned to the starting point, rated subjective safety and attention allocation, and were teleported to the next sector. In two thirds of trials,



Figure 2: Participant Tasks. VR session task (a) starting positions and exemplary tasks for each sector. Excavator task (b) with task objects in turquoise, score, and time displays. Excavator cabin display (c), including assistance system (middle), colors indicate distance to detected worker.

the excavator task overlapped spatially with the participant's task, i.e., occurred in the same sector.

Analysis

Excavator data analysis can be found in Emmerich, et al., 2026 and is not part of this short paper. The study employed a within-subjects experimental design with two independent variables: (1) assistance system presence. (2) workspace overlap, and visual attention allocation as time-varying covariate. Analyses utilized linear mixed-effects models for task completion times and cumulative link mixed models for subjective safety ratings. Random effects structures included by-participant random intercepts and random slopes for all fixed effects. Optimal models were determined via forward selection, preferring simpler models unless significant improvement was proven. Explorative analyses of interaction strategies used averages and standard deviations of visual attention allocation ratio for median splits.

Results

Task Completion Time

The best-fitting model for task completion time (Table 1) included an interaction between task overlap and AS use. Task completion times were higher when tasks overlapped, and this increase was larger when the AS was used (by 0.51 s without AS and 0.90 s with; Figure 3a). More attention on the excavator was also associated with longer completion times, with 0.34 s longer completion times per additional 10% of task time spent focusing on the excavator.

Subjective Safety

The optimal mixed model predicting subjective safety ratings (Table 2) included all first-order interactions.

Table 1: Mixed model predictors of task completion time

Predictors	β	ΔAIC	χ^2 (df)	p
Intercept (I)	5.731			
Task Overlap (T)	0.507	15.0	21.0 (3)	<.001
Assistance System (AS)	-0.042	-1.1	4.9 (3)	.177
Attention Distribution (AD)	3.360	11.6	17.6 (3)	<.001
T x AS	0.398	3.0	12.9 (5)	.048
T x AD		-8.9	1.2 (5)	.948
AS x AD		-3.1	7.0 (5)	.224
T x AS x AD		-14.6	3.4 (9)	.945

Thus, AS effects (Figure 3b) depended on task overlap and attention allocation. Under task overlap, the AS reduced subjective safety, more strongly with more attention on the excavator. Here, AS-related decreases ranged from -0.52 to -1.12 categories, depending on attention levels. Without overlap, AS decreases were smaller (0.26 to -0.60 categories). Conversely, effects of task overlap were stronger with AS and decreased with more attention on the excavator (-0.76 to -0.87 with AS; 0.02 to -0.09 without).

Communication Strategies

Exploratively, participants were classified using averages and standard deviations of visual attention allocated to the excavator, creating four groups of two to three participants (more/less attention on the excavator, more/less consistent across trials). Independent of consistency, participants who paid more attention to the excavator showed a marginally significant tendency toward longer completion times, $\Delta AIC = 0.5$, $\chi^2(3) = 6.519$, $p = 0.089$, regardless of AS, or task overlap (Figure 4). No trends were apparent for consistency, nor in the case of subjective safety.

Discussion

This study demonstrated the utility of a safe and controllable linked-lab simulator setup for investigating safety-critical interactions between construction site workers and excavator operators.

The observed increase in task completion times under task overlap reflects the fundamental challenge of shared workspace coordination. When operator and worker tasks brought them into spatial proximity, workers required extra time to monitor machine movements and negotiate safe passage.

Counterintuitively, completion times increased further with proximity warnings, suggesting that the AS changed operator behavior in ways that hindered

Table 2: Mixed models predictions of subjective safety

Predictors	β	ΔAIC	χ^2 (df)	p
Intercept (I)	7.466			
Task Overlap (T)	-0.093	21.2	27.2 (3)	<.001
Assistance System (AS)	0.257	12.7	18.6 (3)	<.001
Attention Distribution (AD)	-4.258	117.7	123.7 (3)	<.001
T x AS	-0.781	5.0	15.0 (5)	.010
T x AD	0.115	5.73	15.7 (5)	.008
AS x AD	-0.853	5.9	15.9 (5)	.007
T x AS x AD		-16.0	2.0 (9)	.992

rather than supported coordination. This may be due to the AS binding operator capacity, reducing communication effectiveness, or because it promoted more assertive operator behavior, which impaired the co-worker's performance. Here, lower excavator task performance with the AS supports the former explanation, but participants rated the system on average as good (9% poor, 36% OK, 36% good, 18% excellent) and as neither inconsistent (average 2.18 on 1-5 scale) nor cumbersome (average 2.09).

The AS was also associated with lower subjective safety during task overlap, particularly among workers focusing on the excavator. Consistent with the second explanation, workers may have interpreted AS-induced behavioral changes as more aggressive or as more hesitant (e.g., sudden stops), heightening perceived risk. This highlights that assistance system design for excavators must account for downstream effects on interaction partners, as improved operator awareness that reduces behavioral predictability may not yield the promised safety benefits.

Finally, the exploratory analysis hints at meaningful individual differences in monitoring strategies. The marginally significant tendency for high-attention participants to show longer task completion times across conditions may reflect a general cautiousness rather than adaptive situational responding.

Limitations and Future Directions. Several limitations should be acknowledged. The study's modest sample size, predominantly male and young demographic, and, despite training, the lack of excavator experience may limit generalizability.

On the operator side, tracking data mainly supported realistic avatar behavior to enhance communication and immersion. Future work will use in-cabin eye-tracking to analyze joint attention (Thennakoon, et al., 2025) and bilateral gaze dynamics in human-machine-human interactions. The absence of verbal communication, while plausible on noisy construction sites requiring hearing protection, was restrictive. Finally, the game-like excavator task may emphasize performance over safety. More collaborative tasks, like trench work or pipe laying, will be considered.

Conclusion

This study demonstrated that linked-lab simulator setups are a viable and powerful method to investigate

the complex interactions in construction contexts, allowing for the systematic and reproducible development, testing and validation of assistance systems to improve the safety on construction sites. Here, the findings indicate that assistance system use can have unintended effects on the performance and perceived safety of nearby workers who are not the direct users, underscoring the need to consider multiple stakeholders early in system development and evaluation to improve construction site safety.

References

- Bangor, A., Kortum, P., and Miller, J., 2008. An empirical evaluation of the System Usability Scale. *Int J Hum Comput Interact*, 24(6), pp. 574–594.
- Bazilinskyy, P., Kooijman, L., Dodou, D., and De Winter, J. C., 2020. Coupled simulator for research on the interaction between pedestrians and (automated) vehicles. In: *Driving Simulation Conference Europe. Antibes, France*.
- Bruck, L., Haycock, B., and Emadi, A., 2021. A Review of Driving Simulation Technology and Applications. *IEEE Open J Veh Technol*, 2, pp. 1–16.
- DGUV, 2025. *Arbeitsunfallgeschehen 2024*. Deutsche Gesetzliche Unfallversicherung e.V. (DGUV).
- Emmerich, S., Reinhard, R., Schneider, J., Hekele, F., Nicolai, T., and Burger, M., 2026. Coupled Humans-in-the-Loop Co-Simulation for Validation of Safety-Critical Assistance Systems on the Construction Site. In: *Proceedings 9th CVT Symposium*.
- Feldstein, I., Lehsing, C., Dietrich, A., and Bengler, K., 2018. Pedestrian simulators for traffic research. *Int J Hum Factors Model*, 6(4), pp. 250–265.
- Hübner, M., Feierle, A., Rettenmaier, M., and Bengler, K., 2022. External communication of automated vehicles in mixed traffic. *Transport Res F-Traf*, 87, pp. 365–378.
- Lehsing, C., Benz, T., and Bengler, K., 2016. Insights into Interaction-Effects of Human-Human Interaction in Pedestrian Crossing Situations using a linked Simulator Environment. In: *13th IFAC Symposium on Analysis, Design, and Evaluation of Human-Machine Systems HMS 2016*. Vol. 49. 19, pp. 138–143.
- Mahadevan, K., Sanoubari, E., Somanath, S., Young, J. E., and Sharlin, E., 2019. AV-Pedestrian Interaction Design Using a Pedestrian Mixed Traffic Simulator. In: *Proceedings of the 2019 on Designing Interactive Systems Conference*, pp. 475–486.
- Özel, B. and Pekerçli, M., 2025. Construction site hazard recognition via mobile immersive virtual reality. *Autom Constr*, 173.
- Perlman, A., Sacks, R., and Barak, R., 2014. Hazard recognition and risk perception in construction. *Saf Sci*, 64.
- Sun, Y., Albaeaino, G., Gheisari, M., and Eiris, R., 2022. Online site visits using virtual collaborative spaces. *Adv Eng Info*, 53.
- Sun, Y., Bhadaniya, P., Eiris, R., and Gheisari, M., 2026. Online construction site visit. *Comput Educ X Real*, 8.

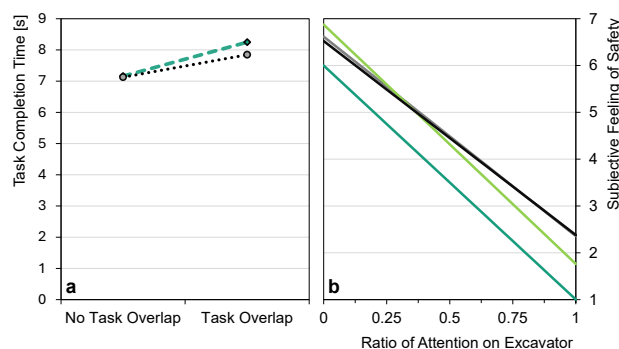


Figure 3: Effects of assistance system use (green = with, black = without) and task-area overlap (dark = overlap, light = no overlap) on (a) task completion time and (b) predicted subjective safety as a function of visual attention allocation.

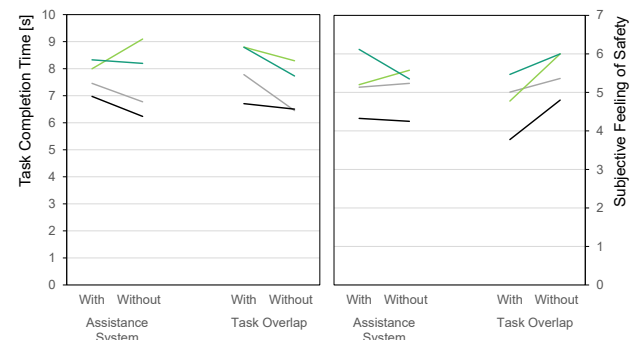


Figure 4: Average task completion times and subjective feelings of safety for participants with high (green) or low (monochrome) levels of attention spent on the excavator, with either more (darker) or less consistency (lighter).

- Thennakoon, K., Abeysinghe, Y., Mahanama, B., Ashok, V., and Jayarathna, S., 2025. Beyond Gaze Overlap. In: *Proceedings of the 2025 IEEE International Conference on Information Reuse and Integration and Data Science (IRI)*.
- Zambare, P. and Liu, Y., 2025. Modeling and Simulation of Pedestrian Behavior in Traffic Environment. *IEEE Access*, 13, pp. 149576–149603.
- Zhang, C., Lu, Y., Xu, R., Ye, X., Shi, Y., and Lu, P., 2017. An educational tool based on virtual construction site visit game. *Mod Appl Sci*, 11(8).

Funding. This project was co-funded by the European Union and the state of Rhineland-Palatinate.

AUTHORS' INDEX

A

Andersson, A.	189
Arab, A.-H.	143
Arnoux, E.	175
Åslund, J.	87

B

Belda-Lois, J.-M.	29
Bergen, M.	133
Blouin, D.	159
Böhle, M.	11
Bos, J. E.	33
Bräunl, T.	185
Brown, C.	83, 191
Burger, M.	197

C

Carrera Akutain, X.	95
Cattaneo, P.	95
Chardonnet, J.-R.	19
Cichon, M.	137
Colombet, F.	55, 113
Comolli, F.	95
Costa Botelho, T.	193

D

De-Oliveira, F.	113
Djukic, M.	187
Duarte, S. P.	193
Duell, S.	103
Düselder, A.	187
Durville, L.	143
Dute, M.	187

E

Eckstein, L.	119, 151
Einhäuser, W.	133
Emmerich, S.	197
Essalih, A.	143

F

Fahmy, O.	47
Fang, Z.	113
Ferreira, S.	193

Fischer, M.	133
Fittkau, N.	171
Fonseca, T.	193

G

Garrett, N.	191
Gausemeier, S.	171
Gobbi, M.	95
Gsellmann, I.	125

H

Hadj Selem, F.	143
Hager, C.	125
Hallak, Y.	159
Heiderich, M.	67
Hekele, F.	197
Holzappel, C.	83

J

Jacumet, R.	103
Jukić, M.	125

K

Kaskaya, S.	47
Kharrazi, S.	87
Kherji, W.	143
Kircher, K.	39
Kling, R.	137
Kolff, M.	67
Kriesten, R.	165

L

Laborie, B.	159
Lebbah, M.	143
Legran, P.	119, 151
Leibold, M.	103
Lilljebjörn, E. G.	87
Lobo, A.	193
Louvenberg, S.	33
Loya Garnica	61

M

Malena, K.	75, 171
Maletzky, A.	125
Mastinu, G.	95
Mateo-Martínez, B.	29

Mignotte, N.	55
Mohellebi, H.	175
Morgand, E.	55
Muik, T.	125

N

Nalda-Tárrega, V.	29
Nenninger, P.	165
Nicolai, T.	197

O

Obradors, R. T.	95
Ochel, L.	39
Olayan, A.	19

P

Pape, K.	75
Pautet, L.	159
Pool, D. M.	67
Popieul, J.-C.	175
Posselt, J.	55
Pretto, P.	125
Prior, M.	125

Q

Qiu, D.	113
---------	-----

R

Rahal, M.-C.	143
Rauch, Y.	165
Rebling, P.	165
Regnier, S.	113, 175
Reinhard, R.	197
Ricciardi, V.	67
Rieff, B.	11
Rinderknecht, S.	47
Rizk, A.	171
Roberts, P.	185
Rosenberger, M.	125
Russ, F.	119

S

Saab, L.	159
Sadat, S.	175
Sander, N.	165
Sandoz, B.	19
Schlueter, M.	83
Schneider, J.	197
Schwienbacher, M.	103

Sentouh, C.	175
Shyrokau, B.	67
Sliacan, J.	39
Smith, E.	191
Sommervogel, M.	113
Souman, J. L.	33

T

Tarlowski, S.	151
Temme, G.	133
Timmermann, J.	75
Trächtler, A.	75, 171

V

Vaillant, E.	175
van der Stelt, M.	185

W

Wagner, S.	103
Weidel, M.	39
Wieschollek, M.	137
Wild, M.	137
Williams, C.	191

Z

Zantner, F.	67
-------------	----

