Perceived Comfort and Safety in Automated Driving based on Physiological Signals

Findings from a Proving Ground Study

Jurjen Scharringa

Perceived Comfort and Safety in Automated Driving based on Physiological Signals

Findings from a Proving Ground Study

by

Jurjen Scharringa

Student number:	4708652	
Project Duration:	October 4 2024 - June 24 2025	
Thesis Supervisors:	Prof. dr. ir. R. Happee,	Delft University of Technology
	Ir. K. Gkentsidis,	Siemens Digital Industries Software
	Ir. ing. M. Sarrazin,	Siemens Digital Industries Software
Thesis Committee:	Prof. dr. ir. R. Happee,	Delft University of Technology
	Dr. Ir. G. Papaioannou,	Delft University of Technology
	Ir. V. Kotian,	Delft University of Technology
	Ir. K. Gkentsidis,	Siemens Digital Industries Software

Cover: Image taken of the Vehicle Under Test during the data acquisition days. Original image extended to current format with Photoshop.

Preface

I am pleased to present my Master's Thesis, which is the final step in my Master's degree in Robotics at the Delft University of Technology. This thesis explores perceived comfort and safety in automated driving and how those perceptions relate to physiological signals, specifically focusing on the Galvanic Skin Response (GSR).

This study is based on experimental data provided by my host company, Siemens Digital Industries Software. I first conducted an analysis to examine the relationship between passengers' self-reported comfort scores and their GSR. Based on the observed correlations, I then explored the potential of the GSR for predicting passenger comfort using a deep learning approach.

I began with a literature review of similar research, focusing on how previous research addressed the challenges of subjective comfort ratings and the use of the GSR. This was followed by a statistical analysis to explore the patterns within the collected dataset. Finally, predictive models were developed and evaluated to assess the extent to which the GSR, both on its own and augmented with vehicle dynamics or perception data, can predict perceived passenger comfort.

I hope that the findings presented in this thesis contribute meaningfully to the growing research field of comfort in automated driving and that fellow researchers can build on insights found in this thesis.

This Master's Thesis is submitted as one of the requirements for the Master's degree in Robotics at the Mechanical Engineering faculty at Delft University of Technology. The presented research was supervised by Prof. Dr. Ir. Riender Happee of Delft University of Technology, and Ir. Konstantinos Gkentsidis and Ir. ing. Mathieu Sarrazin from Siemens Digital Industries Software in Leuven, Belgium.

J.A. Scharringa
Leuven, June 2025

Acknowledgments

This thesis would not have been possible without the support and guidance of many people. First and foremost, I am deeply grateful to my supervisor, Prof. dr. ir. Riender Happee, at Delft University of Technology. Whenever I found myself taking on too much at once and overlooking crucial steps, Prof. Happee would encourage me to pause, reflect and return to the fundamental questions underlying this research. His patient questioning and insightful feedback were essential in shaping the direction of my work.

I also wish to thank my supervisors at Siemens Digital Industries Software, Ir. Konstantinos Gkentsidis and Ir. ing. Mathieu Sarrazin, for their valuable guidance throughout this project and constructive feedback on my thesis. In particular, I am especially grateful to Konstantinos for always making time for a quick discussion whenever necessary and for always responding to my questions, sometimes even during weekends and evenings. My thanks also go to the Siemens colleagues who were involved in conducting the experiment and collecting the dataset that served as the foundation of this thesis, and to my fellow interns at Siemens, whose support and camaraderie made the internship both productive and enjoyable.

Finally, I am very thankful to my friends and family for their support. Your thoughtful discussions and encouragement kept me motivated and sometimes even led to fresh insights.

Summary

While the introduction of automated driving offers various potential benefits in terms of enhanced road safety, improved traffic flow and allowing passengers to engage in non-driving-related tasks, these advantages depend on public acceptance of the technology. To address this, research on automated driving has extended from pure technical feasibility to include passenger comfort as an important factor, as it has been identified as a key factor in improving public acceptance. This thesis contributes to this research, as it combines the study of perceived comfort and safety in automated vehicles and proposes predictive models for subjective passenger experience and objective measures.

This thesis draws on data from an experiment conducted by Siemens Digital Industries Software (SISW). In this Wizard-of-Oz autonomous vehicle study, 32 participants participated as passengers in a ride conducted by a professional driver through a proving ground track featuring five distinct scenarios: a pedestrian crossing without visual obstruction, roadworks, a pedestrian crossing with visual obstruction, a cut-in and a car-following scenario. Each participant completed four laps, alternating between a predefined calm and aggressive driving style to elicit varied responses. After each scenario, participants provided subjective ratings of perceived comfort, safety and overall ride comfort. The Galvanic Skin Response was continuously measured throughout the experiment, alongside vehicle dynamics and perception data, yielding a comprehensive dataset.

Initial correlation analyses indicated significant correlations between all three self-reported subjective ratings and the GSR signal. However, when the data were stratified by style, thereby holding the driving style constant, these associations largely disappeared. These findings suggest that while the GSR is effective in capturing broad changes in passenger comfort, but fails to capture subtle differences.
Subjective ratings further revealed that participants felt more comfortable and safe around scenarios involving a pedestrian than in scenarios involving another vehicle, a pattern supported by corresponding physiological responses. No significant difference emerged between scenarios with versus without visual obstruction, nor between scenarios in which a pedestrian crossed the road and those in which they remained stationary.

A state-of-the-art deep learning model, fed with the phasic and tonic components of the GSR signal, distinguished calm-driven scenarios from aggressive-driven scenarios with an accuracy of 88.61%, underscoring the GSR's potential as an objective physiological marker of comfort. However, predicting subjective comfort and safety ratings proved more challenging. Incorporating vehicle dynamics and perception data yielded marginal gains but failed to achieve satisfactory performance. This shortfall stemmed from several key challenges: a highly imbalanced dataset biased toward positive responses, the inherently subjective nature of perceived comfort and safety and the substantial inter-subject variability of the GSR signal.
To counter these issues, this study combined synthetic oversampling of underrepresented responses with participant-specific fine-tuning. The best-performing configuration relied solely on the phasic and tonic components of the GSR and was fine-tuned for each test participant. Under an exact-match criterion, this configuration reached 58.1% (perceived comfort), 58.4% (perceived safety) and 54.3% (overall ride comfort); when adjacent classes were also accepted as true positives, these rose to 88.5%, 86.5% and 90.1%, respectively.
These results confirm that GSR is a reliable indicator of broad comfort levels but still struggles to resolve finer distinctions. The substantial gain in accuracy through user-adapted training further underscores the high inter-subject variability not only in the GSR signal but also in how individuals perceive comfort, highlighting the deeply personal and subjective nature of comfort and safety assessment.

The key findings of this thesis research are summarized and submitted to the IEEE International Conference on Intelligent Transportation Systems (ITSC) 2025 and await approval. The appendices provide a more detailed account of the research process, including comprehensive explanations, conclusions and visualizations of intermediate results, steps and figures that are not included in the main paper.

Contents

Statement of Purpose

The purpose of this chapter is to frame the scientific paper that forms the core of this thesis around the presented overarching research question:

> How can physiological arousal, measured through Galvanic Skin Response, combined with vehicle dynamics and perception data, be utilized to understand and predict passengers' perceived comfort and safety in automated driving?

The presented paper builds upon an earlier version submitted to the IEEE International Conference on Intelligent Transportation Systems (ITSC) 2025. It has since been revised to include updated findings and a refined methodology. The paper presents the methodology, key findings and interpretation needed to address the overarching research question in a self-contained and publishable format.

While the conference paper format required conciseness, this thesis provides comprehensive supporting analyses and detailed methodological considerations through its appendices. The supplementary materials are structured as follows:

Appendix A presents a comprehensive overview of the experimental setup performed by Siemens Digital Industries Software, including visual documentation. Furthermore, it outlines the data preprocessing pipelines and presents initial data visualizations.

Appendix B focuses on the analytical part of the research question, specifically addressing how we can understand perceived comfort and safety and their relationship to physiological arousal. Through a systematic analysis of the collected data, this appendix examines multiple research sub-questions and provides an expanded discussion of the findings presented in the main paper, incorporating additional statistical tests and visualizations that complement the main results.

Appendix C addresses the predictive modeling aspect of the research question. It provides an examination of the deep learning architecture implemented, investigates various input configurations and addresses challenges encountered during model employment. The analysis includes comprehensive evaluation metrics, presenting results through confusion matrices and Receiver Operating Characteristic curves, extending beyond the core findings presented in the main paper.

Appendix D contains an acknowledgment statement regarding the use of artificial intelligence tools in this research, detailing their specific applications.

Through this structure, the thesis combines the conciseness of a publishable scientific paper while retaining a comprehensive coverage of the material. The main paper provides a focused presentation of the core research, while the appendices offer methodological depth, detailed analyses and experimental considerations, allowing readers to engage with the research at different levels of detail.

1
Scientific Paper

Perceived Comfort and Safety in Automated Driving based on Physiological Signals: Findings from a Proving Ground Study

Jurjen Scharringa[1,2], Konstantinos Gkentsidis[2], Mathieu Sarrazin[2], Riender Happee[1], Karl Janssens[2]

[1]Delft University of Technology, Netherlands
[2]Siemens Digital Industries Software, Leuven, Belgium

Abstract—**With the advancements in automated driving, there is an increasing focus on passenger comfort and safety, fueled by the desire to establish a unique user experience identity among automotive companies. This study investigates the potential of the Galvanic Skin Response (GSR) as a physiological marker for assessing user experience. For this purpose, a test study of 32 participants was performed by collecting GSR measurements and self-reported comfort and safety scores, using a Wizard-of-Oz setup on a closed test-track over repeated laps, alternating between distinct driving styles. Statistical analysis revealed the phasic maximum amplitude and peak count as the features most strongly correlating with both objective driving style and perceived comfort and safety ratings. The GSR measurements were also given as input for a predictive model for classifying the driving style, yielding an accuracy of 88.61%. General performance of the same model for perceived comfort and safety prediction on a five-point Likert scale was, however, notably lower, whereas participant-specific model calibration yielded substantially higher performance. This indicates that while the GSR consistently reflects physiological responses linked to perceived comfort and safety, the signal exhibits a strong inter-subject variability, highlighting the necessity of personalized calibration for accurate passenger-experience assessment.**

Index Terms—**Galvanic Skin Response, Automated Driving, Passenger, Perceived Comfort, Perceived Safety**

I. Introduction

Significant progress has been made in the development of automated driving (AD) over the past years, with AD offering various benefits, such as improved road safety, optimized traffic flow and improved mobility to those hindered from driving, while raising new questions about passenger comfort and safety [1], [2]. Specifically, with the role of the driver shifting to that of the passenger in highly automated driving, defining comfort becomes increasingly complex [3].

Prior to the introduction of autonomous vehicles, studies on automotive comfort were predominantly focused on in-car ergonomic factors such as vibration, noise, temperature and air quality [4], which were largely quantifiable through objective measures. Now, however, complex psychological cognitive and emotional factors are introduced to comfort, such as trust, perceived safety, anxiety and stress, as well as physiological responses such as motion sickness [5]–[7]. While a well-agreed-upon definition of comfort is lacking in the scientific community, there is a broad consensus that comfort is an inherently subjective measure [8], [9].

Therefore, to study this comfort induced by various scenarios, it is necessary to obtain passengers' subjective feelings, which are typically gathered through interviews and scales. While these approaches are well-suited for long-term assessments [10], they suffer from limitations such as low efficiency, delayed feedback and inconsistent results across repeated measures [11]. To enable more consistent and scalable assessments, recent efforts have been made to link subjective perceptions of comfort to objective indicators. These include both vehicle dynamics measures, such as accelerations and jerks [9], as well as physiological signals that reflect emotional and cognitive states [6]. Physiological signals such as the Galvanic Skin Response (GSR), Heart Rate (HR), Heart Rate Variability (HRV) and Electroencephalogram (EEG) offer insights into passengers' physiological state and have been increasingly explored as potential markers for comfort in automated driving context [11].

This study contributes to a more nuanced understanding of comfort by investigating how physiological arousal, measured via the GSR signal, together with vehicle dynamics and perception data, can reflect and predict passenger comfort and perceived safety.

The remainder of this paper has the following structure: Section II provides a brief explanation of the GSR signal, followed by a review of related works. Section III describes the experiment, collected data and processing. Section IV presents the results of both a statistical analysis and predictive modeling. Finally, section V discusses these findings and their limitations, and section VI concludes this research and outlines future work.

II. Related Work

GSR, or Electrodermal Activity (EDA), refers to the change in skin conductance due to activation of sweat glands triggered by the sympathetic nervous system in response to emotional arousal or stress. The signal is measured in microSiemens (μS) using electrodes on the skin, and can be decomposed into a tonic and phasic component [12], [13].

The tonic component represents the baseline of the GSR and is also referred to as the tonic Skin Conductance Level (SCL). It changes slowly over time, varies widely over different subjects and increases when stimulation is introduced and gradually decreases in resting periods. The phasic component captures rapid fluctuations, small waves superimposing the tonic signal, known as Skin Conductance Responses (SCRs). They can occur without an identifiable stimulus (non-specific SCRs), occurring at a frequency of 3/min at rest and 20/min during high arousal, or in response to a stimulus (event-related SCRs) [13], [14].

Given the physiological basis of the GSR and its relation to arousal and stress, researchers have explored its relationship with perceived comfort and related states. Dillen et al. (2020) identified GSR as a significant predictor of passenger comfort and anxiety, using a comfort and anxiety rating scale [6]. Similarly, Meng et al. (2024) found two features derived from the GSR signal to significantly correlate with comfort, measured through a discomfort throttle [11]. Radhakrishnan et al. (2020) also reported a significant positive correlation between the frequency of discomfort button presses and the number of SCRs per minute [15]. In the context of motion sickness, Irmak, Pool and Happee (2020), Schneider et al. (2022), Tan et al. (2022), Wagner-Dougles et al (2024) and Xiang et al (2024) all found significant correlations between GSR metrics and motion sickness, reported through scales [5], [16]–[19]. However, a similar effect was not found by Henry et al. (2023), who did not find a significant correlation with GSR and subjective motion sickness [20]. Other subjective states, such as trust, have also been linked to GSR activity. Walker et al. (2019), Ajenaghughrure, da Costa Sousa, Lamas (2020), and Mühl et al. (2019) reported significant correlations, although Mühl et al. found this effect only in real-world driving, not in driving simulators [2], [21], [22]. Despite these promising outcomes, several studies report inconsistencies. Beggiato et al. (2019) found no significant correlation between GSR and discomfort [8]. Likewise, Niermann and Lüdtke (2020) and Smyth et al. (2021) reported a link between GSR and their respective subjective measure, motion sickness and stress, but questioned the reliability and suggested further research [7], [23].

While many studies report significant correlations between GSR and various subjective states, the findings remain inconsistent across contexts, measures and setups. This paper further investigates GSR as a physiological marker capturing perceived comfort and safety in automated driving.

III. Method

A. Experiment

The experiment was performed on a closed test track at the Griesheim proving ground of the TU Darmstadt, Germany (Figure 1), employing a Wizard-of-Oz Autonomous Vehicle (AV) approach with a human driver controlling the vehicle. An expert human driver was recruited to perform all the maneuvers of this experiment consistently with the assistance of adaptive cruise control functionalities (ACC), ensuring repeatability across all trials. The ego vehicle, i.e. the Vehicle Under Test (VUT) in this experiment was a KIA EV6 car. A second manual-driven vehicle, namely the Global Vehicle Target (GVT) in this test, was also present, interacting with the VUT among the different scenarios.

Following the preparation and the start, five distinct test scenarios were conducted. All scenarios were performed following the specifications of the Euro NCAP test protocol for crash avoidance systems [24]. The first scenario, Car-to-Pedestrian Turning Adult (CPTA), featured a pedestrian crossing without visual obstruction. This was followed by a roadwork scenario, a pedestrian crossing with visual obstruction, Car-to-Pedestrian Nearside Child Obstructed-50 (CPNCO-50), a cut-in maneuver (ACC Cut-in CCR) and lastly a car-following scenario (CCRB). Both pedestrian-related scenarios used a real human as pedestrian, while the last two scenarios involved the presence of the manually driving GVT. Each scenario was confined to a 30-second window. After completing one lap, the sequence was repeated three times, resulting in four total laps driven. The first and third laps were driven calm, and the second and fourth were driven aggressive. These two driving styles, defined by parameters such as velocity, acceleration and steering dynamics, were used to evoke different responses. Their characteristics per scenario are summarized in Table I. During the first and fourth laps, the pedestrian did not cross in the CPNCO-50 scenario, introducing an element of unpredictability.

B. Participant Data

For this experiment, 32 participants (17M, 15F) aged 18 to 82 years ($M = 49.4, SD = 21.1$) participated in the study. During the experiment, both self-reported scores and physiological responses were recorded. Each scenario was followed by a complete stop of the VUT and a questionnaire with three questions in which the participant evaluated the scenario on perceived comfort and on a five-item Likert scale ranging from "very uncomfortable/unsafe" to "very comfortable/safe", with the exception of scenario 4 and 5, which were presented consecutively without intermediate stop. The following three questions were asked:

1) How safe did you feel during the car ride?
2) How safe did you feel interacting with the [pedestrian, roadworks, pedestrian, vehicle][1]?
3) How comfortable did you find the movement of the vehicle?

The first question measures general perceived comfort, focusing on how the vehicle's behavior around the encountered object and scenario affects the passenger. The second question targets object-specific perceived safety. The third question measures the overall ride comfort related to the vehicle's motion. Before the experiment, participants completed a pre-questionnaire assessing their trust in AV and susceptibility to motion sickness. Upon completing the experiment, they reported their willingness to adopt an AV featuring one or both of the presented driving styles.

[1] The specific object corresponds to the scenario as described: pedestrian for the 1st and 3rd scenarios, roadworks for the 2nd, and vehicle for the 4th.

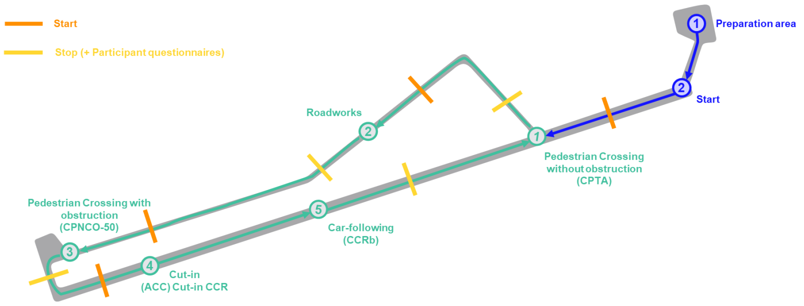

Start
Stop (+ Participant questionnaires)
1
Preparation area
2
Start
Roadworks
2
1
Pedestrian Crossing without obstruction (CPTA)
Pedestrian Crossing with obstruction (CPNCO-50)
3
5
Car-following (CCRb)
4
Cut-in (ACC) Cut-in CCR

Fig. 1. Overview of the five test scenarios, with Table I presenting the driving characteristics.

TABLE I
DRIVING CHARACTERISTICS PER DRIVING STYLE PER SCENARIO.

Scenario / Parameter	Ped. crossing without obstruction		Roadworks		Ped. crossing with obstruction		Cut-in		Car-following	
	Calm	Aggressive	Calm	Aggressive	Calm	Aggressive	Calm	Aggressive	Calm	Aggressive
VUT Target velocity (km/h)	30	50	30	70	50	70	50	70	30	50
GVT Target velocity (km/h)	-	-	-	-	-	-	30	50	-	-
Max. longitudinal acc. (m/s^2)	1.5	5	1	4	2	6	3	4	0.5	1
Max. lateral acc. (m/s^2)	2	7	1	5	2	8	4	8	0.5	0.5
Distance to obj. (m)	-	-	20[a]	10[a]	-	-	-	-	40[b]	12[b]

[a]distance to road construction works, [b]distance to GVT.

The test-vehicle was equipped according to the AV measurement framework developed by Devriendt et al. [25]. The GSR was continuously recorded using a Mind Media Nexus | 10 MKII device via electrodes on the index and middle finger at 32 Hz, while the vehicle dynamics were captured at 200 Hz. One participant's GSR data was lost due to a technical issue, and among the remaining 31 participants, six laps were excluded due to measurement errors (e.g. values exceeding 200μS), resulting in 118 valid laps for analysis. These yielded 590 time series, each corresponding to an instance of one of the five scenarios, used in the subsequent analysis.

C. Signal Processing

1) GSR: Preliminary analysis showed a systematic linear increase in GSR values across consecutive laps for all participants, indicating either temporal sensitivity of the GSR signal, cumulative influence of external factors, or a technical baseline drift – a pattern also reported in prior GSR and comfort-related studies [5], [8], [26]. To correct for this trend, a linear detrending procedure was applied by subtracting a least-squares fitted straight line from the GSR signals, following the same preprocessing approach used in prior work [6]. The signal was then decomposed into its tonic and phasic components using the cvxEDA algorithm implemented in Neurokit2 [27], [28]. For later feature analysis, peak detection is applied to the phasic component using a threshold of 0.03μS [12]. Figure 2 illustrates a decomposed GSR signal for a single participant during two laps for both driving styles, with annotated regions indicating the time intervals corresponding to each scenario, showcasing differences in physiological responses between the two driving styles. Table II shows the extracted features from both the phasic and tonic components of the GSR signal used in further analysis [13], [14]. These features were selected by a lightweight screening by removing one from each highly correlated pair, reducing dimensionality and improving interpretability while preserving informative variance.

2) Vehicle Dynamics: To suppress the high-frequency noise of the linear acceleration measurements, a 5th-order Butterworth low-pass filter with a cut-off frequency of 1 Hz was applied. The jerk was computed via numerical differentiation from the filtered acceleration data. Finally, all VD data were down-sampled to a sampling rate of 32 Hz to match the GSR signal and aligned with the GSR signal using the UNIX timestamps.

3) Perception: Perception was recorded using a forward-facing camera at 10 Hz and interpolated to 32 Hz to match the GSR signal. Time-to-collision (TTC) was computed with the distance to the detected object and the relative speed, while time-headway (THW) was calculated as the distance to the object divided by the VUT's velocity. When no objects were detected, high-threshold padding was applied to indicate no immediate collision risk. Two additional signals were added to the time series: one binary signal indicating whether data points were padded or not, and another encoding the type of detected object.

IV. EXPERIMENTAL ANALYSIS

A. Statistical Analysis

A correlation matrix based on Linear Mixed-Effects (LME) models was built to obtain a comprehensive overview of the relationships among all study variables. For this analysis, all features listed in II were derived over 30-second periods for each event, in each individual, along with the maximum velocity, accelerations, jerk and yaw rate, and minimum distance to

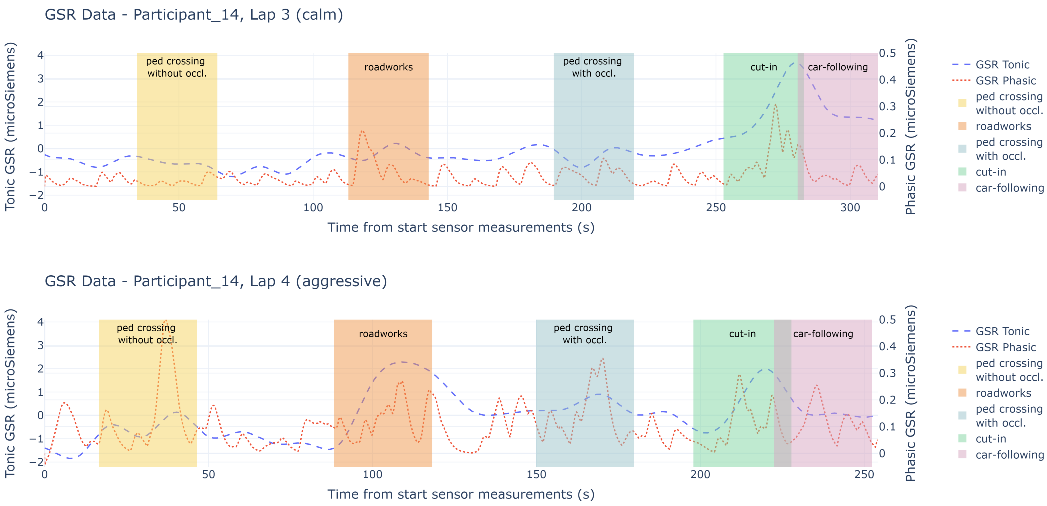


Fig. 2. Time series plot of phasic and tonic components for a single participant during two laps for both driving styles, annotated with regions indicating the time interval of each distinct test scenario.

TABLE II
DESCRIPTION OF GSR FEATURES DERIVED FOR EACH 30-SECOND SCENARIO THAT ARE USED IN FURTHER ANALYSIS.

Features	
Phasic	
maximum	The maximum value of the signal
minimum	The minimum value of the signal
mean td.	Mean time derivative of the signal
slope	Linear trend of the signal
peak count	Amount of phasic peaks in period
rise time	The duration required for the signal to increase from onset to peak
recovery time	The duration required for the signal to decline from its peak towards its baseline
Tonic	Description
mean	The average value of the signal
standard dev.	The variability of the signal
mean td.	Mean time derivative of the signal
skewness	Degree of asymmetry in the signal distribution
kurtosis	Degree of flatness in the signal distribution

the detected object and time-to-collision. Qualitative variables were numerically encoded: driving style was coded as “0” for calm and “1” for aggressive, while questionnaire responses were coded from “1” for “very uncomfortable/safe” to “5” for “very comfortable/safe”.

This statistical analysis aims to assess the pairwise relationship among all variables, including driving style, questionnaire responses, GSR features and VD features. For this purpose, LME models were specifically employed, as they are well-suited to handling the repeated measures structure of this experiment (i.e., four laps times five scenarios per participant). Repeated measures within the same participant typically violate the independence assumption, as one participant might be more expressive by responding “very uncomfortable” and “very comfortable” whereas another might be more nuanced by responding “neutral” and “comfortable” on the same question. Similarly, GSR responses are not independent within individuals. Standard methods (e.g. ANOVA, Spearman’s correlation) rely upon this assumption of independence and are therefore inappropriate for this correlation test, as ignoring this within-subject correlation inflates the Type I error rate. LME models explicitly account for the non-independence of data originating from the same participant by incorporating random effects (e.g. random intercepts and/or slopes per participant). This allows one to capture individual variation in baseline responses while still estimating fixed effects of interest, providing a more robust and accurate framework for statistical analysis.

The results of this statistical analysis are shown in Table III in a correlation matrix. Each cell in this matrix represents the outcome of a separate LME model, fitted with one feature as the dependent variable and the other feature as fixed effect, while including a participant-specific random intercept and slope to account for individual baseline and variability in the strength of the relationship. Because the random effects already capture the inter-subject differences in baseline and sensitivity, no additional scaling of the GSR data was required. To account for the cumulative inflation of Type I error rate due to the multiplicity of the analysis, p-values were corrected using the Benjamini-Hochberg False Discovery Rate procedure, which controls the expected proportion of false discoveries among the rejected hypotheses.

Several strong and statistically significant correlations emerge from the results of Table III. Driving style shows a strong negative correlation with all three questionnaire responses, indicating that participants generally reported lower scores for the scenarios driven in the aggressive laps. The positive correlations between driving style and VD features confirm the expected physical differences in driving styles.

TABLE III

Correlation matrix of all features (driving style, questionnaire responses, GSR, VD and perception). Correlations and regression coefficients were determined using LMEs. Underlined entries mark non-significant correlation ($p > 0.05$).

		1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.	22.	23.	24.
1.	style	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
2.	Q1[a]	-.74	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
3.	Q2[b]	-.69	.77	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
4.	Q3[c]	-.79	.79	.64	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
5.	phasic_max	.65	-.36	-.37	-.43	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
6.	phasic_min	.43	-.31	-.36	-.30	.50	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
7.	phasic_mean_td	$\underline{.08}$	$\underline{.00}$	$\underline{-.00}$	$\underline{-.06}$	.10	$\underline{-.09}$	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
8.	phasic_slope	.21	$\underline{-.08}$	$\underline{-.08}$	-.12	.30	$\underline{.03}$	.64	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
9.	phasic_peak_count	.59	-.47	-.51	-.46	.68	.67	.12	$\underline{.18}$	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
10.	phasic_rise_time	.35	-.18	-.15	-.20	.45	.22	.18	.32	.22	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
11.	phasic_recovery_time	.31	-.16	-.14	-.19	.33	.22	$\underline{.12}$	.28	.40	.77	·	·	·	·	·	·	·	·	·	·	·	·	·	·
12.	tonic_mean	$\underline{.05}$	-.14	-.13	$\underline{-.07}$	.14	.29	-.28	-.28	.17	$\underline{-.05}$	$\underline{.03}$	·	·	·	·	·	·	·	·	·	·	·	·	·
13.	tonic_std	.30	-.22	-.26	-.25	.45	.38	$\underline{.00}$	.17	.33	.21	.16	.19	·	·	·	·	·	·	·	·	·	·	·	·
14.	tonic_mean_td	.30	-.17	-.24	-.19	.49	.38	.17	.29	.37	.35	.24	$\underline{.02}$	.47	·	·	·	·	·	·	·	·	·	·	·
15.	tonic_skewness	$\underline{.05}$	$\underline{-.05}$	$\underline{.02}$	$\underline{-.03}$	$\underline{-.06}$	-.15	.22	.35	$\underline{-.04}$	$\underline{.09}$	$\underline{.03}$	-.20	$\underline{-.05}$	$\underline{-.09}$	·	·	·	·	·	·	·	·	·	·
16.	tonic_kurtosis	$\underline{.02}$	$\underline{-.04}$	$\underline{.03}$	$\underline{-.02}$	$\underline{-.03}$	$\underline{-.07}$	.10	$\underline{.03}$	$\underline{-.03}$	$\underline{.02}$	$\underline{.01}$	$\underline{.01}$	-.24	-.11	.39	·	·	·	·	·	·	·	·	·
17.	vel_max	.61	-.42	-.47	-.44	.56	.46	$\underline{.07}$	.19	.41	.27	.25	.31	.50	.46	$\underline{-.03}$	$\underline{-.04}$	·	·	·	·	·	·	·	·
18.	acc_lon_max	.90	-.56	-.53	-.61	.58	.38	$\underline{.10}$	.24	.38	.32	.29	$\underline{.06}$	.28	.30	$\underline{.07}$	$\underline{.02}$	.67	·	·	·	·	·	·	·
19.	acc_lat_max	.35	-.25	-.36	-.26	.37	.35	$\underline{.09}$	.22	.26	.17	.15	$\underline{.06}$	.46	.43	$\underline{-.03}$	$\underline{-.10}$	.67	.40	·	·	·	·	·	·
20.	jerk_lon_max	.81	-.49	-.41	-.54	.53	.36	$\underline{.10}$	.21	.33	.31	.22	$\underline{-.04}$	.19	.30	$\underline{.05}$	$\underline{.02}$	.47	.83	.15	·	·	·	·	·
21.	jerk_lat_max	.63	-.41	-.44	-.45	.65	.49	.13	.21	.39	.31	.25	.12	.46	.51	$\underline{-.07}$	$\underline{-.09}$	.78	.60	.70	.53	·	·	·	·
22.	yaw_rate_max	.40	-.19	-.13	-.25	.45	.24	.21	.26	.24	.29	.18	-.19	.16	.37	$\underline{-.02}$	$\underline{-.01}$	.34	.36	$\underline{.13}$	.53	.59	·	·	·
23.	d_min	.29	-.32	-.35	-.27	$\underline{.10}$	$\underline{.07}$	$\underline{-.10}$	$\underline{-.03}$	$\underline{.10}$	$\underline{.04}$	$\underline{.08}$	.12	$\underline{.14}$	$\underline{-.02}$	$\underline{.03}$	$\underline{-.00}$	.21	.30	.30	$\underline{.08}$	$\underline{.09}$	-.27	·	·
24.	ttc_min	.31	-.34	-.33	-.26	$\underline{-.11}$	$\underline{-.06}$	-.31	-.22	$\underline{-.01}$	$\underline{-.07}$	$\underline{.01}$	$\underline{.21}$	$\underline{-.05}$	-.24	$\underline{-.01}$	$\underline{.05}$	$\underline{-.02}$	.28	$\underline{-.06}$	$\underline{.15}$	-.27	-.51	.91	·

[a] How safe did you feel during the car ride?

[b] How safe did you feel interacting with the [pedestrian, roadworks, pedestrian, vehicle]?

[c] How comfortable did you find the movement of the vehicle?

Both phasic and tonic features show a systematic correlation with driving style and questionnaire scores, suggesting that the physiological arousal tracks both the nature of driving and the participants' reported scores. The maximum longitudinal acceleration and lateral jerk show the highest correlation among VD features to the GSR features, hinting at the physiological sensitivity of these specific dynamic aspects. Finally, the perception features showed a correlation to the subjective responses, but not to the physiological responses.

When the same analysis was conducted separately for each driving style – thus eliminating any variance in driving style – nearly all observed significant correlations largely disappeared. These findings suggest that the driving style may be the primary factor shaping perceived comfort and safety, with participants showing relatively consistent GSR patterns and comfort ratings within each driving style.

Further analysis using pairwise t-tests (averaging responses per participant to ensure independence) revealed consistent significant decreases in comfort and safety scores when transitioning from calm to aggressive driving style (all $p < 0.001$), with the roadwork scenario exhibiting the most pronounced decrease. These subjective differences were reflected by increased phasic activity across all scenarios, while an elevated tonic activity was only observed in the pedestrian crossing without obstruction and roadwork scenarios. Comparing scenario types, participants reported significantly higher comfort ($p < 0.01$) and safety ($p < 0.001$) scores in pedestrian crossing scenarios compared to those involving the GVT. This observation was only done when comparing these scenarios in general to each other; cross-style comparisons showed lower ratings for the aggressively-driven scenario, regardless of whether it involved the pedestrian or the GVT. Visual obstruction or whether or not the pedestrian crossed the road did not significantly affect subjective or physiological responses. Age-related comparisons revealed that the youngest group (18-34) reported the lowest perceived safety ($p < 0.0001$) and had the highest phasic responses ($p < 0.0001$), potentially reflecting their lower driving experience and a more reactive sympathetic nervous system. No significant differences were found for gender ($p > 0.05$). Lastly, participants who reported a lower trust in AV and those who would only adopt an AV with the calm driving style or would not adopt one at all consistently gave lower subjective scores (all $p < 0.0001$), suggesting a potential pre-existing bias.

B. *Driving Style Modeling*

Table III demonstrates a statistically significant correlation between driving style and GSR signal. To further explore this relationship, the next step is to assess the predictive power of the GSR signal for the driving style classification. Whereas the previous statistical analysis focused on the features extracted from the 30-second time series of each event, the current analysis considers the 30-second time series as a whole. For this purpose, a deep learning (DL) approach was pursued by using a modified version of the Time Evidence Fusion Network (TEFN; Zhan et al., 2024), specifically adapted to the requirements of the time series classification task. TEFN matches or exceeds current state-of-the-art architectures in forecasting accuracy, while relying on far fewer parameters [29], making the model a well-suited choice for this use case, as its parameter efficiency aligns with the constraints of the limited dataset and with its demonstrated effectiveness on time series.

For the binary classification task, the model used both phasic and tonic components of the GSR signal as input for each scenario, with each input channel having a fixed sequence length of 960. Each GSR input channel was z-standardized ($\mu = 0, \sigma^2 = 1$) per participant to reduce inter-subject variability, where the mean and standard deviation were computed across all laps of that participant. Data from

21 participants were used for training, with an additional 5 participants reserved for validation during training and a final 5 participants for testing. Training was conducted for a maximum of 500 epochs with a learning rate of $1e-4$, using a dropout of 0.3 to regularize learning and an early stopping mechanism based on the validation loss with a patience of 50 epochs to prevent overfitting on the limited dataset.

To assess the model's robustness and generalization, despite the limited dataset, a 10-fold cross-validation was employed, ensuring that the performance metrics reflect the variability across different participant groupings and reduce potential bias from a single train-validation-test split.

Across all folds, the model achieves the following performance metrics ($M \pm SD$): 88.61% ± 3.77 accuracy, 87.73% ± 4.75 precision, 89.70% ± 8.86 recall and 88.61% ± 3.98 F1-score. The aggregated Receiver Operating Characteristic (ROC) curve yields an average Area Under the Curve (AUC) of 0.938, indicating a strong discriminative ability to distinguish driving styles.

C. Subjective Score Modeling

The preceding section examined the predictive power of the GSR signal for the driving style classification. Next, the analysis is extended to predicting the perceived comfort and safety responses, shifting the task from binary to multi-class classification, with five classes corresponding to each of the response options in the questionnaire. The same DL approach with the same TEFN architecture is used as in the preceding section, only changing the loss function to a *Cross-Entropy* loss function fit for the multi-class classification task. Each model is evaluated under varying input configurations, consisting of one or a combination of the following data sources:

1) GSR signal: Phasic and tonic components (z-standardized per participant).
2) Vehicle Dynamics (VD): Velocity, longitudinal and lateral acceleration and jerk, and yaw rate.
3) Perception: distance to object, time-to-collision, time-headway and type of object.

A key challenge in this task, however, arose from the skewed distribution of predominantly positive questionnaire responses, as illustrated in Figure 3. This class imbalance complicated direct supervised training on the subjective labels due to the limited exposure to the negative responses. Furthermore, it is notable that various participants report feeling "very comfortable/safe" for some scenarios under the aggressive driving style. This raises the concern of whether these responses were genuine, potentially reflected by low physiological arousal, or whether social desirability or other biases led the participants to give more positive responses despite heightened arousal levels.

To mitigate the class imbalance, a Synthetic Minority Oversampling Technique (SMOTE) approach was employed. By generating synthetic samples of the minority classes by interpolating between existing samples, SMOTE improves the model's performance on the imbalanced dataset [30]. In this study, SMOTE was applied with $k = 5$ nearest neighbors to generate the synthetic samples.

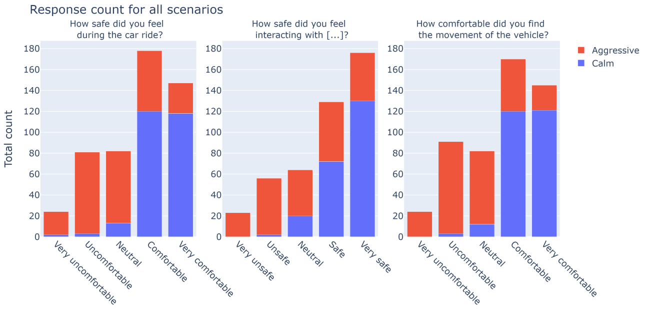


Fig. 3. Distribution of all self-reported scores during the experiment.

To comprehensively evaluate model performance, a two-fold approach is applied. First, performance metrics are reported based on an *exact match* criterion (hard), considering only predictions that perfectly match the true label class. Second, a *near-fit* criterion (soft) is introduced, where predictions within one class of the ground truth (e.g. predicting "comfortable" instead of "very comfortable") are also considered true positive. This soft criterion addresses the inherent subjectivity of self-reported scores, where fine-grained distinctions are difficult for participants to make and less critical in practical applications. Table IV presents the results for both evaluation strategies for each questionnaire item using 10-fold cross-validation.

D. User-Adapted Subjective Score Modeling

The model effectively minimized training loss; however, its relatively low evaluation scores in Table IV indicate limited generalization to data from unseen participants. This outcome is not surprising, considering the inherently subjective nature of perceived comfort and safety and the inter-subject variability in physiological responses.

To mitigate this, a user-adapted modeling approach was subsequently explored. The model trained on the original training set was used as a general model. For each test participant, a copy of this model was created and fine-tuned. During fine-tuning, each participant's data was split into N support scenario pairs (each with a calm and aggressive variant of the same scenario) and $10 - N$ query pairs for evaluation. Fine-tuning involved updating only the final projection layer of the model,- while freezing the rest of the parameters, over 20 epochs at a learning rate of $1e-4$. This approach allowed the model's output mapping to adapt to participant-specific data while preserving the shared feature representations learned during pre-training.

Table V presents the performance results, showing both hard accuracies and soft accuracies across all input configurations for support set sizes ranging from $N = 0$ to $N = 9$ in the fine-tuning process.

TABLE IV
MACRO PERFORMANCE METRICS ($M\% \pm SD\%$) OVER 10-FOLD CROSS VALIDATION IN THE SELF-REPORTED SCORE CLASSIFICATION ON SELF-REPORTED SCORE FOR COMFORT (Q1, Q3) AND PERCEIVED SAFETY (Q2) MODEL BASED ON VARIOUS INPUT CONFIGURATIONS. RESULTS ARE SHOWN AFTER APPLYING SMOTE TO ADDRESS CLASS IMBALANCE.

		Hard				**Soft**			
		Accuracy	**Precision**	**Recall**	**F1 score**	**Accuracy**	**Precision**	**Recall**	**F1 score**
		Q1: How safe did you feel during the car ride?							
Baseline		38.7 ± 5.4	8.2 ± 5.9	20.0 ± 0.0	11.7 ± 1.8	75.5 ± 9.4	49.1 ± 13.5	60.0 ± 0.0	52.0 ± 13.4
Ours	GSR	30.3 ± 4.6	24.4 ± 6.0	26.6 ± 4.8	23.9 ± 5.0	69.7 ± 6.5	61.7 ± 5.6	62.6 ± 5.2	60.0 ± 6.0
	VD	39.4 ± 6.3	35.2 ± 5.6	36.8 ± 6.4	33.4 ± 6.6	82.3 ± 5.1	74.9 ± 7.5	78.0 ± 6.7	74.6 ± 7.4
	GSR+VD	34.5 ± 4.5	33.4 ± 4.4	37.2 ± 3.5	30.4 ± 2.9	80.5 ± 5.1	72.9 ± 7.6	76.4 ± 8.1	72.5 ± 8.1
	GSR+P	35.4 ± 8.1	33.0 ± 7.5	34.3 ± 9.7	29.5 ± 7.5	74.2 ± 8.6	69.4 ± 9.3	72.6 ± 9.7	68.1 ± 10.0
	VD+P	42.2 ± 5.2	38.0 ± 5.6	38.7 ± 4.4	34.7 ± 5.2	81.0 ± 6.8	73.3 ± 8.7	76.9 ± 6.9	73.4 ± 8.7
	GSR+VD+P	41.2 ± 5.4	36.9 ± 6.9	40.4 ± 5.7	35.0 ± 6.5	79.3 ± 6.0	74.1 ± 9.0	77.5 ± 8.1	73.6 ± 9.3
		Q2: How safe did you feel interacting with the ...[a]							
Baseline		44.9 ± 10.5	9.1 ± 2.0	20.0 ±0.0	12.5 ± 1.8	69.9 ± 8.8	36.7 ± 7.7	60.0 ± 0.0	39.6 ± 7.7
Ours	GSR	35.1 ± 5.1	25.2 ± 4.4	25.2 ± 5.0	23.4 ± 4.2	67.5 ± 8.2	60.0 ± 5.9	62.5 ± 6.3	58.6 ± 6.3
	VD	35.6 ± 3.4	29.2 ± 6.3	30.3 ± 5.6	25.8 ± 4.1	77.2 ± 8.2	71.0 ± 8.9	74.6 ± 7.2	70.0 ± 9.0
	GSR+VD	37.3 ± 6.8	31.4 ± 4.2	32.7 ± 6.3	27.7 ± 4.3	75.3 ± 8.2	69.8 ± 8.9	73.4 ± 9.2	68.7 ± 9.2
	GSR+P	37.6 ± 8.4	26.5 ± 5.5	27.3 ± 6.2	24.9 ± 5.3	74.8 ± 3.1	67.4 ± 6.0	71.9 ± 5.5	67.1 ± 3.9
	VD+P	38.1 ± 7.7	26.1 ± 7.4	29.3 ± 8.7	24.0 ± 6.1	74.7 ± 10.2	73.4 ± 8.9	73.7 ± 8.9	69.7 ± 9.6
	GSR+VD+P	37.5 ± 6.3	29.3 ± 6.5	30.5 ± 5.6	26.6 ± 4.6	76.0 ± 7.8	71.9 ± 8.2	74.2 ± 8.4	70.6 ± 8.9
		Q3: How comfortable did you find the movement of the vehicle?							
Baseline		35.3 ± 3.4	7.3 ± 0.8	20.0 ±0.0	10.7 ± 1.0	73.0 ± 13.0	45.2 ± 13.6	60.0 ± 0.0	48.1 ± 13.2
Ours	GSR	28.8 ± 4.1	26.8 ± 5.8	25.5 ± 4.6	22.6 ± 3.9	73.5 ± 4.8	67.0 ± 4.0	70.1 ± 6.1	65.4 ± 5.5
	VD	33.8 ± 6.5	30.9 ± 6.5	32.6 ± 7.2	27.6 ± 76.4	80.7 ± 6.5	73.9 ± 7.6	78.9 ± 8.8	72.8 ± 9.2
	GSR+VD	33.3 ± 3.8	30.8 ± 5.9	32.6 ± 7.2	27.6 ± 4.3	81.5 ± 7.4	76.8 ± 9.1	82.3 ± 8.9	76.3 ± 10.4
	GSR+P	29.2 ± 3.4	27.6 ± 3.0	26.9 ± 4.4	23.8 ± 3.0	73.6 ± 3.4	69.5 ± 3.5	72.6 ± 7.3	66.7 ± 4.0
	VD+P	36.5 ± 4.5	34.4 ± 5.2	33.3 ± 3.9	27.7 ± 4.5	79.6 ± 7.7	75.3 ± 6.5	78.8 ± 8.1	72.7 ± 7.8
	GSR+VD+P	32.2 ± 5.1	32.4 ± 7.8	33.2 ± 7.4	27.6 ± 6.5	79.0 ± 5.7	75.5 ± 7.8	81.9 ± 5.5	73.5 ± 5.5

[a]: [pedestrian, roadworks, pedestrian, vehicle]

V. DISCUSSION

The purpose of this study was to explore the relationship between perceived comfort and safety in highly automated driving and physiological signals, specifically focusing on the GSR, following the results from a proving ground study. Table III presents a comprehensive correlation matrix based on Linear Mixed-Effect (LME) models covering all key variables: driving style, self-reported scores on perceived comfort and safety, GSR features, vehicle dynamics (VD) and perception features. This all-in-one approach revealed several significant correlations across domains. Among the GSR features, the phasic maximum amplitude and peak count emerged as the most robust indicator, showing the strongest correlation with both objective driving style and perceived comfort and safety. GSR activity tends to increase as perceived comfort and safety decrease, an outcome found in similar studies [6], [11], [15]. Furthermore, a stronger observed link for the maximum longitudinal acceleration and lateral jerk hints at a heightened sensitivity of GSR to abrupt vehicle movement.

Comparative analysis of scenarios revealed that car-related scenarios consistently exhibited lower comfort and safety scores, accompanied by elevated GSR responses. Notably, the roadwork scenario demonstrated the most dramatic shift, transitioning from the most comfortable and safe rated scenario under the calm driving style to the least comfortable and safe under the aggressive driving style. These scenarios are characterized by significantly higher vehicle dynamics than the pedestrian crossing scenarios (Table I). This pattern suggests that decreased comfort and increased physiological arousal are either primarily driven by elevated vehicle dynamics, by the perceived risk associated with the presence of another vehicle, or by a combination of both factors.

Building on these findings, a deep learning approach was employed, in parallel with the statistical analysis, for testing the predictive power of the GSR signal on different driving styles. Using the entire time series data as input, the model achieved 88.61% accuracy, indicating that the GSR demonstrates potential as a non-invasive indicator for assessing driving style. To our knowledge, no previous study has leveraged passenger GSR for a driving style classification, making our 88.61% accuracy a benchmark in this domain.

Before extending this framework to predict perceived comfort (Q1), perceived safety (Q2), and overall ride comfort (Q3) on a five-point Likert scale, a previously introduced *near-fit* criterion (soft) was applied to treat adjacent-class predictions as correct, aligning with both participants uncertainty and the practical aim of broadly distinguishing comfort from discomfort.

Despite this soft criterion, general models using GSR alone or augmented with vehicle dynamics (VD) or perception data performed only modestly, slightly outperforming, and occasionally underperforming, a majority-class baseline classifier (Table IV). Introducing user-adapted models, fine-tuned using a partition of the test scenarios, significantly enhanced the results, with improvements of up to 25% in hard metric performances and up to 20% in soft metric performances compared to these general models.

The best configuration was found for a model using only GSR as input and fine-tune training on a support set size of

TABLE V

Hard and soft accuracy metrics ($M\% \pm SD\%$) over 10-fold cross validation for self-reported comfort (Q1, Q3) and perceived safety (Q2) classification using various input configurations and support set sizes (N). Results are shown for different support set sizes ($N = 0$ to $N = 9$) used in fine-tuning the model. SMOTE was applied to address class imbalance. Bold values indicate the best-performing configuration for $N \leq 5$.

Support Pairs	**Input config.**	**Hard accuracy**			**Soft accuracy**		
		Q1	*Q2*[a]	*Q3*	*Q1*	*Q2*[a]	*Q3*
	Baseline	38.7 ± 5.4	44.9 ± 10.5	35.3 ± 3.4	75.5 ± 9.4	69.9 ± 8.8	73.0 ± 13.0
$N = 0$	GSR	30.3 ± 4.6	35.1 ± 5.1	28.8 ± 4.1	69.7 ± 6.5	67.5 ± 8.2	73.5 ± 4.8
	VD	39.4 ± 6.3	35.6 ± 3.4	33.8 ± 6.5	82.3 ± 5.1	77.2 ± 8.2	80.7 ± 6.5
	GSR+VD	34.5 ± 4.5	37.3 ± 6.8	33.3 ± 3.8	80.5 ± 5.1	75.3 ± 8.2	81.5 ± 7.4
	GSR+P	35.4 ± 8.1	37.6 ± 8.4	29.2 ± 3.4	74.2 ± 8.6	74.8 ± 3.1	73.0 ± 3.4
	VD+P	42.2 ± 5.2	38.1 ± 7.7	36.5 ± 4.5	73.3 ± 8.7	74.7 ± 10.2	79.6 ± 7.7
	GSR+VD+P	41.2 ± 5.4	37.5 ± 6.3	32.2 ± 5.1	79.3 ± 6.0	76.0 ± 7.8	79.0 ± 5.7
$N = 1$	GSR	47.6 ± 10.0	55.2 ± 7.7	43.3 ± 8.9	85.8 ± 4.1	83.7 ± 5.1	84.1 ± 9.4
	VD	47.0 ± 8.5	57.5 ± 10.8	40.9 ± 10.0	83.5 ± 3.7	82.1 ± 8.5	79.4 ± 8.3
	GSR+VD	48.1 ± 7.7	57.8 ± 8.1	41.7 ± 8.3	86.3 ± 3.5	85.5 ± 6.4	81.5 ± 8.9
	GSR+P	44.8 ± 9.5	46.4 ± 9.6	44.5 ± 8.4	81.9 ± 4.6	77.1 ± 8.4	80.9 ± 8.7
	VD+P	48.4 ± 9.4	50.5 ± 8.6	43.4 ± 9.9	83.9 ± 4.3	83.4 ± 7.0	80.8 ± 7.8
	GSR+VD+P	49.5 ± 8.6	50.5 ± 8.6	43.8 ± 9.2	86.5 ± 4.4	82.3 ± 6.7	80.3 ± 8.9
$N = 2$	GSR	52.2 ± 6.3	56.5 ± 7.0	45.7 ± 5.0	84.7 ± 5.3	84.7 ± 6.1	82.6 ± 3.9
	VD	54.8 ± 5.4	56.3 ± 6.8	47.9 ± 9.2	85.5 ± 3.3	84.2 ± 6.2	85.3 ± 4.5
	GSR+VD	56.1 ± 7.3	59.3 ± 9.2	48.0 ± 8.7	89.9 ± 3.1	89.2 ± 4.1	87.8 ± 4.0
	GSR+P	49.7 ± 13.1	57.2 ± 8.4	46.0 ± 8.4	86.6 ± 4.1	81.9 ± 9.7	81.8 ± 8.7
	VD+P	46.8 ± 6.1	48.7 ± 12.1	45.6 ± 8.2	83.5 ± 5.1	85.7 ± 5.7	84.7 ± 8.9
	GSR+VD+P	47.6 ± 10.2	47.8 ± 11.6	44.6 ± 8.3	84.4 ± 3.9	86.4 ± 4.5	83.7 ± 7.8
$N = 3$	GSR	56.1 ± 6.1	60.9 ± 7.9	52.2 ± 6.2	86.3 ± 3.9	86.1 ± 5.2	86.1 ± 7.0
	VD	53.4 ± 6.6	57.0 ± 8.1	50.0 ± 5.8	86.2 ± 4.0	90.0 ± 3.9	86.0 ± 6.6
	GSR+VD	52.5 ± 10.8	58.5 ± 7.6	50.8 ± 5.8	88.4 ± 4.0	90.1 ± 4.2	87.0 ± 4.1
	GSR+P	50.1 ± 7.4	53.4 ± 8.9	47.6 ± 6.1	84.2 ± 5.2	82.3 ± 10.0	84.4 ± 5.5
	VD+P	46.4 ± 7.5	48.7 ± 11.2	41.9 ± 7.7	83.2 ± 6.1	84.0 ± 8.4	82.3 ± 8.8
	GSR+VD+P	47.5 ± 7.4	48.6 ± 12.1	42.2 ± 7.5	83.6 ± 7.8	86.1 ± 6.9	84.3 ± 8.0
$N = 4$	**GSR**	**58.1 ± 5.3**	**58.4 ± 8.9**	**54.3 ± 8.4**	**88.5 ± 4.1**	**86.5 ± 4.7**	**90.1 ± 6.1**
	VD	53.3 ± 4.8	56.8 ± 12.8	49.2 ± 5.8	86.9 ± 4.3	88.9 ± 5.4	87.2 ± 4.8
	GSR+VD	50.8 ± 8.3	57.5 ± 10.8	48.3 ± 7.6	88.4 ± 7.6	92.3 ± 4.6	85.6 ± 4.6
	GSR+P	50.7 ± 8.0	51.6 ± 8.8	50.2 ± 8.3	82.3 ± 7.3	85.5 ± 7.5	85.8 ± 7.8
	VD+P	46.2 ± 7.7	48.5 ± 9.3	46.0 ± 8.2	82.6 ± 6.0	85.8 ± 6.6	80.4 ± 6.8
	GSR+VD+P	46.5 ± 8.8	49.7 ± 8.2	47.3 ± 8.3	85.4 ± 6.9	84.1 ± 6.1	84.2 ± 6.3
$N = 5$	GSR	56.2 ± 7.1	58.7 ± 8.4	54.2 ± 8.3	88.1 ± 5.4	87.9 ± 6.1	89.8 ± 6.2
	VD	50.0 ± 4.5	53.9 ± 13.1	45.2 ± 7.6	88.0 ± 3.3	86.3 ± 7.4	86.3 ± 4.2
	GSR+VD	46.7 ± 5.5	56.1 ± 14.0	46.5 ± 4.8	88.1 ± 4.7	87.0 ± 6.6	87.6 ± 3.6
	GSR+P	47.4 ± 7.1	53.5 ± 10.7	50.4 ± 8.3	79.6 ± 8.9	87.4 ± 5.2	86.8 ± 7.1
	VD+P	43.7 ± 7.6	46.8 ± 11.2	45.0 ± 6.8	81.7 ± 7.4	80.3 ± 11.2	81.3 ± 11.9
	GSR+VD+P	45.1 ± 8.9	42.5 ± 7.2	46.8 ± 7.7	84.6 ± 6.5	81.4 ± 9.8	87.9 ± 6.6
$N = 6$	GSR	56.4 ± 7.1	61.5 ± 11.0	63.5 ± 6.3	86.1 ± 5.6	86.0 ± 8.1	94.2 ± 4.0
	VD	47.5 ± 6.4	53.0 ± 11.0	44.2 ± 6.9	88.1 ± 3.9	86.3 ± 6.1	84.2 ± 3.9
	GSR+VD	39.7 ± 3.9	55.8 ± 12.8	47.9 ± 4.7	87.7 ± 5.9	83.2 ± 7.9	87.8 ± 4.3
	GSR+P	48.2 ± 8.6	57.7 ± 10.2	48.3 ± 8.1	83.8 ± 8.1	86.9 ± 4.8	86.6 ± 6.1
	VD+P	44.6 ± 5.3	52.4 ± 11.9	40.6 ± 5.7	82.6 ± 5.3	81.3 ± 8.6	79.4 ± 8.9
	GSR+VD+P	49.5 ± 6.4	47.9 ± 8.4	44.7 ± 6.1	82.9 ± 7.2	83.4 ± 11.4	85.9 ± 7.7
$N = 7$	GSR	57.7 ± 8.4	61.6 ± 10.3	57.5 ± 6.0	85.9 ± 7.0	84.9 ± 11.1	93.5 ± 4.3
	VD	40.9 ± 7.5	48.8 ± 12.0	46.1 ± 6.6	88.0 ± 5.8	82.9 ± 11.6	80.9 ± 4.8
	GSR+VD	32.8 ± 7.9	52.7 ± 10.2	46.6 ± 7.5	85.2 ± 6.5	87.5 ± 7.5	82.6 ± 5.7
	GSR+P	42.7 ± 13.3	59.6 ± 8.8	46.1 ± 12.7	80.8 ± 6.7	86.5 ± 7.3	86.1 ± 7.2
	VD+P	43.1 ± 9.5	52.3 ± 14.5	39.4 ± 9.0	79.3 ± 7.7	80.7 ± 7.4	79.9 ± 12.8
	GSR+VD+P	46.9 ± 10.6	49.5 ± 10.9	35.8 ± 9.2	81.6 ± 10.4	80.0 ± 15.3	83.7 ± 9.7
$N = 8$	GSR	60.2 ± 11.2	65.2 ± 8.8	58.9 ± 9.6	83.8 ± 8.5	89.0 ± 6.4	92.2 ± 5.2
	VD	35.3 ± 11.3	43.3 ± 14.9	38.6 ± 5.8	85.9 ± 6.8	80.0 ± 10.1	78.2 ± 9.1
	GSR+VD	33.0 ± 7.9	49.3 ± 16.8	46.3 ± 10.9	83.2 ± 12.0	77.3 ± 11.2	85.2 ± 6.4
	GSR+P	47.3 ± 11.0	58.1 ± 11.1	48.5 ± 11.4	78.5 ± 7.4	82.8 ± 11.0	85.9 ± 8.8
	VD+P	50.2 ± 11.0	42.8 ± 11.8	41.6 ± 9.2	78.1 ± 7.2	83.8 ± 10.3	83.8 ± 10.2
	GSR+VD+P	51.0 ± 12.6	49.0 ± 14.4	37.3 ± 9.6	78.8 ± 9.6	85.5 ± 9.8	80.9 ± 6.8
$N = 9$	GSR	58.7 ± 12.9	-	57.9 ± 9.2	82.2 ± 10.5	-	92.3 ± 5.8
	VD	37.8 ± 12.3	-	40.6 ± 6.6	82.1 ± 8.0	-	77.8 ± 9.3
	GSR+VD	34.5 ± 8.7	-	47.2 ± 9.1	81.8 ± 14.0	-	83.9 ± 5.6
	GSR+P	45.3 ± 11.8	-	47.8 ± 8.9	77.2 ± 8.5	-	84.8 ± 9.2
	VD+P	49.3 ± 12.0	-	40.7 ± 8.7	79.8 ± 9.0	-	81.9 ± 8.9
	GSR+VD+P	46.5 ± 10.2	-	36.8 ± 10.2	75.4 ± 9.5	-	78.9 ± 9.3

[a]: Support size to this question is limited to $N = 8$ as this question was not asked during laps 1, 4.

$N = 4$, achieving 58.1%, 58.4% and 54.3% hard accuracies and 88.5%, 86.5% and 90.1% soft accuracies for Q1, Q2 and Q3, respectively (Table V). While results for $N > 5$ support pairs were computed, only configurations with $N \leq 5$ were considered for the final configuration selection to ensure reliable performance assessment with the limited test set (10 scenario pairs). This configuration outperformed all general models and other personalized configurations combining VD and perception data, which tended to overfit when fine-tuned. This demonstrates that, while a one-size-fits-all model has limited capacity, participant-specific calibration robustly unlocks the GSR's full predictive power. This need for personalization aligns with the observed individual differences in subjective and physiological responses, particularly among the younger participants who reported lower perceived safety scores and exhibited heightened phasic activity.

Contextualizing these results with existing literature, the soft metric accuracy exceeds previous binary classification approaches for trust (78.2%, [31]), motion sickness (77%, [26]), comfort (71.9%, [32]) and stress (73%, [33]). In hard metric accuracy, it outperforms a 4-class classification on comfort (55.99%, [34]). The hard metric performance exceeds that of existing 4-class comfort classification (55.99%, [34]), but falls short of a 4-class motion sickness classifier (86%, [17]) and a 10-class comfort classifier, with a similar soft metric allowance (92.4%, [35]). Notably, these studies relied on multiple physiological input modalities, whereas the proposed configuration uses GSR as sole physiological marker. Furthermore, the presented approach advances state-of-the-art methodologies by introducing a participant-specific fine-tuning, bridging the gap between general models [17], [26], [31]–[35] and single-participant trained models [36].

Before pursuing the deep learning approach, a traditional feature-based machine learning approach using the GSR featurs from Table II was evaluated. Random Forests, XGBoost and Support Vector Machines were applied to both driving style and subjective comfort predictions. Driving style accuracy ranged from 64% to 71%, while subjective ratings reached 26% to 42%. Despite appearing comparable to Table IV, these models largely defaulted to majority-class predictions. Due to the underperformance on the more separable driving style task, this approach was not further pursued.

The GSR preprocessing pipeline was systematically evaluated across different approaches. While an initial baseline drift correction explored tonic component z-standardization [16], [37], this only addressed inter-lap drift. A linear detrending approach [6] was ultimately adopted to correct both inter- and intra-lap drift. For signal decomposition, cvxEDA [28] and Braithwaite's high-pass filtering [14] showed similar correlation patterns, with cvxEDA finally selected for its stronger physiological foundation compared to Braithwaite's purely numerical filtering approach. The scaling strategy evolved through three iterations: per-lap min-max normalization, which effectively removed relative differences between the calm and aggressive driven laps; per-participant min-max normalization, yielding 82.87% driving style classification accuracy; and finally per-participant z-standardization, further improving driving style classification accuracy to 88.61% and boosting comfort and safety prediction accuracies by 3-8%.

Based on these findings, this study recommends a preprocessing pipeline of (i) linear detrending of the raw GSR signal, (ii) cvxEDA decomposition, and (iii) participant-wise z-standardization. While this requires both low- and high-arousal data per user, this aligns with the necessary user-adapted fine-tuning for comfort and safety models. Alternatively, collecting a brief low-arousal baseline could enable real-time driving style classification, as deviations from this baseline indicate elevated arousal or anomalous events.

Two additional limitations are worth noting. First, while the exclusive use of the GSR as a physiological marker demonstrates the ability of a minimalistic approach, future work could benefit from incorporating additional physiological signals to potentially enhance predictive capabilities. A second limitation lies in the controlled proving ground environment, which may not fully capture real-world uncertainties and trust issues that passengers would experience, likely contributing to the imbalanced dataset favoring positive responses and potentially limiting generalizability.

Future research directions should focus on expanding the experimental scope. This includes scenarios that isolate perception and vehicle dynamics effects, specifically scenarios with identical vehicle dynamics but varying perception contexts, could help understand the relative impact of perception versus vehicle dynamics factors on passenger comfort. Additional valuable scenarios could include traffic jams, interaction with cyclists, parking maneuvers, and roundabout interactions. Furthermore, incorporating repeated measurements from the same participants over multiple days could provide insights into the temporal stability of GSR responses and individual reaction patterns. Ultimately, collecting real-world data on public roads would provide the most authentic insights into passenger comfort, while simultaneously posing significant challenges for analysis due to the unique nature of each driving situation and the difficulty in establishing a comparable baseline.

VI. Conclusion

Significant correlations emerged between GSR and all targets, including objective driving style, questionnaire responses and vehicle dynamics features, confirming the signal's broad sensitivity. GSR cleanly separates the calm from the aggressive driving style in our proving ground study, making it a reliable trigger for real-time style adjustment. While general models showed modest performance in predicting perceived comfort and safety, the introduction of participant-specific calibration substantially improved predictive accuracy. This highlights that physiological responses to automated driving are inherently individual, and accounting for these personal differences through personalized fine-tuning significantly enhances the GSR's predictive power. Future research should focus on implementing efficient personalization strategies in real-world applications and exploring more diverse driving conditions.

Acknowledgment

This research is supported by VLAIO (Flanders Innovation & Entrepreneurship Agency) within the BECAREFUL (Belgium consortium for enhanced safety & comfort perception of future autonomous vehicles) project, no. HBC.2021.0939 and by the European Union's Horizon Europe research and innovation program under grant agreement No. 101076754 – project AIthena.

References

[1] L. Vasile, N. Dinkha, B. Seitz, C. Däsch, and D. Schramm, "Comfort and safety in conditional automated driving in dependence on personal driving behavior," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 4, pp. 772–784, 2023.

[2] K. Mühl, C. Strauch, C. Grabmaier, S. Reithinger, A. Huckauf, and M. Baumann, "Get ready for being chauffeured: Passenger's preferences and trust while being driven by human and automation," *Human Factors*, vol. 62, no. 8, pp. 1322–1338, 2020, pMID: 31498656.

[3] C. Peng, C. Wei, A. Solernou, M. Hagenzieker, and N. Merat, "User comfort and naturalness of automated driving: The effect of vehicle kinematics and proxemics on subjective response," OSF Preprints, 2023.

[4] M. C. G. da Silva, "Measurements of comfort in vehicles," *Measurement Science and Technology*, vol. 13, no. 6, pp. R41–R60, 2002.

[5] T. Irmak, D. M. Pool, and R. Happee, "Objective and subjective responses to motion sickness: The group and the individual," *Experimental Brain Research*, vol. 239, no. 2, pp. 515–531, 2021.

[6] N. Dillen, M. Ilievski, E. Law, L. E. Nacke, K. Czarnecki, and O. Schneider, "Keep calm and ride along: Passenger comfort and anxiety as physiological responses to autonomous driving styles," in *CHI Conference on Human Factors in Computing Systems*, 2020.

[7] D. Niermann and A. Lüdtke, "Measuring driver discomfort in autonomous vehicles," in *Intelligent Human Systems Integration 2020*, 2020, pp. 52–58.

[8] M. Beggiato, F. Hartwich, and J. Krems, "Using smartbands, pupillometry and body motion to detect discomfort in automated driving," *Frontiers in Human Neuroscience*, vol. 12, p. 338, 2018.

[9] H. Bellem, B. Thiel, M. Schrauf, and J. F. Krems, "Comfort in automated driving: Preferences for different driving styles and dependence on personality traits," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 55, pp. 90–100, 2018.

[10] K. N. de Winkel, T. Irmak, R. Happee, and B. Shyrokau, "Standards for passenger comfort in automated vehicles: Acceleration and jerk," *Applied Ergonomics*, vol. 106, p. 103881, 2023.

[11] H. Meng, X. Zhao, J. Chen, B. Wang, and Z. Yu, "Physiological representation of passenger cognitive comfort in overtaking scenarios," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 102, pp. 241–259, 2024.

[12] W. Boucsein, *Electrodermal Activity*. Boston, MA: Springer, 2012.

[13] M. E. Dawson, A. M. Schell, and D. L. Filion, "The electrodermal system," in *Handbook of Psychophysiology*, 3rd ed. Cambridge: Cambridge University Press, 2007, pp. 157–181.

[14] J. Braithwaite, G. Watson, D. Jason, R. Jones, and M. Rowe, "A guide for analysing electrodermal activity (eda) skin conductance responses (scrs) for psychological experiments," Birmingham, UK, 2015.

[15] V. Radhakrishnan, N. Merat, T. Louw, M. G. Lenné, R. Romano, E. Paschalidis, F. Hajiseyedjavadi, C. Wei, and E. R. Boer, "Measuring drivers' physiological response to different vehicle controllers in highly automated driving," *Information*, vol. 11, no. 8, p. 390, 2020.

[16] E. N. Schneider, B. Buchheit, P. Flotho, M. J. Bhamborae, F. I. Corona-Strauss, F. Dauth, M. Alayan, and D. J. Strauss, "Electrodermal responses to driving maneuvers in a motion sickness inducing real-world driving scenario," *IEEE Transactions on Human-Machine Systems*, vol. 52, no. 5, pp. 994–1003, 2022.

[17] R. Tan, W. Li, F. Hu, X. Xiao, S. Li, Y. Xing, H. Wang, and D. Cao, "Motion sickness detection for intelligent vehicles: A wearable-device-based approach," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems*, 2022, pp. 4355–4362.

[18] L. Wagner-Douglas, P. Seiwert, N. Schierhorst, A. Kirmas, N. Hennes, G. Voß, K. Rewitz, A. Mertens, D. Müller, V. Nitsch, and L. Eckstein, "Detection of motion sickness in participants through subjective and objective measurement," in —, 2024.

[19] X. Xiang, J. Zeng, X. Ding, S. Li, W. Liao, and C. Feng, "Research on vehicle comfort testing and evaluation based on the characterization of passenger motion sickness degree," in *3rd International Conference on Biomedical and Intelligent Systems*, 2024.

[20] Éléonore H. Henry, C. Bougard, C. Bourdin, and L. Bringoux, "Car sickness in real driving conditions: Effect of lateral acceleration and predictability reflected by physiological changes," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 97, pp. 123–139, 2023.

[21] F. Walker, J. Wang, M. H. Martens, and W. B. Verwey, "Gaze behaviour and electrodermal activity: Objective measures of drivers' trust in automated vehicles," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 64, pp. 401–412, 2019.

[22] I. B. Ajenaghughrure, S. C. D. C. Sousa, and D. Lamas, "Risk and trust in artificial intelligence technologies: A case study of autonomous vehicles," in *2020 13th International Conference on Human System Interaction*, 2020, pp. 118–123.

[23] J. Smyth, S. Birrell, R. Woodman, and P. Jennings, "Exploring the utility of EDA and skin temperature as individual physiological correlates of motion sickness," *Applied Ergonomics*, vol. 92, p. 103315, 2021.

[24] Euro NCAP, "Crash avoidance: Frontal collisions protocol," Brussels, 2024, version 0.9.

[25] H. Devriendt, M. Sarrazin, T. D'hondt, K. Gkentsidis, and K. Janssens, "A multimodal sensor setup for in situ comparison of driving dynamics, physiological responses and passenger comfort in autonomous vehicles," in *Intelligent Human Systems Integration 2025*, 2025, pp. —.

[26] O. E. Shodipe and R. S. Allison, "Modelling the relationship between the objective measures of car sickness," in *2023 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*. Regina, SK, Canada: IEEE, 2023, pp. 570–575.

[27] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, and S. H. A. Chen, "NeuroKit2: A python toolbox for neurophysiological signal processing," *Behavior Research Methods*, vol. 53, no. 4, pp. 1689–1696, 2021.

[28] A. Greco, G. Valenza, A. Lanatà, E. Scilingo, and L. Citi, "cvxEDA: A convex optimization approach to electrodermal activity processing," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 4, pp. 797–804, 2016.

[29] T. Zhan, Y. He, Y. Deng, Z. Li, W. Du, and Q. Wen, "Time evidence fusion network: Multi-source view in long-term time series forecasting," arXiv preprint arXiv:2405.06419, 2024.

[30] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, p. 321–357, Jun. 2002.

[31] I. B. Ajenaghughrure, S. C. D. C. Sousa, and D. Lamas, "Psychophysiological modelling of trust in technology: Comparative analysis of algorithm ensemble methods," in *2021 IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*. Herl'any, Slovakia: IEEE, 2021, pp. 000 161–000 168.

[32] H. Su and Y. Jia, "Study of human comfort in autonomous vehicles using wearable sensors," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 11 490–11 504, 2022.

[33] P. Zontone, A. Affanni, R. Bernardini, L. D. Linz, A. Piras, and R. Rinaldo, "Stress evaluation in simulated autonomous and manual driving through the analysis of skin potential response and electrocardiogram signals," *Sensors*, vol. 20, no. 9, p. 2494, 2020.

[34] S. Peng, X. Zhang, W. Zhu, and R. Dou, "Comfort of autonomous vehicles incorporating quantitative indices for passenger feeling," *Journal of Shanghai Jiaotong University (Science)*, vol. 29, no. 6, pp. 1063–1070, 2024.

[35] M. P. Cieslak, "Towards accurate ride comfort evaluation using biometric measurements and neural networks," Master's thesis, Coventry University, 2019.

[36] D. Niermann and A. Lüdtke, "Predicting vehicle passenger stress based on sensory measurements," in *Intelligent Systems and Applications*, K. Arai, S. Kapoor, and R. Bhatia, Eds., vol. 1252. Cham: Springer International Publishing, 2021, pp. 303–314, advances in Intelligent Systems and Computing.

[37] A. C. Stephenson, I. Eimontaite, P. Caleb-Solly, P. L. Morgan, T. Khatun, J. Davis, and C. Alford, "Effects of an unexpected and expected event on older adults' autonomic arousal and eye fixations during autonomous driving," *Frontiers in Psychology*, vol. 11, no. 571961, 2020.

A

Method

This chapter contains a more in-depth description of the experiment conducted by Siemens Digital Industries Software, which forms the basis of this experiment, along with details on the data collected and its preprocessing steps.

A.1. Experiment

The experiment took place at the Griesheim proving ground of the Technical University of Darmstadt, Germany. A Wizard-of-Oz autonomous vehicle (AV) methodology was employed, in which a human driver, obscured behind a black panel, controlled the vehicle, while the participant, seated next to the driver, perceived the vehicle as operating autonomously. An expert driver was recruited to perform all maneuvers of this experiment consistently, aided by adaptive cruise control functionalities to enhance repeatability.
The test vehicle, referred to as Vehicle Under Test (VUT), was a KIA EV6. A second vehicle, driven manually and referred to as the Global Target Vehicle, interacted with the VUT in different scenarios.

After preparation and start, five distinct test scenarios were conducted per lap. An overview of the test track and layout of the scenarios is illustrated in Figure A.1. All scenarios were performed following the specifications of the Euro NCAP test protocol for crash avoidance systems, and will be further discussed subsequently [9]. Each scenario was confined to a 30-second window.
After the five test scenarios, one lap was completed; this sequence was then repeated three more times, yielding a total of four laps. The first lap and the third lap were driven in a "calm" driving style, whereas the second lap and fourth lap were driven in an "aggressive" driving style. These two driving styles, defined by parameters such as velocity, acceleration, steering dynamics or distance to an object, were used to evoke different responses.

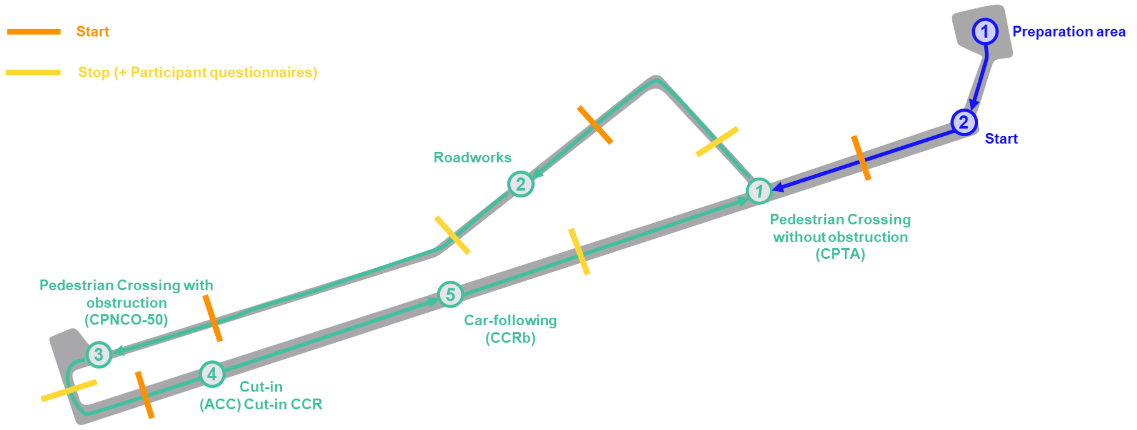

Start
Stop (+ Participant questionnaires)
1
Preparation area
2
Start
Roadworks
2
1
Pedestrian Crossing without obstruction (CPTA)
Pedestrian Crossing with obstruction (CPNCO-50)
3
5
Car-following (CCRb)
4
Cut-in (ACC) Cut-in CCR

Figure A.1: Overview of the five test scenarios present in the experiment.

Pedestrian Crossing without obstruction (CPTA)

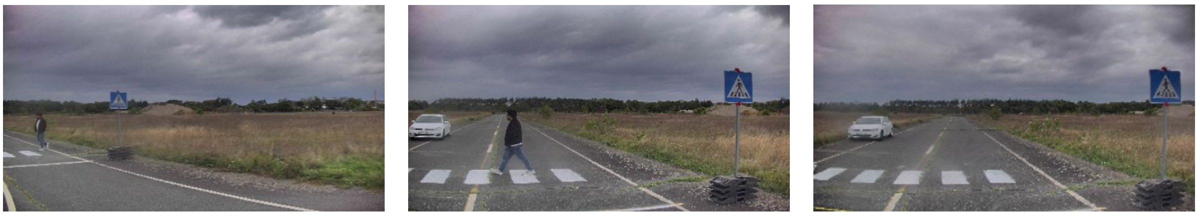


Figure A.2: Sequence of images illustrating the pedestrian crossing without obstruction (CPTA) scenario.

The first scenario involved the *Pedestrian Crossing without obstruction* – the Car-to-Pedestrian Turning Adult (CPTA) scenario as defined by NCAP. A pedestrian crossed the road at a pedestrian crossing without any visual obstruction. The driving style differences are summarized in Table A.1, where the main variations are in the approach speed of the VUT and the intensity of the braking for the pedestrian.

Parameter	Calm	Aggressive
VUT Target velocity (km/h)	30	50
GVT Target velocity (km/h)	-	-
Max. lon. acc. (m/s^2)	1.5	5
Max. lat. acc. (m/s^2)	2	7
Distance to obj. (m)	-	-

Table A.1: Driving characteristics for the Pedestrian Crossing Without Obstruction scenario

Roadworks

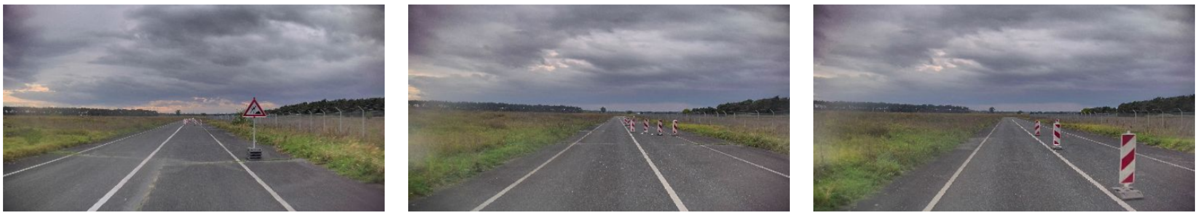


Figure A.3: Sequence of images illustrating the roadworks scenario.

The second scenario was the *Roadworks* scenario, in which the VUT navigated past roadwork markers. Table A.2 outlines the driving style characteristics. Here, the aggressive driving style was characterized by significantly higher speeds and closer proximity to the markers, with a sharper lateral movement when switching lanes.

Parameter	Calm	Aggressive
VUT Target velocity (km/h)	30	70
GVT Target velocity (km/h)	-	-
Max. lon. acc. (m/s^2)	1	4
Max. lat. acc. (m/s^2)	1	5
Distance to obj. (m)	20[a]	10[a]

[a] distance to road construction works.

Table A.2: Driving characteristics for the Roadworks scenario

Pedestrian Crossing with obstruction (CPNCO-50)

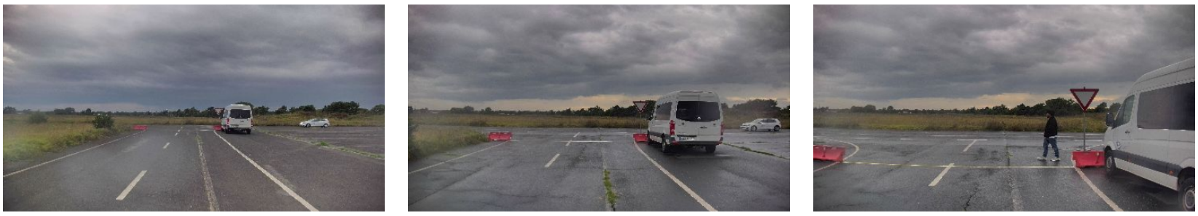


Figure A.4: Sequence of images illustrating the pedestrian crossing with obstruction (CPNCO-50) scenario.

The third scenario is the *Pedestrian Crossing with obstruction*, corresponding to the Car-to-Pedestrian Nearside Child Obstructed (CPNCO-50) protocol by NCAP. Here, a parked van blocked the participant's view of the pedestrian. In lap 1 and 4 the pedestrian did not cross, while in lap 2 and 3 they did, adding unpredictability to the scenario. Table A.3 presents the driving characteristics, highlighting that the aggressive style features a faster approach and more forceful braking.

Parameter	Calm	Aggressive
VUT Target velocity (km/h)	50	70
GVT Target velocity (km/h)	-	-
Max. lon. acc. (m/s^2)	2	6
Max. lat. acc. (m/s^2)	2	8
Distance to obj. (m)	-	-

Table A.3: Driving characteristics for the Pedestrian Crossing With Obstruction scenario

Cut-in (CCR)

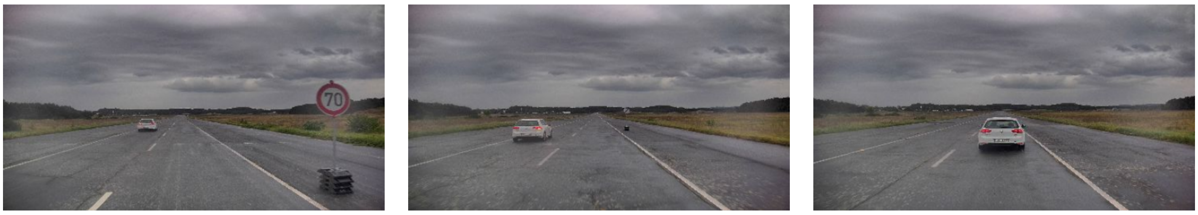


Figure A.5: Sequence of images illustrating the cut-in (CCR) scenario.

The fourth scenario incorporated a *Cut-in* maneuver, where the GVT merged in front of the VUT, forcing the VUT to perform a braking maneuver. Table A.4 shows that the primary difference between the calm and aggressive driving style were the VUT's target velocity and the magnitude of deceleration during the maneuver.

Parameter	Calm	Aggressive
VUT Target velocity (km/h)	50	70
GVT Target velocity (km/h)	30	50
Max. lon. acc. (m/s^2)	3	4
Max. lat. acc. (m/s^2)	4	8
Distance to obj. (m)	-	-

Table A.4: Driving characteristics for the Cut-in scenario

Car-following (CCRb)

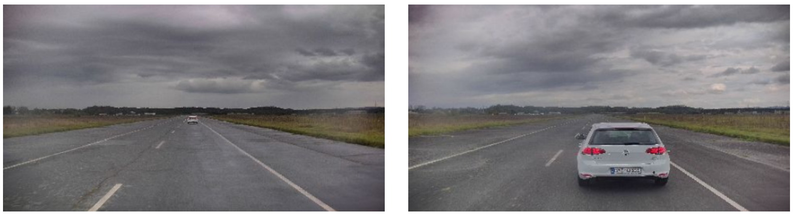


Figure A.6: Images illustrating the car-following (CCRb) scenario under different driving styles (left: calm, right: aggressive).

The final scenario was a *Car-following* scenario, in which the VUT followed the GVT along a straight path. Table A.5 presents the driving style characteristics, showing that in the calm driving style, the VUT maintained a significantly larger distance to the GVT compared to the aggressive driving style. Accelerations remained relatively similar for both styles.

Parameter	Calm	Aggressive
VUT Target velocity (km/h)	30	50
GVT Target velocity (km/h)	-	-
Max. lon. acc. (m/s^2)	0.5	1
Max. lat. acc. (m/s^2)	0.5	0.5
Distance to obj. (m)	40[b]	12[b]

[b] distance to GVT.

Table A.5: Driving characteristics for the Car-following scenario

A.2. Participant Data

For the experiment, 32 participants (17M, 15F) aged 18 to 82 years ($M = 49.2, SD = 21.1$) were involved. Figure A.7 summarizes the demographics of the participant group.

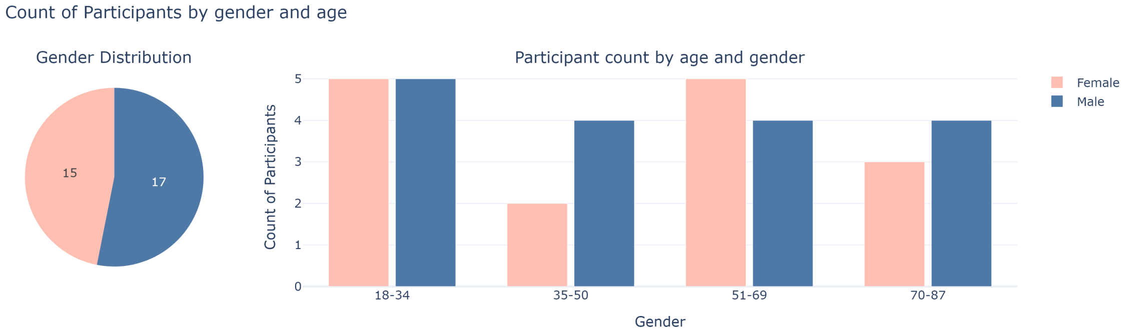


Figure A.7: Demographic distribution of participants by gender and age.

A.2.1. Self-Reported Scores

As seen in Figure A.1, each scenario was followed by a complete stop of the VUT and a questionnaire with three questions in which the participants evaluated the scenario on perceived comfort, safety and overall ride comfort on a five-point Likert scale ranging from “very uncomfortable/unsafe” to “very comfortable/safe”. For scenario 4 – Cut-in – and scenario 5 – Car-following – however, there was only one combined questionnaire administered after both scenarios; these responses will be used for the data analysis of both scenarios.

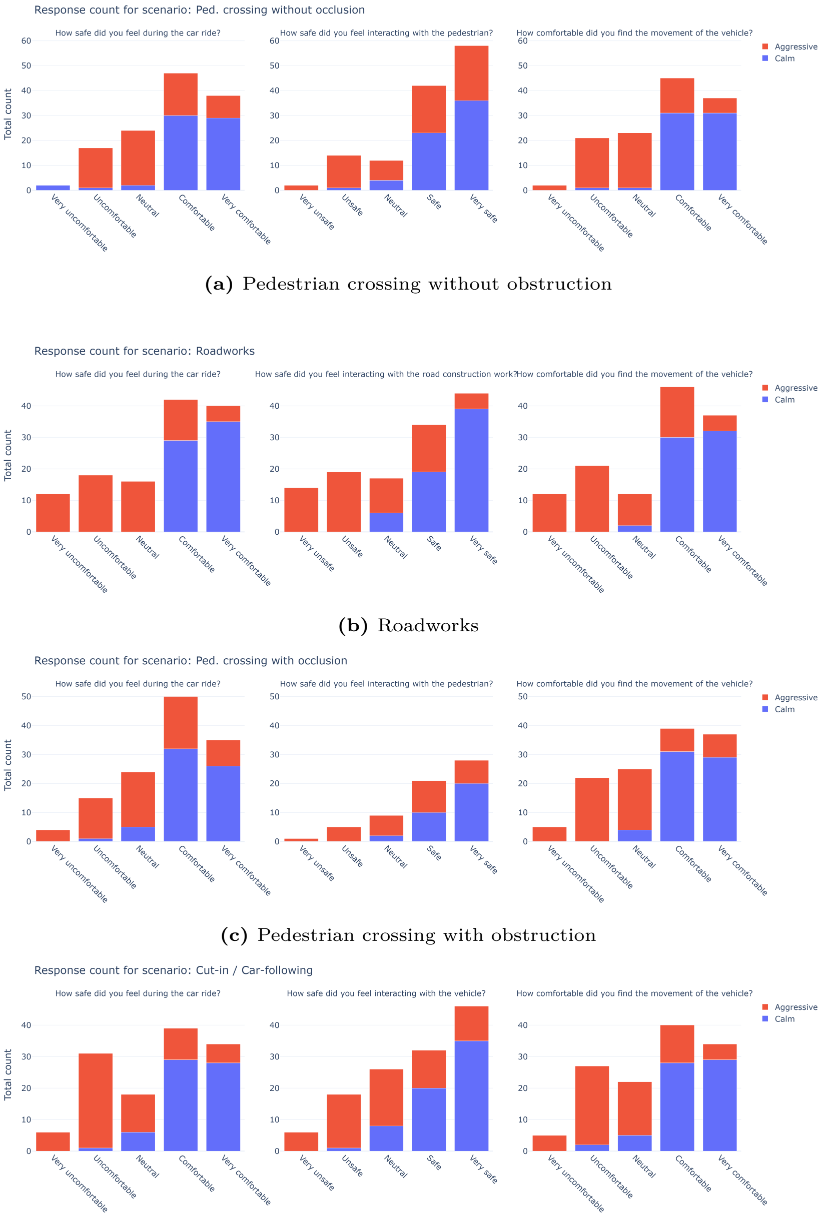
The following three questions were asked:

1. How safe did you feel during the car ride?

2. How safe did you feel interacting with the [pedestrian, roadworks, pedestrian, vehicle]?
3. How comfortable did you find the movement of the vehicle?

The first question measures general perceived comfort in how the vehicle behaves around the encountered object and in the scenario, and the second targets object-specific perceived safety and the third measures overall ride comfort from the vehicle's general motion.
The questionnaire was originally administered in German and translated into English for this research. Although the English translation of the first question uses the term "safe", the original German version focuses on the comfort of how the vehicle interacts with its surroundings. Consequently, while the phrasing in English may suggest a safety-related question, the actual responses reflect perceived comfort.

The distributions of the responses for the pedestrian crossing without obstruction, roadworks, pedestrian crossing with obstruction and cut-in / car-following scenario are shown in Figure A.8.



(a) Pedestrian crossing without obstruction

(b) Roadworks

(c) Pedestrian crossing with obstruction

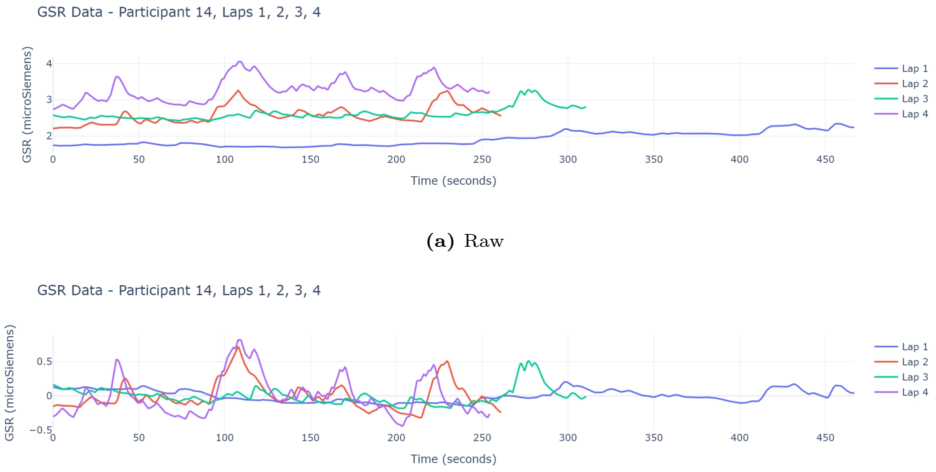
(d) Cut-in / car-following

Figure A.8: Distributions of subjective questionnaire responses for all scenarios, segmented by driving style (calm vs. aggressive). Responses cover the three questions on perceived comfort (Q1), perceived safety (Q2) and overall ride comfort (Q3).

A.2.2. Galvanic Skin Response

The GSR was continuously recorded each lap using the Mind Media Nexus | MKII device, recording skin conductivity in μS at 32 Hz. One participant's GSR data was lost due to a technical issue, and among the remaining 31 participants, six laps were excluded due to measurement errors (e.g., values exceeding 200 μS), resulting in 118 valid laps for analysis.

Preliminary data visualizations revealed an intriguing trend: the GSR increased progressively with each lap for all participants, as illustrated in Figure A.9(a) for a single representative participant. This trend either suggests that the GSR is highly susceptible to time-related effects or that external factors influenced the measurements. A similar observation has been reported by Irmak, Pool, and Happee [15], who found that the GSR correlates more strongly with elapsed time than motion sickness. Beggiato, Hartwich, and Krems [2] and Shodipe and Allison [27] also made this observation and attributed this trend to an increase in ambient temperature and shifts in electrode-skin interfaces over time, respectively. Other possible explanations for this trend could involve factors such as the accumulation of moisture between electrodes and skin or gradual changes in the conductive gel's consistency over time. This temporal trend poses a potential challenge for data interpretation. For instance, when analyzing the correlation between driving style and GSR, the results might be biased. A significant correlation would likely be found since laps 2 and 4, both driven aggressively, occurred later in the experiment. However, this may not necessarily reflect a genuine relationship between driving style and GSR, but rather an artefact of the GSR increase over time. To address this issue, a linear detrending procedure was applied to the GSR on a per-lap basis. Figure A.9(b) shows the results of this procedure.
This approach has also been used by Dillen et al. [8]. A comparable outcome can be achieved through baseline correction, a method employed by Smyth et al. [28], Gabrielli et al. [10], Morris, Erno, and Pilcher [20] and Xiang et al. [34]. This involves recording approximately 2-5 minutes of baseline GSR data prior to the experiment for each participant, which is then used to correct the subsequent measurements. Alternative approaches include isolating only the phasic component of the GSR, done by Henry et al. [12] and Perello-March et al. [23], and detrending the tonic component, as employed by Niermann, Trende, and Luedtke [21] and Schneider et al. [25].



(a) Raw

(b) Linearly detrended

Figure A.9: Time series plot of the raw and linearly detrended GSR for a single participant during all four laps.

The GSR data is processed using the Neurokit2 Python toolkit [18]. The GSR signal is decomposed into the phasic and tonic components via the convex optimization-based cvxEDA algorithm [11]. In cvxEDA, the observed skin conductance y, given an N-sample long signal, is modeled as:

$$y = r + t + \epsilon \tag{A.1}$$

Where r is the phasic component representing a time series of Skin Conductance Responses (SCRs), t is the slowly-varying tonic baseline and ϵ is an additive noise term. r, t and ϵ are N-long column vectors, and ϵ is an independent and identically distributed sequence of zero-average Gaussian random variables with variance σ^2 that represent modeling and measuring errors.
The algorithm is built on the following four assumptions, reproduced from Greco et al. [11]:

1. SCRs are preceded by bursts from the sudomotor nerves controlling the sweat glands. These bursts are temporally discrete episodes, i.e. SCRs are generated by a neural signal that is sparse and non-negative because of the nature of a nerve activity.
2. The relationship between the number of sweat glands recruited and the amplitude of a firing burst is linear. Moreover, the output response of the system depends only on the instant where the nerve input is applied. Stated otherwise, the timecourse of a single SCR induced by a neural burst is not influenced by previous ones, even when their SCRs overlap. In the light of these considerations it is reasonable to characterize the system as linear time-invariant.
3. The sweat diffusion process has a subject-specific impulse response function (IRF) which is relatively stable for all SCRs from the same subject.
4. The phasic activity is superimposed to a slowly varying tonic activity with spectrum below 0.05 Hz, i.e. whose information content can be represented by samples spaced every 10s (e.g., by 10-s averages).

The following steps describe the decomposition algorithm as described by Greco et al. [11]. The tonic component is modeled as the sum of cubic B-spline functions, an offset and a linear term:

$$t = B\ell + Cd \tag{A.2}$$

Where the cubic B-spline functions have equally spaced knots every 10s, as per assumption 4. B contains the cubic B-spline basis functions, and ℓ is the vector of the spline weights. C is a $N \times 2$ matrix with $C_{i,1} = 1$ and $C_{i,2} = i/N$ and d is a 2×1 vector with the offset and slope for the linear trend.
Under assumptions 2 and 3, the shape of a single SCR is modeled using the biexponential Bateman impulse response function :

$$h(\tau) = (e^{(-\frac{\tau}{\tau_0})} - e^{(-\frac{\tau}{\tau_1})})u(\tau) \tag{A.3}$$

With τ_0 and τ_1 the slow and fast time constants and $u(\tau)$ the step function. This Bateman function models the diffusion of sweat through the gland ducts [11]. Greco et al. [11] then use Laplace to transform the Bateman function to:

$$\mathcal{L}\{h(\tau)\} = \frac{1}{s + \tau_0^{-1}} - \frac{1}{s + \tau_1^{-1}} \tag{A.4}$$

With τ_0^{-1} and τ_1^{-1} being the poles of this second-order linear time-invariant system. To approximate the continuous Laplace domain expression into a discrete-time domain, a bilinear transform using $s = \frac{2}{\delta}\frac{z-1}{z+1}$ is used with sampling interval δ, resulting in the following Autoregressive Moving Average (ARMA) model:

$$H(z) = \frac{(1 + z^{-1})^2}{\psi + \theta z^{-1} + \zeta z^{-2}} \tag{A.5}$$

With the coefficients ψ, θ and ζ:

$$\psi = \frac{(\tau_1^{-1}\delta + 2)(\tau_0^{-1}\delta + 2)}{\tau_1^{-1}\delta^2 - \tau_0^{-1}\delta^2}$$

$$\theta = \frac{2\tau_1^{-1}\delta^2 - 8}{\tau_1^{-1}\delta^2 - \tau_0^{-1}\delta^2}$$

$$\zeta = \frac{(\tau_1^{-1}\delta - 2)(\tau_0^{-1}\delta - 2)}{\tau_1^{-1}\delta^2 - \tau_0^{-1}\delta^2}$$

The authors represent the ARMA model in matrix form as:

$$q = A^{-1}p, \quad r = Mq \tag{A.6}$$

Where p represents the sudomotor nerve activity and q is an auxiliary variable used to find p indirectly. M and A both tridiagonal matrices with $M_{i,i} = M_{i,i-2} = 1$, $M_{i,i-1} = 2$, and $A_{i,i} = \psi$, $A_{i,i-1} = \theta$ and $A_{i,i-2} = \zeta$ for $3 \leq i \leq N$. As a result, Equation A.1 is written as the following observation model:

$$y = Mq + B\ell + Cd + \epsilon \tag{A.7}$$

Given Equation A.7, Greco et al. [11] follows a probabilistic formulation using the Maximum a Posteriori (MAP) estimation to estimate the parameters q, ℓ and d that represent the phasic driver, tonic spline coefficients and drift for the measured signal y:

$$[q, \ell, d] = \arg\max_{q,\ell,d} P[q, \ell, d|y] \tag{A.8}$$

By assuming independence between q, ℓ and d and applying Bayes' theorem, the following is obtained:

$$P(q, l, d \mid y) \propto P(y \mid q, l, d) \cdot P(q) \cdot P(l) \cdot P(d) \tag{A.9}$$

With $P[y|q, \ell, d]$ being the likelihood of observing a specific skin conductance given the model's parameters and $P[q]$, $P[\ell]$ and $P[d]$ the prior probabilities of the parameters. $P[y]$ is omitted by the authors as it plays no role in the optimization. The authors emphasize that, since their model relies solely on the definition of priors, no preprocessing of the GSR signal, such as filtering, is required.

To model the sudomotor nerve activity (p) that represents the input of the system, as per assumption 1, the authors use a Poisson distribution $p_i \sim \text{Pois}(\lambda\delta)$ where λ is the average number of spikes per time unit δ. The Poisson distribution is replaced with an exponential distribution of the same mean to relax the constraint $p_i \in \mathbb{N}$ to $p_i \geq 0$, resulting in $P[q]$ becoming:

$$P[q] = \prod_{i=1}^{N} \frac{1}{\lambda\delta} e^{-\frac{p_i}{\lambda\delta}} \propto \prod_{i=1}^{N} \exp\left(-(\lambda\delta)^{-1}(Aq)_i\right). \tag{A.10}$$

Following assumption 4, the tonic component is then modeled under the assumption of a uniform frequency spectrum in the band $0 - 0.05$ Hz. With equally spaced spline knots at intervals of $\Delta = 10$ seconds, the corresponding sampling frequency is exactly twice the upper frequency bound. Under this setup, the spline coefficients ℓ_i can be assumed independent and identically distributed. At each knot, the amplitude is then modeled as normally distributed, resulting in the following prior $P[\ell]$:

$$P[\ell] = \prod_{i=1}^{Q} \frac{1}{\sqrt{2\pi}\,\sigma_\ell} \exp\left(-\frac{1}{2}\frac{\ell_i^2}{\sigma_\ell^2}\right), \tag{A.11}$$

The authors assume for the drift coefficients d uninformative priors and thus drop $P[d]$ from further analysis.

The likelihood term $P[y|q, \ell, d]$ follows directly from A.7 and $\epsilon \sim \mathcal{N}(0, \sigma^2)$ in:

$$P[y \mid q, \ell, d] = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{((Mq + B\ell + Cd - y)_i)^2}{2\sigma^2}\right). \tag{A.12}$$

By replacing the priors and taking the logarithm, the authors get:

$$\ln P[q, \ell, d \mid y] = -\frac{1}{2\sigma^2}\sum_{i=1}^{N}((Mq + B\ell + Cd - y)_i)^2 - \frac{1}{\lambda\delta}\sum_{i=1}^{N}(Aq)_i - \frac{1}{2\sigma_\ell^2}\sum_{i=1}^{Q}\ell_i^2 + \text{const.} \tag{A.13}$$

After multiplying by σ^2, substituting $\alpha = \sigma^2/(\lambda\delta)$ and $\gamma = \sigma^2/\sigma_\ell^2$, Equation A.13 is rewritten as a constrained minimization problem that the authors present as the core of their algorithm:

$$\begin{aligned} \min_{q,\ell,d} \quad & \frac{1}{2}\|Mq + B\ell + Cd - y\|_2^2 + \alpha\|Aq\|_1 + \frac{\gamma}{2}\|\ell\|_2^2 \\ \text{subject to} \quad & Aq \geq 0. \end{aligned} \tag{A.14}$$

The optimization problem is then rewritten into standard Quadratic Programming (QP) form and solved using a publicly available sparse-QP solver. After finding the optimal $[q, \ell, d]$ the tonic component can be derived from Equation A.2, and the phasic component can be found with $p = Aq$.

Several other algorithms have been proposed to decompose the GSR into its phasic and tonic components. Braithwaite's frequency-filtering method applied high-pass and low-pass filters to roughly partition fast SCRs from slow tonic baseline drifts. The filter filters forward and backwards to avoid phase distortion, making it a non-causal filter [4]. Continuous Decomposition Analysis (CDA) uses a parametric Batemun function as impulse response function and iterative deconvolution to continuously estimate the phasic driver and tonic trend [3]. SparsEDA, however, performs a non-negative sparse deconvolution of the GSR signal to jointly recover a highly sparse phasic driver and a smooth tonic baseline. The tonic component is parameterized via a first-order Taylor series expansion, and both signals are extracted through a convex ℓ_1 regularized optimization [13].
While literature suggests that sparsEDA excels at isolating the phasic component, its tonic component is overly flattened, obscuring meaningful shifts in baseline arousal [17]. Braithwaite's frequency-filtering methods are computationally efficient but fail to disentangle overlapping SCRs and can distort the true SCR shape. Moreover, CDA is not yet available in Python, limiting its practical usability. For these reasons, this study used cvxEDA for GSR decomposition. Its phasic component modeling is comparable to that of SparsEDA, while the tonic component more reliably captures baseline fluctuations [13]. By explicitly modeling both phasic activity and baseline dynamics, cvxEDA offers a physiologically grounded and accurate decomposition into phasic and tonic components.

Figure A.10 presents a decomposed GSR signal for a representative participant, captured during two laps representing both driving styles. Annotated regions indicate the time intervals corresponding to each scenario.

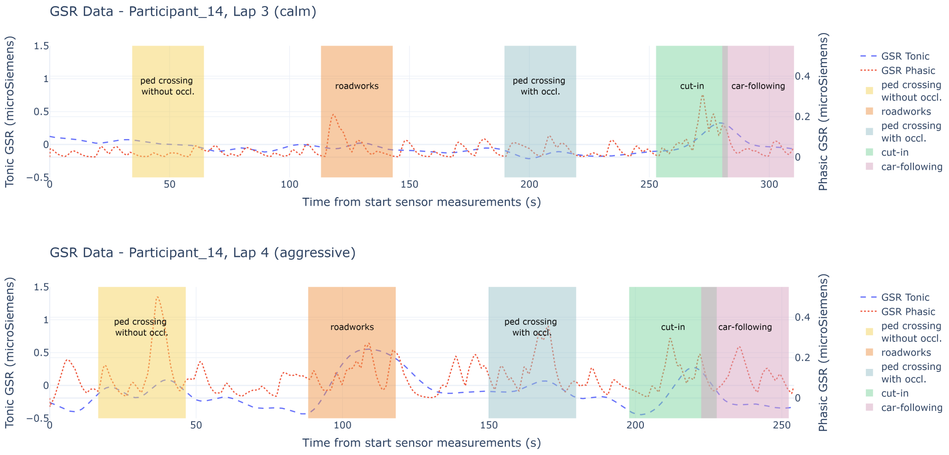


Figure A.10: Time series plot of phasic and tonic components for a single participant during two laps for both driving styles (top: calm, bottom: aggressive), annotated with regions indicating the time interval of each distinct test scenario.

Figures A.11 to A.16 depict the phasic and tonic components for every scenario and participant throughout the experiment. To enhance visual clarity, each signal is z-standardized ($\mu = 0, \sigma^2 = 1$) on a per-participant basis, allowing for visual comparison across conditions.

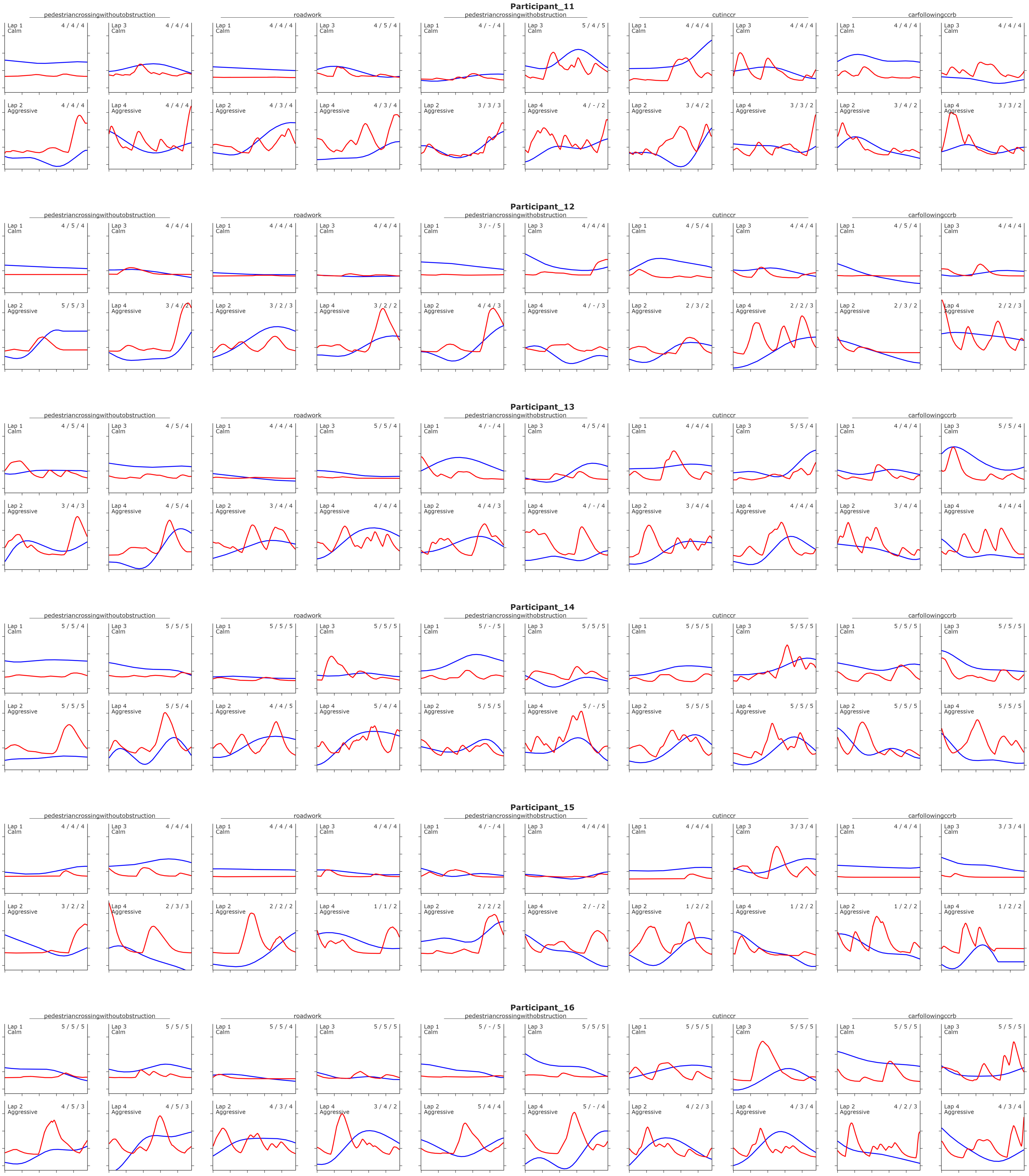


Figure A.11: Phasic (red) and tonic (blue) components of the GSR signal for all scenarios and all laps of participants 11-16. Signals were standardized on a per-participant basis. Lap number, driving style and Q1, Q2 and Q3 scores (1 = very uncomfortable/unsafe, 5 = very comfortable/safe) are annotated.

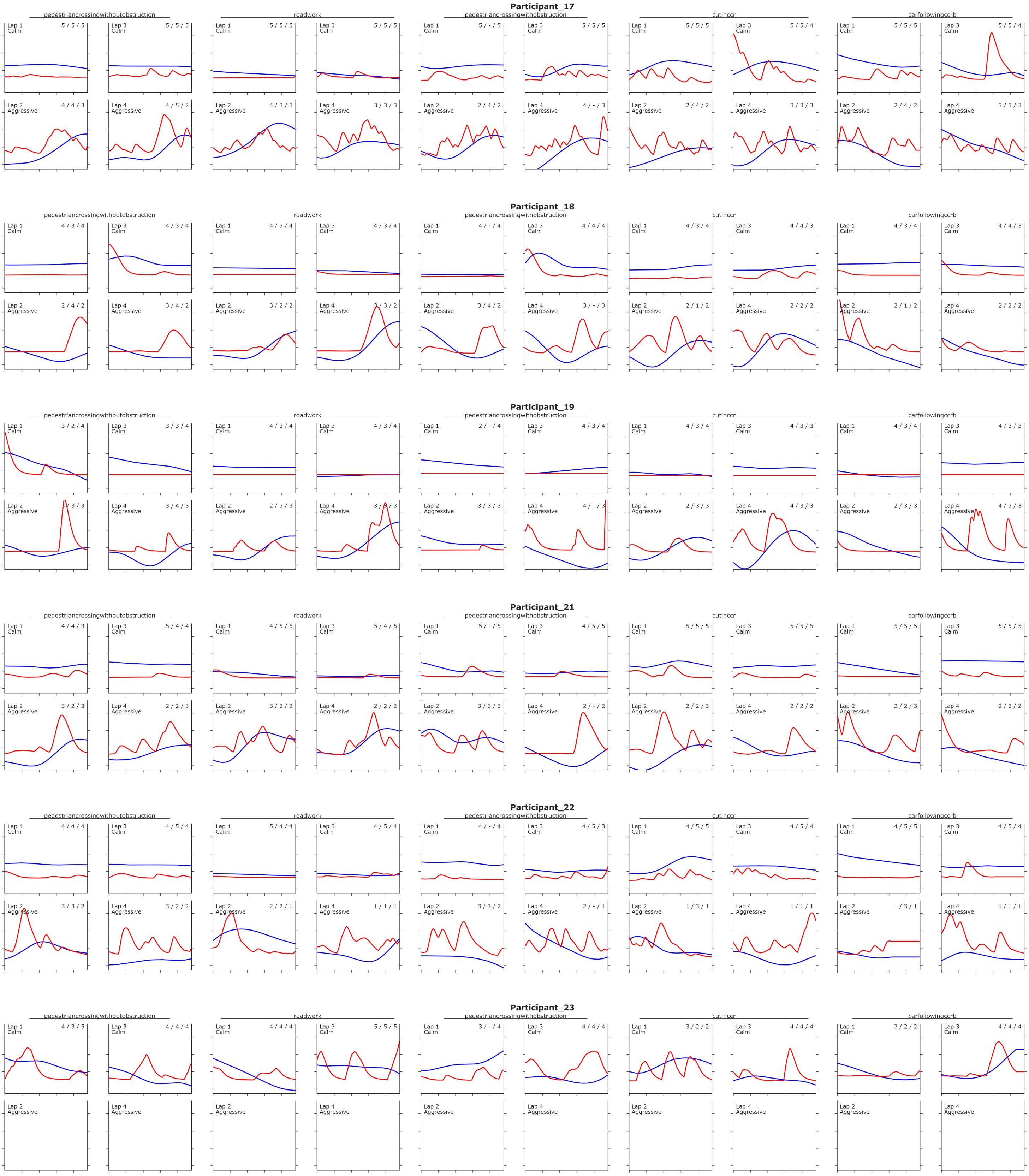


Figure A.12: Phasic (red) and tonic (blue) components of the GSR signal for all scenarios and all laps of participants 17-23. Signals were standardized on a per-participant basis. Lap number, driving style and Q1, Q2 and Q3 scores (1 = very uncomfortable/unsafe, 5 = very comfortable/safe) are annotated.

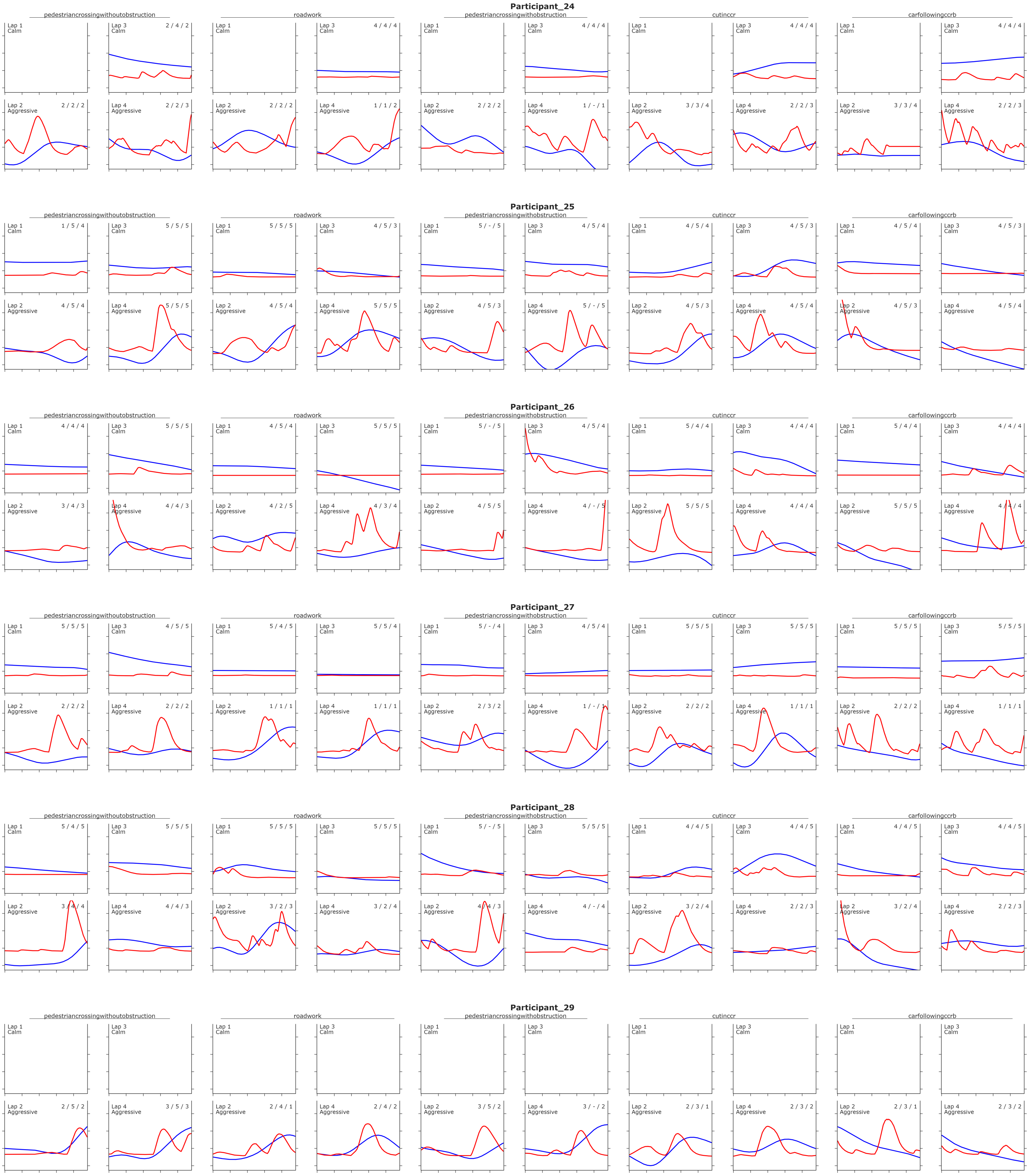


Figure A.13: Phasic (red) and tonic (blue) components of the GSR signal for all scenarios and all laps of participants 24-29. Signals were standardized on a per-participant basis. Lap number, driving style and Q1, Q2 and Q3 scores (1 = very uncomfortable/unsafe, 5 = very comfortable/safe) are annotated.

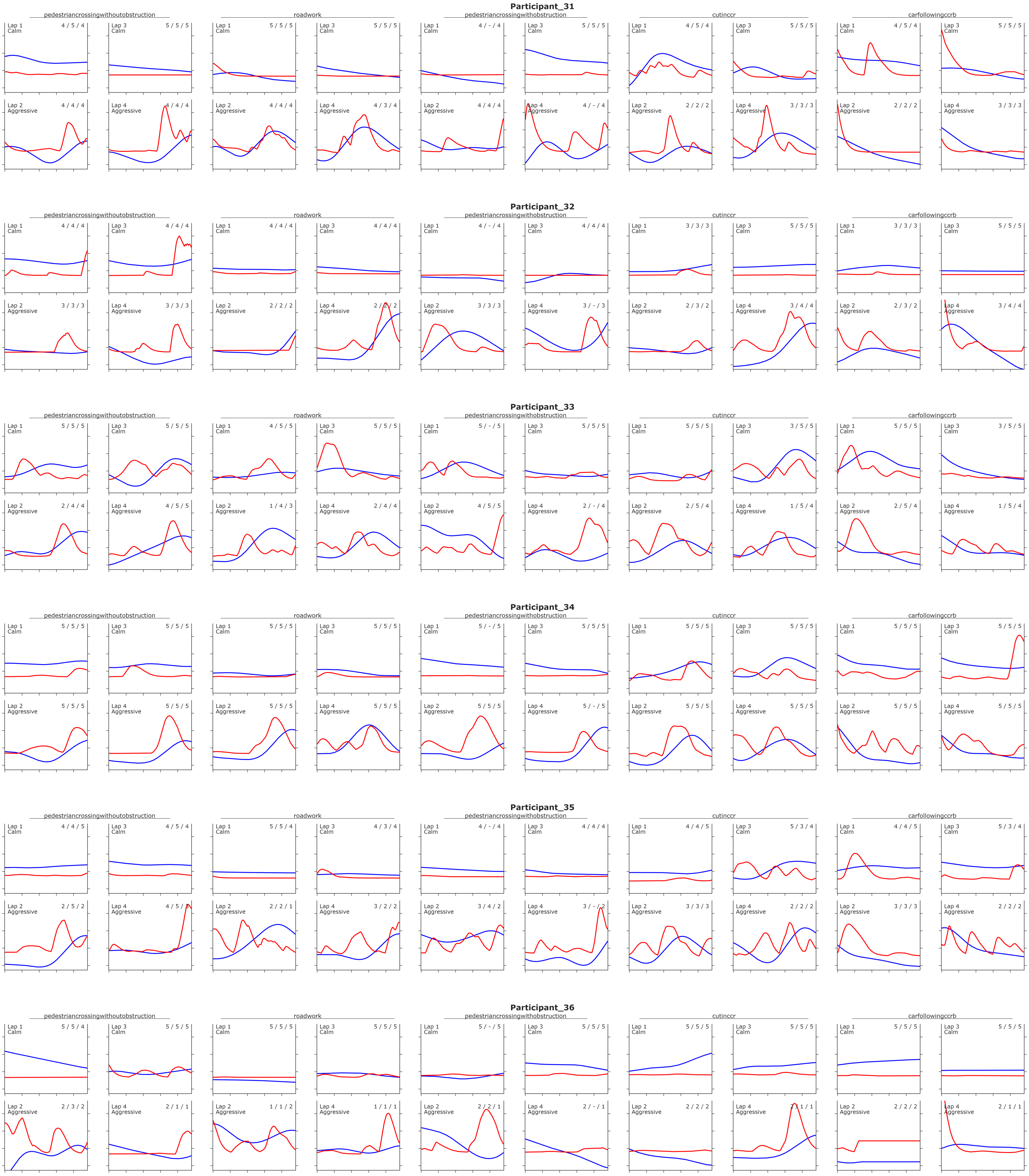


Figure A.14: Phasic (red) and tonic (blue) components of the GSR signal for all scenarios and all laps of participants 31-36. Signals were standardized on a per-participant basis. Lap number, driving style and Q1, Q2 and Q3 scores (1 = very uncomfortable/unsafe, 5 = very comfortable/safe) are annotated.

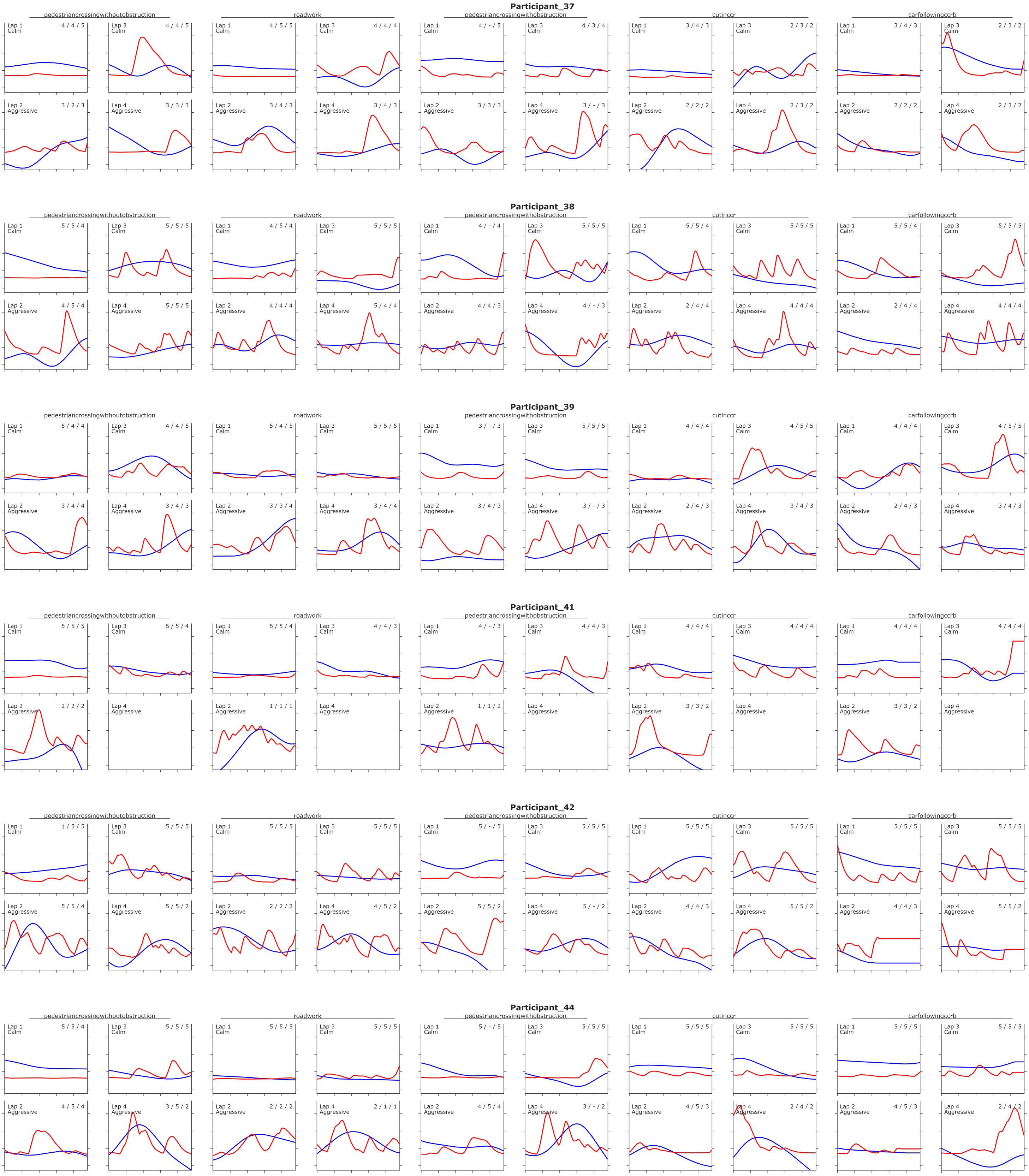


Figure A.15: Phasic (red) and tonic (blue) components of the GSR signal for all scenarios and all laps of participants 37-44. Signals were standardized on a per-participant basis. Lap number, driving style and Q1, Q2 and Q3 scores (1 = very uncomfortable/unsafe, 5 = very comfortable/safe) are annotated.

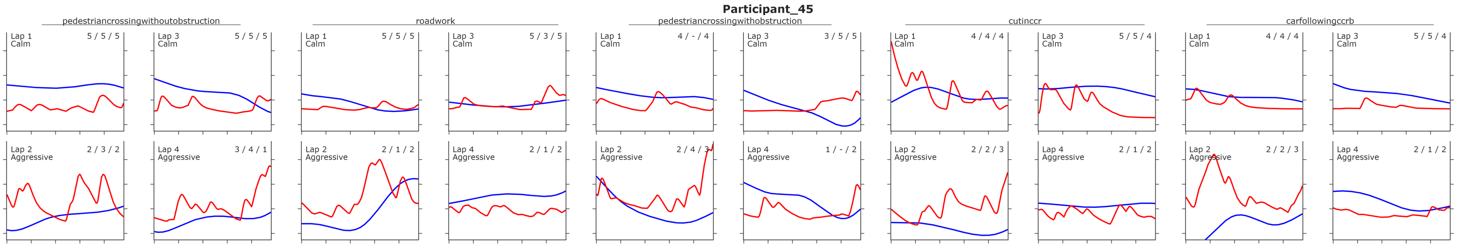


Figure A.16: Phasic (red) and tonic (blue) components of the GSR signal for all scenarios and all laps of participants 45. Signals were standardized on a per-participant basis. Lap number, driving style and Q1, Q2 and Q3 scores (1 = very uncomfortable/unsafe, 5 = very comfortable/safe) are annotated.

A.2.3. Vehicle Dynamics

Vehicle dynamics are recorded at 200 Hz using accelerometers, with the AV measurement framework established by Devriendt et al. [7]. Data included the vehicle's velocity, longitudinal and lateral acceleration and yaw rate. To reduce high-frequency noise in the linear acceleration data, a 5th-order Butterworth low-pass filter with a cutoff frequency of 1 Hz was applied. Jerk was then calculated through numerical differentiation of the filtered data. Finally, all vehicle dynamics data were downsampled to 32 Hz to match the GSR signal and aligned using UNIX timestamps. This resulted in a time series with 960 points and 6 signal features.

To visualize the vehicle dynamics across scenarios, Figures A.17 to A.21 present the mean and variance of each signal, separated by driving style.

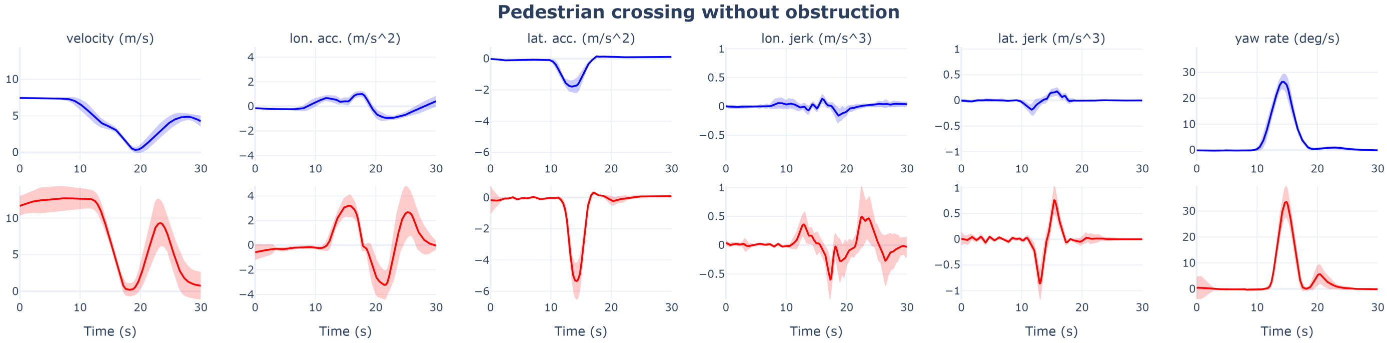


Figure A.17: Mean and variance (shaded) of vehicle dynamics in the Pedestrian crossing without obstruction scenario, highlighting differences between the calm (blue) and aggressive (red) driving style.

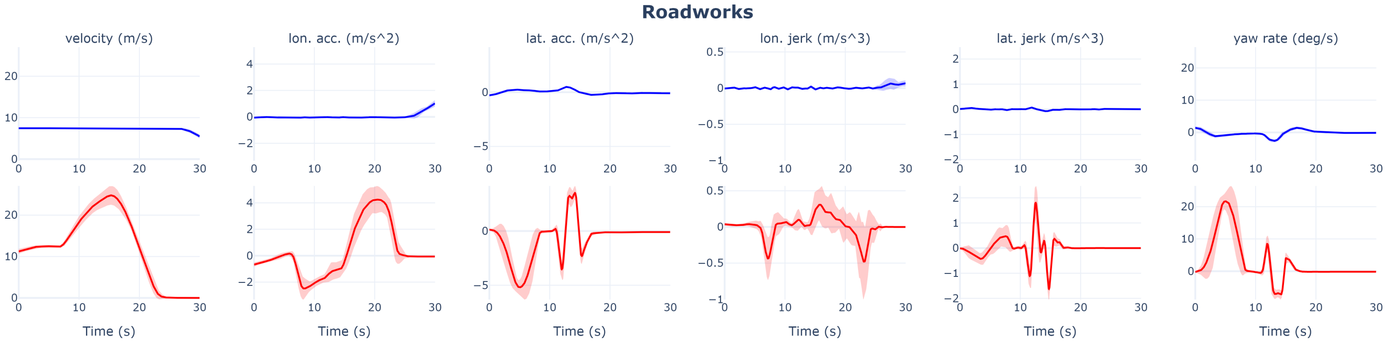


Figure A.18: Mean and variance (shaded) of vehicle dynamics in the Roadworks scenario, highlighting differences between the calm (blue) and aggressive (red) driving style.

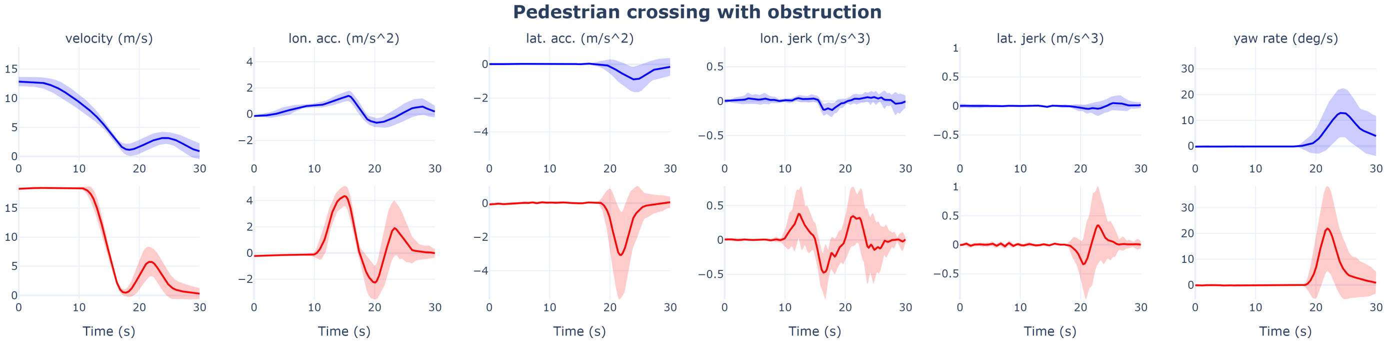


Figure A.19: Mean and variance (shaded) of vehicle dynamics in the Pedestrian crossing with obstruction scenario, highlighting differences between the calm (blue) and aggressive (red) driving style.

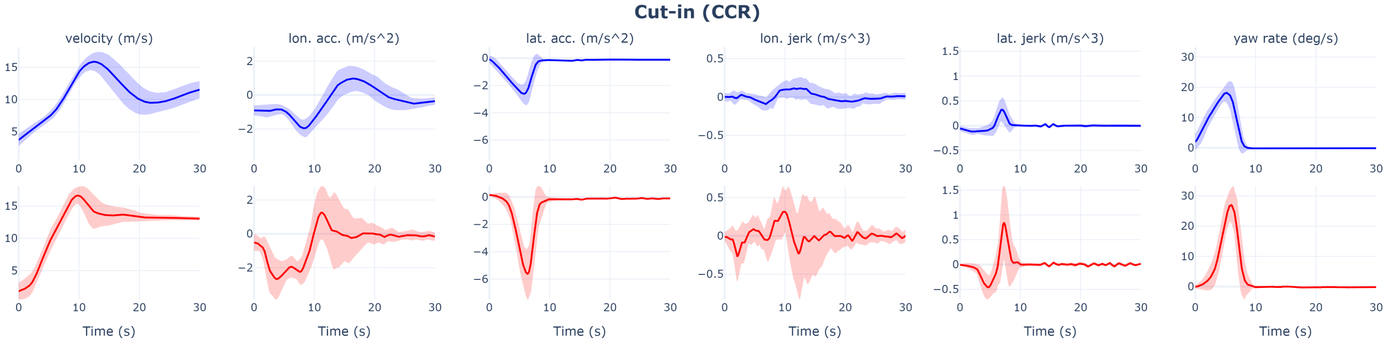


Figure A.20: Mean and variance (shaded) of vehicle dynamics in the Cut-in scenario, highlighting differences between the calm (blue) and aggressive (red) driving style.

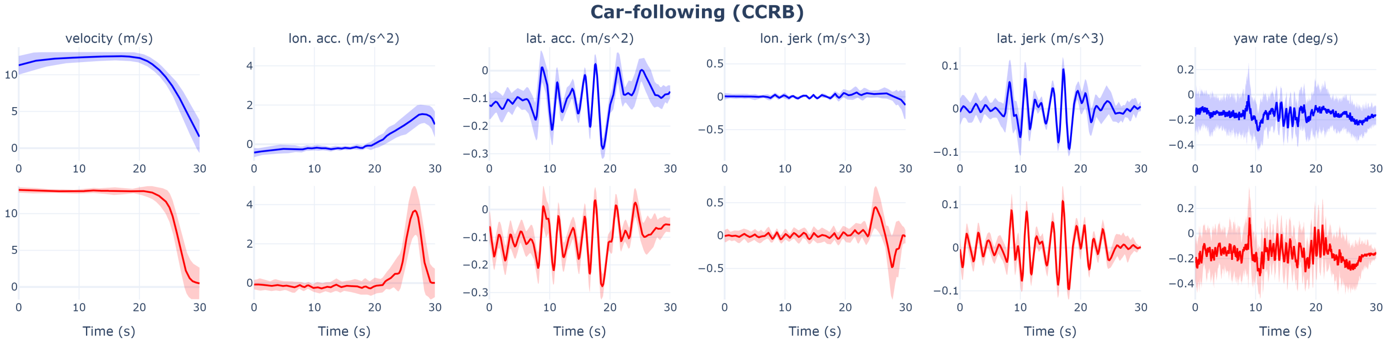


Figure A.21: Mean and variance (shaded) of vehicle dynamics in the Car-following scenario, highlighting differences between the calm (blue) and aggressive (red) driving style.

A.2.4. Perception

Perception data were captured by a forward-facing camera mounted on the VUT's roof rack. This data was recorded at 10 Hz and upsampled to 32 Hz through interpolation to synchronize with the GSR signal. Using the distance to detected objects and the relative velocity, the Time-to-Collision (TTC) was calculated. The Time Headway (THW) was computed by dividing the distance to the object by the vehicle's velocity.
When no objects were detected at the beginning or end of the scenario, padding was applied: distance, TTC, and THW values were set to high thresholds (80, 30, 10, respectively) to indicate no immediate collision risk. Additionally, a binary signal was added to the time series to mark whether a data point was padded or not, and a categorical signal was added to encode the type of detected object, with 0 for no object, 1 for pedestrians, 2 for cars, and 3 for roadwork markers. This yielded a time series with 960 points and 5 signal features.

In several scenarios, the perception pipeline did not produce any detections. Specifically, 185 out of 590 scenarios had no detected objects. In these cases, the perception data were populated with a constant placeholder of -1 to prevent mismatching format errors.

Figures A.22 to A.26 present the perception signals, without the missing values, with the mean and variance of each signal, separated by driving style.

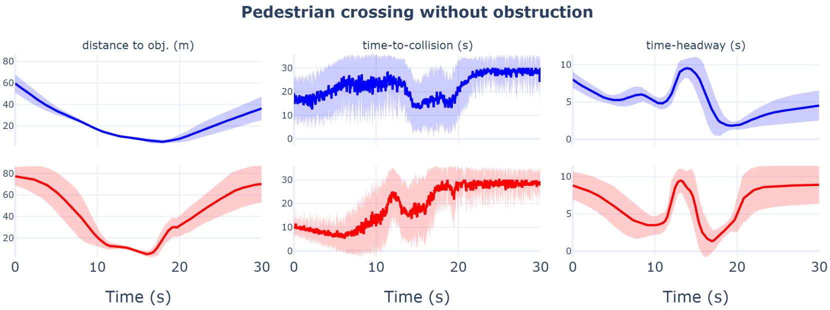


Figure A.22: Mean and variance (shaded) of perception signals in the Pedestrian crossing without obstruction scenario, highlighting differences between the calm (blue) and aggressive (red) driving style.

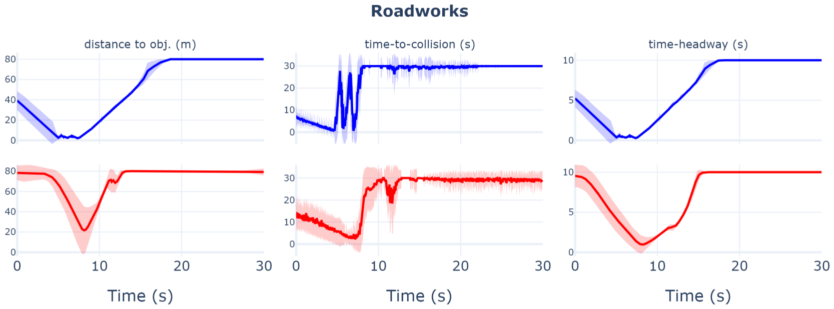


Figure A.23: Mean and variance (shaded) of perception signals in the Roadworks scenario, highlighting differences between the calm (blue) and aggressive (red) driving style.

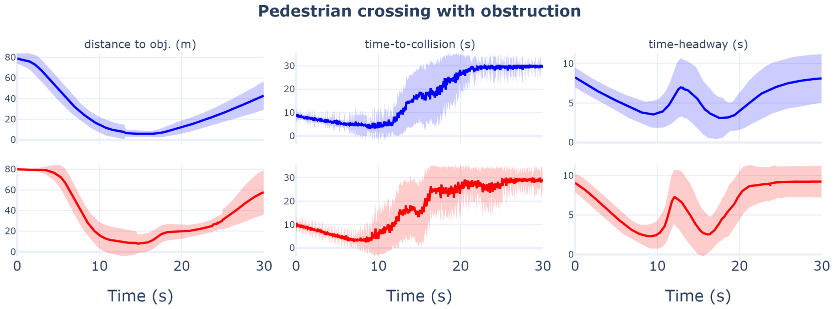


Figure A.24: Mean and variance (shaded) of perception signals in the Pedestrian crossing with obstruction scenario, highlighting differences between the calm (blue) and aggressive (red) driving style.

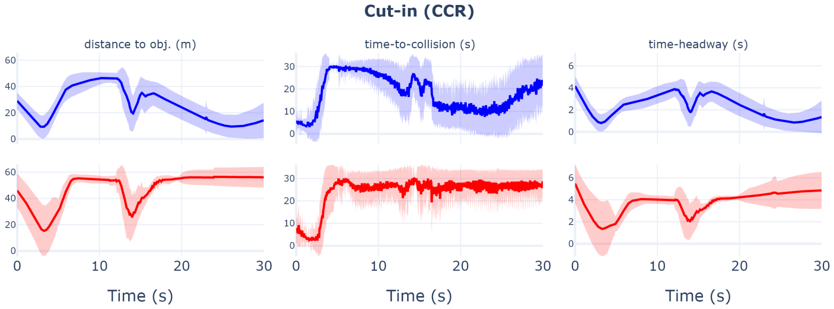


Figure A.25: Mean and variance (shaded) of perception signals in the Cut-in scenario, highlighting differences between the calm (blue) and aggressive (red) driving style.

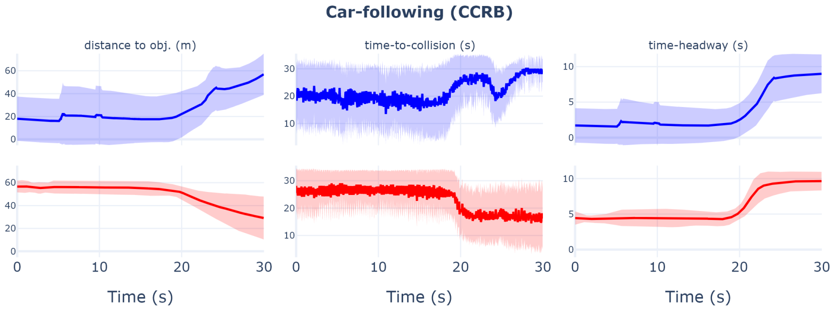


Figure A.26: Mean and variance (shaded) of perception signals in the Car-following scenario, highlighting differences between the calm (blue) and aggressive (red) driving style.

A.2.5. Dataset Size and Composition

In total, this study collected physiological recordings from 31 participants, alongside synchronized measurements of vehicle dynamics and perception-related signals. Each participant completed four laps with five scenarios per lap, resulting in 620 raw time series recordings (31 × 4 × 5). Due to technical issues and measurement errors, six of these laps were excluded, yielding a final dataset of 590 valid sequences. Each sequence has a fixed duration of 30 seconds and was sampled at 32 Hz, resulting in 960 time steps per signal.

B

Data Analysis

This chapter addresses the understanding half of the central research question:

> How can physiological arousal, measured through Galvanic Skin Response, combined with vehicle dynamics and perception data, be utilized to **understand** and predict passengers' perceived comfort and safety in automated driving?

A summary of these findings is included in the scientific paper presented in Chapter 1. While the scientific paper provides a concise overview of the key findings, this chapter presents a detailed account of how these key findings were found. Through feature extraction, correlation analyses, visualizations and pairwise tests, this chapter examines how the GSR, vehicle dynamics and perception influence the perceived comfort and safety scores.

To understand the interplay between comfort, safety, physiological arousal, vehicle dynamics and perception, the following five subquestions were formulated:

1. Did the calm and aggressive driving style elicit distinct perceived comfort and safety scores and physiological (GSR) reactions?
2. To what extent does the GSR signal reflect changes in perceived comfort, safety and overall ride comfort?
3. Which of the three input modalities, GSR, vehicle dynamics or perception, explains the largest share of variance in perceived comfort and safety?
4. Which of the objective signals exerts the strongest influence on the GSR signal?
5. In what ways do scenario characteristics (e.g., the presence of another vehicle versus a pedestrian or visibility) and passenger demographics (age, gender, trust in automation) influence the relationships identified in the previous questions?

With these subquestions, this chapter not only aims to quantify statistical relationships among physiological, vehicle dynamics and perception data, but also explores the underlying dynamics of perceived comfort and safety. It examines how different scenarios and passenger characteristics shape these experiences, and investigates the GSR signal itself and its key factors driving its variation.

To structure this analysis, the chapter first introduces the specific features extracted from the GSR signal, vehicle dynamics and perception data. It then outlines the statistical methodology, including correlation analyses and pairwise comparisons. The results section begins by presenting correlation matrices and key visualizations that illustrate the relationship between GSR and perceived comfort and safety. This is followed by comparisons across scenarios and participant demographics, revealing how contextual and personal factors shape these comfort and safety scores and physiological responses. The chapter concludes with a discussion that addresses each of the subquestions, followed by a conclusion that integrates the findings and reflects on how to understand perceived comfort and safety.

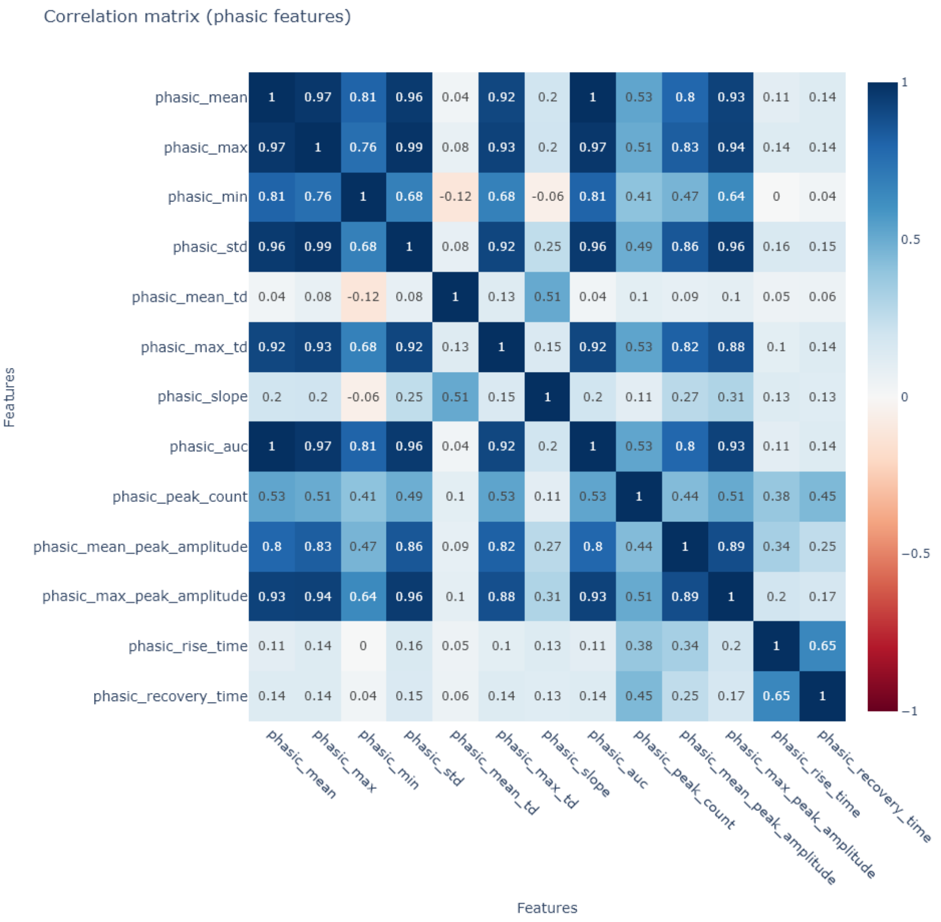
B.1. Feature Extraction and Selection

The statistical analysis was conducted on a feature basis. Each feature was computed over 30-second windows for each scenario, lap and participant. From the physiological data, both the phasic and tonic components were used to extract features. For vehicle dynamics, features were based on velocity, longitudinal and lateral acceleration, jerk and yaw rate. For perception-related signals, features were derived from the distance to the object, object type, time-to-collision (TTC) and time-headway (THW). For the physiological and vehicle dynamics features, the mean, maximum, minimum and standard deviation were calculated. For the perception features, the minimum values of distance, TTC and THW were extracted, and the mean and maximum time derivatives were computed to capture the dynamics of the scenario. Additional physiological features are listed in Table B.1.

	Feature Description
Phasic features	
peak count	Amount of SCRs in period.
mean peak ampl.	Mean amplitude of present SCRs.
max. peak ampl.	Maximum amplitude of present SCRs
rise time	Time interval between SCR onset and peak.
recovery time	Time interval between SCR peak and 50% recovery point.
mean td.	Mean time derivative of the signal.
maximum td.	Maximum time derivative of the signal.
slope	Linear trend of the signal.
AUC	Area Under the Curve of the signal.
Tonic features	
mean td.	Mean time derivative of the signal.
maximum td.	Maximum time derivative of the signal.
range	Range between the minimum and maximum of the signal.
slope	Linear trend of the signal.
skewness	Degree of asymmetry in the signal distribution.
kurtosis	Degree of flatness in the signal distribution.

Table B.1

This, however, results in a total of 59 candidate features. Such a large set of features reduces the clarity of the analysis and increases the risk of false positives, as each additional predictor adds to the number of hypothesis tests. To improve interpretability and lower the Type I error rate, a pairwise Pearson correlation filter (threshold $|r \geq 0.8|$) is applied for each separate signal. This lightweight screening removes one feature from each highly correlated pair, thereby reducing dimensionality and improving the interpretability without sacrificing informative variance. Figure B.1 to B.4 displays the resulting correlation matrices for each signal.



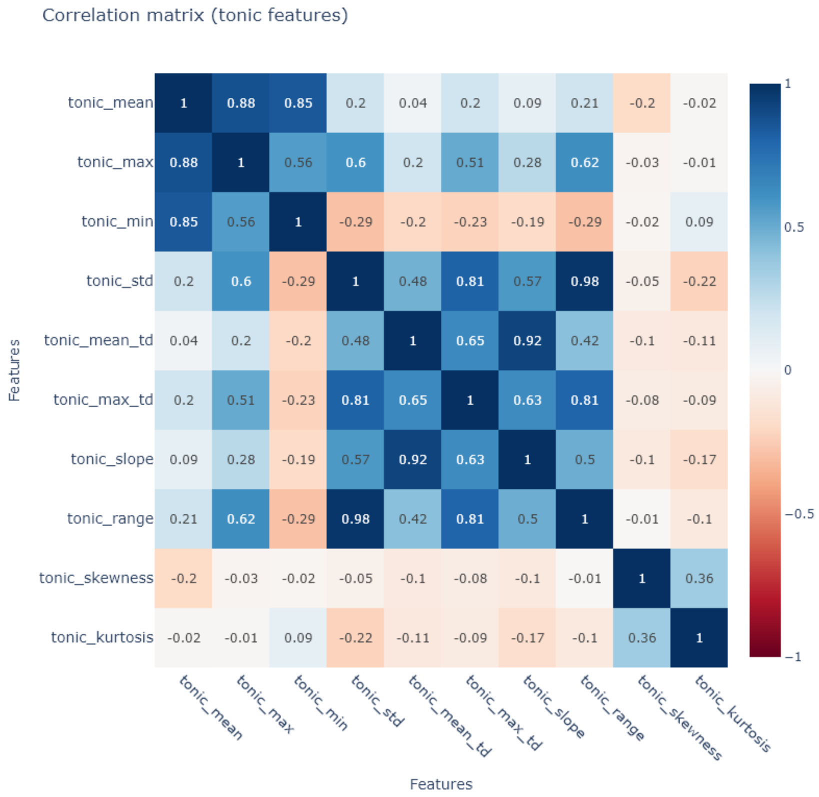
	phasic_mean	phasic_max	phasic_min	phasic_std	phasic_mean_td	phasic_max_td	phasic_slope	phasic_auc	phasic_peak_count	phasic_mean_peak_amplitude	phasic_max_peak_amplitude	phasic_rise_time	phasic_recovery_time
phasic_mean	1	0.97	0.81	0.96	0.04	0.92	0.2	1	0.53	0.8	0.93	0.11	0.14
phasic_max	0.97	1	0.76	0.99	0.08	0.93	0.2	0.97	0.51	0.83	0.94	0.14	0.14
phasic_min	0.81	0.76	1	0.68	-0.12	0.68	-0.06	0.81	0.41	0.47	0.64	0	0.04
phasic_std	0.96	0.99	0.68	1	0.08	0.92	0.25	0.96	0.49	0.86	0.96	0.16	0.15
phasic_mean_td	0.04	0.08	-0.12	0.08	1	0.13	0.51	0.04	0.1	0.09	0.1	0.05	0.06
phasic_max_td	0.92	0.93	0.68	0.92	0.13	1	0.15	0.92	0.53	0.82	0.88	0.1	0.14
phasic_slope	0.2	0.2	-0.06	0.25	0.51	0.15	1	0.2	0.11	0.27	0.31	0.13	0.13
phasic_auc	1	0.97	0.81	0.96	0.04	0.92	0.2	1	0.53	0.8	0.93	0.11	0.14
phasic_peak_count	0.53	0.51	0.41	0.49	0.1	0.53	0.11	0.53	1	0.44	0.51	0.38	0.45
phasic_mean_peak_amplitude	0.8	0.83	0.47	0.86	0.09	0.82	0.27	0.8	0.44	1	0.89	0.34	0.25
phasic_max_peak_amplitude	0.93	0.94	0.64	0.96	0.1	0.88	0.31	0.93	0.51	0.89	1	0.2	0.17
phasic_rise_time	0.11	0.14	0	0.16	0.05	0.1	0.13	0.11	0.38	0.34	0.2	1	0.65
phasic_recovery_time	0.14	0.14	0.04	0.15	0.06	0.14	0.13	0.14	0.45	0.25	0.17	0.65	1

Figure B.1: Correlation matrix of phasic features. Values represent Pearson correlation coefficients between feature pairs. Highly correlated pairs ($|r \geq 0.8|$) are flagged for redundancy removal.

The correlation matrix in Figure B.1 clearly shows that many features are highly correlated, indicating a redundancy within the original set of features. The features retained after filtering are:

- Phasic max.
- Phasic min.
- Phasic mean td.
- Phasic slope
- Phasic peak count
- Phasic rise time
- Phasic recovery time

Discarding a total of 6 phasic features.



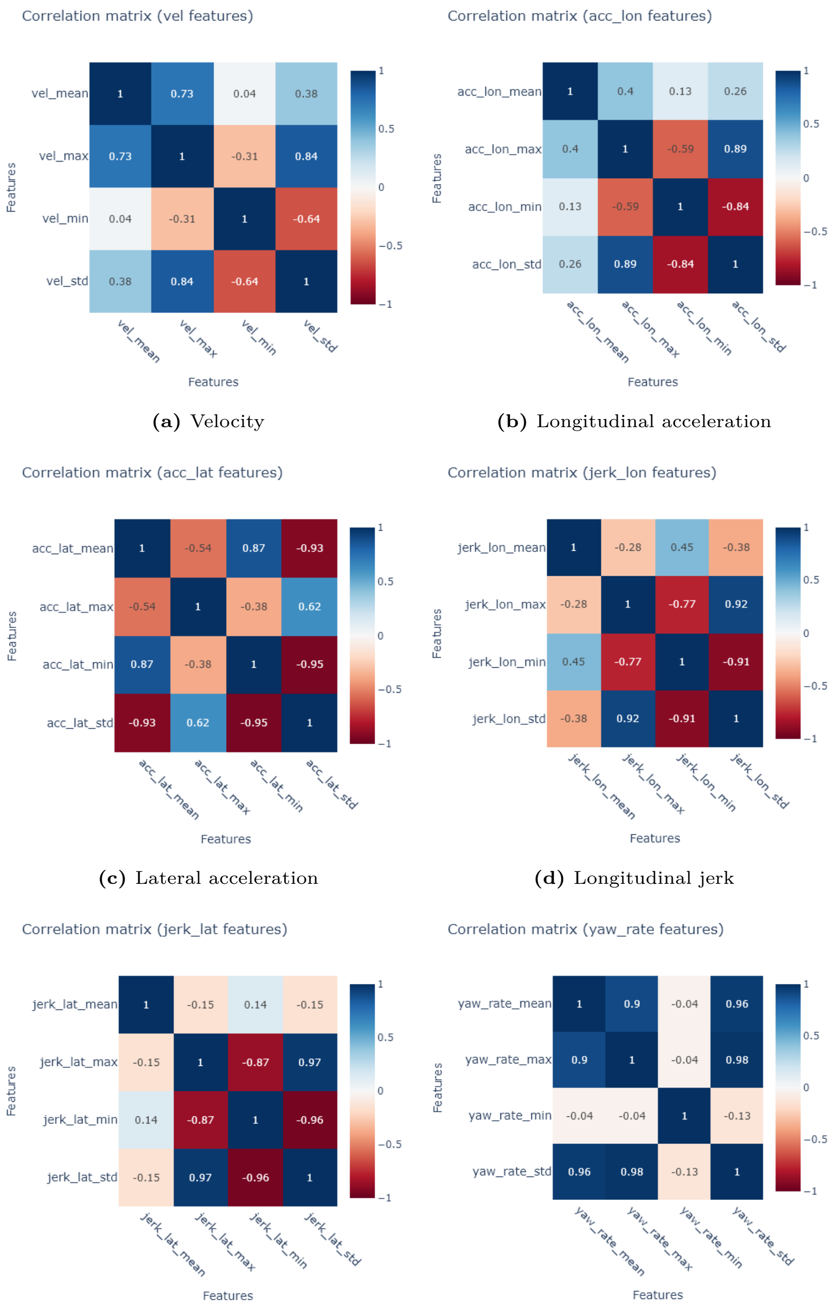
	tonic_mean	tonic_max	tonic_min	tonic_std	tonic_mean_td	tonic_max_td	tonic_slope	tonic_range	tonic_skewness	tonic_kurtosis
tonic_mean	1	0.88	0.85	0.2	0.04	0.2	0.09	0.21	-0.2	-0.02
tonic_max	0.88	1	0.56	0.6	0.2	0.51	0.28	0.62	-0.03	-0.01
tonic_min	0.85	0.56	1	-0.29	-0.2	-0.23	-0.19	-0.29	-0.02	0.09
tonic_std	0.2	0.6	-0.29	1	0.48	0.81	0.57	0.98	-0.05	-0.22
tonic_mean_td	0.04	0.2	-0.2	0.48	1	0.65	0.92	0.42	-0.1	-0.11
tonic_max_td	0.2	0.51	-0.23	0.81	0.65	1	0.63	0.81	-0.08	-0.09
tonic_slope	0.09	0.28	-0.19	0.57	0.92	0.63	1	0.5	-0.1	-0.17
tonic_range	0.21	0.62	-0.29	0.98	0.42	0.81	0.5	1	-0.01	-0.1
tonic_skewness	-0.2	-0.03	-0.02	-0.05	-0.1	-0.08	-0.1	-0.01	1	0.36
tonic_kurtosis	-0.02	-0.01	0.09	-0.22	-0.11	-0.09	-0.17	-0.1	0.36	1

Figure B.2: Correlation matrix of tonic features. Values represent Pearson correlation coefficients between feature pairs. Highly correlated pairs ($|r \geq 0.8|$) are flagged for redundancy removal.

This correlation matrix in B.2 reveals several strong linear relationships, specifically among those derived from temporal dynamics. After filtering, the following features were retained:

- Tonic mean
- Tonic std.
- Tonic mean td.
- Tonic skewness
- Tonic kurtosis

Discarding a total of 5 tonic features.



(a) Velocity

(b) Longitudinal acceleration

(c) Lateral acceleration

(d) Longitudinal jerk

(e) Lateral jerk

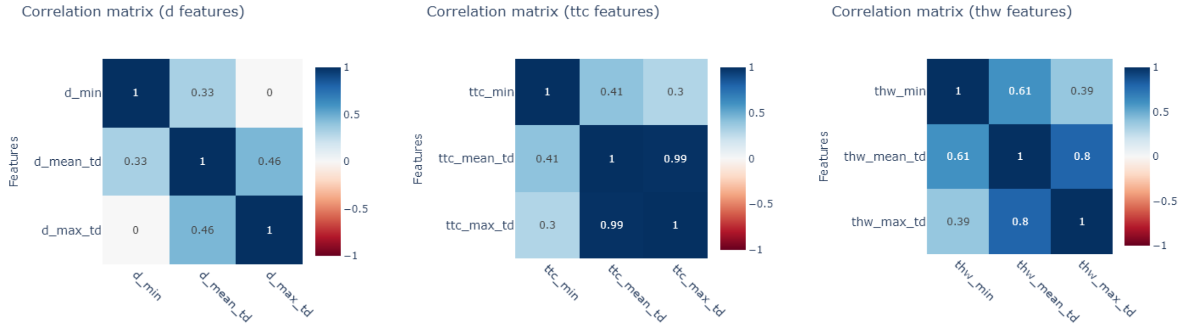
(f) Yaw rate

Figure B.3: Correlation matrix of vehicle dynamics features. Values represent Pearson correlation coefficients between feature pairs. Highly correlated pairs ($|r \geq 0.8|$) are flagged for redundancy removal.

Across all vehicle dynamics, there are various strong relationships. After applying the correlation-based filter, the remaining features for each signal are:

- **Velocity:** mean, max., min.
- **Lon. acc.:** mean, max.
- **Lat acc.:** max., min.
- **Lon. jerk:** mean., max.
- **Lat. jerk:** mean., max.
- **Yaw rate:** max., min.

Discarding 11 vehicle dynamics-related features.



(a) Distance **(b)** Time-to-collision **(c)** Time-headway

Figure B.4: Correlation matrix of perception features. Values represent Pearson correlation coefficients between feature pairs. Highly correlated pairs ($|r \geq 0.8|$) are flagged for redundancy removal.

The following features were kept after filtering for each signal:

- **d:** min., mean td., max. td.
- **TTC:** min., max. td.,
- **THW:** min., max. td.,

Discarding 2 perception-related features.

The total number of features has now been reduced from 59 to 35 for the subsequent statistical analysis.

B.2. Method

Linear Mixed-Effect (LME) models were employed to perform the statistical analysis aimed at finding significant correlations between features. LMEs were chosen for their ability to account for repeated measures within participants. Each participant experiences every scenario four times, once during each lap, resulting in repeated observations for both the subjective responses and the physiological responses. These repeated measures tend to correlate within the same participant, as individuals often have a consistent response style. For example, one individual might consistently give extreme responses, such as“very uncomfortable” and “very comfortable”, whereas another might respond more moderately with ratings like “neutral” and “comfortable”. Similarly, GSR responses are not independent within individuals either. This consistency reflects each individual's baseline or personality, and thus creates a within-subject correlation. Therefore, the data violates the assumption of independent observations. Standard statistical tests, such as ANOVA or Spearman's Rank Correlation test, however, rely on this independence. By assuming independence, these tests ignore the fact that observations from the same subject are more likely to be similar. This underestimation leads to an increased likelihood of declaring a result as significant, and thus inflating the Type I error rate. LME models explicitly account for the non-independence of observations from the same participant by incorporating random intercepts and/or slopes per participant.

For this analysis, LME models were implemented in Python using statsmodels' mixed linear model (`smf.mixedlm`). This model assumes a linear relationship between variables while accounting for the repeated measures structure with the formula:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0i} + u_{1i} x_{ij} + \epsilon_{ij}$$

With y_{ij} representing the j^{th} outcome feature for participant i, x_{ij} the predictor feature, β_0 and β_1 the fixed effects, u_{0i} and u_{1i} the random intercept and slope for participant i and finally ϵ_{ij} the error term. The random intercept u_{0i} allows each participant to have their own baseline, while the random slope u_{1i} accounts for individual differences in how strongly the predictor x_{ij} influences outcome y_{ij}. By incorporating these participant-specific random intercepts, the model explicitly accounts for within-subject correlation in the repeated measures. This eliminates the need for per-participant scaling of the GSR signal and enhances analysis on subjective ratings by capturing participants' variability via random effects.

Qualitative variables such as driving style and questionnaire responses were numerically encoded to be included in the LME. The driving style was encoded as "0" for "calm" and "1" for "aggressive", while the questionnaire responses were encoded from "1" for "very uncomfortable/unsafe" to "5" for "very comfortable/safe".

To interpret each individual feature contribution, one LME model is generated for each predictor feature and each outcome feature. All predictor and outcome values are standardized across the population, allowing a direct comparison of outcome correlation coefficients. However, when performing multiple statistical tests, the probability of obtaining false-positive results increases with each test. As each test has a chance of incorrectly rejecting the null-hypothesis, many tests cumulatively inflate the overall Type I error rate. To address this multiple comparison problem, the Benjamini-Hochberg False Discovery Rate (FDR) was applied to correct the computed p-values. The Benjamini-Hochberg procedure ranks all p-values and selects a threshold that controls the expected proportion of false positives among the rejected hypotheses. This procedure is less conservative, and therefore more powerful, compared to the alternative Bonferroni correction when conducting a large number of tests.

To explore the effects of specific experimental settings, a complementary statistical analysis was conducted. This analysis assessed whether significant differences existed between events by evaluating selected features. Features measured twice for each participant under identical conditions were first averaged to ensure independence of data and meet the test assumptions. Pairwise comparisons were then performed using a paired t-test after confirming normality with the Shapiro-Wilk test, both implemented in the Python SciPy package. For within-subject comparisons across all five scenarios, a repeated-measures ANOVA was applied from the statsmodels package and for between-subject measures, such as age group or self-reported trust towards automated driving, a one-way ANOVA was used from SciPy. These tests help determine if significant differences exist and which condition produces the highest average values.

During analysis, data from several participants were excluded for the following reasons:

- Participant 19: All data were excluded due to a near-flat GSR signal, indicating either a lack of physiological responsiveness or potential measurement error, while showing clear variability in subjective responses. Such non-responsiveness is consistent with the phenomenon of electrodermal non-responding [33].
- Participants 23 and 29: Data from both participants were excluded entirely as data from one of the two driving styles, either calm or aggressive, was missing due to measurement errors. The lack of comparative data obstructs the analysis, as the analysis involves a within-subject comparison across driving styles.
- Participant 34: All subjective responses were excluded. This participant rated every scenario as "very comfortable" or "very safe" while showing clear GSR variability, suggesting either a potential misunderstanding of the questionnaire or biased response behavior. GSR and driving style data, however, were retained.

B.3. Results

B.3.1. Correlation Analysis

The correlation test is done in eight stages. First, all extracted features are analyzed together across both driving styles in the experiment; these results are shown in Table B.2. Next, the same analysis was performed separately for the calm driving style and the aggressive driving style, with results presented in Tables B.3 and B.4, respectively. Finally, each scenario is analyzed separately across both driving styles. The corresponding results are presented in Tables B.5 to B.9.

		1. style	2. Q1[a]	3. Q2[b]	4. Q3[c]	5. phasic_max	6. phasic_min	7. phasic_mean_td	8. phasic_slope	9. phasic_peak_count	10. phasic_rise_time	11. phasic_recovery_time	12. tonic_mean	13. tonic_std	14. tonic_mean_td	15. tonic_skewness	16. tonic_kurtosis	17. vel_mean	18. vel_max	19. vel_min	20. acc_lon_mean	21. acc_lon_max	22. acc_lat_max	23. acc_lat_min	24. jerk_lon_mean	25. jerk_lon_max	26. jerk_lat_mean	27. jerk_lat_max	28. yaw_rate_max	29. yaw_rate_min	30. d_min	31. d_mean_td	32. d_max_td	33. ttc_min	34. ttc_max_td	35. thw_min	36. thw_max_td
1.	style	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
2.	Q1[a]	**-.74**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
3.	Q2[b]	**-.69**	**.77**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
4.	Q3[c]	**-.79**	**.79**	**.64**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
5.	phasic_max	**.65**	**-.36**	**-.37**	**-.43**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
6.	phasic_min	**.43**	**-.31**	**-.36**	**-.30**	**.50**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
7.	phasic_mean_td	.08	.00	-.00	-.06	.10	-.09	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
8.	phasic_slope	**.21**	-.08	-.08	**-.12**	**.30**	.03	**.64**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
9.	phasic_peak_count	**.52**	**-.37**	**-.40**	**-.37**	**.58**	**.58**	**.13**	**.24**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
10.	phasic_rise_time	**.33**	**-.19**	**-.16**	**-.20**	**.42**	**.20**	**.18**	**.33**	**.42**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
11.	phasic_recovery_time	**.32**	**-.17**	**-.15**	**-.20**	**.37**	**.22**	.11	**.28**	**.48**	**.63**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
12.	tonic_mean	.05	**-.14**	**-.13**	-.07	**.14**	**.29**	**-.28**	**-.28**	**.20**	-.05	.03	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
13.	tonic_std	**.30**	**-.22**	**-.26**	**-.25**	**.45**	**.38**	.00	**.17**	**.42**	**.21**	**.19**	**.19**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
14.	tonic_mean_td	**.30**	**-.17**	**-.24**	**-.19**	**.49**	**.38**	**.17**	**.29**	**.48**	**.34**	**.28**	.02	**.47**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
15.	tonic_skewness	.05	-.05	.02	-.03	-.06	**-.15**	**.22**	**.35**	-.03	.10	.05	**-.20**	-.05	-.09	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
16.	tonic_kurtosis	.02	-.04	.03	-.02	-.03	-.07	**.10**	.03	-.02	.02	.00	.01	**-.24**	-.11	**.39**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
17.	vel_mean	**.54**	**-.42**	**-.42**	**-.40**	**.41**	**.37**	-.08	.01	**.39**	**.15**	**.19**	**.48**	**.38**	**.25**	-.05	-.03	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
18.	vel_max	**.61**	**-.42**	**-.47**	**-.44**	**.56**	**.46**	.07	**.19**	**.52**	**.25**	**.29**	**.31**	**.50**	**.46**	-.03	-.04	**.74**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
19.	vel_min	**-.47**	**.29**	**.26**	**.31**	**-.27**	**-.14**	-.10	**-.16**	**-.20**	**-.18**	**-.17**	.06	-.08	-.07	**-.12**	-.08	-.01	**-.33**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
20.	acc_lon_mean	**.23**	-.08	**-.10**	**-.14**	.00	-.06	**.17**	**.22**	-.01	-.02	.05	**-.20**	-.10	**-.12**	**.17**	.08	-.13	.11	**-.50**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
21.	acc_lon_max	**.90**	**-.56**	**-.53**	**-.61**	**.58**	**.38**	.10	**.24**	**.48**	**.29**	**.30**	.06	**.28**	**.30**	.07	.02	**.53**	**.67**	**-.54**	**.40**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
22.	acc_lat_max	**.35**	**-.25**	**-.36**	**-.26**	**.37**	**.35**	.09	**.22**	**.35**	**.14**	**.18**	.06	**.46**	**.43**	-.03	-.10	**.33**	**.67**	**-.19**	**.24**	**.40**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
23.	acc_lat_min	**-.65**	**.37**	**.33**	**.43**	**-.66**	**-.42**	**-.18**	**-.28**	**-.49**	**-.35**	**-.29**	.01	**-.37**	**-.49**	.05	.06	**-.30**	**-.62**	**.33**	.10	**-.58**	**-.39**	·	·	·	·	·	·	·	·	·	·	·	·	·	·
24.	jerk_lon_mean	**-.38**	**.26**	**.26**	**.29**	**-.18**	**-.15**	-.02	-.01	**-.17**	**-.12**	**-.17**	-.02	-.04	-.07	-.03	-.04	-.02	**-.20**	**.39**	-.06	**-.35**	.04	**.29**	·	·	·	·	·	·	·	·	·	·	·	·	·
25.	jerk_lon_max	**.81**	**-.49**	**-.41**	**-.54**	**.53**	**.36**	.10	**.21**	**.43**	**.28**	**.24**	-.04	**.19**	**.30**	.05	.02	**.35**	**.47**	**-.40**	**.22**	**.83**	**.15**	**-.66**	**-.29**	·	·	·	·	·	·	·	·	·	·	·	·
26.	jerk_lat_mean	-.07	**.37**	**.52**	**.23**	**-.22**	**-.21**	.16	.07	**-.26**	-.06	-.06	**-.48**	**-.46**	**-.26**	.08	.07	**-.88**	**-.89**	.14	**.69**	-.07	**-.49**	**.22**	.02	-.04	·	·	·	·	·	·	·	·	·	·	·
27.	jerk_lat_max	**.63**	**-.41**	**-.44**	**-.45**	**.65**	**.49**	**.13**	**.21**	**.51**	**.29**	**.28**	.12	**.46**	**.51**	-.07	-.09	**.47**	**.78**	**-.29**	-.02	**.60**	**.70**	**-.87**	**-.20**	**.53**	**-.14**	·	·	·	·	·	·	·	·	·	·
28.	yaw_rate_max	**.40**	**-.19**	**-.13**	**-.25**	**.45**	**.24**	**.21**	**.26**	**.33**	**.30**	**.22**	**-.19**	**.16**	**.37**	-.02	-.01	**-.15**	**.34**	**-.42**	-.01	**.36**	.13	**-.84**	**-.30**	**.53**	-.00	**.59**	·	·	·	·	·	·	·	·	·
29.	yaw_rate_min	**-.30**	**.22**	**.33**	**.22**	**-.33**	**-.34**	-.04	**-.17**	**-.30**	-.10	**-.14**	-.07	**-.42**	**-.38**	.05	.11	**-.33**	**-.60**	.05	**-.19**	**-.33**	**-.97**	**.31**	-.05	-.09	**.16**	**-.64**	-.04	·	·	·	·	·	·	·	·
30.	d_min	**.29**	**-.32**	**-.35**	**-.27**	.10	.07	-.10	-.03	.13	.03	.10	**.12**	.14	-.02	.03	-.00	**.34**	**.21**	**-.21**	**.18**	**.30**	**.30**	.08	-.06	.08	-.08	.09	**-.27**	**-.27**	·	·	·	·	·	·	·
31.	d_mean_td	**-.18**	.09	.05	.10	-.04	-.02	-.02	.04	-.09	**-.14**	-.12	.00	.05	-.02	-.05	**-.18**	**.20**	-.10	**.46**	**-.16**	**-.22**	.10	**.14**	**.40**	**-.19**	.09	-.03	**-.39**	**-.18**	-.10	·	·	·	·	·	·
32.	d_max_td	-.05	-.05	-.01	-.05	.03	.17	.01	-.01	.11	.04	.00	.12	.03	.08	-.05	-.08	**.22**	.06	**.42**	**-.70**	-.10	**-.15**	**-.27**	.04	.01	-.03	.17	.07	.09	**-.19**	**.34**	·	·	·	·	·
33.	ttc_min	**.31**	**-.34**	**-.33**	**-.26**	-.11	-.06	**-.31**	**-.22**	-.04	-.11	-.03	.21	-.05	**-.24**	-.01	.05	**.25**	-.02	**-.43**	.35	**.28**	-.06	**.45**	-.26	.15	-.00	**-.27**	**-.51**	.05	**.91**	-.32	**-.22**	·	·	·	·
34.	ttc_max_td	**-.14**	.10	.11	.09	-.02	.03	.13	.02	-.01	.04	-.03	-.07	-.02	.02	.02	-.06	**-.21**	**-.14**	.05	-.16	**-.12**	-.04	-.06	-.06	-.10	.14	-.01	.15	.05	**-.63**	-.01	.01	**-.72**	·	·	·
35.	thw_min	**.22**	**-.23**	**-.23**	**-.22**	-.08	-.08	**-.18**	**-.16**	-.05	-.05	.03	.09	-.10	**-.22**	.04	.09	.13	-.06	**-.29**	**.27**	**.23**	-.08	**.31**	**-.28**	.04	-.00	**-.21**	**-.33**	.08	**.75**	**-.40**	**-.28**	**.58**	**-.23**	·	·
36.	thw_max_td	.01	-.00	.05	-.03	.07	.03	.12	.10	.07	.06	.04	-.05	-.03	.02	.03	-.03	.07	.03	.06	-.05	.03	**-.18**	-.16	-.01	.14	.03	.05	.18	**.16**	-.23	.18	**.86**	**-.16**	.02	-.25	·

[a] How safe did you feel during the car ride?

[b] How safe did you feel interacting with the [pedestrian, roadworks, pedestrian, vehicle]?

[c] How comfortable did you find the movement of the vehicle?

Table B.2: Correlation matrix of all features (driving style, questionnaire responses, GSR, VD, and Perception). Correlation and regression coefficients were determined using separate LMEs for each feature pair. Bold values denote a significant correlation ($p < 0.05$) and underlined values denote non-significant correlations ($p > 0.05$).

Following Table B.2, the following observations can be made with regard to the driving style:

1. **Driving style vs. questionnaire responses:** A strong correlation between driving style and all three questionnaire responses can be found ($\beta = -0.69$ to -0.79). This shows that the "aggressive" driving style consistently made the participants feel less comfortable and less safe during the various scenarios.
2. **Driving style vs. phasic component:** The driving style also shows a strong correlation with sympathetic burst-related phasic features such as the maximum and amount of SCRs. As these are related to physiological arousal, this suggests that the "aggressive" driving style does increase arousal consistently over all participants.
3. **Driving style vs tonic component:** Baseline arousal, as represented by the tonic component, shows more variability with the "aggressive" driving style, as can be seen by the significant correlation with the standard deviation and mean time derivative. Despite being significant, these features show a lower regression coefficient than those from the phasic component, indicating that the phasic component is more responsive to rapid and high-intensity movements, making it the primary physiological marker for moment-to-moment observations.
4. **Driving style vs vehicle dynamics & perception:** As the vehicle dynamics and perception features are inherently embedded in the driving style, these strong correlations simply confirm that the "aggressive" driving style is indeed driven very differently than the "calm" driving style.

Turning to the participants' self-reported scores on perceived comfort and safety, the following becomes apparent:

1. **Questionnaire responses vs. phasic:** Lower comfort/safety scores align with larger, more frequent sympathetic bursts, reflected in the phasic maximum amplitude ($\beta = -0.36$ to -0.43) and peak count ($\beta = -0.37$ to -0.40). The phasic signal, therefore, provides a sensitive, physiological marker for moment-to-moment discomfort.
2. **Questionnaire responses vs. tonic:** Baseline arousal, measured b tonic std., and mean td., also rises as comfort and perceived safety decrease, though the effect is weaker ($\beta = -0.22$ to -0.26, $\beta = -0.17$ to -0.24). The tonic component captures a slower background tension, rather than the acute responses as captured by the phasic component.
3. **Questionnaire response vs vehicle dynamics & perception:** Almost every motion or perception metric correlates with the scores, yet the strongest correlations come from longitudinal-related motions, particularly the maximum acceleration and maximum jerk, while distance-keeping related features (TTC, THW) contribute less. This suggests that the participants assess comfort and perceived safety primarily based on how the vehicle moves rather than its proximity to other objects.

Many of these same features show an even stronger correlation with the driving style. Since driving style also has the highest correlation to the questionnaire responses, this raises the concern that the correlations may reflect the differences in driving style, rather than being directly linked to the questionnaire responses. To address this concern, the same correlation analysis is done for data recorded only during the "calm" or "aggressive" driving style, these results are listed in Tables B.3 and B.4, respectively, and will be discussed later.

Focusing on the phasic component highlights the following:

1. **Phasic component vs. vehicle dynamics:** All phasic features, especially phasic max. and peak count, are responsive to motion minima and maxima. Only the mean td. fails occasionally in finding a significant correlation to these extremes. This pattern reinforces that abrupt, high-intensity maneuvers are primary triggers of sympathetic bursts.
2. **Phasic component vs. perception:** The phasic mean td. and slope are significantly correlated to the minimum TTC and THW. The phasic rise time is significantly correlated to the mean td. of the distance to the object.

Finally, regarding the tonic component, the following observations can be made:

1. **Tonic component vs. vehicle dynamics:** Baseline arousal also rises with motion extremes, mirroring the phasic results. Shape metrics like skewness and kurtosis remain non-informative, however. The slightly larger coefficient for the tonic std. and mean td. suggest that the sustained motion load shapes the tonic level more than isolated bursts.
2. **Tonic component vs. perception:** Except for a modest correlation between tonic mean td. and both TTC and THW, perception metrics show no influence to the tonic activity. Therefore, baseline arousal is only marginally sensitive to environmental cues.

Among all physiological features, the phasic maximum amplitude and peak count show the strongest correlation to the subjective comfort metrics Q1, Q2, and Q3, making them the most indicative GSR-based features. As these correlations are statistically significant, they can be meaningfully interpreted with perceived comfort and safety.

Furthermore, both phasic and tonic features prioritize motion intensity and extremes over direction or environmental cues: longitudinal and lateral extremes elicit comparable correlation coefficients, while perception features contribute minimally. Correlations among vehicle dynamics and perception features mainly confirm that the “calm” and “aggressive” driving styles differ as intended; a pairwise comparison between them adds little explanatory power for passenger state.

		1. style	2. Q1[a]	3. Q2[b]	4. Q3[c]	5. phasic_max	6. phasic_min	7. phasic_mean_td	8. phasic_slope	9. phasic_peak_count	10. phasic_rise_time	11. phasic_recovery_time	12. tonic_mean	13. tonic_std	14. tonic_mean_td	15. tonic_skewness	16. tonic_kurtosis	17. vel_mean	18. vel_max	19. vel_min	20. acc_lon_mean	21. acc_lon_max	22. acc_lat_max	23. acc_lat_min	24. jerk_lon_mean	25. jerk_lon_max	26. jerk_lat_mean	27. jerk_lat_max	28. yaw_rate_max	29. yaw_rate_min	30. d_min	31. d_mean_td	32. d_max_td	33. ttc_min	34. ttc_max_td	35. thw_min	36. thw_max_td
1.	style	-	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
2.	Q1[a]	-	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
3.	Q2[b]	-	**.47**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
4.	Q3[c]	-	**.47**	**.27**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
5.	phasic_max	-	-.02	-.09	-.02	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
6.	phasic_min	-	-.08	-.06	-.07	**.58**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
7.	phasic_mean_td	-	.02	-.02	.01	-.21	**-.31**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
8.	phasic_slope	-	-.01	-.05	.04	-.21	**-.19**	**.54**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
9.	phasic_peak_count	-	.03	-.10	.03	**.61**	**.64**	.06	.08	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
10.	phasic_rise_time	-	.11	.02	.07	**.39**	**.24**	.11	**.21**	**.37**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
11.	phasic_recovery_time	-	.12	-.01	.06	**.37**	**.24**	.03	.13	**.51**	**.85**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
12.	tonic_mean	-	-.13	-.05	-.09	**.33**	**.34**	**-.24**	**-.26**	**.18**	.05	.06	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
13.	tonic_std	-	-.13	-.06	-.11	**.47**	**.33**	-.03	-.08	**.29**	**.18**	.13	**.32**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
14.	tonic_mean_td	-	-.04	-.06	.01	**.34**	**.33**	.07	.06	**.31**	**.23**	**.18**	.07	.19	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
15.	tonic_skewness	-	-.12	-.02	-.12	-.16	**-.18**	**.24**	**.46**	-.05	.01	.01	-.08	-.02	-.08	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
16.	tonic_kurtosis	-	-.11	-.04	-.07	.01	.06	.07	.01	.02	.03	.08	.13	-.09	-.07	**.22**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
17.	vel_mean	-	-.01	-.08	-.04	**.22**	**.22**	-.11	-.11	**.12**	.06	.03	**.42**	**.28**	**.21**	-.05	.02	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
18.	vel_max	-	-.07	-.06	-.06	**.32**	**.33**	-.08	-.08	**.21**	**.14**	.13	**.41**	**.33**	**.32**	-.06	.11	**.72**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
19.	vel_min	-	.08	-.01	-.00	.06	.05	-.11	-.12	-.01	-.03	-.03	.05	.09	.11	-.12	-.10	**.35**	-.07	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
20.	acc_lon_mean	-	.03	.04	.03	**-.32**	**-.39**	**.17**	.12	**-.21**	**-.15**	-.09	**-.16**	**-.27**	**-.42**	**.18**	-.04	**-.33**	**-.30**	**-.47**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
21.	acc_lon_max	-	-.09	-.07	-.05	**.23**	**.24**	-.08	-.05	**.16**	.13	.06	**.33**	**.19**	**.14**	.00	.02	**.46**	**.59**	**-.32**	.05	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
22.	acc_lat_max	-	.09	.02	.05	**-.19**	**-.22**	-.04	-.05	**-.18**	**-.17**	**-.15**	**-.33**	**-.17**	**-.22**	-.04	-.11	**-.42**	**-.73**	**.50**	.10	**-.55**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
23.	acc_lat_min	-	.06	.00	.01	**-.33**	**-.37**	.07	.07	**-.25**	**-.17**	**-.17**	-.13	**-.26**	**-.40**	**.14**	-.14	.04	**-.47**	**.17**	**.60**	-.11	**.42**	·	·	·	·	·	·	·	·	·	·	·	·	·	·
24.	jerk_lon_mean	-	.07	.01	.06	-.08	-.06	-.04	-.06	-.07	-.10	-.11	.07	-.05	**-.18**	.00	-.07	**.39**	-.11	**.28**	.08	.08	.11	.50	·	·	·	·	·	·	·	·	·	·	·	·	·
25.	jerk_lon_max	-	-.04	-.02	.02	**.25**	.35	-.14	-.09	**.24**	.11	.07	**.18**	**.17**	**.36**	-.11	.07	**.15**	**.40**	-.02	**-.54**	**.41**	**-.34**	**-.74**	-.23	·	·	·	·	·	·	·	·	·	·	·	·
26.	jerk_lat_mean	-	.05	.07	.00	-.21	**-.24**	.09	.05	-.14	-.04	.04	-.27	**-.16**	-.23	.01	-.05	**-.41**	**-.99**	**.23**	**.42**	**-.60**	**.48**	**.56**	.09	**-.57**	·	·	·	·	·	·	·	·	·	·	·
27.	jerk_lat_max	-	-.06	-.02	-.02	**.37**	**.39**	-.11	-.13	**.25**	**.16**	**.16**	**.21**	**.32**	**.47**	**-.16**	.13	**.25**	**.63**	.02	**-.67**	**.17**	**-.43**	**-.91**	**-.37**	**.61**	-.14	·	·	·	·	·	·	·	·	·	·
28.	yaw_rate_max	-	-.06	.04	.02	.13	**.16**	.04	.02	**.13**	.13	.12	-.10	.05	**.20**	-.06	.12	**-.53**	.09	**-.51**	**-.19**	-.02	**-.30**	**-.81**	**-.60**	**.41**	**-.18**	**.58**	·	·	·	·	·	·	·	·	·
29.	yaw_rate_min	-	-.11	-.03	-.06	.13	**.17**	.06	.08	**.16**	.13	.12	**.20**	.11	.14	.06	**.15**	.12	**.52**	**-.69**	-.03	**.45**	**-.89**	**-.47**	**-.19**	**.35**	**-.29**	**.37**	**.50**	·	·	·	·	·	·	·	·
30.	d_min	-	.01	-.12	-.20	-.09	-.09	.16	.02	-.09	-.07	-.08	-.05	-.01	-.12	.09	-.02	**-.90**	-.33	-.47	.12	-.41	-.07	.14	-.50	-.14	-.00	**-.16**	.66	.15	·	·	·	·	·	·	·
31.	d_mean_td	-	.05	-.06	-.00	**-.17**	-.14	-.02	-.04	-.11	**-.24**	**-.21**	.02	-.06	**-.17**	-.03	**-.18**	**.40**	**-.20**	**.45**	.12	-.09	**.35**	**.71**	**.51**	**-.33**	**.35**	**-.47**	**-.85**	**-.49**	-.13	·	·	·	·	·	·
32.	d_max_td	-	-.08	-.17	-.03	.06	.08	.02	-.07	.13	.14	.03	.13	.18	.14	-.01	-.02	.44	**.28**	**.75**	**-.55**	.07	-.11	-.07	.08	.01	.07	.17	**-.29**	-.01	-.08	**.34**	·	·	·	·	·
33.	ttc_min	-	.06	-.03	-.12	-.18	-.15	-.03	.05	-.13	-.09	-.08	-.09	-.04	**-.18**	.06	-.03	-.73	-.23	**-.56**	.27	.06	-.12	.07	-.12	-.07	-.00	-.17	.41	**.23**	**.57**	-.26	-.18	·	·	·	·
34.	ttc_max_td	-	-.09	.00	-.08	.04	.05	-.03	.00	-.06	.10	-.03	-.06	-.01	.04	.08	-.07	-.14	-.16	-.06	-.08	-.04	.02	-.02	.01	.02	.10	-.05	.09	.05	-.31	-.04	-.05	**-.80**	·	·	·
35.	thw_min	-	-.01	-.07	-.13	-.08	-.22	.05	.02	-.10	.02	-.05	-.33	-.19	-.22	.08	.19	**-.83**	-.25	**-.75**	.54	-.06	-.08	-.06	**-.72**	**-.17**	-.06	-.16	**.96**	.16	**.67**	**-.96**	-.22	**.76**	-.11	·	·
36.	thw_max_td	-	-.11	-.12	-.02	.04	.10	.03	.02	.12	.15	.03	.09	.06	.06	.02	-.01	**.29**	.19	.18	-.31	.08	-.15	.08	.01	-.01	.08	.14	-.21	-.06	-.08	**.29**	**.91**	-.17	-.03	-.18	·

[a] How safe did you feel during the car ride?

[b] How safe did you feel interacting with the [pedestrian, roadworks, pedestrian, vehicle]?

[c] How comfortable did you find the movement of the vehicle?

Table B.3: Correlation matrix of all features recorded during the "calm" driving style (questionnaire responses, GSR, VD, and Perception). Correlation and regression coefficients were determined using separate LMEs for each feature pair. Bold values denote a significant correlation ($p < 0.05$) and underlined values denote non-significant correlations ($p > 0.05$).

		1. style	2. Q1[a]	3. Q2[b]	4. Q3[c]	5. phasic_max	6. phasic_min	7. phasic_mean_td	8. phasic_slope	9. phasic_peak_count	10. phasic_rise_time	11. phasic_recovery_time	12. tonic_mean	13. tonic_std	14. tonic_mean_td	15. tonic_skewness	16. tonic_kurtosis	17. vel_mean	18. vel_max	19. vel_min	20. acc_lon_mean	21. acc_lon_max	22. acc_lat_max	23. acc_lat_min	24. jerk_lon_mean	25. jerk_lon_max	26. jerk_lat_mean	27. jerk_lat_max	28. yaw_rate_max	29. yaw_rate_min	30. d_min	31. d_mean_td	32. d_max_td	33. ttc_min	34. ttc_max_td	35. thw_min	36. thw_max_td
1.	style	-	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
2.	Q1[a]	-	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
3.	Q2[b]	-	**.58**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
4.	Q3[c]	-	**.59**	**.41**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
5.	phasic_max	-	.11	.04	.01	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
6.	phasic_min	-	-.04	-.06	.01	**.30**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
7.	phasic_mean_td	-	.03	08	.01	**.14**	**-.14**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
8.	phasic_slope	-	.10	.06	.01	**.36**	-.08	**.63**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
9.	phasic_peak_count	-	-.04	**-.16**	-.02	**.26**	**.46**	.09	**.14**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
10.	phasic_rise_time	-	.03	.10	.03	**.29**	.03	**.21**	**.41**	.02	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
11.	phasic_recovery_time	-	.05	.09	.02	**.16**	.07	.10	**.29**	**.18**	**.40**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
12.	tonic_mean	-	-.06	-.08	-.05	.01	**.40**	**-.34**	**-.40**	**.21**	**-.32**	-.11	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
13.	tonic_std	-	-.07	**-.18**	-.08	**.34**	**.28**	-.06	.11	**.28**	.11	.04	**.17**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
14.	tonic_mean_td	-	.05	-.09	-.01	**.47**	**.29**	**.18**	**.34**	**.40**	**.28**	**.21**	-.13	**.55**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
15.	tonic_skewness	-	.01	.10	.05	-.11	**-.28**	**.22**	**.37**	-.12	**.14**	.01	**-.33**	**-.16**	**-.15**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
16.	tonic_kurtosis	-	.01	.08	.01	-.11	-.10	.13	.08	-.08	-.00	-.06	-.08	**-.28**	-.11	**.48**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
17.	vel_mean	-	**-.24**	**-.31**	-.09	.01	**.25**	**-.22**	**-.23**	**.22**	**-.18**	-.01	**.67**	**.34**	.05	**-.15**	-.10	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
18.	vel_max	-	-.07	**-.25**	-.06	**.28**	**.31**	.06	**.14**	**.36**	-.01	.13	**.31**	**.49**	**.41**	-.11	**-.15**	**.56**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
19.	vel_min	-	**-.14**	-.07	-.03	.01	.16	-.18	-.18	.15	.04	-.05	**.24**	.11	.05	-.16	-.09	.33	-.07	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
20.	acc_lon_mean	-	**.12**	.04	.02	**-.15**	**-.16**	**.15**	**.20**	-.08	-.08	.01	**-.27**	**-.14**	-.08	**.16**	.12	**-.37**	.09	**-.73**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
21.	acc_lon_max	-	.05	-.04	-.01	.03	-.05	.08	**.21**	.04	-.03	.09	-.11	-.00	.06	.11	.00	-.13	**.28**	**-.47**	**.57**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
22.	acc_lat_max	-	-.06	**-.23**	-.05	**.24**	**.26**	.06	**.18**	**.31**	.06	.12	**.15**	**.54**	**.48**	-.10	-.11	**.40**	**.84**	**-.21**	**.19**	**.27**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
23.	acc_lat_min	-	-.09	-.04	-.01	**-.51**	**-.22**	**-.22**	**-.26**	**-.24**	**-.27**	**-.14**	**.14**	**-.25**	**-.44**	.12	**.18**	.14	**-.38**	-.06	**.27**	.01	**-.26**	·	·	·	·	·	·	·	·	·	·	·	·	·	·
24.	jerk_lon_mean	-	-.00	-.03	.02	.15	.03	.07	.13	.17	**.19**	.01	-.08	**.28**	**.31**	-.05	-.09	.06	**.22**	**.20**	.05	-.04	**.44**	**-.36**	·	·	·	·	·	·	·	·	·	·	·	·	·
25.	jerk_lon_max	-	**.13**	**.19**	.06	.04	-.02	.09	**.14**	-.01	.07	-.02	**-.28**	**-.15**	.05	.08	.00	**-.46**	**-.17**	-.10	**.16**	**.41**	**-.23**	**-.28**	.01	·	·	·	·	·	·	·	·	·	·	·	·
26.	jerk_lat_mean	-	**.37**	**.71**	.10	-.35	**-.58**	.35	.30	**-.40**	.08	-.02	**-.66**	**-.81**	**-.51**	.33	.24	**-.76**	**-.81**	**-.80**	**.78**	**.47**	**-.77**	.18	**-.32**	**.91**	·	·	·	·	·	·	·	·	·	·	·
27.	jerk_lat_max	-	-.02	**-.16**	-.04	**.46**	**.35**	.11	**.17**	**.34**	**.14**	**.15**	**.16**	**.44**	**.48**	**-.17**	**-.20**	**.29**	**.73**	-.04	**-.17**	.11	**.68**	**-.77**	.11	.04	-.14	·	·	·	·	·	·	·	·	·	·
28.	yaw_rate_max	-	**.16**	**.15**	.03	**.42**	.09	**.28**	**.32**	**.17**	**.27**	**.14**	**-.32**	.06	**.33**	-.03	-.12	**-.44**	**.17**	.03	-.07	.05	.01	**-.89**	.12	**.42**	.12	**.53**	·	·	·	·	·	·	·	·	·
29.	yaw_rate_min	-	.07	**.24**	.05	**-.25**	**-.29**	-.01	**-.14**	**-.31**	-.05	-.10	**-.20**	**-.52**	**-.45**	.12	.11	**-.44**	**-.85**	**.19**	-.15	**-.24**	**-.99**	**.25**	**-.18**	**.24**	**.23**	**-.68**	.00	·	·	·	·	·	·	·	·
30.	d_min	-	**-.17**	-.14	-.05	-.12	-.04	-.16	-.16	.04	-.07	.06	**.22**	.04	-.12	.04	.01	.41	.07	-.08	.12	.05	**.22**	**.45**	.08	**-.31**	-.08	-.15	**-.59**	**-.24**	·	·	·	·	·	·	·
31.	d_mean_td	-	-.10	-.09	-.05	**.22**	**.19**	-.07	.07	**.22**	**.17**	**.22**	.02	**.34**	**.32**	-.10	**-.19**	**.24**	**.29**	**.37**	**-.44**	**-.22**	**.28**	**-.44**	**.22**	-.01	-.14	**.41**	**.28**	**-.27**	-.01	·	·	·	·	·	·
32.	d_max_td	-	-.08	.01	-.06	**.20**	**.25**	-.02	-.06	.15	.08	-.00	.11	-.00	.10	-.10	-.10	.12	-.01	**.58**	**-.71**	**-.33**	**-.23**	**-.46**	-.05	.10	-.12	**.28**	**.39**	.16	**-.32**	**.35**	·	·	·	·	·
33.	ttc_min	-	-.16	-.11	-.06	**-.48**	-.17	**-.33**	**-.47**	-.20	-.30	-.11	.32	-.16	**-.56**	-.03	.07	.27	-.37	-.07	.19	-.01	-.17	**.73**	-.14	-.23	.01	**-.75**	**-.87**	.16	.23	-.61	**-.42**	·	·	·	·
34.	ttc_max_td	-	.09	.05	.04	.17	.15	.12	.04	.12	**.09**	.09	-.03	.11	.20	-.11	-.13	-.23	.07	.17	-.19	-.05	.01	**-.64**	-.04	.02	.10	.43	**.54**	-.01	**-.87**	**.09**	.14	**-.95**	·	·	·
35.	thw_min	-	**-.12**	-.04	-.05	**-.36**	**-.22**	**-.24**	**-.31**	**-.20**	**-.20**	-.09	**.22**	**-.25**	**-.40**	.04	.11	**.21**	**-.34**	**-.18**	**.21**	.02	**-.21**	**.82**	-.15	**-.30**	.03	**-.56**	**-.83**	.17	**.67**	**-.54**	**-.43**	**.54**	**-.26**	·	·
36.	thw_max_td	-	.10	**.22**	.03	.15	.03	**.16**	.09	-.02	.27	.03	**-.20**	-.18	.04	.06	.00	-.46	-.16	.15	-.31	.01	-.28	**-.54**	-.06	**.33**	.00	.08	**.87**	.35	**-.68**	.08	**.92**	**-.22**	.14	**-.72**	·

[a] How safe did you feel during the car ride?

[b] How safe did you feel interacting with the [pedestrian, roadworks, pedestrian, vehicle]?

[c] How comfortable did you find the movement of the vehicle?

Table B.4: Correlation matrix of all features recorded during the "aggressive" driving style (questionnaire responses, GSR, VD, and Perception). Correlation and regression coefficients were determined using separate LMEs for each feature pair. Bold values denote a significant correlation ($p < 0.05$) and underlined values denote non-significant correlations ($p > 0.05$).

Tables B.3 and B.4 present the correlation matrices after statistically hiding the driving style constant. With the driving style variance removed, virtually all phasic and tonic features fall to non-significance when compared to the questionnaire responses, except for the peak count and tonic standard deviation for Q2. A similar effect can be seen for the vehicle dynamics and perception features, except for a few correlating ones in the "aggressive" analysis.
This pattern underscores that the differences in perceived comfort and safety are primarily caused by the driving style itself. While physiological, vehicle dynamics and perception features can distinguish between broadly different comfort levels, such as those between the calm and aggressive driving style, they are largely insensitive to more subtle variations. An interesting trend emerges from Table B.4: The negative correlations between the velocity-related features and Q1/Q2 suggest that scenarios with higher driving speeds were perceived as less comfortable and safe. In contrast, the positive correlation with lateral jerk indicates that scenarios involving more lateral movement were rated more positively. Additionally, a negative correlation with lateral jerk can be found with GSR-related features, suggesting that more lateral movement corresponds to a calmer physiological state, implying that participants experienced these scenarios subjectively differently but also physiologically. The significant correlations observed under the "aggressive" driving style between vehicle dynamics and GSR-related features indicate that participants' physiological responses varied across scenarios, reflecting variations in how scenarios are perceived and processed on a physiological level.

The correlation analyses done in Table B.2 to B.4 are repeated for each of the five scenarios separately. Tables B.5 to B.9 represent these results.

		1. style	2. Q1[a]	3. Q2[b]	4. Q3[c]	5. phasic_max	6. phasic_min	7. phasic_mean_td	8. phasic_slope	9. phasic_peak_count	10. phasic_rise_time	11. phasic_recovery_time	12. tonic_mean	13. tonic_std	14. tonic_mean_td	15. tonic_skewness	16. tonic_kurtosis	17. vel_mean	18. vel_max	19. vel_min	20. acc_lon_mean	21. acc_lon_max	22. acc_lat_max	23. acc_lat_min	24. jerk_lon_mean	25. jerk_lon_max	26. jerk_lat_mean	27. jerk_lat_max	28. yaw_rate_max	29. yaw_rate_min	30. d_min	31. d_mean_td	32. d_max_td	33. ttc_min	34. ttc_max_td	35. thw_min	36. thw_max_td
1.	style	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
2.	Q1[a]	**-.57**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
3.	Q2[b]	**-.41**	**.64**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
4.	Q3[c]	**-.65**	**.67**	**.52**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
5.	phasic_max	**.88**	**-.43**	**-.35**	**-.53**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
6.	phasic_min	**.56**	-.17	-.19	**-.31**	**.65**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
7.	phasic_mean_td	**.35**	-.19	-.19	**-.29**	.23	-.09	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
8.	phasic_slope	**.59**	**-.31**	-.20	**-.34**	**.42**	-.02	**.64**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
9.	phasic_peak_count	**.39**	.01	-.19	-.24	**.40**	**.35**	.15	**.19**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
10.	phasic_rise_time	**.47**	-.05	-.10	-.23	**.40**	**.19**	.14	**.23**	.19	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
11.	phasic_recovery_time	**.33**	.00	-.04	-.07	**.30**	.20	.01	.16	**.46**	**.79**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
12.	tonic_mean	**-.30**	.24	**.29**	.18	-.03	**.25**	-.08	**-.25**	.03	-.20	-.14	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
13.	tonic_std	**.45**	-.04	-.04	-.27	**.42**	**.48**	.07	.08	**.26**	.16	.09	**.32**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
14.	tonic_mean_td	**.71**	**-.36**	**-.26**	**-.44**	**.45**	**.33**	.15	**.36**	**.30**	.24	.23	**-.33**	.23	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
15.	tonic_skewness	.13	-.15	.04	-.13	.10	-.05	**.30**	**.44**	.01	-.00	-.03	.00	-.11	.04	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
16.	tonic_kurtosis	.10	-.16	-.03	-.14	.19	-.06	.22	.16	.07	-.02	-.08	-.02	-.15	-.01	.28	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
17.	vel_mean	**.99**	**-.51**	**-.43**	**-.64**	**.66**	**.39**	**.28**	**.46**	**.36**	**.45**	**.32**	**-.24**	**.32**	**.52**	.11	.10	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
18.	vel_max	**.99**	**-.50**	**-.43**	**-.65**	**.66**	**.40**	**.27**	**.45**	**.35**	**.46**	**.32**	**-.24**	**.33**	**.53**	.10	.08	**.98**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
19.	vel_min	-.51	.25	.16	**.37**	**-.33**	**-.29**	-.02	-.09	**-.21**	**-.28**	-.21	.15	-.04	**-.34**	-.06	.02	**-.39**	-.51	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
20.	acc_lon_mean	**.98**	**-.55**	**-.46**	**-.66**	**.64**	**.38**	**.24**	**.43**	**.35**	**.42**	**.29**	**-.25**	**.31**	**.49**	.09	.09	**.97**	**.98**	**-.50**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
21.	acc_lon_max	**.98**	**-.50**	**-.43**	**-.66**	**.65**	**.40**	**.30**	**.46**	**.36**	**.44**	**.32**	-.21	**.33**	**.56**	.10	.10	**.99**	**.95**	**-.45**	**.87**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
22.	acc_lat_max	.77	**-.41**	**-.37**	**-.56**	**.47**	**.28**	.21	**.39**	**.30**	**.33**	**.26**	-.18	**.30**	**.42**	.07	.06	.74	.75	**-.40**	**.65**	**.78**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
23.	acc_lat_min	**-.94**	**.48**	**.50**	**.69**	-.68	**-.42**	**-.27**	**-.46**	**-.39**	**-.49**	**-.36**	.18	**-.34**	**-.53**	-.11	-.09	**-.93**	**-.97**	**.52**	**-.91**	**-.97**	**-.82**	·	·	·	·	·	·	·	·	·	·	·	·	·	·
24.	jerk_lon_mean	-.08	-.05	-.04	.08	.03	-.13	.20	.08	-.01	-.06	-.12	-.11	-.01	-.08	.13	.18	.08	-.05	**.80**	.06	-.02	-.15	.05	·	·	·	·	·	·	·	·	·	·	·	·	·
25.	jerk_lon_max	.98	**-.54**	**-.52**	**-.74**	**.56**	**.38**	**.23**	.39	**.35**	**.35**	**.24**	-.19	**.31**	.50	.09	.10	**.94**	**.93**	**-.45**	.92	.97	**.79**	**-.98**	.05	·	·	·	·	·	·	·	·	·	·	·	·
26.	jerk_lat_mean	.12	.29	.02	-.11	-.00	.16	-.04	-.02	-.88	-.03	-.35	.17	.96	.08	-.07	-.08	-.07	.13	-.02	-.36	.09	.10	-.18	.28	-.09	·	·	·	·	·	·	·	·	·	·	·
27.	jerk_lat_max	**.94**	**-.47**	**-.47**	**-.69**	**.66**	**.38**	**.29**	**.48**	**.37**	**.46**	**.34**	-.16	**.32**	**.50**	.14	.11	**.95**	**.95**	**-.50**	**.90**	**.95**	**.86**	**-.98**	.00	**.88**	.08	·	·	·	·	·	·	·	·	·	·
28.	yaw_rate_max	**.75**	-.22	**-.21**	**-.47**	**.59**	**.37**	**.25**	**.38**	**.39**	**.43**	**.38**	-.06	**.31**	**.40**	.17	.10	**.85**	**.76**	**-.59**	**.77**	**.68**	**.59**	**-.90**	.07	**.78**	.06	**.91**	·	·	·	·	·	·	·	·	·
29.	yaw_rate_min	-.38	**.45**	**.41**	**.43**	**-.23**	-.14	-.09	-.13	-.13	-.16	-.09	**.19**	-.14	**-.22**	-.10	-.02	**-.37**	-.39	**.22**	**-.39**	**-.40**	**-.30**	**.36**	-.08	**-.47**	.14	**-.31**	**-.30**	·	·	·	·	·	·	·	·
30.	d_min	-.36	**.35**	**.37**	.25	**-.23**	-.04	-.20	**-.23**	-.23	-.03	-.04	.18	-.09	-.14	-.11	-.12	-.45	-.39	-.27	**-.46**	-.35	-.19	.24	-.28	**-.33**	.15	-.25	-.15	.02	·	·	·	·	·	·	·
31.	d_mean_td	**.60**	**-.45**	**-.40**	**-.50**	.22	.11	.08	**.22**	.16	.23	.20	-.27	.06	.23	.03	-.10	**.58**	**.58**	-.25	**.63**	**.54**	**.40**	**-.41**	.18	**.61**	-.01	**.40**	.19	-.21	**-.57**	·	·	·	·	·	·
32.	d_max_td	.64	-.31	-.22	**-.34**	**.31**	.04	.13	**.27**	.23	**.39**	.22	**-.25**	.20	**.45**	.03	.04	.66	**.63**	**-.32**	**.61**	**.67**	.59	**-.57**	.19	.61	-.09	.68	**.52**	**-.48**	**-.37**	**.49**	·	·	·	·	·
33.	ttc_min	**-.64**	**.37**	**.38**	**.53**	**-.39**	-.10	**-.27**	**-.32**	-.35	**-.32**	-.19	**.30**	-.07	**-.40**	-.15	-.15	**-.74**	**-.67**	.25	**-.67**	**-.65**	**-.51**	**.63**	-.29	**-.63**	.12	**-.65**	**-.61**	.21	**.57**	-.33	**-.55**	·	·	·	·
34.	ttc_max_td	-.20	.08	.07	.03	-.13	.03	-.02	-.05	**-.20**	**-.26**	-.25	.13	-.00	**-.23**	.11	-.05	**-.29**	-.22	.49	**-.34**	-.20	-.14	.17	-.16	**-.30**	.20	-.16	-.18	**.18**	-.03	-.11	-.24	.17	·	·	·
35.	thw_min	-.58	.39	**.36**	**.43**	-.28	-.09	**-.21**	**-.28**	**-.28**	-.19	-.12	**.25**	-.13	-.27	-.10	-.07	**-.68**	-.60	-.09	**-.72**	**-.58**	-.40	.48	**-.30**	**-.61**	.13	**-.47**	-.34	.16	**.91**	**-.68**	**-.56**	**.77**	.06	·	·
36.	thw_max_td	**.82**	**-.44**	**-.35**	**-.55**	**.38**	.16	**.25**	**.32**	.22	**.47**	**.32**	**-.26**	.26	**.46**	.05	-.02	**.88**	**.89**	**-.40**	**.78**	**.82**	**.72**	**-.80**	.13	.79	-.09	**.88**	**.75**	-.31	**-.39**	**.37**	**.79**	**-.56**	-.26	**-.54**	·

[a] How safe did you feel during the car ride?

[b] How safe did you feel interacting with the pedestrian?

[c] How comfortable did you find the movement of the vehicle?

Table B.5: Correlation matrix of all features recorded during the pedestrian crossing without obstruction scenario (driving style, questionnaire responses, GSR, VD, and Perception). Correlation and regression coefficients were determined using separate LMEs for each feature pair. Bold values denote a significant correlation ($p < 0.05$) and underlined values denote non-significant correlations ($p > 0.05$).

		1. style	2. Q1[a]	3. Q2[b]	4. Q3[c]	5. phasic_max	6. phasic_min	7. phasic_mean_td	8. phasic_slope	9. phasic_peak_count	10. phasic_rise_time	11. phasic_recovery_time	12. tonic_mean	13. tonic_std	14. tonic_mean_td	15. tonic_skewness	16. tonic_kurtosis	17. vel_mean	18. vel_max	19. vel_min	20. acc_lon_mean	21. acc_lon_max	22. acc_lat_max	23. acc_lat_min	24. jerk_lon_mean	25. jerk_lon_max	26. jerk_lat_mean	27. jerk_lat_max	28. yaw_rate_max	29. yaw_rate_min	30. d_min	31. d_mean_td	32. d_max_td	33. ttc_min	34. ttc_max_td	35. thw_min	36. thw_max_td
1.	style	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
2.	Q1[a]	**-.71**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
3.	Q2[b]	**-.79**	**.86**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
4.	Q3[c]	**-.69**	**.90**	**.80**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
5.	phasic_max	**.96**	**-.66**	**-.70**	**-.71**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
6.	phasic_min	**.85**	**-.58**	**-.68**	**-.61**	**.72**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
7.	phasic_mean_td	.25	-.14	-.11	-.19	.16	-.05	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
8.	phasic_slope	**.53**	**-.32**	**-.35**	**-.28**	**.44**	**.27**	**.41**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
9.	phasic_peak_count	**.83**	**-.75**	**-.80**	**-.68**	.67	**.60**	.16	**.35**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
10.	phasic_rise_time	**.58**	**-.42**	**-.44**	**-.37**	**.52**	**.36**	**.23**	.46	**.35**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
11.	phasic_recovery_time	.52	**-.35**	**-.35**	**-.37**	.48	**.40**	.21	**.35**	**.58**	**.98**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
12.	tonic_mean	**.54**	**-.34**	**-.35**	**-.36**	**.39**	**.45**	**-.24**	-.09	**.45**	.01	.14	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
13.	tonic_std	**.95**	**-.67**	**-.61**	**-.68**	**.79**	**.67**	**.24**	**.43**	**.66**	**.58**	**.56**	**.39**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
14.	tonic_mean_td	**.93**	**-.67**	**-.69**	**-.63**	**.70**	**.58**	**.33**	**.60**	**.67**	**.60**	**.49**	.25	**.75**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
15.	tonic_skewness	.01	-.07	-.02	-.04	.01	-.18	**.43**	**.44**	-.05	.11	.07	**-.50**	.01	.15	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
16.	tonic_kurtosis	-.11	.16	.11	.12	-.08	.01	.17	.05	.01	-.01	-.05	-.20	-.12	.07	.23	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
17.	vel_mean	**.98**	**-.72**	**-.72**	**-.69**	**.72**	**.64**	.18	**.40**	**.63**	**.52**	**.49**	**.41**	**.71**	**.70**	.01	-.08	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
18.	vel_max	**.95**	**-.73**	**-.72**	**-.70**	**.72**	**.64**	.19	**.41**	**.64**	**.51**	**.50**	**.40**	**.71**	**.70**	.02	-.08	**.96**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
19.	vel_min	**-.95**	**.71**	**.70**	**.68**	**-.71**	**-.63**	-.19	**-.41**	**-.62**	**-.54**	**-.49**	**-.40**	**-.70**	**-.71**	-.02	.09	**-.99**	**-.98**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
20.	acc_lon_mean	**.97**	**-.73**	**-.72**	**-.69**	**.71**	**.62**	.18	**.40**	**.62**	**.53**	**.47**	**.40**	**.69**	**.70**	.02	-.08	**.98**	**.99**	**-.93**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
21.	acc_lon_max	**.99**	**-.72**	**-.72**	**-.68**	**.71**	**.60**	.18	**.44**	**.60**	**.52**	**.44**	**.39**	**.69**	**.70**	.02	-.11	**.98**	**.99**	**-.99**	**.98**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
22.	acc_lat_max	**.96**	**-.73**	**-.72**	**-.71**	**.71**	**.64**	.19	**.41**	**.64**	**.50**	**.46**	**.40**	**.70**	**.70**	.02	-.06	**.95**	**.97**	**-.95**	**.95**	**.95**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
23.	acc_lat_min	**-.98**	**.73**	**.71**	**.69**	**-.72**	**-.64**	-.18	**-.40**	**-.64**	**-.53**	**-.49**	**-.41**	**-.71**	**-.70**	-.01	.08	**-.99**	**-.99**	**.99**	**-.99**	**-.96**	**-.97**	·	·	·	·	·	·	·	·	·	·	·	·	·	·
24.	jerk_lon_mean	**-.76**	**.61**	**.58**	**.59**	**-.55**	**-.43**	-.07	**-.35**	**-.51**	**-.41**	**-.37**	**-.33**	**-.51**	**-.52**	-.08	-.05	-.74	**-.75**	**.77**	**-.79**	-.74	**-.73**	**.75**	·	·	·	·	·	·	·	·	·	·	·	·	·
25.	jerk_lon_max	**.94**	**-.81**	**-.75**	**-.74**	**.69**	**.53**	.17	**.44**	**.63**	**.49**	**.41**	.36	**.69**	**.71**	.04	-.08	**.94**	**.95**	**-.94**	**.99**	**.98**	**.99**	**-.96**	**-.62**	·	·	·	·	·	·	·	·	·	·	·	·
26.	jerk_lat_mean	**-.96**	**.70**	**.71**	**.68**	**-.68**	**-.61**	-.19	**-.40**	**-.62**	**-.50**	**-.44**	**-.41**	**-.68**	**-.67**	-.02	.09	**-.97**	**-.96**	**.96**	**-.96**	**-.93**	**-.93**	**.96**	**.68**	**-.83**	·	·	·	·	·	·	·	·	·	·	·
27.	jerk_lat_max	**.98**	**-.72**	**-.73**	**-.70**	.67	**.64**	.18	**.38**	**.63**	**.46**	**.47**	**.41**	**.68**	**.67**	-.01	-.04	**.99**	**.99**	**-.97**	**.97**	**.92**	**.97**	**-.99**	**-.69**	**.85**	**-.93**	·	·	·	·	·	·	·	·	·	·
28.	yaw_rate_max	**.96**	**-.72**	**-.71**	**-.69**	**.72**	**.64**	.19	**.40**	**.64**	**.53**	**.49**	**.41**	**.71**	**.70**	.01	-.07	**.97**	**.99**	**-.99**	**.99**	**.96**	**.97**	**-.98**	**-.72**	**.84**	**-.95**	**.93**	·	·	·	·	·	·	·	·	·
29.	yaw_rate_min	**-.98**	**.74**	**.72**	**.72**	**-.70**	**-.64**	-.19	**-.40**	**-.64**	**-.49**	-.43	**-.40**	**-.69**	**-.69**	-.01	.06	**-.97**	**-.99**	**.99**	**-.99**	**-.94**	**-.98**	**.95**	**.71**	**-.86**	**.95**	**-.94**	**-.97**	·	·	·	·	·	·	·	·
30.	d_min	.53	**-.56**	**-.58**	**-.49**	**.39**	**.34**	**.36**	**.34**	**.45**	**.27**	.28	.06	**.41**	.47	.28	.10	.54	.55	-.53	.52	**.52**	**.56**	-.54	**-.34**	**.51**	**-.50**	.51	.54	**-.56**	·	·	·	·	·	·	·
31.	d_mean_td	**-.91**	**.64**	**.69**	**.57**	**-.65**	**-.54**	**-.30**	**-.41**	**-.62**	-.68	**-.52**	**-.28**	**-.61**	**-.67**	-.05	.03	**-.90**	**-.89**	**.90**	**-.91**	**-.88**	**-.89**	**.90**	**.56**	**-.77**	**.85**	**-.83**	**-.90**	**.88**	**-.43**	·	·	·	·	·	·
32.	d_max_td	**-.64**	**.45**	**.51**	.45	**-.46**	**-.33**	-.06	**-.21**	**-.42**	-.27	-.24	-.24	**-.45**	**-.42**	-.05	.08	**-.64**	**-.64**	**.65**	-.64	-.62	**-.63**	**.64**	**.47**	**-.53**	**.63**	-.60	-.64	.63	**-.40**	.59	·	·	·	·	·
33.	ttc_min	**.72**	**-.59**	**-.67**	**-.60**	**.52**	**.47**	**.36**	**.46**	**.62**	**.35**	**.53**	.20	.55	**.65**	.25	.05	**.73**	**.73**	**-.72**	**.74**	**.70**	**.74**	-.75	**-.50**	**.77**	**-.67**	**.74**	**.74**	**-.76**	**.81**	**-.67**	**-.53**	·	·	·	·
34.	ttc_max_td	**-.96**	**.95**	**.97**	**.94**	**-.96**	**-.96**	-.96	**-.95**	**-.94**	**-.93**	-.97	-.16	-.94	**-.97**	-.17	-.66	**-.94**	**-.96**	**.98**	**-.99**	**-.99**	**-.98**	**.98**	**.94**	**-.96**	**.96**	**-.99**	**-.98**	**.94**	**-.95**	**.95**	**.96**	**-.94**	·	·	·
35.	thw_min	**.43**	**-.40**	-.42	**-.38**	**.32**	**.27**	**.31**	**.29**	**.39**	.21	.26	.01	**.34**	**.40**	.26	.12	**.44**	**.45**	**-.43**	**.42**	**.43**	**.46**	**-.44**	**-.26**	**.44**	**-.42**	**.42**	**.44**	-.47	**.96**	**-.35**	**-.29**	.95	**-.17**	·	·
36.	thw_max_td	**-.67**	**.47**	**.60**	**.46**	**-.50**	**-.36**	-.02	**-.19**	**-.51**	-.18	-.21	-.22	**-.45**	**-.46**	-.02	.09	**-.66**	**-.65**	**.66**	**-.66**	**-.62**	**-.64**	**.67**	**.43**	**-.50**	**.62**	-.60	**-.67**	**.67**	-.34	**.57**	**.98**	**-.42**	.33	**-.34**	·

[a] How safe did you feel during the car ride?

[b] How safe did you feel interacting with the roadworks?

[c] How comfortable did you find the movement of the vehicle?

Table B.6: Correlation matrix of all features recorded during the roadworks scenario (driving style, questionnaire responses, GSR, VD, and Perception). Correlation and regression coefficients were determined using separate LMEs for each feature pair. Bold values denote a significant correlation ($p < 0.05$) and underlined values denote non-significant correlations ($p > 0.05$).

		1. style	2. Q1[a]	3. Q2[b]	4. Q3[c]	5. phasic_max	6. phasic_min	7. phasic_mean_td	8. phasic_slope	9. phasic_peak_count	10. phasic_rise_time	11. phasic_recovery_time	12. tonic_mean	13. tonic_std	14. tonic_mean_td	15. tonic_skewness	16. tonic_kurtosis	17. vel_mean	18. vel_max	19. vel_min	20. acc_lon_mean	21. acc_lon_max	22. acc_lat_max	23. acc_lat_min	24. jerk_lon_mean	25. jerk_lon_max	26. jerk_lat_mean	27. jerk_lat_max	28. yaw_rate_max	29. yaw_rate_min	30. d_min	31. d_mean_td	32. d_max_td	33. ttc_min	34. ttc_max_td	35. thw_min	36. thw_max_td
1.	style	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
2.	Q1[a]	**-.58**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
3.	Q2[b]	**-.51**	**.82**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
4.	Q3[c]	**-.66**	**.76**	**.69**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
5.	phasic_max	**.89**	**-.51**	**-.64**	**-.61**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
6.	phasic_min	**.82**	**-.46**	**-.59**	**-.59**	**.70**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
7.	phasic_mean_td	.25	-.11	-.05	-.21	**.33**	.14	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
8.	phasic_slope	**.37**	-.22	-.11	**-.27**	**.37**	.17	**.62**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
9.	phasic_peak_count	**.41**	**-.46**	-.07	**-.58**	**.43**	**.44**	.16	.21	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
10.	phasic_rise_time	**.29**	-.22	.06	**-.26**	**.33**	**.22**	**.24**	**.27**	.12	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
11.	phasic_recovery_time	**.27**	-.17	.16	**-.28**	**.30**	.18	.19	**.25**	**.67**	**.82**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
12.	tonic_mean	.20	-.21	**-.48**	-.23	.19	**.26**	.03	.01	.19	.15	.10	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
13.	tonic_std	.15	.01	.27	-.14	**.26**	.15	.16	.08	.17	-.03	-.02	-.19	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
14.	tonic_mean_td	**.42**	-.21	-.27	**-.29**	**.42**	**.39**	**.26**	.21	**.27**	**.41**	**.29**	.08	-.05	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
15.	tonic_skewness	-.00	.05	-.16	.11	.05	-.09	-.03	**.27**	.01	.05	.13	-.07	-.04	.07	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
16.	tonic_kurtosis	-.00	.10	.05	.09	.04	-.15	.09	-.06	-.04	-.07	.12	-.04	.08	-.08	**.53**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
17.	vel_mean	**.92**	**-.56**	**-.51**	**-.66**	**.66**	**.61**	.18	**.27**	**.41**	**.30**	**.25**	.15	.13	**.30**	-.02	-.01	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
18.	vel_max	**.91**	**-.56**	**-.50**	**-.66**	**.68**	**.61**	.19	**.27**	**.41**	**.30**	**.27**	.16	.12	**.30**	-.01	.00	**.99**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
19.	vel_min	-.37	.14	.34	.16	**-.28**	-.23	-.17	-.19	-.15	-.01	.01	-.18	-.20	-.13	-.09	-.12	-.36	-.35	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
20.	acc_lon_mean	**.86**	**-.54**	**-.51**	**-.65**	**.60**	.59	.17	**.26**	**.39**	**.28**	**.27**	.18	.11	**.28**	-.03	.01	**.90**	**.98**	**-.26**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
21.	acc_lon_max	**.98**	**-.57**	**-.51**	**-.67**	**.66**	**.59**	.17	**.29**	**.43**	**.34**	**.26**	.14	.14	**.31**	.01	-.03	**.97**	**.95**	**-.25**	**.86**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
22.	acc_lat_max	**.75**	**-.59**	**-.39**	**-.58**	**.52**	.35	.11	**.25**	**.36**	**.22**	.10	-.00	**.21**	.29	-.04	-.03	**.78**	**.77**	-.18	**.79**	**.71**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
23.	acc_lat_min	**-.84**	**.61**	**.52**	**.68**	**-.59**	**-.49**	-.18	**-.24**	**-.40**	**-.24**	-.21	-.12	-.18	**-.26**	.06	.02	**-.86**	**-.82**	.24	**-.83**	**-.80**	**-.76**	·	·	·	·	·	·	·	·	·	·	·	·	·	·
24.	jerk_lon_mean	-.17	.04	.17	.01	-.12	-.06	-.02	-.10	-.11	-.05	-.08	-.14	-.02	-.04	-.08	-.08	-.16	-.17	**.62**	**-.39**	-.15	-.23	**.30**	·	·	·	·	·	·	·	·	·	·	·	·	·
25.	jerk_lon_max	**.90**	**-.56**	**-.53**	**-.66**	**.65**	**.58**	.16	**.25**	**.41**	**.24**	.22	.13	.16	**.32**	.02	.06	**.98**	**.97**	**-.27**	**.88**	**.97**	**.67**	**-.87**	-.13	·	·	·	·	·	·	·	·	·	·	·	·
26.	jerk_lat_mean	.34	**-.36**	-.24	**-.31**	.28	.27	.09	.15	.19	.09	.20	.14	-.01	.12	-.04	-.01	.53	.60	-.07	**.90**	.30	.29	-.54	-.30	.41	·	·	·	·	·	·	·	·	·	·	·
27.	jerk_lat_max	**.79**	**-.63**	**-.58**	**-.70**	**.59**	**.48**	.18	**.27**	**.41**	**.26**	**.22**	.10	.18	**.30**	-.04	-.02	.82	**.79**	**-.23**	.81	**.77**	**.75**	**-.99**	-.26	**.81**	**.47**	·	·	·	·	·	·	·	·	·	·
28.	yaw_rate_max	**.82**	**-.56**	**-.48**	**-.64**	**.59**	**.52**	.15	.23	**.38**	**.26**	.23	.12	.18	**.28**	-.07	-.03	**.84**	**.82**	-.18	**.84**	**.77**	**.73**	**-.90**	-.22	**.79**	**.45**	**.91**	·	·	·	·	·	·	·	·	·
29.	yaw_rate_min	-.37	.33	.42	.04	**-.22**	**-.16**	.19	.06	-.09	-.09	.06	.00	**-.17**	-.16	-.01	.01	-.31	-.29	.24	-.28	-.26	**-.99**	**.87**	.11	-.67	-.09	**-.98**	-.48	·	·	·	·	·	·	·	·
30.	d_min	-.39	**.34**	.03	**.30**	-.24	-.24	-.06	-.06	-.32	-.15	-.32	-.11	.00	-.11	.08	-.01	**-.41**	-.44	.04	**-.47**	**-.37**	-.27	**.41**	**.31**	-.37	**-.41**	-.40	-.36	-.03	·	·	·	·	·	·	·
31.	d_mean_td	.33	-.27	-.28	**-.33**	.15	-.04	.03	.09	.01	-.06	.17	.08	-.16	-.15	-.03	.16	**.36**	**.40**	-.18	**.56**	.28	.29	**-.49**	-.33	**.37**	**.46**	**.41**	**.44**	-.01	**-.41**	·	·	·	·	·	·
32.	d_max_td	.38	-.31	-.28	-.28	**.29**	**.26**	.15	.16	.46	.08	.03	.12	**.17**	.18	-.00	-.09	.38	.37	-.15	**.36**	.35	.35	**-.46**	-.22	.35	.18	.43	**.35**	-.06	-.16	.03	·	·	·	·	·
33.	ttc_min	-.29	.20	.00	.10	-.35	-.19	-.09	-.14	-.22	-.20	-.27	-.11	-.01	-.15	-.15	-.09	-.22	-.29	.02	-.16	-.25	-.30	.31	.29	-.30	-.04	-.28	-.22	.10	**.89**	**-.32**	**-.46**	·	·	·	·
34.	ttc_max_td	.61	-.44	-.31	-.46	.80	.22	.25	.15	.29	.19	.56	.12	-.03	.16	-.04	.05	.37	.53	-.06	.55	.24	.57	-.90	-.79	.45	.18	.90	.55	-.04	-.86	**.65**	.71	**-.94**	·	·	·
35.	thw_min	**-.86**	.63	.25	**.55**	-.38	-.40	-.18	-.15	**-.52**	-.25	-.42	-.23	.08	-.16	.13	.01	**-.81**	**-.88**	.03	**-.84**	**-.73**	**-.70**	**.95**	.18	**-.82**	-.53	**-.86**	**-.79**	.00	**.95**	-.60	**-.42**	.42	.04	·	·
36.	thw_max_td	**.41**	-.53	-.52	-.40	.35	**.23**	.17	**.30**	**.35**	.20	.17	.04	**.24**	**.25**	.06	-.02	**.78**	**.46**	-.13	.62	**.63**	.31	-.58	-.23	**.40**	.10	**.39**	**.88**	-.10	-.17	-.14	**.76**	-.17	.54	-.63	·

[a] How safe did you feel during the car ride?

[b] How safe did you feel interacting with the pedestrian?

[c] How comfortable did you find the movement of the vehicle?

Table B.7: Correlation matrix of all features recorded during the pedestrian crossing with obstruction scenario (driving style, questionnaire responses, GSR, VD, and Perception). Correlation and regression coefficients were determined using separate LMEs for each feature pair. Bold values denote a significant correlation ($p < 0.05$) and underlined values denote non-significant correlations ($p > 0.05$).

		1. style	2. Q1[a]	3. Q2[b]	4. Q3[c]	5. phasic_max	6. phasic_min	7. phasic_mean_td	8. phasic_slope	9. phasic_peak_count	10. phasic_rise_time	11. phasic_recovery_time	12. tonic_mean	13. tonic_std	14. tonic_mean_td	15. tonic_skewness	16. tonic_kurtosis	17. vel_mean	18. vel_max	19. vel_min	20. acc_lon_mean	21. acc_lon_max	22. acc_lat_max	23. acc_lat_min	24. jerk_lon_mean	25. jerk_lon_max	26. jerk_lat_mean	27. jerk_lat_max	28. yaw_rate_max	29. yaw_rate_min	30. d_min	31. d_mean_td	32. d_max_td	33. ttc_min	34. ttc_max_td	35. thw_min	36. thw_max_td
1.	style	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
2.	Q1[a]	**-.71**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
3.	Q2[b]	**-.58**	**.82**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
4.	Q3[c]	**-.66**	**.83**	**.75**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
5.	phasic_max	**.76**	**-.51**	**-.45**	**-.49**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
6.	phasic_min	**.44**	**-.36**	**-.35**	**-.33**	**.41**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
7.	phasic_mean_td	.09	-.11	-.15	-.16	.01	.01	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
8.	phasic_slope	-.02	-.03	-.12	-.04	-.08	.20	**.53**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
9.	phasic_peak_count	**.25**	**-.30**	**-.29**	-.16	.21	.32	.12	.13	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
10.	phasic_rise_time	.21	-.10	-.10	-.07	**.30**	**.25**	.11	.19	-.08	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
11.	phasic_recovery_time	.14	.03	.03	.02	**.27**	**.20**	-.03	.02	.04	**.77**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
12.	tonic_mean	-.12	-.03	.07	.14	.10	.19	-.20	-.24	.20	.14	.14	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
13.	tonic_std	.18	-.15	-.15	-.09	**.31**	.17	-.01	-.00	**.24**	.17	**.31**	.09	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
14.	tonic_mean_td	.02	-.09	-.06	.04	.12	.18	.14	.08	**.31**	.19	.17	.12	**.65**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
15.	tonic_skewness	.14	-.17	-.21	-.24	-.00	.08	.03	**.36**	.02	.04	-.09	**-.36**	.10	.09	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
16.	tonic_kurtosis	-.28	.12	.12	.07	**-.33**	-.09	-.04	-.02	-.05	-.12	-.19	.14	**-.30**	-.24	-.04	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
17.	vel_mean	**.90**	**-.63**	**-.57**	**-.60**	.50	**.26**	.04	.00	**.21**	.23	.15	-.11	.18	.03	.11	-.08	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
18.	vel_max	**.42**	**-.36**	**-.41**	**-.39**	**.30**	.22	-.01	.02	.12	.04	.06	-.13	.18	.07	.13	-.09	.45	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
19.	vel_min	-.66	**.52**	**.49**	**.55**	**-.48**	-.20	.01	.06	-.10	-.07	-.15	.03	-.14	.02	-.06	.14	**-.70**	**-.64**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
20.	acc_lon_mean	**-.94**	**.67**	**.62**	**.68**	**-.44**	**-.27**	.00	.03	-.17	-.11	-.15	.08	-.15	-.02	-.13	.19	-.85	**-.59**	**.92**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
21.	acc_lon_max	**.80**	**-.58**	**-.54**	**-.60**	**.47**	**.29**	-.08	-.10	.20	.17	.17	-.08	.16	.07	.19	-.15	**.69**	**.49**	**-.64**	**-.74**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
22.	acc_lat_max	**.93**	-.76	**-.89**	**-.90**	.75	**.35**	.06	-.07	.19	.13	.33	-.08	.20	.04	.06	-.17	**.93**	**.89**	**-.94**	**-.96**	.91	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
23.	acc_lat_min	**-.82**	**.66**	**.64**	**.63**	**-.53**	**-.25**	-.05	.04	-.20	-.14	-.15	-.00	-.12	-.03	-.05	.18	-.76	**-.48**	**.85**	**.87**	**-.70**	**-.56**	·	·	·	·	·	·	·	·	·	·	·	·	·	·
24.	jerk_lon_mean	**-.25**	.13	.13	.19	**-.20**	-.10	.02	.01	.07	-.06	-.17	.04	.01	.04	-.05	**.26**	-.09	-.22	**.46**	**.31**	-.21	-.26	**.38**	·	·	·	·	·	·	·	·	·	·	·	·	·
25.	jerk_lon_max	**.85**	**-.67**	**-.50**	**-.61**	**.50**	**.28**	-.10	-.10	.19	.18	.14	-.06	.15	.04	.13	-.19	**.77**	**.45**	**-.65**	**-.82**	**.89**	**.43**	**-.82**	-.21	·	·	·	·	·	·	·	·	·	·	·	·
26.	jerk_lat_mean	-.43	.33	**.35**	**.38**	**-.45**	-.18	-.00	.01	-.11	-.02	.09	.04	-.24	.02	-.05	-.01	**-.56**	**-.52**	**.79**	.49	**-.40**	**-.61**	**.47**	.04	**-.37**	·	·	·	·	·	·	·	·	·	·	·
27.	jerk_lat_max	**.73**	**-.57**	**-.65**	**-.57**	**.50**	**.26**	.06	-.05	.17	.14	.18	-.01	.10	.06	.02	**-.23**	.68	**.54**	**-.87**	**-.81**	**.68**	**.59**	**-.85**	**-.43**	**.66**	**-.50**	·	·	·	·	·	·	·	·	·	·
28.	yaw_rate_max	**.84**	**-.65**	**-.63**	**-.62**	.52	**.27**	.07	-.02	.20	.14	.13	-.01	.15	.03	.04	-.16	**.78**	**.47**	**-.82**	**-.86**	**.68**	**.58**	**-.97**	**-.31**	**.68**	**-.59**	**.86**	·	·	·	·	·	·	·	·	·
29.	yaw_rate_min	-.96	.26	.93	.14	-.52	**-.39**	-.01	-.05	-.06	-.45	-.39	-.10	-.01	-.00	.06	.11	-.96	-.24	.52	**.59**	-.55	**-.96**	.84	.26	-.68	.18	**-.93**	-.53	·	·	·	·	·	·	·	·
30.	d_min	**.61**	**-.41**	-.30	-.22	**.33**	.14	.13	.08	.27	.14	.23	.06	.19	.25	-.06	-.09	**.61**	-.00	**-.48**	**-.53**	**.37**	**.57**	**-.61**	-.05	**.52**	-.18	**.78**	**.50**	-.49	·	·	·	·	·	·	·
31.	d_mean_td	.67	**-.60**	**-.40**	**-.53**	**.46**	.18	-.10	-.05	.27	**.33**	**.24**	.04	.17	.17	.02	-.22	.45	.23	**-.40**	**-.46**	.72	**.19**	**-.53**	-.28	**.65**	-.08	**.46**	**.52**	-.05	**.30**	·	·	·	·	·	·
32.	d_max_td	**.45**	-.32	-.21	-.27	.27	.18	.17	.13	.07	.30	.24	-.09	.08	.15	.10	-.17	.17	.09	-.31	-.33	.35	.18	**-.38**	**-.41**	.35	-.01	**.37**	**.31**	-.05	.10	.33	·	·	·	·	·
33.	ttc_min	**.40**	-.30	-.18	-.21	**.28**	.18	.07	.00	.07	.18	.24	.02	.17	**.17**	.03	.03	**.58**	.28	**-.50**	**-.48**	**.39**	**.24**	**-.51**	-.17	**.49**	-.37	.54	**.49**	-.09	**.67**	.38	.06	·	·	·	·
34.	ttc_max_td	.22	**-.21**	-.34	-.28	.19	.20	-.01	-.03	.08	.07	.14	.00	-.01	.07	-.14	-.08	.06	-.24	.18	.07	-.04	.00	-.03	.03	.25	.03	.03	-.02	.08	.63	-.02	-.07	-.09	·	·	·
35.	thw_min	**.60**	-.37	-.22	-.16	.33	.13	.09	.03	.25	**.12**	.19	.02	.17	.20	-.08	-.03	**.59**	-.02	**-.39**	**-.46**	.34	**.52**	**-.56**	-.04	**.51**	-.17	**.72**	**.45**	-.32	**.88**	.20	-.03	**.55**	-.07	·	·
36.	thw_max_td	.04	-.12	-.02	-.04	.10	.09	.09	.06	.01	.24	.16	-.07	-.01	.06	.04	-.27	-.24	-.09	-.01	.03	.06	-.10	-.04	-.32	.01	.31	.08	-.00	.06	-.18	.18	**.97**	-.24	.18	-.27	·

[a] How safe did you feel during the car ride?

[b] How safe did you feel interacting with the vehicle?

[c] How comfortable did you find the movement of the vehicle?

Table B.8: Correlation matrix of all features recorded during the cut-in scenario (driving style, questionnaire responses, GSR, VD, and Perception). Correlation and regression coefficients were determined used separate LMEs for each feature pair. Bold values denote a significant correlation ($p < 0.05$) and underlined values denote non-significant correlations ($p > 0.05$).

		1. style	2. Q1[a]	3. Q2[b]	4. Q3[c]	5. phasic_max	6. phasic_min	7. phasic_mean_td	8. phasic_slope	9. phasic_peak_count	10. phasic_rise_time	11. phasic_recovery_time	12. tonic_mean	13. tonic_std	14. tonic_mean_td	15. tonic_skewness	16. tonic_kurtosis	17. vel_mean	18. vel_max	19. vel_min	20. acc_lon_mean	21. acc_lon_max	22. acc_lat_max	23. acc_lat_min	24. jerk_lon_mean	25. jerk_lon_max	26. jerk_lat_mean	27. jerk_lat_max	28. yaw_rate_max	29. yaw_rate_min	30. d_min	31. d_mean_td	32. d_max_td	33. ttc_min	34. ttc_max_td	35. thw_min	36. thw_max_td
1.	style	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
2.	Q1[a]	**-.71**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
3.	Q2[b]	**-.58**	**.82**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
4.	Q3[c]	**-.66**	**.83**	**.75**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
5.	phasic_max	**.53**	**-.38**	**-.38**	**-.39**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
6.	phasic_min	**.54**	**-.43**	**-.42**	**-.29**	**.32**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
7.	phasic_mean_td	-.26	.18	.16	.19	**-.37**	-.12	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
8.	phasic_slope	**-.44**	**.31**	**.33**	**.32**	**-.49**	-.25	**.66**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
9.	phasic_peak_count	**.31**	-.33	-.36	-.10	**.24**	.32	-.07	-.11	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
10.	phasic_rise_time	.15	-.03	-.01	.07	**.26**	.09	-.11	-.07	.08	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
11.	phasic_recovery_time	.21	-.17	-.06	-.00	.15	.16	-.04	.04	**.48**	**0.98**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
12.	tonic_mean	-.15	.01	.01	.16	.06	.07	-.16	-.24	-.01	.11	-.05	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
13.	tonic_std	.06	-.11	-.12	-.07	.16	.01	-.05	-.05	.15	-.14	-.20	-.05	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
14.	tonic_mean_td	-.17	.18	.11	.12	.02	.09	.23	.07	-.03	.04	-.03	.15	**-.48**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
15.	tonic_skewness	-.07	.12	.16	.20	-.18	-.07	.08	**.38**	-.06	.07	.03	-.19	.18	**-.31**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
16.	tonic_kurtosis	.15	-.06	-.02	.04	-.10	.12	.13	.14	-.02	.09	.01	.04	-.21	-.01	.23	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
17.	vel_mean	**.64**	**-.52**	**-.49**	-.43	**.27**	.22	-.16	**-.28**	**.27**	.04	.12	-.01	.01	-.03	-.19	-.01	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
18.	vel_max	**.76**	**-.61**	**-.53**	**-.54**	**.37**	**.36**	-.14	**-.31**	**.33**	.15	.12	-.01	-.00	-.02	-.03	.12	**.65**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
19.	vel_min	**-.74**	**.41**	**.49**	**.60**	**-.33**	**-.34**	.07	.17	**-.30**	-.14	-.11	.13	.04	-.03	-.02	-.01	**-.57**	**-.66**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
20.	acc_lon_mean	**.99**	**-.56**	**-.53**	**-.56**	**.30**	**.35**	-.18	**-.25**	.33	.17	.17	-.09	.01	-.10	-.05	.12	**.66**	**.69**	**-.59**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
21.	acc_lon_max	**.91**	**-.63**	**-.53**	**-.61**	**.35**	**.38**	-.15	-.23	**.28**	.18	.26	-.10	.03	-.07	-.03	.05	.49	**.61**	**-.51**	**.73**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
22.	acc_lat_max	.08	.01	-.01	.07	-.03	.03	.18	.05	.13	-.04	.06	-.10	-.03	.00	-.02	-.05	.06	-.00	.14	-.07	.05	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
23.	acc_lat_min	-.10	.08	.08	.10	.08	-.04	-.00	.11	-.08	.03	-.01	-.03	.12	-.00	.16	.07	-.12	-.08	.05	-.10	-.06	-.02	·	·	·	·	·	·	·	·	·	·	·	·	·	·
24.	jerk_lon_mean	**-.83**	**.57**	**.56**	**.62**	**-.28**	**-.33**	.06	.15	**-.19**	-.10	-.17	.10	-.05	.04	.03	-.10	**-.37**	**-.44**	**.54**	**-.70**	**-.87**	-.01	.03	·	·	·	·	·	·	·	·	·	·	·	·	·
25.	jerk_lon_max	**.82**	**-.61**	**-.51**	**-.58**	**.29**	**.37**	-.13	-.16	**.32**	.15	.27	-.11	.02	-.07	-.04	.01	**.50**	**.51**	**-.40**	**.64**	**.91**	.16	-.01	**-.64**	·	·	·	·	·	·	·	·	·	·	·	·
26.	jerk_lat_mean	**-.41**	**.36**	**.31**	**.28**	-.19	-.17	.04	.17	-.13	.01	.07	.02	-.07	.06	.00	-.04	-.19	-.18	.25	**-.34**	**-.34**	.01	.02	**.25**	**-.31**	·	·	·	·	·	·	·	·	·	·	·
27.	jerk_lat_max	.16	-.13	-.16	-.15	.04	.14	-.06	-.10	.18	-.15	.10	-.06	.02	-.10	-.07	.01	.24	.20	.04	.08	.12	**.42**	-.24	-.06	.12	-.05	·	·	·	·	·	·	·	·	·	·
28.	yaw_rate_max	**.29**	-.11	-.15	-.22	**.30**	**.23**	-.15	-.11	.14	.12	-.02	.02	-.00	-.07	.06	.02	**.25**	**.28**	**-.24**	**.35**	**.36**	-.15	.02	**-.30**	**.25**	-.06	-.02	·	·	·	·	·	·	·	·	·
29.	yaw_rate_min	**-.50**	**.32**	**.36**	**.37**	**-.34**	**-.24**	.17	.09	**-.26**	-.20	-.29	.18	-.03	.02	.05	.06	**-.39**	**-.24**	**.31**	**-.49**	-.53	-.26	.08	**.35**	**-.55**	.21	-.10	**-.35**	·	·	·	·	·	·	·	·
30.	d_min	**.57**	**-.62**	**-.63**	**-.54**	.28	.16	-.06	-.22	.04	.11	.15	-.17	.03	-.03	-.16	-.02	**.46**	.36	-.33	.44	**.70**	.06	-.16	**-.49**	**.76**	**-.30**	.16	.03	**-.44**	·	·	·	·	·	·	·
31.	d_mean_td	**-.82**	**.67**	**.54**	**.63**	**-.41**	-.35	.11	**.27**	-.10	**-.28**	**-.33**	-.00	.06	.07	.10	-.23	-.45	**-.49**	**.35**	-.62	**-.78**	-.06	.20	**.61**	**-.66**	**.40**	-.11	-.24	**.41**	**-.51**	·	·	·	·	·	·
32.	d_max_td	**-.38**	.12	.03	.13	-.11	-.12	.01	.09	-.17	-.18	-.27	-.04	.01	-.11	-.03	-.05	**-.34**	-.13	.02	-.33	**-.32**	-.20	-.02	.15	**-.33**	.39	-.11	-.10	**.27**	-.26	**.45**	·	·	·	·	·
33.	ttc_min	**.98**	**-.81**	**-.90**	**-.75**	**.53**	.34	-.15	**-.38**	.17	**.12**	.18	-.18	.04	-.07	-.15	.00	**.83**	.57	-.54	.61	**.98**	.10	-.17	**-.92**	**0.98**	**-.30**	.29	.07	**-.69**	0.97	-.68	-.40	·	·	·	·
34.	ttc_max_td	**-.74**	.38	.30	.50	**-.32**	-.22	.07	.21	-.01	-.00	-.15	-.05	-.03	.11	.12	-.01	**-.82**	-.29	.59	-.65	**-.91**	-.03	.11	.25	**-.93**	**.16**	-.11	.03	.65	**-.94**	.43	.39	**-.94**	·	·	·
35.	thw_min	**.97**	**-.76**	**-.69**	**-.77**	**.44**	.31	-.07	**-.38**	.07	.09	.21	-.13	.03	-.06	-.12	.06	**.37**	**.58**	**-.43**	**.61**	**.88**	.06	-.17	**-.69**	**.71**	**-.40**	.08	.17	**-.46**	**.76**	**-.91**	**-.31**	**.54**	**-.37**	·	·
36.	thw_max_td	**-.38**	.19	.08	.15	-.11	-.11	.01	.09	-.10	-.13	-.21	.02	.04	-.11	.05	-.04	**-.36**	-.12	.00	-.31	**-.33**	-.23	.01	.16	**-.33**	.38	-.14	-.10	**.25**	-.30	**.46**	**.78**	-.23	.09	**-.37**	·

[a] How safe did you feel during the car ride?
[b] How safe did you feel interacting with the vehicle?
[c] How comfortable did you find the movement of the vehicle?

Table B.9: Correlation matrix of all features recorded during the car-following scenario (driving style, questionnaire responses, GSR, VD, and Perception). Correlation and regression coefficients were determined used separate LMEs for each feature pair. Bold values denote a significant correlation ($p < 0.05$) and underlined values denote non-significant correlations ($p > 0.05$).

Tables B.5 to B.9 reveal how the different scenarios evoke different responses, both subjective and physiological:

1. For all scenarios, a significant negative correlation can be found between style and questionnaire response, indicating that for each scenario, the aggressive driving style was rated more negatively. The effect is largest in the roadwork scenario ($\beta = -0.69$ to -0.79) and smallest in the ped. crossing without obstruction ($\beta = -0.41$ to -0.65). This underscores that the same style change is felt with varying intensity across scenarios.
2. Similarly, GSR-related features show different correlation coefficients between scenarios. The roadwork scenario shows the highest correlation coefficients and the car-following the lowest.
3. The phasic maximum amplitude remains the most robust indicator among GSR features of both driving style and subjective responses across all scenarios. In contrast, phasic peak count loses significance in several scenarios.
4. The tonic component shows very high correlation coefficients with both the driving style and subjective scores in the roadwork scenario, but becomes insignificant in the ped. crossing with obstruction, cut-in, and car-following and marginally significant in the ped. crossing with obstruction.
5. The correlation coefficients between driving style and vehicle dynamics approach $|\beta| \approx 1.00$ as these vehicle dynamics are intrinsic to the definition of aggressive versus calm driving.
6. Similarly, in those scenarios where the perception characteristics are salient (roadworks and car-following), a high correlation coefficient can be found.
7. Furthermore, a strong correlation ($|\beta| > 0.5$) exists between vehicle dynamics and GSR in the ped. crossing with and without obstruction and the roadwork scenario, suggesting larger motion differences provoke stronger physiological responses.

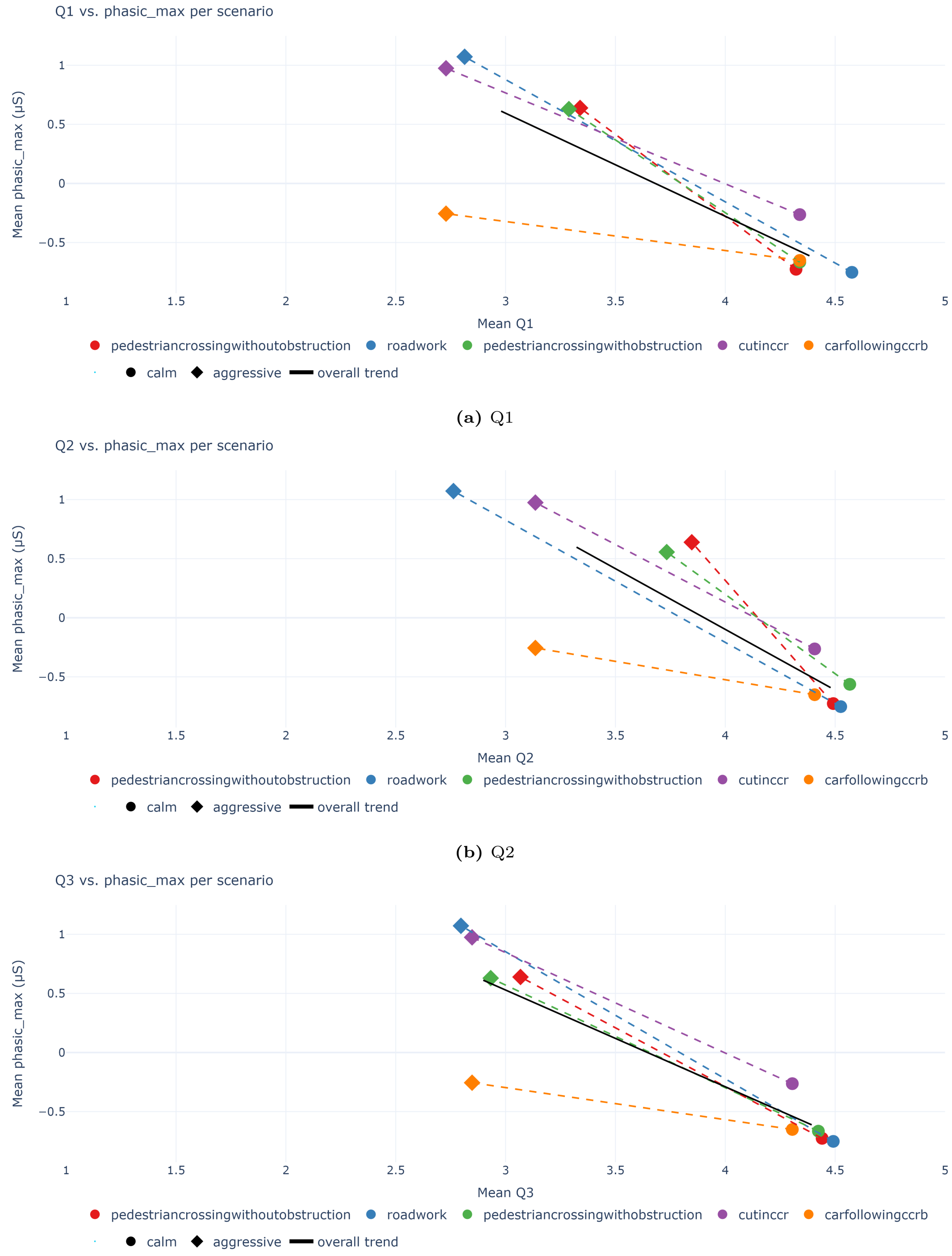
Note:

> For the roadworks scenario in Table B.6, the maximum time derivative of the TTC correlates at $|\beta| \approx 0.95 - 0.99$ with almost every variable. This near-unity inflation is most likely a numerical artefact: when the VUT passes the roadwork cones, the TTC jumps from ≈ 5 seconds to ≈ 25 seconds, creating identical extreme slopes across trials. Therefore, this observation should be treated with caution or omitted from analysis.

This scenario-specific analysis indicates that the different scenarios elicit distinct subjective and physiological responses. While the aggressive driving consistently lowers perceived comfort and safety, the magnitude of both the subjective and physiological responses varies considerably across scenarios. Scenarios involving higher vehicle dynamics tend to provoke stronger reactions, hinting that motion intensity plays a key role in shaping both perceived and physiological discomfort.

To complement the statistical correlation analysis, a series of visualizations was created to illustrate further and interpret the observed relationships. These visualizations focus on three key physiological features, phasic maximum, peak count and tonic standard deviation, selected based on their strong correlation with both driving style and subjective responses (see Table B.2). The resulting plots aim to provide intuitive insights into the strength, direction and patterns behind the correlations, helping to better understand how the GSR relates across participants and conditions.

Figures B.5, B.6, and B.7 visualize a scatter of the average subjective response (Q1, Q2, Q3) and the average phasic maximum, peak count and tonic standard deviation per scenario per driving style condition. The dashed lines link each scenario's calm and aggressive points, highlighting the magnitude of change between driving styles. Scenario averages reveal scenario-specific patterns, and the overall trend represents the average relationship across all scenarios and participants.

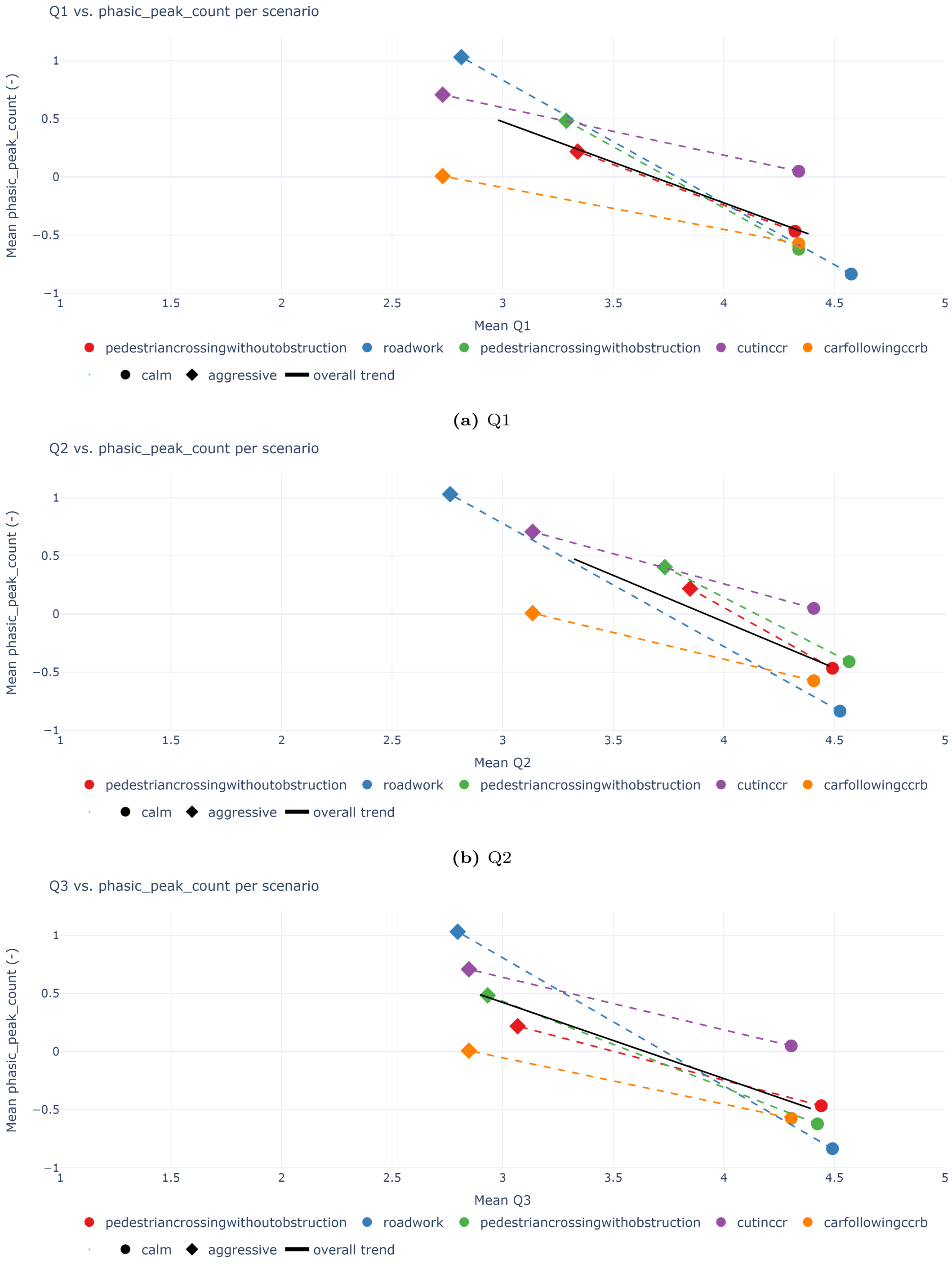


(a) Q1

(b) Q2

(c) Q3

Figure B.5: Scatter of mean phasic maximum and mean subjective responses (Q1, Q2, Q3) per scenario and driving style. Calm (circle) and aggressive (diamonds) are joined by dashed lines; the solid black line shows the overall trend. The phasic maximum is standardized participant-wise.

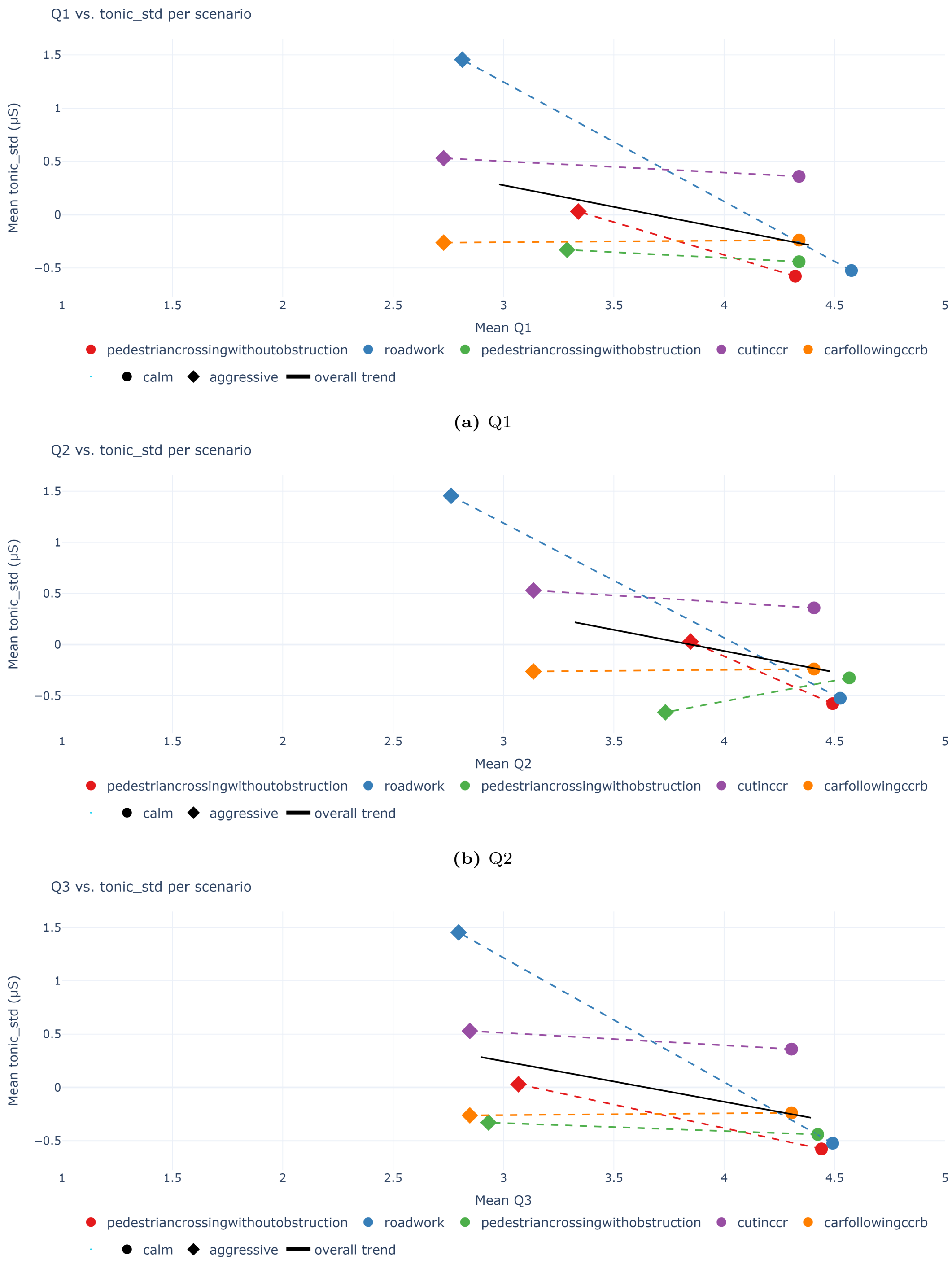


(a) Q1

(b) Q2

(c) Q3

Figure B.6: Scatter of mean phasic peak count and mean subjective responses (Q1, Q2, Q3) per scenario and driving style. Calm (circle) and aggressive (diamonds) are joined by dashed lines; the solid black line shows the overall trend. The phasic peak count is standardized participant-wise.



(a) Q1

(b) Q2

(c) Q3

Figure B.7: Scatter of mean tonic standard deviation and mean subjective responses (Q1, Q2, Q3) per scenario and driving style. Calm (circle) and aggressive (diamonds) are joined by dashed lines; the solid black line shows the overall trend. The tonic standard deviation is standardized participant-wise.

Analysis of Figures B.5, B.6 and B.7 shows that across all participants and scenarios, there is a negative linear relationship between the key features and comfort scores. This indicates that when GSR increases, the participants generally felt less comfortable, although individual results may vary.
When comparing the figures, it is clear that the phasic component has a stronger connection to feelings of discomfort than the tonic component.

Several other interesting patterns emerge from the figures:

1. The roadwork scenario shows dramatic contrasts: during calm driving, it is rated on average as the most comfortable and safe scenario, but during aggressive driving, it is perceived as one of the most uncomfortable and unsafe.
2. Pedestrian crossings, both with and without obstruction, are seen as the most comfortable and safe during aggressive driving, with especially little difference in how safe participants felt between driving conditions (Q2).
3. When it comes to overall ride comfort, all scenarios receive similar ratings on average.

The roadwork scenario also stands out with the steepest slope across key features, suggesting that both physiological and subjective responses changed substantially between calm and aggressive driving styles.

Figure B.8 further visualizes the relationship between the key GSR features, phasic maximum, peak count and tonic standard deviation, and the participants' self-reported comfort and safety scores. Each jittered black dot is a single observation, colored dashed lines trace the scenario-specific mean response at each comfort/safety level, and the solid black line depicts the overall mean trend across all scenarios and participants.

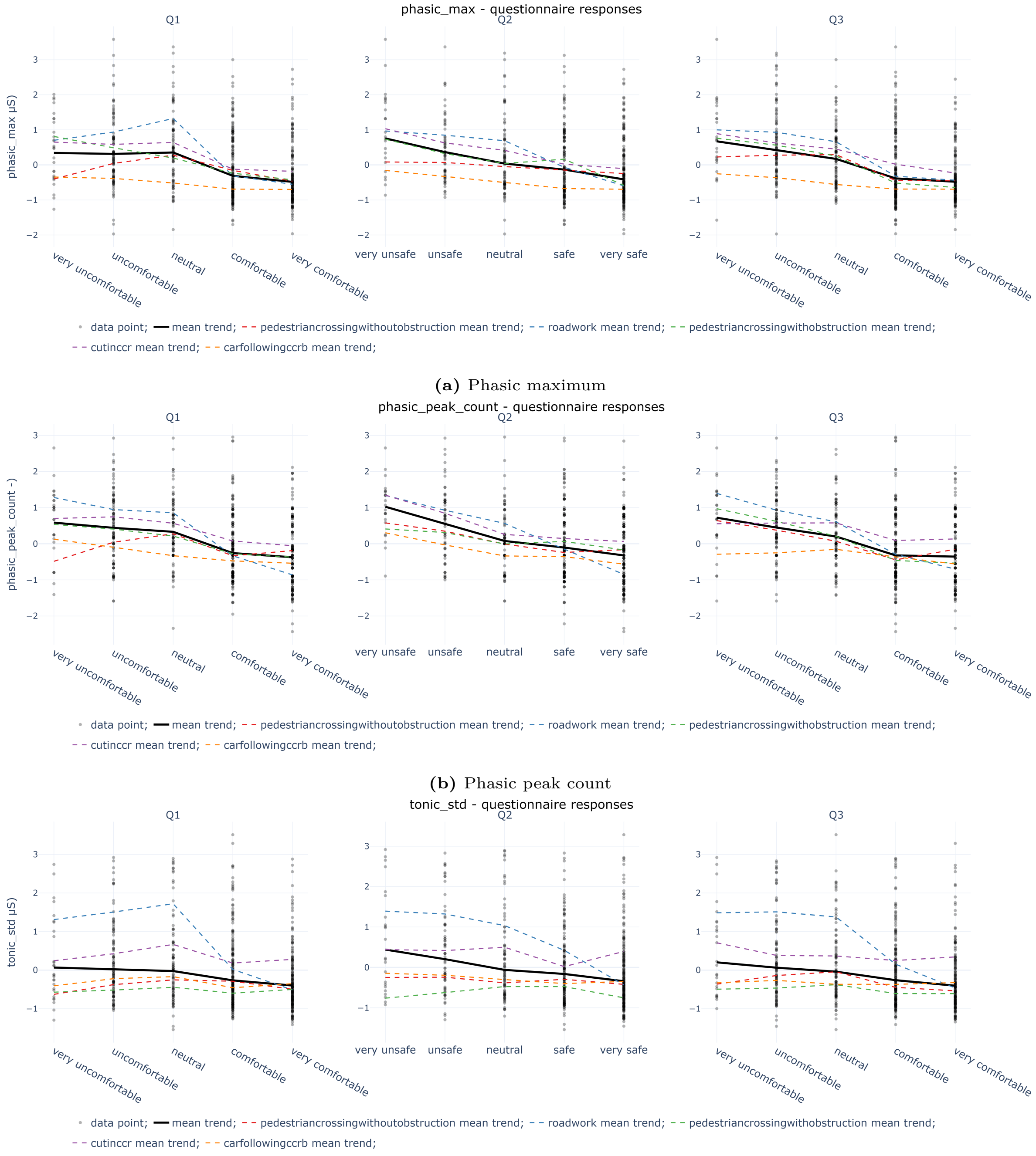
Consistent with Figures B.5-B.7, there is a clear negative association: as self-reported comfort and safety scores decrease, GSR activity increases. Moreover, the phasic features exhibit steeper slopes than the tonic feature, indicating once again a stronger link to self-reported moment-to-moment discomfort.

Several scenario-specific patterns emerge:

1. The roadwork scenario elicits the highest overall physiological arousal and the most pronounced slopes, indicating large shifts in both physiological and subjective responses.
2. The cut-in scenario produces the next greatest GSR response and moderately steep trends.
3. Contrary, the pedestrian crossing scenarios and car-following scenario show relatively flat slopes, reflecting only modest variations in GSR across comfort and safety levels.

Taken together, these results reaffirm that, on average, GSR increases with decreased comfort and safety scores reported by the participants. Both physiological and subjective responses vary systematically per scenario. In particular, the roadwork scenario is associated with lower comfort and safety scores and the highest overall GSR arousal.

These observations of Figure B.5 to B.8 are consistent with the statistical results obtained from the LME models, which confirm the same patterns in both subjective as physiological responses across all scenarios and per scenario.



(a) Phasic maximum

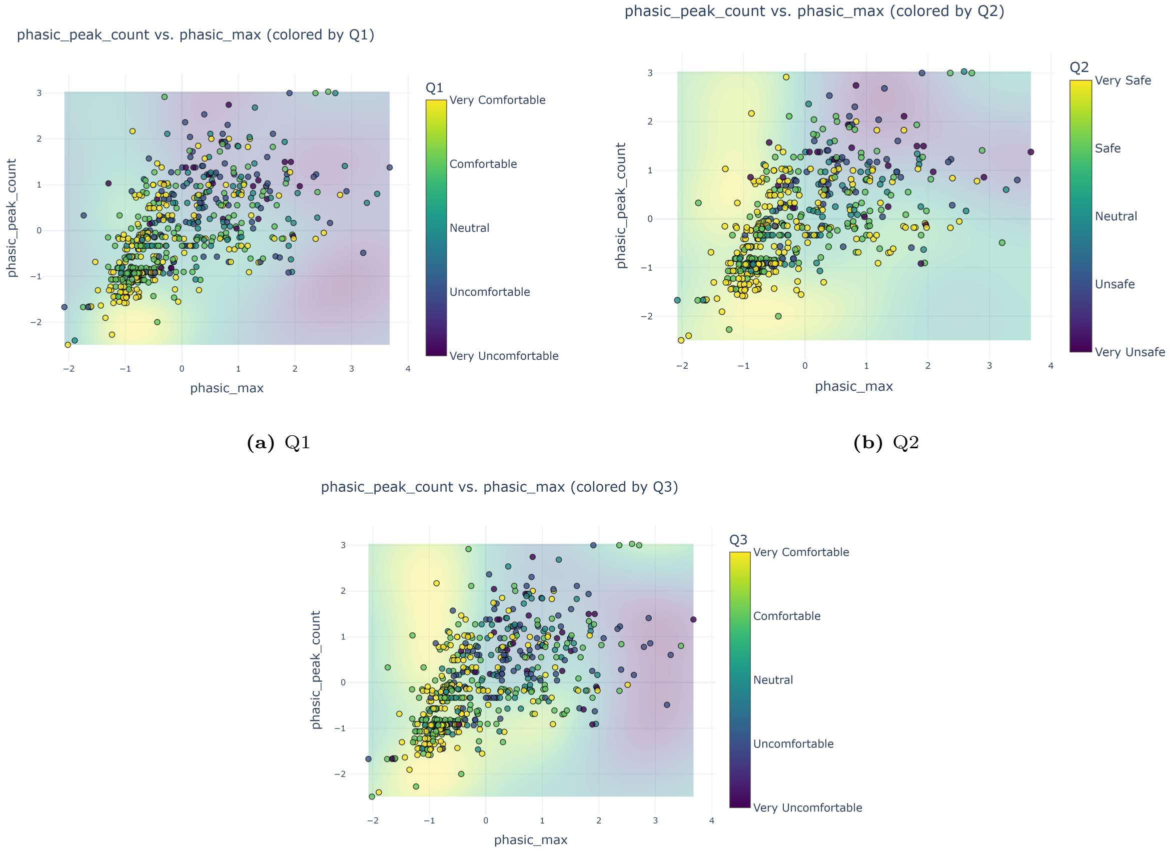
(b) Phasic peak count

(c) Tonic standard deviation

Figure B.8: Phasic maximum (**a**), peak count (**b**) and tonic standard deviation (**c**) jittered as a function of the subjective responses (Q1, Q2, Q3). Individual data points (jittered transparent black dots), scenario-specific mean trends (colored dashed lines), and the overall mean trend (black line) are shown for each questionnaire item.

With the phasic maximum amplitude and peak count identified as the most indicative features of passenger state, Figure B.9 visualizes all recorded data points, plotting these two GSR-based indicators against each other and coloring each point by the participants' corresponding subjective score.

A Support Vector Machine (SVM) is fitted to outline in the feature space where each subjective score (Q1, Q2, Q3) predominates. Overall, lower comfort and safety scores are generally associated with stronger phasic responses, despite some local exceptions.



(a) Q1

(b) Q2

(c) Q3

Figure B.9: Scatter cross plot of phasic maximum amplitude and peak count, colored by subjective ratings (Q1, Q2, Q3). Background shading represents SVM-derived decision regions for each subjective rating. Both phasic features are z-standardized on a per-participant basis.

B.3.2. Comparative Analysis

The dataset obtained by the diverse experiment allows for a wide range of pairwise comparisons between subjective responses and physiological measures across scenarios, different obstructions, demographic variables and pre-questionnaire responses related to trust in automated driving, willingness to adapt automated driving and likelihood of motion sickness.

The three questionnaire responses are compared and three physiological-related features. The three features chosen to represent the physiological responses are the phasic maximum, peak count and tonic mean time derivative. The two phasic features showed the highest correlations with the subjective scores and vehicle dynamics data. The tonic mean time derivative was the only physiological-related feature to show a correlation with perception data.

Scenarios

Figure B.10 presents the comparisons between each scenario.

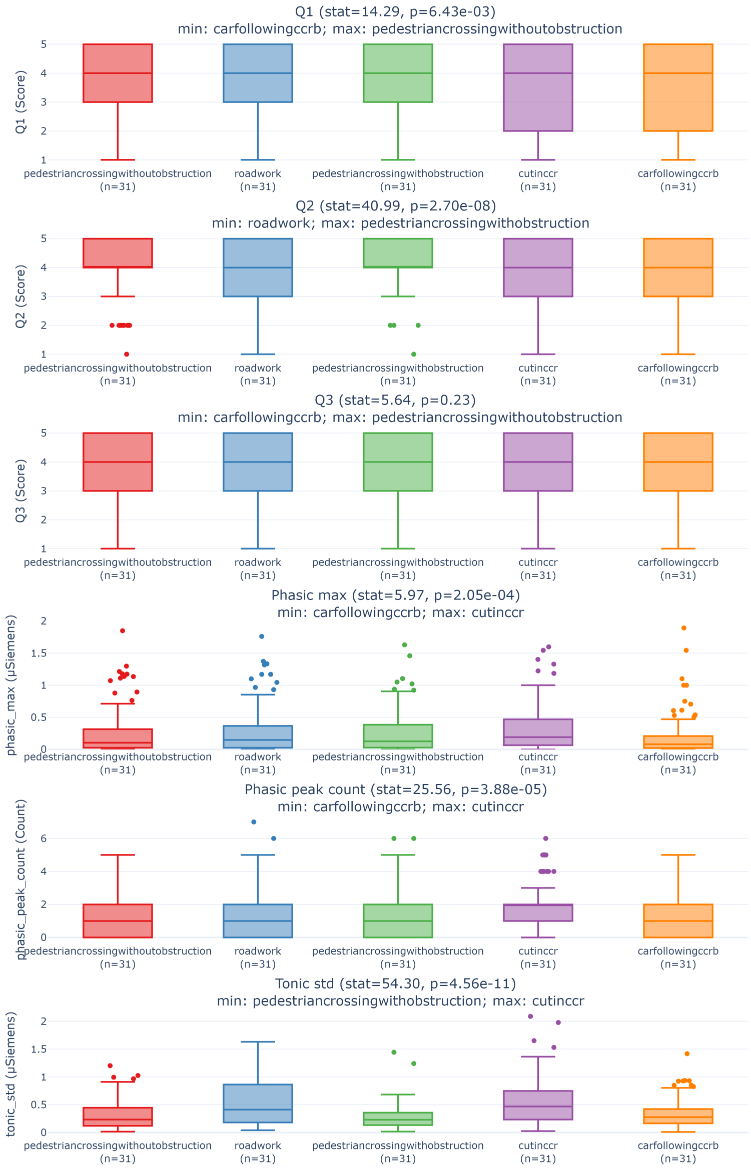


Figure B.10: Boxplots comparing subjective and physiological responses across all scenarios. Statistical significance was assessed by using a repeated-measures ANOVA. Test statistics and p-values are indicated in each subplot title. Sample size per group (n) reflects the number of unique participants in this grouping.

This figure shows statistically significant differences across scenarios for perceived comfort (Q1) and safety (Q2), but not for overall ride comfort (Q3), suggesting that environmental elements influence perceived comfort and safety, but not overall ride comfort. The *carfollowingccrb/cutinccr* scenario was rated as least comfortable, and the *roadwork* scenario as least safe. Physiologically, significant differences were also observed. The *cutinccr* scenario consistently triggered the highest sympathetic arousal, followed by *roadwork*. Notably, an inconsistency emerged; despite being perceived as the least comfortable, *carfollowingccrb* showed the lowest phasic response. Furthermore, pedestrian-related scenarios were generally perceived as more comfortable and safe, which is also reflected in the physiological data.

Object-related

With various different obstruction settings in the experiment, pairwise comparisons can be performed to gain insights into how different objects or settings influence passenger state. Figure B.11 shows the boxplot of the differences between scenarios that involve a car (the cut-in and car-following scenarios) and a pedestrian (pedestrian crossing without and with visual obstruction scenarios).

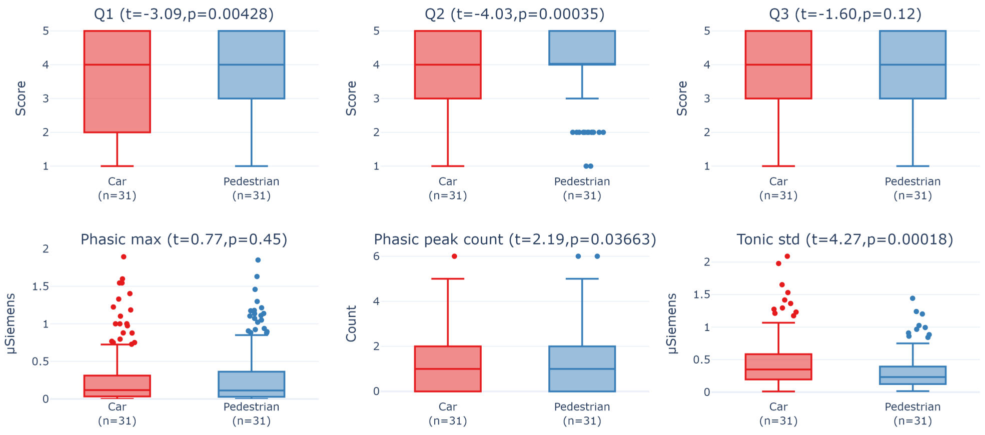


Figure B.11: Boxplots comparing subjective and physiological responses between scenarios involving a car or a pedestrian. Statistical significance was assessed using paired t-tests. Test statistics and p-values are indicated in each subplot title. Sample size per group (n) reflects the number of unique participants in this grouping.

This figure reveals strong, significant differences for perceived comfort (Q1) and perceived safety (Q2) between scenarios involving cars and pedestrians, with higher comfort and safety scores reported in the pedestrian-related scenarios. No significant difference was observed in overall ride comfort (Q3).
On the physiological side, a significant increase in phasic peak count and tonic standard deviation indicates greater sympathetic nervous system activation during car-related scenarios. However, no significant difference was found for the phasic maximum.
This suggests that car-related scenarios are perceived more uncomfortable, as evidenced by both subjective and physiological responses. A likely explanation lies in the higher speeds typically associated with these scenarios or the increased risk they pose, since collisions, especially at such speeds, tend to have more severe consequences for passengers. Consequently, participants may have exhibited stronger responses under these conditions.

Configuration	Q1		Q2		Q3		Phasic max.		Phasic peak count		Tonic std.	
	t-stat	p-value	t-stat	p-value	t-stat	p-value	t-stat	p-value	t-stat	p-value	t-stat	p-value
Car, calm vs. Ped., calm	0.24	=0.81	-1.72	=0.09	-0.93	=0.35	1.64	=0.11	2.82	<0.05	4.44	<0.05
Car, aggressive vs. Ped., aggressive	-4.00	<0.05	-3.64	<0.05	-1.40	=0.15	-1.01	=0.32	0.23	=0.82	2.25	<0.05
Car, calm vs. Ped., aggressive	6.11	<0.05	3.25	<0.05	6.91	<0.05	-5.79	<0.05	-5.4	<0.05	1.35	=0.29
Car, aggressive vs. Ped., calm	-7.12	<0.05	-6.37	<0.05	-8.04	<0.05	3.78	<0.05	7.57	<0.05	4.88	<0.05

Table B.10: Comparison of subjective and physiological responses between scenarios involving a car or a pedestrian for each driving style configuration. Paired t-tests were done for statistical analysis.

According to Table B.10, during the calm driving there were no significant differences in participants' comfort and safety ratings between the car- and pedestrian-related scenarios. Under aggressive driving, however, participants reported higher comfort and perceived safety around the pedestrian.
As expected, comparisons involving different driving styles, regardless of the object, generally resulted in significant differences, highlighting the strong influence of the driving style on passenger perception and state.

The second comparison, shown in Figure B.12, shows the differences between the pedestrian without obstruction and with visual obstruction.

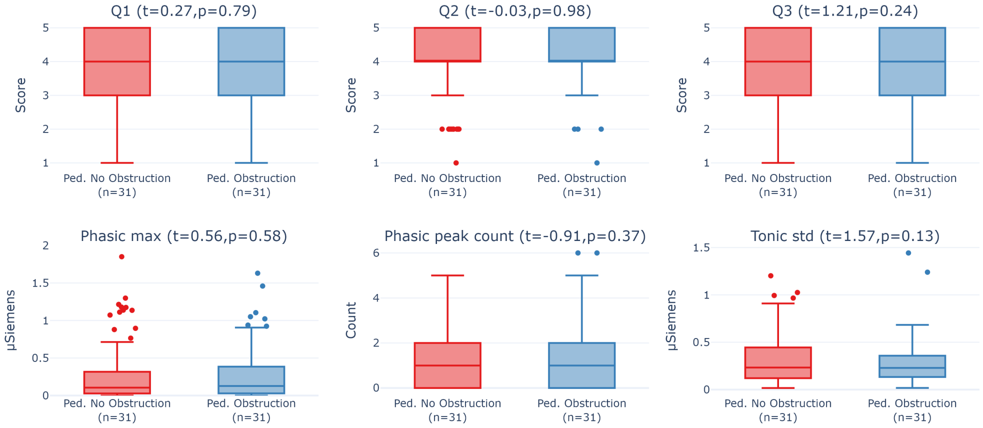


Figure B.12: Boxplots comparing subjective and physiological responses between the pedestrian crossing scenarios with and without visual obstruction. Statistical significance was assessed using paired t-tests. Test statistics and p-values are indicated in each subplot title. Sample size per group (n) reflects the number of unique participants in this grouping.

This figure shows no statistically significant difference between scenarios with and without visual obstruction for the pedestrian crossing across all subjective and physiological responses. This indicates that the presence of a visual obstruction in the pedestrian-related scenarios did not meaningfully alter the passenger's state.

Table B.11 presents the result of the same comparison with the driving style as an extra factor to distinguish between configurations.

	Q1		**Q2**		**Q3**		**Phasic max.**		**Phasic peak count**		**Tonic std.**	
Configuration	t-stat	p-value	t-stat	p-value	t-stat	p-value	t-stat	p-value	t-stat	p-value	t-stat	p-value
No obs., calm vs. Obs. calm	-0.38	=0.70	-0.96	=0.34	-0.14	=0.89	-0.69	=0.49	0.60	=0.55	-0.84	=0.41
No obs., aggressive vs Obs. aggressive	0.75	=0.46	0.78	=0.44	1.02	=0.32	1.39	=0.17	-1.91	=0.07	3.27	<0.05
No obs., calm vs. Obs. aggressive	3.80	<0.05	3.86	<0.05	6.50	<0.05	-5.50	<0.05	-5.39	<0.05	-1.68	=0.10
No obs., aggressive vs. Obs. calm	-6.29	<0.05	-3.61	<0.05	-7.84	<0.05	3.71	<0.05	6.24	<0.05	3.58	<0.05

Table B.11: Comparison of subjective and physiological responses between the pedestrian crossing scenarios with and without visual obstruction (Obs.) for each driving style configuration. Paired t-tests were done for statistical analysis.

This table indicates that the presence or absence of a visual obstruction in the pedestrian crossing did not significantly affect participants' perceived comfort or safety, regardless of driving style. Similarly, no significant differences were observed in the physiological features, except for the tonic standard deviation.
Contrary, comparisons involving different driving styles generally yielded significant differences across both subjective and physiological measures. Reinforcing the finding that driving style plays a dominant role in shaping the passengers' state.

Thirdly, in the pedestrian crossing without obstruction, the pedestrian did not cross the road in laps 1 and 4 but did so in laps 2 and 3. Figure B.13 shows the different responses in this setting.

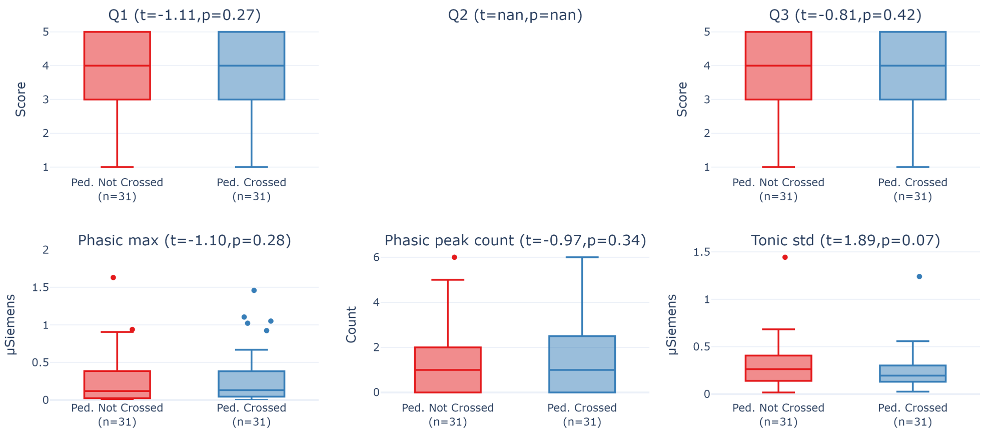


Figure B.13: Boxplots comparing subjective and physiological responses between pedestrian crossing with visual obstruction, where the pedestrian either crossed or did not. Statistical significance was assessed using paired t-tests. Test statistics and p-values are indicated in each subplot title. Sample size per group (n) reflects the number of unique participants in this grouping.
Note: Q2 was omitted in laps where the pedestrian did not cross, as this question was not applicable. As a result, no statistical test could be performed, and the plot is excluded.

No significant differences were found across any of the subjective or physiological measures, suggesting that whether the participant crossed or not had little to no impact on the passengers' state. However, the tonic standard deviation approached statistical significance ($p = 0.07$), potentially indicating a subtle increase in baseline arousal when the pedestrian did not cross, possibly due to heightened attention or anticipation. This effect, however, remains marginal and not significant.

Results of the same comparison, now extended to include driving style as an additional factor to distinguish the conditions, are shown in Table B.12.

	Q1		**Q2**		**Q3**		**Phasic max.**		**Phasic peak count**		**Tonic std.**	
Configuration	t-stat	p-value	t-stat	p-value	t-stat	p-value	t-stat	p-value	t-stat	p-value	t-stat	p-value
Not crossed, calm vs. Crossed calm	-0.62	=0.54	-	-	-0.44	=0.66	-0.76	=0.45	-1.22	=0.23	-1.07	=0.29
Not crossed aggressive vs. Crossed aggressive	-0.12	=0.90	-	-	-0.21	=0.83	-0.31	=0.76	0.46	=0.64	3.00	<0.05
Not crossed, calm vs. Crossed aggressive	3.68	<0.05	-	-	7.39	<0.05	-2.49	<0.05	-3.74	<0.05	0.28	=0.78
Not crossed, aggressive vs. Crossed calm	-4.51	<0.05	-	-	-5.93	<0.05	2.88	<0.05	2.20	<0.05	2.00	=0.05

Table B.12: Comparison of subjective and physiological responses between the pedestrian crossing with visual obstruction scenario, where the pedestrian either crossed or did not for each driving style configuration. Paired t-tests were done for statistical analysis.

This table again shows that when driving style is held constant, no significant differences emerge between the configurations where the pedestrian crossed or did not cross. This indicates that the environmental variation has minimal impact on the participant's subjective and physiological responses.
Comparisons across different driving styles consistently result in significant differences again, further reinforcing the conclusion that driving style is the dominant factor influencing passenger perception and physiological state.

Demographics
With a wide range of ages and a balanced gender distribution, pairwise comparison across demographic groups may offer additional insights into how different segments of the population respond to automated driving. Figure B.14 shows the pairwise comparison between genders.

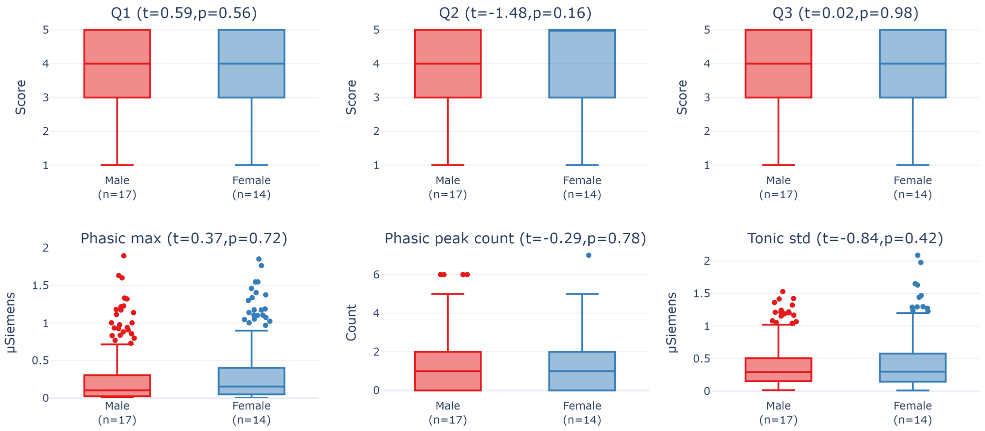


Figure B.14: Boxplots comparing subjective and physiological responses between genders for all scenarios. Statistical significance was assessed using paired t-tests. Test statistics and p-values are indicated in each subplot title. Sample size per group (n) reflects the number of unique participants in this grouping.

In this figure, no significant differences can be observed between genders, suggesting that gender did not meaningfully influence how participants perceived or physiologically responded to the automated driving experiment.

Figure B.15 shows the comparisons across age groups.

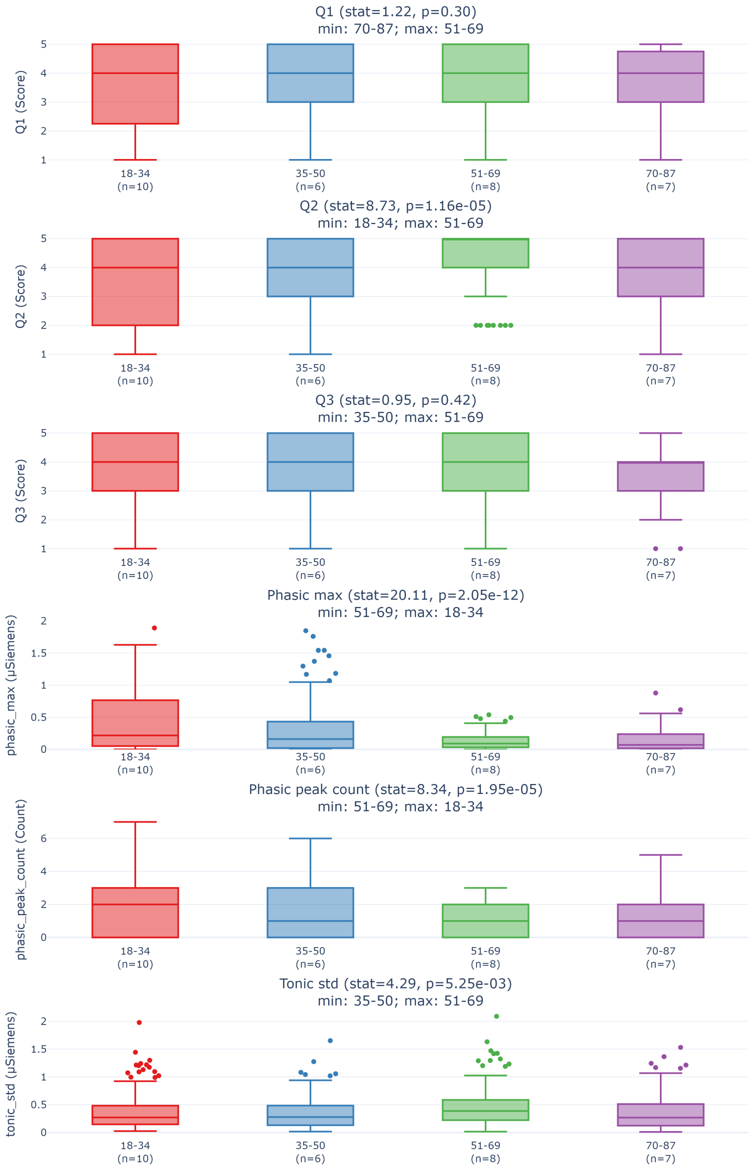


Figure B.15: Boxplots comparing subjective and physiological responses across age groups. Statistical significance was assessed by using a one-way ANOVA. Test statistics and p-values are indicated in each subplot title. Sample size per group (n) reflects the number of unique participants in this grouping.

This figure shows that, among the self-reported scores, only perceived safety (Q2) differs significantly between age groups, with the youngest participants rating scenarios as least safe. This could reflect their limited driving experience, leading to a heightened perception of risk. The physiological responses also vary significantly by age. The two younger groups exhibit increased phasic activity, suggesting a more reactive sympathetic nervous system, possibly indicating a quicker or more intense response to stimuli.

However, given the small group sizes ($n = 6 - 10$), the reliability of these statistical differences should be interpret with caution. While p-values fall below 0.05, the limited sample size reduces the statistical power.

Pre-Questionnaire and Post-Questionnaire

Figures B.16, B.17 and B.18 show the comparisons between the pre- and post-questionnaire questions related to likeliness to motion sickness, trust in automated driving and willingness to adopt a self-driving car.

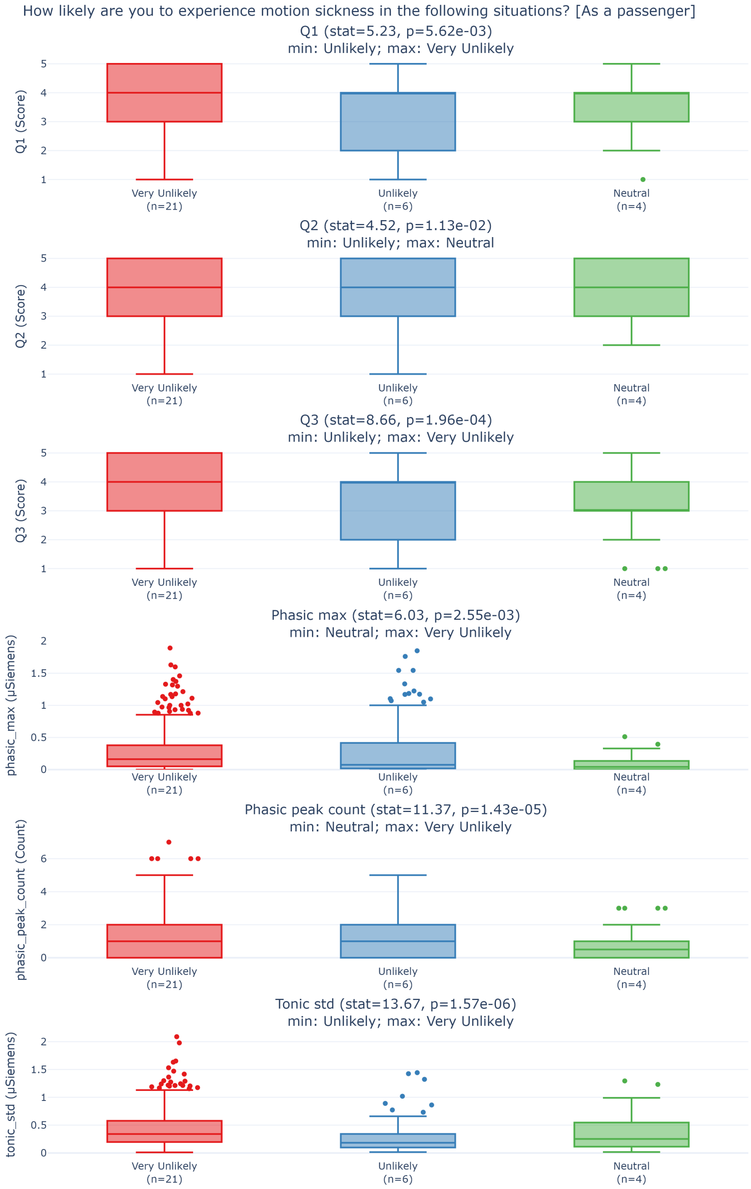


Figure B.16: Boxplots comparing subjective and physiological responses across pre-questionnaire responses. Statistical significance was assessed by using a one-way ANOVA. Test statistics and p-values are indicated in each subplot title. Sample size per group (n) reflects the number of unique participants in this grouping.

In this figure there are significant differences across all measurements; however, no consistent interpretable pattern emerges. Notably, participants who responded “Very unlikely” showed the highest phasic activity, contrary to expectations, as greater physiological arousal might be expected in those more prone to motion sickness.

This unexpected trend, along with the overall lack of a clear pattern, may be due to the small and uneven group sizes. As such, these results should be interpreted with caution as they may reflect random variation rather than meaningfully differences.

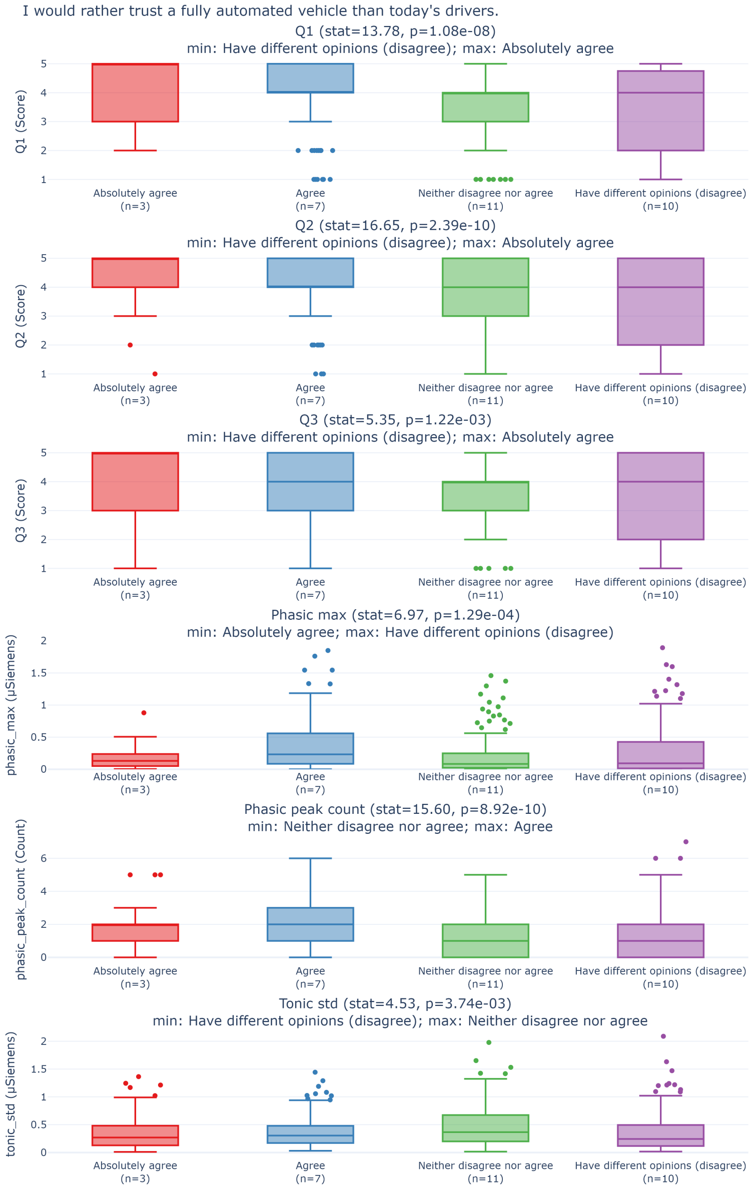


Figure B.17: Boxplots comparing subjective and physiological responses across pre-questionnaire responses. Statistical significance was assessed by using a one-way ANOVA. Test statistics and p-values are indicated in each subplot title. Sample size per group (n) reflects the number of unique participants in this grouping.

This figure shows that participants who disagreed with the presented statement consistently reported lower subjective scores. This may reflect a form of bias or scepticism towards automated vehicles, which could have shaped their overall experience during the experiment.

Physiologically, this group also exhibit higher feature values, suggesting they may have been more stressed or alert, potentially because of their negative bias to automated driving.

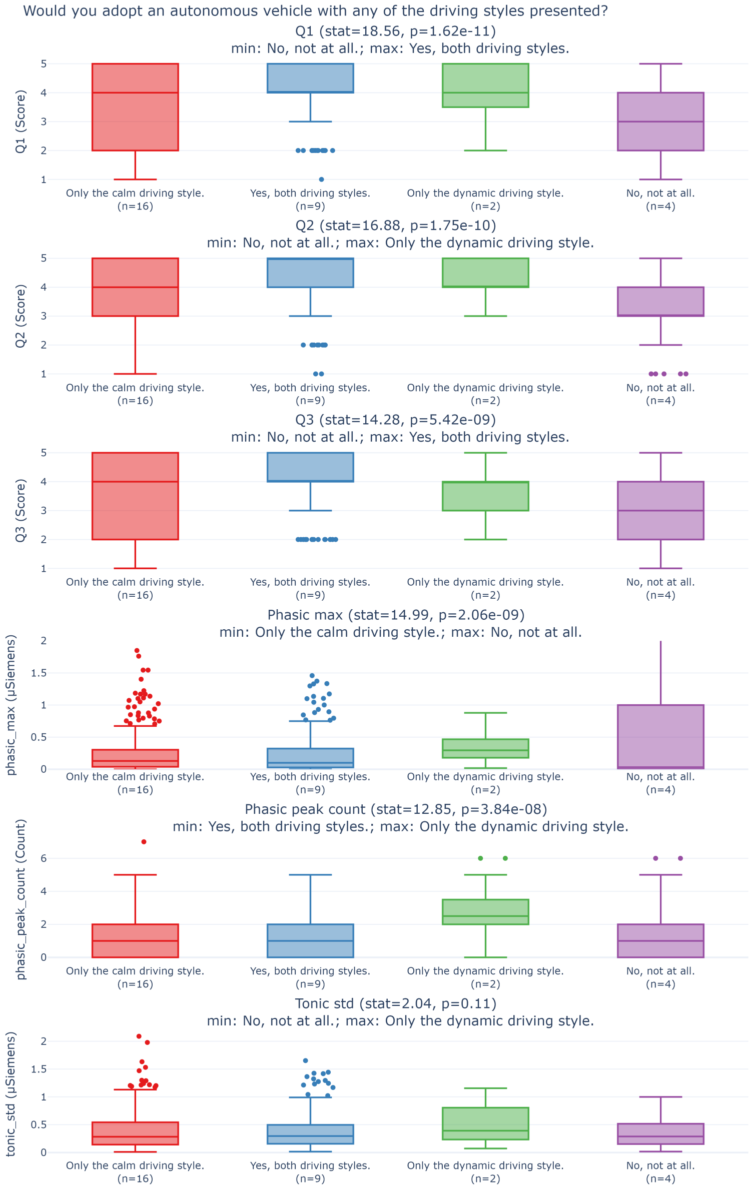


Figure B.18: Boxplots comparing subjective and physiological responses across pre-questionnaire responses. Statistical significance was assessed by using a one-way ANOVA. Test statistics and p-values are indicated in each subplot title. Sample size per group (n) reflects the number of unique participants in this grouping.

This figure reveals that participants who indicated they would adopt autonomous vehicles only with a calm driving style, or not at all, reported lower comfort and perceived safety scores. These lower rating may have influenced their post-questionnaire response, reflecting critique towards automated driving. No clear pattern is observed in the phasic component and no significant differences were found in baseline arousal, suggesting that these subjective perceptions were not strongly mirrored in the sympathetic nervous system activity.

To the post-questionnaire question: *How would you describe your driving style?* out of four options, only two options were chosen:

- Defensive driving style: ”I prioritize safety, keep my distance and follow traffic rules diligently.”
- Confident driving style: ”I feel safe behind the wheel, make quick decisions and am in control.”

Figure B.19 shows a pairwise comparison between the segments of participants who choose either the first or the second response.

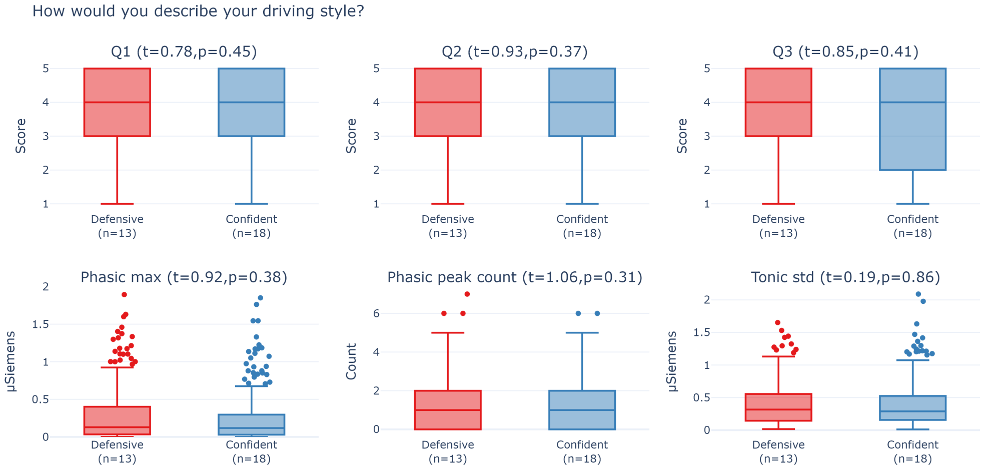


Figure B.19: Boxplots comparing subjective and physiological responses between participants' self-reported, own driving styles. Statistical significance was assessed using paired t-tests. Test statistics and p-values are indicated in each subplot title. Sample size per group (n) reflects the number of unique participants in this grouping.

No statistically significant differences were observed across any of the measures. Although participants identifying as "confident" driving tended to report slightly lower overall ride comfort scores (Q3), this trend was not significant.
These findings suggest that individuals' self-reported own driving style does not meaningfully influence their perceived comfort, safety or physiological arousal as passengers.

B.4. Discussion

This chapter aimed to understand passengers' perceived comfort and safety in automated driving using the Galvanic Skin Response, combined with vehicle dynamics and perception data. In doing so, five subquestions will be answered in this section using the data gathered for this study:

1. **Did the calm and aggressive driving style elicit distinct perceived comfort and safety scores and physiological (GSR) responses?**

   To answer this question, the following observations were made:

   - Significant correlations emerged between all three subjective score metrics and the driving style, where the aggressive driving style consistently resulted in lower scores in perceived comfort, safety and overall ride comfort (Table B.2: 1-4).
   - Significant correlations emerged between the majority of the GSR-related features and the driving style, where the aggressive driving style consistently led to increased GSR activity. The best features found to indicate the driving style in the GSR signal are (Table B.2: 1, 5-16):
     - Phasic maximum amplitude
     - Phasic peak count
   - The effect of driving style on subjective responses and GSR-related features is illustrated in Figures B.5, B.6 and B.7.

   Together, these findings confirm that the two driving styles elicited distinct subjective and physiological responses in this experiment.

2. **To what extent does the GSR signal reflect changes in perceived comfort, safety and overall ride comfort?**

   The following observations were made regarding this question:

- Significant correlations emerged between all three subjective score metrics and various GSR-related features, where lower comfort and safety scores were reflected by an elevated GSR activity (Table B.2 2-16).
- Stronger correlations were found for phasic features than tonic features, indicating that the short-term fluctuations driven by the sympathetic nervous system more accurately reflect passengers' emotional state. The strongest correlations were found for:
  - Phasic maximum amplitude
  - Phasic peak count
- This effect is observable in Figure B.8, which shows an average increase in GSR feature value for a decrease in comfort and safety scores.
- This relationship is further illustrated in Figure B.9 in which a clear trend emerges: lower subjective ratings tend to cluster in regions of increased phasic activity, reinforcing the link between the GSR and perceived discomfort.
- However, the significant correlations largely disappear when the analysis is isolated per driving style configuration (Tables B.3, B.4 5-16).

These findings illustrate that the GSR broadly reflects changes in perceived comfort and safety, particularly when comparing the distinct driving styles in this experiment. However, within a single driving style configuration, where the subjective differences are more subtle, the GSR signal does not capture these finer nuances.

These findings are consistent with prior studies: one demonstrated that the GSR is a significant predictor of passenger comfort and anxiety, measured through comfort and anxiety rating [8]; one that identified the phasic maximum amplitude as a key indicator of discomfort measured through a continuous discomfort throttle [19]; and one that highlighted the phasic peak count as a key indicator of discomfort measured through the amount of discomfort button presses per minute [24].

3. **Which of the three input modalities, GSR, vehicle dynamics or perception, explains the largest share of variance in perceived comfort and safety?**

   Table B.2 reveals the following order magnitude of correlation for input modalities:
   1. Longitudinal movement (maximum acceleration, jerk)
   2. Velocity features (maximum)
   3. Lateral movement (maximum acceleration, jerk)
   4. Phasic component features (maximum, peak count)
   5. Perception features (minimum distance to object, TTC)
   6. Tonic component features (standard deviation)
   7. Directional movement (yaw rate minimum, maximum)

4. **Which of the objective signals exerts the strongest influence on the GSR signal?**

   Using the phasic maximum amplitude and peak count as key features of the GSR, the following order of magnitude of correlation to the GSR signal emerges from Table B.2:
   1. Lateral movement (minimum acceleration, maximum jerk)
   2. Longitudinal movement (maximum acceleration, maximum jerk)
   3. Velocity (maximum)
   4. Directional movement (yaw rate minimum, maximum)
   5. Perception cues (non-significant)

   These results suggest that GSR is particularly sensitive to abrupt vehicle motion, with the strongest responses linked to extreme values in lateral and longitudinal movements. It also suggests that the GSR is negligibly influenced by perception-related cues; however, this outcome should be interpreted with caution, as it is unclear whether it reflects true insensitivity or simply stems from the incomplete perception data.

   A similar study reported that GSR increased proportionally with the magnitude of longitudinal acceleration and jerk. However, it also found an elevated GSR in the presence of a lead vehicle and proximity, which was not observed in this experiment [8]. Furthermore, a significant correlation between phasic component features and lateral acceleration was found [12], and a quick reaction of GSR to strong braking [29].

5. **In what ways do scenario characteristics (e.g., VUT target velocity, the presence of another vehicle versus a pedestrian or visibility) and passenger demographics (age, gender, trust in automation) influence the relationships identified in the previous questions?**

   Comparing subjective and physiological responses across scenarios reveals the following:
   - Comparing the scenarios in general reveals:
     - The cut-in/car-following scenario received a significantly lower perceived comfort (Q1) score (both scenarios were evaluated using the same questionnaire), closely followed by the roadwork scenario.
     - The roadwork scenario received a significantly lower perceived safety score (Q2), followed by the cut-in/car-following.
     - Both pedestrian crossing scenarios received the highest perceived comfort (Q1) and safety (Q2) scores.
     - No significant differences in subjective ratings emerged in the overall ride comfort (Q3) between scenarios.
     - These patterns in perceived comfort and safety were mirrored by the physiologic responses, with the highest physiological arousal measured in the roadwork scenario, followed by the cut-in scenario, while the lowest was observed in the car-following and pedestrian crossing scenarios.
   - Exploring more in-depth the differences between driving styles reveals:
     - The largest differences in subjective and physiological scores between driving styles were found for the roadwork scenario. This scenario featured the most pronounced variation in driving behavior, with target velocity increasing from 30 km/h (calm) to 70 km/h (aggressive), alongside up to fourfold higher longitudinal acceleration and fivefold higher lateral acceleration.
     - The cut-in scenario showed the second-largest difference in physiological arousal between driving styles. Despite only modest acceleration differences, the aggressive condition again involved a higher target velocity of 70 km/h, likely contributing to the observed increase in GSR activity
     - The car-following scenario showed the smallest difference in physiological response between driving styles. In this scenario, the main variation was the reduced distance to the lead vehicle, suggesting that changes in vehicle dynamics, rather than proximity alone, play a more decisive role in passenger arousal.
     - The pedestrian crossing scenarios showed a clear increase in physiological arousal under aggressive driving, though subjective ratings remained relatively stable across driving styles. This may reflect more subtle differences in vehicle dynamics in these scenarios.

   Pairwise comparison in scenario-specific environmental cues reveals the following observations:
   - Scenarios involving the GVT scored significantly lower in perceived comfort (Q1) and safety (Q2) scores than scenarios involving a pedestrian. This effect is also observable in physiological responses, with higher physiological arousal for the scenarios involving the GVT (Figure B.11).
   - No significant differences emerged in subjective and physiological responses between the pedestrian without visual obstruction and with visual obstruction (Figure B.12).
   - No significant differences emerged in subjective and physiological responses in the pedestrian crossing with visual obstruction in the laps where the pedestrian crossed the road or not (Figure B.13).
   - Regardless of whether the environmental context produced a significant effect, in pairwise comparison the scenario driven aggressively was consistently rated less comfortable and safe, and showed higher physiological arousal, hinting that vehicle dynamics outweigh perception cues in shaping passenger responses (Table B.10, B.11, B.12).

Summarizing these findings, the data suggest that passengers' perceived comfort and safety in this experiment are primarily influenced by vehicle dynamics. This effect is consistently mirrored in the physiological responses, with higher GSR activity observed in scenarios characterized by more intense acceleration or speed. Similarly, a prior study found significantly more discomfort in rural driving environments compared to urban settings, possibly due to the higher average velocities typically encountered in rural areas [24]. Moreover, another study found that bad weather conditions, such as rain, correlated with an increase in phasic peaks [26]. These prior studies, taken together with the present findings, further underscore the situational nature of both subjective comfort and physiological arousal.

Exploring passenger demographics reveals:

- No significant differences emerged in subjective and physiological responses between male and female participants (Figure B.14).
- Significant differences emerged in subjective and physiological responses between age groups; the youngest participants reported lower perceived safety, possibly reflecting a lack of driving experience. This group also exhibited higher phasic activity, suggesting a more reactive sympathetic nervous system (Figure B.15).
- No consistent pattern of significant differences was found between groups split by likelihood of motion sickness (Figure B.16).
- Participants reporting low trust in AVs in the pre-questionnaire gave significantly lower comfort and safety ratings, potentially reflecting a pre-existing bias. This was accompanied by higher physiological arousal (Figure B.17).
- Participants who stated they would only adopt AVs with the calm driving style, or not at all, gave significantly lower comfort and safety ratings, possibly reflecting a negative overall evaluation of the experience. No clear pattern emerged in the physiological response, however, as different features peaked in opposing groups (Figure B.18).
- No significant differences emerged in subjective and physiological responses between participants who described their driving style as defensive versus confident (Figure B.19).

Taken together, these demographic findings highlight substantial inter-individual variability in both subjective and physiological responses. Factors such as age and pre-existing attitude towards AV influenced not only self-reported comfort and safety scores, but also GSR patterns. This underscores a key challenge in using the GSR for comfort and safety assessment: both GSR and subjective ratings are highly sensitive to inter-subject differences, making it difficult to generalize findings across diverse populations.

Further research could strengthen the findings of this analysis. While this work showed that GSR is particularly responsive to vehicle dynamics and less so to perception-related cues, this, however, may partly reflect the structure of the scenarios used. Specifically, scenarios involving the GVT typically featured more intense vehicle dynamics than those with the pedestrian.

To disentangle the effects of vehicle dynamics from perception, a future experiment should include scenario pairs that match in vehicle dynamic profiles, but vary in perceptual content. For example, a road crossing that is approached with a consistent target velocity, but is varied by the type of road user crossing the path, such as a pedestrian, cyclist, car or truck.

Furthermore, expanding the scenario set to cover a broader range of driving contexts, such as roundabouts, traffic jams, urban streets or parking maneuvers, as well as incorporating different weather conditions like clear skies, rain, or fog, could provide deeper insights into how comfort and physiological arousal vary across real-world scenarios.

Lastly, increasing the overall participant sample size would enhance the statistical power of the demographic analyses and support more robust conclusions. As all participants in this study were German, expanding the sample to include individuals from different nationalities could also provide insights into how cultural background influences the experience of automated driving.

B.5. Conclusion

This chapter showed that the GSR can be a useful signal for assessing passenger comfort and perceived safety in automated driving. Clear differences emerged in GSR signals between calm and aggressive driving styles, with aggressive driving consistently linked to higher GSR activity and lower subjective ratings. The GSR signal, especially the phasic component's maximum amplitude and peak count, captures broad comfort differences, but it struggles to detect more subtle within-style variations.
Still, the GSR can be used to assess comfort when interpreted alongside self-reported scores: strong agreement between high GSR activity and low comfort ratings strengthens the evidence that a scenario, or driving style, truly undermines passenger comfort. In contrast, when the two diverge, the GSR can be used as an objective cross-check, helping to identify inattentive, biased or inconsistent self-reports. This use makes the GSR a valuable tool for understanding passenger comfort beyond subjective feedback alone.

Leveraging this cross-validation, the analysis pinpointed the scenario features that most consistently provoked discomfort. The strongest subjective and GSR responses occurred in scenarios with intense vehicle dynamics, namely high acceleration, jerk, and speed, while perception-related cues such as the presence of a pedestrian or visual obstruction had minimal effect. This suggests that the perceived comfort and safety, and the GSR are most sensitive to dynamic driving inputs from this experiment.

Scenario and demographic analyses further underline the influence of both contextual and individual factors. This variability poses a challenge for generalizing GSR-based comfort and safety assessment across diverse scenarios and passenger groups.

C

Predictive Modeling

This chapter focuses on the predictive component of the research question in this study:

> How can physiological arousal, measured through Galvanic Skin Response, combined with vehicle dynamics and perception data, be utilized to understand and **predict** passengers' perceived comfort and safety in automated driving?

While the scientific paper written in Chapter 1 presents a concise summary of the results, this chapter offers a more detailed account of the intermediate steps and performance evaluations.

Building on the correlation analyses in Appendix B, which showed that the GSR correlates with both objective driving characteristics and subjective comfort ratings, this chapter now evaluates how well this signal can predict those outcomes. On that basis, the following subquestions were derived:

1. How accurately can the GSR alone predict objective driving style (calm vs. aggressive)?
2. How accurately can the GSR alone predict perceived comfort, perceived safety and overall ride comfort?
3. To what extent do vehicle dynamics and/or perception data improve GSR-based models for subjective comfort metrics?
4. What challenges arise when predicting subjective ratings, and how can these challenges be addressed to improve model performance?

Contrary to Appendix B, which relied on extracted features per scenario, this chapter adopts a Deep Learning (DL) approach by retaining the full 30-second time series signals and applying models directly to the signals.

The chapter is structured as follows: first, the chosen model architecture is presented, followed by evaluations of GSR-based models for driving style classification and subjective comfort predictions. Through this evaluation, the challenges that emerge are explicitly identified, together with mitigation techniques. Finally, in a discussion, each subquestion is addressed, and a conclusion is drawn that integrates the findings of this chapter.

C.1. Proposed Architecture

The proposed architecture, outlined in this chapter, is the Time Evidence Fusion Network (TEFN), developed by Zhan et al. [35]. The TEFN model, originally designed for multivariate time series forecasting, is adapted in this work to perform a classification task. A novel backbone is proposed that achieves comparable performance to state-of-the-art methods while maintaining a significantly lower complexity and reduced training time [35]. Figure C.1 presents the original overall structure of the TEFN architecture.

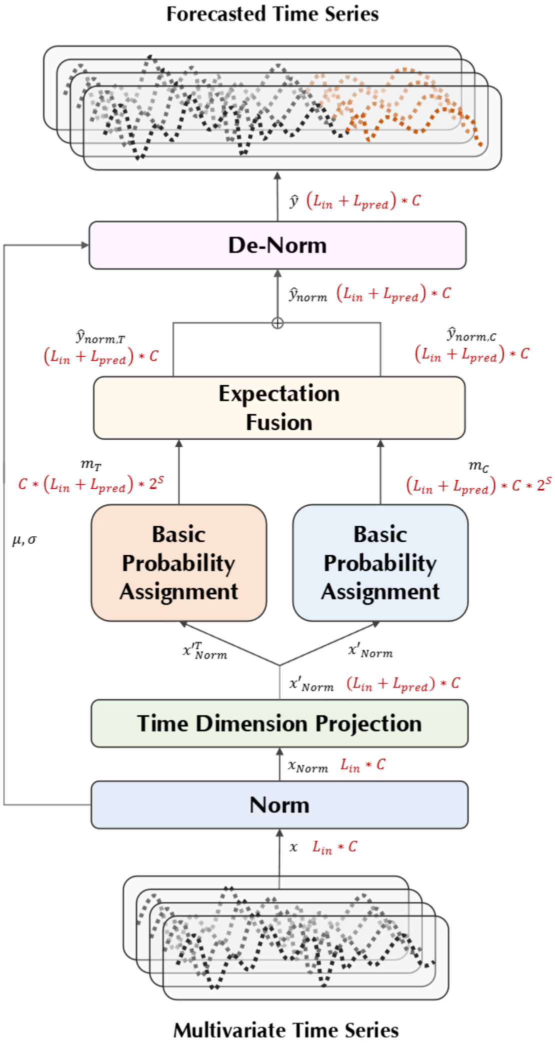

Forecasted Time Series
$\hat{y}$ $(L_{in} + L_{pred}) * C$
De-Norm
$\hat{y}_{norm}$ $(L_{in} + L_{pred}) * C$
$\hat{y}_{norm,T}$
$(L_{in} + L_{pred}) * C$
$\hat{y}_{norm,C}$
$(L_{in} + L_{pred}) * C$
Expectation Fusion
m_T
$C * (L_{in} + L_{pred}) * 2^S$
m_C
$(L_{in} + L_{pred}) * C * 2^S$
μ, σ
Basic Probability Assignment
Basic Probability Assignment
x'^T_{Norm}
x'_{Norm}
x'_{Norm} $(L_{in} + L_{pred}) * C$
Time Dimension Projection
x_{Norm} $L_{in} * C$
Norm
x $L_{in} * C$
Multivariate Time Series

Figure C.1: The overall structure of TEFN for a time series forecasting task. Adapted from *Time Evidence Fusion Network: Multi-source View in Long-Term Time Series Forecasting* by Zhan et al. [35].

Input time series goes through the following modules and the following processes for the classification task: [35]

1. **Normalization**: The input is first normalized by calculating the mean μ and variance σ^2. These μ and σ^2 values are then passed to the final de-normalization to transform the output back to its original map. This normalization reduces the impact of outliers, promotes faster convergence, and enhances stability during training.
2. **Time Dimension Projection**: Next, a basic linear projection layer is typically used to transform the normalized time series x_{norm} of length L_{in} into a sequence x' with length $L_{in}+L_{pred}$. However, this step is omitted for classification tasks, as no prediction of future values is required.

3. **Basic Probability Assignment**: TEFN represents the uncertainty and ambiguity of time series through Evidence Theory. Each dimension of the time series is described by a mass function defined on the power set of 2^S of a finite sample space S, where the choice of S depends on the task at hand. For classification, S is the set of class labels, so S equals the number of classes.

   This representation allows TEFN to capture so-called fuzzy characteristics of the input by assigning a belief distribution, rather than a single point, to each observation. The belief masses are generated with a learnable fuzzy membership function:

   $$m_{D,i,j,k} = \mu(x_{Norm,i,j}) = w_{D,j,k} * x_{Norm,i,j} + b_{D,j,k} \tag{C.1}$$

   Where:
   - $D \in \{T, C\}$ selects the evidence source-time axis T or channel axis C;
   - i indexes the time-step and j the channel;
   - k indexes the F in 2^S, i.e., an individual class label in classification.

   The parameters $w_{D,j,k}$ and $b_{D,j,k}$. which represent the slope and intercept, respectively, are learned during training.

   The resulting vector $m_{D,i,j}$ contains non-negative entries that sum to one, forming the Basic Probability Assignment (BPA) for that data point.

4. **Expectation Fusion**: TEFN handles multivariate time series data by generating separate BPAs for the time dimension (T) and the channel dimension (C). This results in two parallel mass distributions: m_T and m_C.
   These are fused using an expectation fusion approach, which involves multiplying each mass distribution by a learned weight y and summing the results, yielding a single fused mass function.

   The TEFN applies this method of expectation fusion and deliberately avoids the Dempster-Shafer Rule (DSR) as its computational complexity is significantly higher, and the DSR is sensitive to extreme distributions, which can cause a single distribution to dominate the fused result.

5. **Classification**: Unlike the original mode, which uses de-normalization to map the output back to its original values (C.1), the classification variant does not require this step. Instead, the tensor produced by the Expectation Fusion, shaped $L \times C$ with L the input sequence length and C the number of channels, is first passed through a sigmoid activation. The result then flattened into a vector of length $L\,C$ and fed into a fully connected layer mapping $\mathbb{R}^{LC} \to \mathbb{R}^{n_{\text{classes}}}$.

The TEFN architecture was selected as it matches or exceeds current state-of-the-art architectures, while relying on far fewer parameters [35]. This makes the model a well-suited choice for this study, as its parameter efficiency aligns with the constraints of a limited dataset and as it has demonstrated to be effective with time series. TEFN explicitly models both intra-channel and inter-channel dependencies by applying its BPA modules separately along the time axis and the channel axis, followed by a fusion step. This structure allows the TEFN to learn interactions between signals, which can be especially beneficial for GSR data where the phasic and tonic components often interplay.

To mitigate the risk of overfitting, given the dataset size, a dropout was introduced after the Expectation Fusion block. In addition, an early stopping mechanism was implemented during training to further regularize the learning process and avoid overfitting.

C.2. Driving Style

For this task, driving style classification was formulated as a binary classification problem. The Time Evidence Fusion Network (TEFN) was trained using the Galvanic Skin Response (GSR) data as input, consisting of two input features: the phasic and tonic components. Both phasic and tonic components are standardized within each participant before being used as input, ensuring that the mode can robustly account for the inter-subject variability in these components. The model was configured with the hyperparameters listed in Table C.1.

Hyperparameter	Value
Input sequence length	960
Dropout	0.3
Batch size	64
Optimizer	AdamW
Loss function	Binary Cross-Entropy
Learning rate	$1e-4$
Epochs	500
Patience	50

Table C.1: Hyperparameters used for training the TEFN [35] model in the driving style classification task.

To ensure subject-independent evaluation, the dataset was split participant-wise into 21 participants for training, 5 for validation and 5 for final testing. Additionally, to assess the model's robustness and generalization capabilities given the limited dataset size, a 10-fold cross-validation was performed. Model performance was evaluated using accuracy, precision, recall and F1 score. Table C.2 reports the mean, maximum, minimum and standard deviations of these metrics across the 10 folds.

Metric	Mean (%)	Max. (%)	Min. (%)	Std. (%)
Accuracy	88.61	94.74	81.18	3.77
Precision	87.73	95.45	77.27	4.75
Recall	89.70	97.77	68.89	8.86
F1 score	88.61	94.62	80.95	3.98

Table C.2: Performance metrics over 10-fold cross-validation in the driving style classification model based on GSR data (phasic and tonic components).

Figure C.2 and C.3 present an aggregated confusion matrix and ROC-curve plot, respectively.

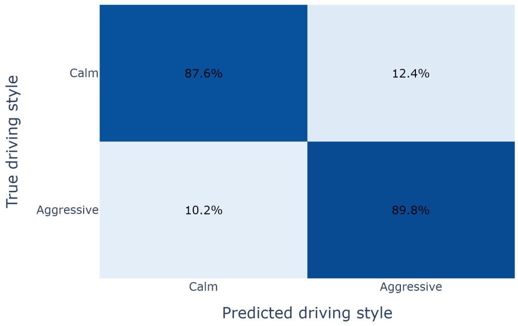


Figure C.2: Aggregated confusion matrix over 10-fold cross-validation for the driving style classification using GSR data (phasic and tonic components).

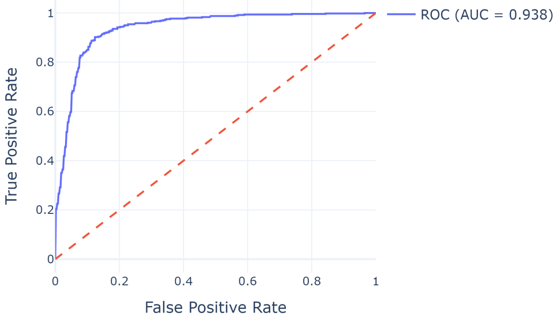


Figure C.3: Aggregated Receiver Operating Characteristic (ROC) curve for the driving style classification model using GSR data (phasic and tonic components), averaged over 10-fold cross-validation. Area Under Curve (AUC) = 0.879

To further explore the predictive power of the GSR signal, the training, validation, and testing processes were repeated using the following three input configurations: the raw, undecomposed GSR signal, the phasic component, and the tonic component. Tables C.3, C.4 and C.5 present the performance metrics for each configuration across a 10-fold cross-validation, respectively.

Metric	Mean (%)	Max. (%)	Min. (%)	Std. (%)
Accuracy	82.42	93.00	74.74	5.73
Precision	83.05	92.16	72.34	6.90
Recall	81.54	96.00	68.89	8.86
F1 score	81.99	93.07	73.91	6.26

Table C.3: Performance metrics over 10-fold cross-validation in the driving style classification model based on GSR data (undecomposed, raw GSR signal).

Metric	Mean (%)	Max. (%)	Min. (%)	Std. (%)
Accuracy	86.31	91.58	77.65	4.64
Precision	85.48	93.47	72.22	6.52
Recall	87.6	94.00	80.00	4.95
F1 score	86.38	91.30	77.11	4.70

Table C.4: Performance metrics over 10-fold cross-validation in the driving style classification model based on GSR data (phasic component).

Metric	Mean (%)	Max. (%)	Min. (%)	Std. (%)
Accuracy	81.14	90.00	70.00	7.21
Precision	82.99	91.11	69.23	7.64
Recall	77.82	94.00	62.00	9.67
F1 score	80.17	90.00	67.39	8.08

Table C.5: Performance metrics over 10-fold cross-validation in the driving style classification model based on GSR data (tonic component).

The results of the driving style classification task reveal the following noteworthy observations:

1. The model's solid performance (Table C.2) highlights the discriminative potential of the GSR signal for distinguishing driving style, offering evidence that physiological responses can capture meaningful behavioral patterns.
2. Precision (87.73%) and recall (89.70%) are fairly balanced, suggesting the model performs reasonably well in identifying both calm and aggressive driving styles without significant bias toward one class. However, the aggregated confusion matrix shows a slightly higher false positive rate for aggressive predictions compared to calm predictions, suggesting that some data recorded under the aggressive driving style may resemble data from calmer states.
3. The high AUC value of 0.938 from the ROC curve indicates that the model has strong discriminative power.
4. However, the relatively high standard deviations and wide ranges between minimum and maximum scores indicate that performance varies substantially across participant splits, suggesting that certain splits are more favorable than others, likely due to the inter-subject variability.
5. Combining the phasic and tonic components as input results in the highest performance (Table C.2), closely followed by only using the phasic component (Table C.4). Only using the tonic component, or directly using the raw, undecomposed GSR signal, performs slightly worse (Tables C.5, C.3, respectively). These results indicate that the phasic component provides the strongest predictive power, with the tonic component further enhancing its effectiveness.

C.3. Perceived Comfort and Safety

The preceding task employed a binary classification setup; however, the current formulation requires a 5-class multiclass classification problem. The objective is to predict self-reported scores on comfort and safety, given in a 5-point Likert scale. Model training, validation and testing are conducted separately for each questionnaire item to account for the item-specific response characteristics.

The TEFN model architecture remains unchanged, but is evaluated under varying input configurations, consisting of one or a combination of the following data sources:

- GSR signal: Phasic and tonic components.
- Vehicle Dynamics (VD): Velocity, longitudinal and lateral acceleration, jerk and yaw rate.
- Perception: distance to object, time-to-collision, time-headway and type of object.

Both phasic and tonic components are z-standardized per participant to enable the model to handle the inter-subject variability of the signal effectively.

To comprehensively evaluate each model's performance, a two-fold evaluation approach is applied. First, performance metrics are reported on an *exact match* criterion (hard), considering predictions that perfectly match the true label class as true positives. Second, a *near match* criterion (soft) is introduced, where predictions within one class of the ground truth (e.g predicting "comfortable" instead of "very comfortable") are also considered as true positives. The soft metrics account for the difficulty that participants face in making fine-grained distinctions between adjacent comfort levels, especially those without prior experience in such experiments. They also better reflect the needs of practical applications, where capturing broad differences between comfort and discomfort is more valuable than distinguishing subtle variations.

The soft metrics addressed the inherent subjectivity of the self-reported scores, where making fine-grained distinctions is difficult for participants to make and less critical in practical application.

Table C.6 lists the hyperparameters used in this task. These are identical to the ones listed in Table C.1, with the exception of the loss function. Here, a Cross-Entropy loss function is applied, as it is well-suited for the model's multi-class classification task.

Hyperparameter	**Value**
Input sequence length	960
Dropout	0.3
Batch size	64
Optimizer	AdamW
Loss function	Cross-Entropy
Learning rate	$1e-4$
Epochs	500
Patience	50

Table C.6: Hyperparameters used for training the TEFN [35] model in the perceived comfort and safety classification task.

As illustrated in Figure A.8, the distribution of the self-reported scores is highly skewed toward positive responses. This presents a key challenge for classification due to two primary factors.

First, the class imbalance can hinder the model's ability to learn meaningful patterns for the minority classes. A model directly trained on this data is likely to converge on predicting the dominant classes, as doing so minimize the loss while effectively ignoring underrepresented cases. This imbalance reduces the model's discrimination power, particularly for detecting discomfort and feeling unsafe, as these responses are the underrepresented responses.

Secondly, the validity of certain responses may be questionable. Various participants reported feeling "very comfortable" even during scenarios driven under the aggressive driving configuration. This inconsistency raises concerns about whether these participants were genuine, potentially reflected in their low physiological arousal, or whether their answers were influenced by social desirability bias or misunderstanding. These biases distort the ground truth labels used for supervised learning, which misleads the model and degrades its performance.

C.3.1. Baseline Learning

Despite these challenges, a first training trial is conducted. The results are presented in Table C.7.

		Hard				**Soft**			
		Accuracy	**Precision**	**Recall**	**F1 score**	**Accuracy**	**Precision**	**Recall**	**F1 score**
		Q1: How safe did you feel during the car ride?							
Baseline		38.7 ± 5.4	8.2 ± 5.9	20.0 ± 0.0	11.7 ± 1.8	75.5 ± 9.4	49.1 ± 13.5	60.0 ± 0.0	52.0 ± 13.4
Ours	GSR	33.0 ± 5.4	29.6 ± 7.5	27.4 ± 5.4	24.4 ± 8.4	77.1 ± 5.6	73.4 ± 7.8	67.9 ± 7.2	66.5 ± 9.2
	VD	41.2 ± 3.3	34.3 ± 5.4	37.6 ± 5.2	33.0 ± 4.9	81.7 ± 4.5	81.9 ± 7.0	79.6 ± 7.1	78.1 ± 7.5
	GSR+VD	37.0 ± 6.3	30.7 ± 6.6	33.7 ± 7.3	28.7 ± 5.9	80.1 ± 4.6	77.6 ± 7.4	74.2 ± 7.9	72.6 ± 8.0
	GSR+P	37.1 ± 5.2	29.8 ± 9.1	30.5 ± 6.6	26.9 ± 8.0	77.8 ± 6.4	76.7 ± 13.0	71.6 ± 9.0	70.5 ± 10.7
	VD+P	41.9 ± 4.3	40.1 ± 8.0	39.3 ± 4.6	35.1 ± 5.9	79.0 ± 3.6	78.8 ± 6.2	75.0 ± 5.7	73.6 ± 7.0
	GSR+VD+P	39.6 ± 4.2	35.0 ± 8.1	33.7 ± 7.0	31.6 ± 7.1	80.8 ± 2.4	77.5 ± 8.1	75.2 ± 6.7	74.6 ± 7.2
		Q2: How safe did you feel interacting with the ...[a]							
Baseline		44.9 ± 10.5	9.1 ± 2.0	20.0 ±0.0	12.5 ± 1.8	69.9 ± 8.8	36.7 ± 7.7	60.0 ± 0.0	39.6 ± 7.7
Ours	GSR	41.5 ± 6.0	23.8 ± 4.2	25.1 ± 1.7	21.8 ± 2.5	71.6 ± 7.7	65.6 ± 11.2	55.3 ± 8.1	54.1 ± 10.2
	VD	40.3 ± 6.8	22.0 ± 5.9	26.4 ± 4.9	21.7 ± 5.0	73.8 ± 6.8	73.7 ± 10.5	61.0 ± 5.6	60.9 ± 5.5
	GSR+VD	41.8 ± 5.9	23.4 ± 3.2	25.7 ± 1.9	22.1 ± 2.6	74.1 ± 8.7	76.2 ± 11.1	63.8 ± 7.8	64.5 ± 9.9
	GSR+P	41.8 ± 6.8	24.9 ± 9.0	25.1 ± 4.6	21.3 ± 4.2	74.3 ± 4.8	66.0 ± 10.0	59.7 ± 4.9	57.9 ± 5.7
	VD+P	41.3 ± 7.8	23.8 ± 6.3	25.4 ± 3.9	22.4 ± 3.8	74.6 ± 2.9	70.6 ± 8.2	62.6 ± 5.3	61.9 ± 6.4
	GSR+VD+P	40.7 ± 7.9	22.3 ± 4.6	24.3 ± 3.7	21.5 ± 3.4	74.5 ± 5.1	72.7 ± 9.2	62.6 ± 6.4	62.4 ± 6.7
		Q3: How comfortable did you find the movement of the vehicle?							
Baseline		35.3 ± 3.4	7.3 ± 0.8	20.0 ±0.0	10.7 ± 1.0	73.0 ± 13.0	45.2 ± 13.6	60.0 ± 0.0	48.1 ± 13.2
Ours	GSR	30.8 ± 4.8	22.3 ± 3.4	28.0 ± 5.5	22.2 ± 3.7	78.1 ± 3.5	78.9 ± 6.8	75.3 ± 7.4	74.3 ± 6.8
	VD	36.8 ± 4.8	31.6 ± 4.5	30.9 ± 3.8	28.8 ± 4.5	82.5 ± 3.8	81.9 ± 6.6	81.4 ± 5.7	78.7 ± 6.4
	GSR+VD	33.2 ± 4.4	27.6 ± 4.7	28.7 ± 5.1	25.2 ± 4.6	81.0 ± 4.0	82.7 ± 3.7	79.9 ± 7.4	78.2 ± 7.0
	GSR+P	32.4 ± 2.5	24.8 ± 3.5	26.8 ± 4.3	22.9 ± 2.6	77.0 ± 6.1	79.7 ± 5.8	73.6 ± 8.4	71.7 ± 9.7
	VD+P	36.5 ± 5.8	30.6 ± 5.8	31.4 ± 5.2	28.6 ± 5.5	80.7 ± 6.0	81.5 ± 5.8	79.5 ± 7.1	75.9 ± 7.4
	GSR+VD+P	34.5 ± 4.3	28.3 ± 5.6	29.0 ± 3.9	25.8 ± 3.7	79.2 ± 5.0	79.1 ± 9.3	77.7 ± 9.0	75.1 ± 9.8

[a] [pedestrian, roadworks, pedestrian, vehicle]

Table C.7: Macro performance metrics ($M\% \pm SD\%$) over 10-fold cross validation in the self-reported score classification on self-reported score for comfort (Q1, Q3) and perceived safety (Q2) model based on various input configurations.

Across all three subjective scores for both hard and soft metrics, the models only marginally exceed or fail to exceed the baseline accuracy. This shortfall is almost certainly driven by the severe class imbalance in the self-reported scores, since the baseline classifier can achieve deceptively high accuracies by always predicting the majority class. Nevertheless, certain multimodal input configurations yield clear improvements in precision and F1 score over the baseline, suggesting that while overall accuracies remain constrained, the multimodal input improves the model's ability to distinguish under-represented classes.

This pattern holds for both hard and soft metrics, indicating that combining GSR with vehicle dynamics or perception data can sharpen the model's performance. Yet, the three-modality input configuration does not consistently outperform the dual-modality models, which may indicate that the added complexity makes the underlying patterns harder to learn or leads to overfitting. No single input configuration dominates across all three questions; augmenting GSR data with vehicle dynamics or perception data does yield better results than GSR alone, but only by a few percentages.

These results motivate the need for targeted strategies to address the class imbalance.

C.3.2. Balanced Training

To tackle the class imbalance problem, a Synthetic Minority Oversampling Technique (SMOTE) approach, developed by Chawla et al. [5], can be employed. This approach generates synthetic samples of minority classes by interpolating between existing samples, improving the model's performance on the imbalanced dataset [5]. This oversampling is applied exclusively to the training set to ensure that the model learns to distinguish minority classes, while the validation and test sets remain untouched. Results of the second training trial with the SMOTE approach are listed in Table C.8.

		Hard				**Soft**			
		Accuracy	**Precision**	**Recall**	**F1 score**	**Accuracy**	**Precision**	**Recall**	**F1 score**
		Q1: How safe did you feel during the car ride?							
Baseline		38.7 ± 5.4	8.2 ± 5.9	20.0 ± 0.0	11.7 ± 1.8	75.5 ± 9.4	49.1 ± 13.5	60.0 ± 0.0	52.0 ± 13.4
Ours	GSR	30.3 ± 4.6	24.4 ± 6.0	26.6 ± 4.8	23.9 ± 5.0	69.7 ± 6.5	61.7 ± 5.6	62.6 ± 5.2	60.0 ± 6.0
	VD	39.4 ± 6.3	35.2 ± 5.6	36.8 ± 6.4	33.4 ± 6.6	82.3 ± 5.1	74.9 ± 7.5	78.0 ± 6.7	74.6 ± 7.4
	GSR+VD	34.5 ± 4.5	33.4 ± 4.4	37.2 ± 3.5	30.4 ± 2.9	80.5 ± 5.1	72.9 ± 7.6	76.4 ± 8.1	72.5 ± 8.1
	GSR+P	35.4 ± 8.1	33.0 ± 7.5	34.3 ± 9.7	29.5 ± 7.5	74.2 ± 8.6	69.4 ± 9.3	72.6 ± 9.7	68.1 ± 10.0
	VD+P	42.2 ± 5.2	38.0 ± 5.6	38.7 ± 4.4	34.7 ± 5.2	81.0 ± 6.8	73.3 ± 8.7	76.9 ± 6.9	73.4 ± 8.7
	GSR+VD+P	41.2 ± 5.4	36.9 ± 6.9	40.4 ± 5.7	35.0 ± 6.5	79.3 ± 6.0	74.1 ± 9.0	77.5 ± 8.1	73.6 ± 9.3
		Q2: How safe did you feel interacting with the ...[a]							
Baseline		44.9 ± 10.5	9.1 ± 2.0	20.0 ±0.0	12.5 ± 1.8	69.9 ± 8.8	36.7 ± 7.7	60.0 ± 0.0	39.6 ± 7.7
Ours	GSR	35.1 ± 5.1	25.2 ± 4.4	25.2 ± 5.0	23.4 ± 4.2	67.5 ± 8.2	60.0 ± 5.9	62.5 ± 6.3	58.6 ± 6.3
	VD	35.6 ± 3.4	29.2 ± 6.3	30.3 ± 5.6	25.8 ± 4.1	77.2 ± 8.2	71.0 ± 8.9	74.6 ± 7.2	70.0 ± 9.0
	GSR+VD	37.3 ± 6.8	31.4 ± 4.2	32.7 ± 6.3	27.7 ± 4.3	75.3 ± 8.2	69.8 ± 8.9	73.4 ± 9.2	68.7 ± 9.2
	GSR+P	37.6 ± 8.4	26.5 ± 5.5	27.3 ± 6.2	24.9 ± 5.3	74.8 ± 3.1	67.4 ± 6.0	71.9 ± 5.5	67.1 ± 3.9
	VD+P	38.1 ± 7.7	26.1 ± 7.4	29.3 ± 8.7	24.0 ± 6.1	74.7 ± 10.2	73.4 ± 8.9	73.7 ± 8.9	69.7 ± 9.6
	GSR+VD+P	37.5 ± 6.3	29.3 ± 6.5	30.5 ± 5.6	26.6 ± 4.6	76.0 ± 7.8	71.9 ± 8.2	74.2 ± 8.4	70.6 ± 8.9
		Q3: How comfortable did you find the movement of the vehicle?							
Baseline		35.3 ± 3.4	7.3 ± 0.8	20.0 ±0.0	10.7 ± 1.0	73.0 ± 13.0	45.2 ± 13.6	60.0 ± 0.0	48.1 ± 13.2
Ours	GSR	28.8 ± 4.1	26.8 ± 5.8	25.5 ± 4.6	22.6 ± 3.9	73.5 ± 4.8	67.0 ± 4.0	70.1 ± 6.1	65.4 ± 5.5
	VD	33.8 ± 6.5	30.9 ± 6.5	32.6 ± 7.2	27.6 ± 6.4	80.7 ± 6.5	73.9 ± 7.6	78.9 ± 8.8	72.8 ± 9.2
	GSR+VD	33.3 ± 3.8	30.8 ± 5.9	32.6 ± 7.2	27.6 ± 4.3	81.5 ± 7.4	76.8 ± 9.1	82.3 ± 8.9	76.3 ± 10.4
	GSR+P	29.2 ± 3.4	27.6 ± 3.0	26.9 ± 4.4	23.8 ± 3.0	73.6 ± 3.4	69.5 ± 3.5	72.6 ± 7.3	66.7 ± 4.0
	VD+P	36.5 ± 4.5	34.4 ± 5.2	33.3 ± 3.9	27.7 ± 4.5	79.6 ± 7.7	75.3 ± 6.5	78.8 ± 8.1	72.7 ± 7.8
	GSR+VD+P	32.2 ± 5.1	32.4 ± 7.8	33.2 ± 7.4	27.6 ± 6.5	79.0 ± 5.7	75.5 ± 7.8	81.9 ± 5.5	73.5 ± 5.5

[a] [pedestrian, roadworks, pedestrian, vehicle]

Table C.8: Macro performance metrics ($M\% \pm SD\%$) over 10-fold cross validation in the self-reported score classification on self-reported score for comfort (Q1, Q3) and perceived safety (Q2) model based on various input configurations. Results are shown after applying SMOTE to address class imbalance.

The introduction of SMOTE leaves the overall accuracies largely unchanged, and in some cases, slightly reduced, but yields substantial gains in precision, recall and F1 score across nearly all input configurations. This confirms that synthetic oversampling effectively mitigates the class imbalance and enables the model to improve on identifying minority classes, rather than defaulting to the majority class. This holds true for both the hard and soft metrics.

Notably, the accuracies across Q2 drop under SMOTE, yet precision, recall and F1 improve a lot versus the results from Table C.7. This suggests that the non-SMOTE models were mostly learning to predict the dominant classes.

GSR-only models remain the weakest performers in both Table C.8 and C.7, whereas vehicle dynamics consistently outperforms GSR alone and shows even stronger performances when combined with the perception input. This pattern emerges both in the hard and soft metrics.

A second strategy to address class imbalance involves incorporating class weights into the loss function. Weights were calculated as follows:

$$w_c = \frac{\sum_{i=1}^{n} n_i}{C \cdot n_c} \tag{C.2}$$

With:

- w_c: the weight for class c.
- n_i: the number of samples in class i.
- C: the total number of classes (= 5).
- n_c: the number of samples in class c.

Table C.9 presents the results of the third training trial, where the *Cross-Entropy* loss function was modified to include these weights.

		Hard				**Soft**			
		Accuracy	**Precision**	**Recall**	**F1 score**	**Accuracy**	**Precision**	**Recall**	**F1 score**
		Q1: How safe did you feel during the car ride?							
Baseline		38.7 ± 5.4	8.2 ± 5.9	20.0 ± 0.0	11.7 ± 1.8	75.5 ± 9.4	49.1 ± 13.5	60.0 ± 0.0	52.0 ± 13.4
Ours	GSR	31.2 ± 5.3	27.0 ± 4.9	28.7 ± 4.6	24.3 ± 3.2	71.3 ± 5.4	64.8 ± 5.3	65.3 ± 5.9	62.7 ± 5.8
	VD	39.1 ± 4.2	37.3 ± 3.6	39.3 ± 4.6	34.3 ± 3.2	79.8 ± 7.1	74.1 ± 7.8	75.6 ± 9.6	72.4 ± 9.3
	GSR+VD	33.4 ± 6.1	29.0 ± 6.3	32.9 ± 6.9	26.9 ± 7.1	75.6 ± 6.6	71.8 ± 8.2	70.8 ± 8.5	67.5 ± 8.5
	GSR+P	35.1 ± 7.7	32.0 ± 9.1	34.7 ± 9.2	29.3 ± 8.9	73.9 ± 8.6	67.5 ± 8.2	70.9 ± 9.9	66.5 ± 9.8
	VD+P	40.9 ± 6.3	37.2 ± 5.6	39.1 ± 6.7	34.5 ± 5.9	78.2 ± 7.8	72.1 ± 8.7	75.0 ± 9.1	70.9 ± 10.2
	GSR+VD+P	38.2 ± 5.8	34.4 ± 6.1	37.8 ± 3.1	32.8 ± 5.1	78.8 ± 6.7	72.6 ± 9.0	75.7 ± 8.4	71.1 ± 9.1
		Q2: How safe did you feel interacting with the ...[a]							
Baseline		44.9 ± 10.5	9.1 ± 2.0	20.0 ±0.0	12.5 ± 1.8	69.9 ± 8.8	36.7 ± 7.7	60.0 ± 0.0	39.6 ± 7.7
Ours	GSR	33.0 ± 7.7	24.4 ± 3.4	26.7 ± 4.2	22.3 ± 3.2	67.9 ± 6.1	64.7 ± 7.2	67.8 ± 9.5	62.4 ± 8.0
	VD	32.5 ± 5.6	28.3 ± 5.1	31.3 ± 5.4	24.7 ± 3.6	74.5 ± 9.3	70.3 ± 7.4	73.5 ± 9.3	68.5 ± 9.3
	GSR+VD	37.2 ± 8.1	28.1 ± 7.9	32.3 ± 4.2	25.6 ± 5.5	70.2 ± 10.5	67.7 ± 9.8	72.0 ± 9.5	65.7 ± 11.6
	GSR+P	35.2 ± 7.8	25.9 ± 6.3	27.8 ± 5.5	24.3 ± 5.4	73.3 ± 7.2	66.3 ± 8.6	69.4 ± 6.7	64.3 ± 8.1
	VD+P	33.8 ± 8.5	29.4 ± 8.3	30.9 ± 4.1	25.5 ± 6.1	74.6 ± 8.7	69.1 ± 10.5	72.2 ± 8.5	67.1 ± 8.5
	GSR+VD+P	34.2 ± 8.3	28.7 ± 7.5	31.6 ± 5.8	25.2 ± 6.9	74.5 ± 8.1	69.0 ± 9.5	72.6 ± 8.0	66.2 ± 9.1
		Q3: How comfortable did you find the movement of the vehicle?							
Baseline		35.3 ± 3.4	7.3 ± 0.8	20.0 ±0.0	10.7 ± 1.0	73.0 ± 13.0	45.2 ± 13.6	60.0 ± 0.0	48.1 ± 13.2
Ours	GSR	30.0 ± 5.4	25.3 ± 4.9	28.0 ± 6.2	21.6 ± 3.9	74.4 ± 9.0	72.0 ± 6.9	74.1 ± 9.6	68.8 ± 10.0
	VD	34.5 ± 5.6	31.9 ± 8.2	33.9 ± 6.5	28.2 ± 5.0	80.0 ± 10.0	75.9 ± 8.8	80.5 ± 7.1	73.5 ± 11.3
	GSR+VD	33.5 ± 5.6	28.8 ± 4.4	33.5 ± 5.4	26.7 ± 4.7	80.0 ± 8.7	75.3 ± 80.0	79.8 ± 6.4	73.4 ± 10.1
	GSR+P	32.0 ± 6.6	31.6 ± 4.5	31.0 ± 7.0	27.1 ± 6.2	73.9 ± 8.8	72.6 ± 6.5	73.5 ± 8.4	68.5 ± 9.6
	VD+P	35.8 ± 6.4	34.7 ± 4.8	35.1 ± 6.3	30.4 ± 6.1	80.5 ± 8.1	75.7 ± 7.3	81.0 ± 5.6	73.5 ± 8.8
	GSR+VD+P	35.1 ± 6.0	28.8 ± 6.1	30.4 ± 4.4	26.8 ± 4.4	77.5 ± 8.1	74.0 ± 10.2	76.1 ± 9.5	69.9 ± 11.7

[a] [pedestrian, roadworks, pedestrian, vehicle]

Table C.9: Macro performance metrics ($M\% \pm SD\%$) over 10-fold cross validation in the self-reported score classification on self-reported score for comfort (Q1, Q3) and perceived safety (Q2) model based on various input configurations. Results are shown with weights added in the loss function for class imbalance.

Comparing the two approaches for mitigating class imbalance, the SMOTE-based approach (Table C.8) yields comparable performances in the hard metrics across all evaluation criteria. However, it shows slightly higher scores in the soft performance metrics. More importantly, the SMOTE-based approach results in substantially lower standard deviations in the soft metrics, indicating greater robustness and consistency.

Because the GSR contributes little on its own and sometimes fails to boost performance when fused with other signals, despite the expectation that physiological arousal should align with subjective reports grounded by the correlations found in B.2, the validity of the data should be checked.

C.3.3. Validity-Screened Training

To address the issue of potential data invalidity, a similar analytical approach is employed as introduced in Appendix B, utilizing Linear Mixed-Effect (LME) models. In this case, however, each analysis is applied on a per-participant basis to identify inconsistencies between self-reported scores and physiological responses.

Participants are flagged as invalid for further analysis based on the following two criteria:

- If there is no significant correlation between driving and their self-reported scores, yet a correlation between driving style and GSR features exists, the participant will be considered unreliable. This suggests that the participant experienced heightened physiological arousal but failed to reflect this in their questionnaire responses.
- The reverse scenario is also considered: if a participant shows a significant correlation between driving style and self-reported scores, but no corresponding correlation with GSR features, this may indicate a true lack of measurable physiological response, so-called non-responders [33], or a potential issue in data acquisition, and is thus flagged as invalid data.

Additionally, as mentioned in section A.2.2, several recorded laps were discarded due to data acquisition errors. Participants affected by this are reviewed. If a participant only has usable data from a single driving configuration (e.g., only calm or aggressive), they are excluded from further analysis. Without exposure to both driving styles, it becomes impossible to capture within-subject differences in response, which are essential for learning meaningful patterns between input features and outcomes.

Table C.10 shows which participants were excluded from further analysis and for which reasons.

Participant	Justification
14, 25, 26, 42	No significant correlation between driving style and subjective ratings but clear physiological responses to driving behavior, suggesting unreliable self-reporting.
19	Minimal GSR response across all conditions, indicating non-responsiveness [33].
23	Missing data for the aggressive driving style; prevents comparison across driving styles.
29	Missing data for the calm driving style; prevents comparison across driving styles.
34	Uniform self-reports (always "very comfortable/safe"), no within-subject variability.

Table C.10: Participants excluded from further analysis with justification based on participant-level validation of subjective and physiological responses.

Following the participant-level data validation, 23 participants remain eligible for further analysis. Of these, 15 are used for training, 4 for validation, and 4 are reserved for final training. As the class distribution remains skewed toward positive responses, the SMOTE approach is again employed to mitigate the class imbalance. Results from this fourth training trial, combining SMOTE with the participant screening procedure, are presented in Table C.11.

		Hard				**Soft**			
		Accuracy	**Precision**	**Recall**	**F1 score**	**Accuracy**	**Precision**	**Recall**	**F1 score**
		Q1: How safe did you feel during the car ride?							
Baseline		41.2 ± 2.8	8.3 ± 0.6	20.0 ± 0.0	11.7 ± 0.6	72.6 ± 4.0	32.0 ± 1.0	60.0 ± 0.0	35.0 ± 0.8
Ours	GSR	29.7 ± 2.6	26.1 ± 3.2	26.2 ± 3.3	23.2 ± 2.8	64.4 ± 3.6	56.2 ± 2.6	59.6 ± 5.2	55.1 ± 4.2
	VD	36.5 ± 2.0	31.6 ± 2.6	33.7 ± 2.3	30.2 ± 2.2	71.9 ± 2.5	64.3 ± 3.8	67.4 ± 4.6	63.0 ± 4.8
	GSR+VD	34.6 ± 3.2	30.0 ± 2.7	32.0 ± 3.1	28.5 ± 2.5	70.4 ± 2.8	63.1 ± 3.6	66.7 ± 5.3	61.8 ± 4.7
	GSR+P	33.3 ± 3.6	30.2 ± 4.0	30.6 ± 4.2	28.1 ± 3.3	68.3 ± 3.6	59.9 ± 3.6	64.6 ± 6.3	59.4 ± 5.0
	VD+P	36.6 ± 2.2	31.4 ± 2.9	33.2 ± 3.0	29.8 ± 2.3	73.3 ± 2.9	64.1 ± 4.2	68.6 ± 5.6	64.0 ± 5.4
	GSR+VD+P	32.6 ± 3.5	28.8 ± 4.1	31.6 ± 4.7	26.9 ± 3.4	71.7 ± 3.3	63.8 ± 3.4	69.0 ± 5.7	63.2 ± 5.2
		Q2: How safe did you feel interacting with the ...[a]							
Baseline		51.9 ± 3.9	10.9 ± 1.5	20.0 ± 0.0	14.3 ± 1.7	72.4 ± 4.8	34.8 ± 4.0	40.0 ± 0.0	37.6 ± 4.1
Ours	GSR	32.7 ± 2.5	22.7 ± 1.7	23.5 ± 5.9	20.7 ± 2.5	61.1 ± 5.1	54.1 ± 4.7	56.6 ± 8.9	51.4 ± 7.5
	VD	32.9 ± 4.1	24.0 ± 1.9	27.3 ± 5.7	22.4 ± 3.0	68.1 ± 3.8	58.6 ± 3.0	64.6 ± 7.1	57.9 ± 4.4
	GSR+VD	34.5 ± 6.3	24.2 ± 3.0	26.1 ± 5.5	21.7 ± 2.1	68.1 ± 3.0	58.9 ± 3.6	63.3 ± 8.1	57.3 ± 5.6
	GSR+P	34.3 ± 3.9	24.2 ± 2.3	26.5 ± 5.5	22.9 ± 2.8	67.0 ± 4.5	59.6 ± 4.5	62.4 ± 9.9	57.3 ± 6.9
	VD+P	34.2 ± 5.4	27.3 ± 2.5	28.5 ± 6.1	24.1 ± 2.5	69.2 ± 5.5	61.1 ± 5.9	66.6 ± 7.1	59.8 ± 6.5
	GSR+VD+P	35.6 ± 4.4	25.9 ± 3.9	28.8 ± 6.0	23.7 ± 3.4	69.6 ± 3.5	61.2 ± 3.6	35.1 ± 7.6	59.5 ± 4.6
		Q3: How comfortable did you find the movement of the vehicle?							
Baseline		37.7 ± 2.0	7.5 ± 0.4	20.0 ± 0.0	10.9 ± 0.4	65.4 ± 3.0	30.4 ± 0.7	40.0 ± 0.0	33.7 ± 0.5
Ours	GSR	27.6 ± 2.5	24.5 ± 2.9	23.1 ± 3.0	21.7 ± 2.3	68.5 ± 5.5	61.8 ± 5.5	62.7 ± 6.5	60.1 ± 6.5
	VD	36.4 ± 3.7	31.9 ± 4.1	33.4 ± 4.1	29.7 ± 4.6	77.3 ± 2.2	70.3 ± 3.3	78.0 ± 3.7	71.3 ± 3.2
	GSR+VD	32.4 ± 2.1	28.5 ± 2.7	29.7 ± 3.2	26.7 ± 1.9	76.5 ± 2.6	69.2 ± 3.9	75.3 ± 4.5	69.5 ± 3.8
	GSR+P	32.6 ± 3.3	28.0 ± 3.4	27.0 ± 3.1	25.5 ± 3.0	71.0 ± 3.4	63.9 ± 4.3	66.9 ± 4.4	63.5 ± 5.0
	VD+P	37.3 ± 2.3	34.4 ± 2.6	35.4 ± 3.3	31.1 ± 2.6	77.0 ± 2.4	71.4 ± 3.8	79.1 ± 3.7	71.7 ± 4.3
	GSR+VD+P	33.9 ± 2.2	31.5 ± 4.5	31.4 ± 3.5	27.0 ± 2.8	74.9 ± 3.9	69.5 ± 3.9	77.2 ± 5.0	69.6 ± 5.4

[a] [pedestrian, roadworks, pedestrian, vehicle]

Table C.11: Macro performance metrics ($M\% \pm SD\%$) over 10-fold cross validation in the self-reported score classification on self-reported score for comfort (Q1, Q3) and perceived safety (Q2) model based on various input configurations. Results are shown after applying SMOTE to address class imbalance and excluding invalid participant data.

A comparison between the results from Table C.8, after applying only SMOTE, and those obtained after both SMOTE and participant screening reveals a clear overall decline in performance. For most input configurations and performance metrics, a better score was obtained by only applying SMOTE than by applying SMOTE and the screening procedure. Notably, the accuracy of the baseline classifier has also improved in Table C.11, particularly for Q2, where it increased significantly ($\sim 15\%$), suggesting an even stronger class imbalance within the test sets. In contrast, the model's accuracy has not improved proportionally, and while it still outperforms the baseline in terms of precision, recall and F1 score, the margin is smaller than in the previous configuration with SMOTE only (C.11). Similar for Q1 and Q3, no improvement can be observed; in fact, various metrics show a decline of up to 5% for hard metrics and 10% for soft metrics compared to the SMOTE-only approach.

These findings may indicate that the model previously benefited from patterns in participant data that were excluded, potentially due to false correlations or overfitting to unreliable labels. Alternatively, the reduction in available training, validation and test data resulting from the participant screening may also have significantly compromised the model's performance. An interesting trend that emerged in Table C.11 is, however, the substantially smaller standard deviations of the results compared to those in Table C.8. This suggests that the removal of the data did contribute to a more uniform dataset.

To summarize the key insights across all four tables (C.7, C.8, C.9, C.11), several consistent findings emerge:

1. The high standard deviations across all performance metrics and input configurations in the 10-fold cross-validation indicate that performance is dependent on the specific participant split, similar to the driving style classification task. This suggests that some folds yield much better predictions than others.
2. Across Q1 (perceived comfort), Q2 (perceived safety), and Q3 (overall ride comfort), performance metrics remain largely consistent across all input and training configurations. While minor differences of a few percentages exist, no single subjective measure stands out as easier or harder to predict. Suggesting that the model performs similarly across all subjective dimensions.
3. Most model configurations are outperformed by the baseline classifier in terms of accuracy. However, this baseline simply predicts the majority class, inflating its accuracy due to the class imbalance. This underscores the importance of evaluating the models using other informative metrics as well, such as precision, recall and F1 score, that better reflect true performance, especially on minority classes.
4. The SMOTE-only training configuration emerges as the most effective, consistently outperforming the baseline training (i.e., no resampling or participant screening) in terms of precision, recall and F1 score, highlighting its strength in handling class imbalance. It also outperforms the SMOTE-with-screening approach across all metrics, suggesting that the reduced dataset size from participant exclusion may have negatively impacted performance.
5. Incorporating vehicle dynamics (VD) or perception (P) data consistently improves performance, supporting the idea that driving characteristics play a significant role in perceived comfort and safety, and that GSR alone is insufficient for robust predictions.
6. In some cases, input configurations excluding GSR, using vehicle dynamics and perception or only vehicle dynamics, outperformed those that included it. This suggests that the GSR may introduce participant-specific signals that hinder the model's ability to generalize to unseen participants. In contrast, objective inputs like the vehicle dynamics may provide more consistent patterns for predicting specific comfort levels across participants.
7. Combining GSR, vehicle dynamics, and perception does not always yield the best result, possibly due to increased model complexity or faulty perception data.
8. The use of soft metrics provides valuable insights, capturing cases where the model's predictions closely align with the participant's reported scores. This approach acknowledges the blurry boundaries in self-reported comfort and perceived safety levels and emphasizes practical relevance over strict categorical correctness.

C.3.4. User-Adapted Training

Following the preceding section, where a general model was trained, validated and tested across participants, the results, while informative, revealed limited performance consistency. Although the model successfully reduced training loss, its relatively low evaluation scores suggest poor generalization to data from unseen participants. This, however, is not entirely unexpected, given the inherently subjective nature of comfort and safety ratings and the inter-subject variability in physiological responses.

To address these limitations, this section explores a personalized modeling approach. A copy of the trained general model from the preceding section is fine-tuned separately for each participant. Fine-tuning is performed by splitting the data from one test participant into N support scenario pairs, where each pair consists of one calm and one aggressive variant of the same scenario (e.g., calm-roadwork, aggressive-roadwork) and $10 - N$ query scenario pairs. Thus, for $N = 1$, the model is fine-tuned on one such support scenario pair, and for $N = 3$ on three support pairs. The remaining $10 - N$ query pairs are used for evaluation.

In this case, fine-tuning can be considered as a form of targeted, participant-specific training. Essentially, a short, secondary training phase that adapts the model to a single participants' data. In this process, the model is updated over a limited number of epochs ($E_{\text{support}} = 20$) only using the selected support pairs for this training. The same learning rate ($LR_{\text{support}} = 1e - 4$) as used for the general model training is applied during fine-tuning.
The following fine-tuning strategy is explored: the entire body of the model is frozen, effectively disabling the learning capabilities of shared feature representation. Only the last fully connected layer remains trainable, allowing the model to adapt its output mapping to participant-specific characteristics while maintaining the general features learned during pre-training.

These design strategies aim to capture participant-specific patterns in how physiological, vehicle dynamics and perception data relate to perceived comfort and safety, with the goal to improve prediction accuracy on unseen conditions for that participant.

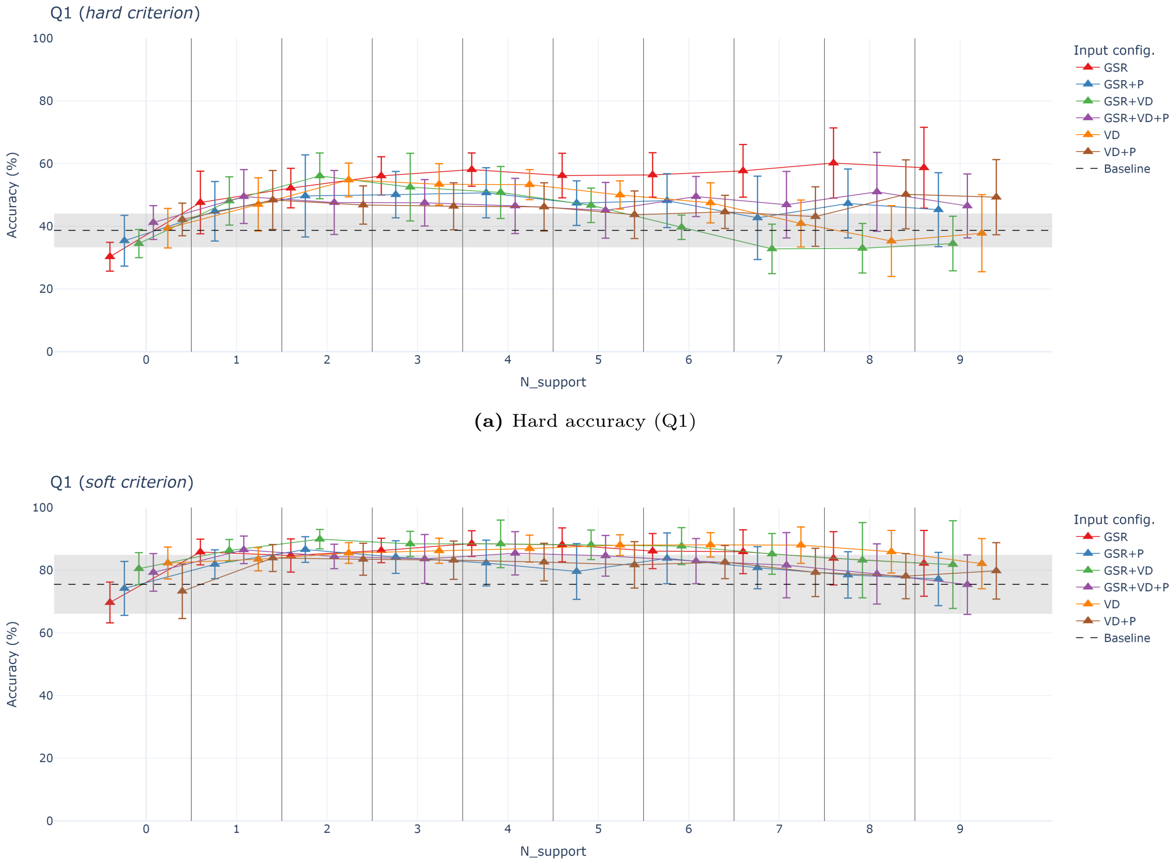
Fine-tuning is performed using the general models previously trained with the SMOTE-only approach. For each of the 10 cross-validation folds, the same five participants held out for testing before, are now each assigned a personal model. This model is fine-tuned individually using varying numbers of support scenario pairs ($N = 0$ to $N = 9$). $N = 0$ corresponds to the unadapted general model, and $N = 9$ is the maximum allowed support size, as at least one query pair is required for evaluation. For Q2, the maximum support size is limited to $N = 8$, as this question was not asked during laps 1 and 4 of the pedestrian crossing with obstruction scenario of the experiment, resulting in fewer available scenario pairs.
Each personalized model is evaluated using both the hard (exact match) and soft (one-off match) criteria. Within each fold, performance metrics are averaged across the five test participants to obtain a fold-level score. Table C.12 presents the hard and soft accuracies across all input configurations for $N = 0$ (general model) to $N = 9$. The results presented are obtained using the fine-tuning strategy in which only the model's final projection layer is updated, while all other parameters remain frozen.

Support Pairs	Input config.	Hard accuracy			Soft accuracy		
		Q1	*Q2*^{a}	*Q3*	*Q1*	*Q2*^{a}	*Q3*
	Baseline	38.7 ± 5.4	44.9 ± 10.5	35.3 ± 3.4	75.5 ± 9.4	69.9 ± 8.8	73.0 ± 13.0
$N = 0$	GSR	30.3 ± 4.6	35.1 ± 5.1	28.8 ± 4.1	69.7 ± 6.5	67.5 ± 8.2	73.5 ± 4.8
	VD	39.4 ± 6.3	35.6 ± 3.4	33.8 ± 6.5	82.3 ± 5.1	77.2 ± 8.2	80.7 ± 6.5
	GSR+VD	34.5 ± 4.5	37.3 ± 6.8	33.3 ± 3.8	80.5 ± 5.1	75.3 ± 8.2	81.5 ± 7.4
	GSR+P	35.4 ± 8.1	37.6 ± 8.4	29.2 ± 3.4	74.2 ± 8.6	74.8 ± 3.1	73.0 ± 3.4
	VD+P	42.2 ± 5.2	38.1 ± 7.7	36.5 ± 4.5	73.3 ± 8.7	74.7 ± 10.2	79.6 ± 7.7
	GSR+VD+P	41.2 ± 5.4	37.5 ± 6.3	32.2 ± 5.1	79.3 ± 6.0	76.0 ± 7.8	79.0 ± 5.7
$N = 1$	GSR	47.6 ± 10.0	55.2 ± 7.7	43.3 ± 8.9	85.8 ± 4.1	83.7 ± 5.1	84.1 ± 9.4
	VD	47.0 ± 8.5	57.5 ± 10.8	40.9 ± 10.0	83.5 ± 3.7	82.1 ± 8.5	79.4 ± 8.3
	GSR+VD	48.1 ± 7.7	57.8 ± 8.1	41.7 ± 8.3	86.3 ± 3.5	85.5 ± 6.4	81.5 ± 8.9
	GSR+P	44.8 ± 9.5	46.4 ± 9.6	44.5 ± 8.4	81.9 ± 4.6	77.1 ± 8.4	80.9 ± 8.7
	VD+P	48.4 ± 9.4	50.5 ± 8.6	43.4 ± 9.9	83.9 ± 4.3	83.4 ± 7.0	80.8 ± 7.8
	GSR+VD+P	49.5 ± 8.6	50.5 ± 8.6	43.8 ± 9.2	86.5 ± 4.4	82.3 ± 6.7	80.3 ± 8.9
$N = 2$	GSR	52.2 ± 6.3	56.5 ± 7.0	45.7 ± 5.0	84.7 ± 5.3	84.7 ± 6.1	82.6 ± 3.9
	VD	54.8 ± 5.4	56.3 ± 6.8	47.9 ± 9.2	85.5 ± 3.3	84.2 ± 6.2	85.3 ± 4.5
	GSR+VD	56.1 ± 7.3	59.3 ± 9.2	48.0 ± 8.7	89.9 ± 3.1	89.2 ± 4.1	87.8 ± 4.0
	GSR+P	49.7 ± 13.1	57.2 ± 8.4	46.0 ± 8.4	86.6 ± 4.1	81.9 ± 9.7	81.8 ± 8.7
	VD+P	46.8 ± 6.1	48.7 ± 12.1	45.6 ± 8.2	83.5 ± 5.1	85.7 ± 5.7	84.7 ± 8.9
	GSR+VD+P	47.6 ± 10.2	47.8 ± 11.6	44.6 ± 8.3	84.4 ± 3.9	86.4 ± 4.5	83.7 ± 7.8
$N = 3$	GSR	56.1 ± 6.1	60.9 ± 7.9	52.2 ± 6.2	86.3 ± 3.9	86.1 ± 5.2	86.1 ± 7.0
	VD	53.4 ± 6.6	57.0 ± 8.1	50.0 ± 5.8	86.2 ± 4.0	90.0 ± 3.9	86.0 ± 6.6
	GSR+VD	52.5 ± 10.8	58.5 ± 7.6	50.8 ± 5.8	88.4 ± 4.0	90.1 ± 4.2	87.0 ± 4.1
	GSR+P	50.1 ± 7.4	53.4 ± 8.9	47.6 ± 6.1	84.2 ± 5.2	82.3 ± 10.0	84.4 ± 5.5
	VD+P	46.4 ± 7.5	48.7 ± 11.2	41.9 ± 7.7	83.2 ± 6.1	84.0 ± 8.4	82.3 ± 8.8
	GSR+VD+P	47.5 ± 7.4	48.6 ± 12.1	42.2 ± 7.5	83.6 ± 7.8	86.1 ± 6.9	84.3 ± 8.0
$N = 4$	**GSR**	**58.1 ± 5.3**	**58.4 ± 8.9**	**54.3 ± 8.4**	**88.5 ± 4.1**	**86.5 ± 4.7**	**90.1 ± 6.1**
	VD	53.3 ± 4.8	56.8 ± 12.8	49.2 ± 5.8	86.9 ± 4.3	88.9 ± 5.4	87.2 ± 4.8
	GSR+VD	50.8 ± 8.3	57.5 ± 10.8	48.3 ± 7.6	88.4 ± 7.6	92.3 ± 4.6	85.6 ± 4.6
	GSR+P	50.7 ± 8.0	51.6 ± 8.8	50.2 ± 8.3	82.3 ± 7.3	85.5 ± 7.5	85.8 ± 7.8
	VD+P	46.2 ± 7.7	48.5 ± 9.3	46.0 ± 8.2	82.6 ± 6.0	85.8 ± 6.6	80.4 ± 6.8
	GSR+VD+P	46.5 ± 8.8	49.7 ± 8.2	47.3 ± 8.3	85.4 ± 6.9	84.1 ± 6.1	84.2 ± 6.3
$N = 5$	GSR	56.2 ± 7.1	58.7 ± 8.4	54.2 ± 8.3	88.1 ± 5.4	87.9 ± 6.1	89.8 ± 6.2
	VD	50.0 ± 4.5	53.9 ± 13.1	45.2 ± 7.6	88.0 ± 3.3	86.3 ± 7.4	86.3 ± 4.2
	GSR+VD	46.7 ± 5.5	56.1 ± 14.0	46.5 ± 4.8	88.1 ± 4.7	87.0 ± 6.6	87.6 ± 3.6
	GSR+P	47.4 ± 7.1	53.5 ± 10.7	50.4 ± 8.3	79.6 ± 8.9	87.4 ± 5.2	86.8 ± 7.1
	VD+P	43.7 ± 7.6	46.8 ± 11.2	45.0 ± 6.8	81.7 ± 7.4	80.3 ± 11.2	81.3 ± 11.9
	GSR+VD+P	45.1 ± 8.9	42.5 ± 7.2	46.8 ± 7.7	84.6 ± 6.5	81.4 ± 9.8	87.9 ± 6.6
$N = 6$	GSR	56.4 ± 7.1	61.5 ± 11.0	63.5 ± 6.3	86.1 ± 5.6	86.0 ± 8.1	94.2 ± 4.0
	VD	47.5 ± 6.4	53.0 ± 11.0	44.2 ± 6.9	88.1 ± 3.9	86.3 ± 6.1	84.2 ± 3.9
	GSR+VD	39.7 ± 3.9	55.8 ± 12.8	47.9 ± 4.7	87.7 ± 5.9	83.2 ± 7.9	87.8 ± 4.3
	GSR+P	48.2 ± 8.6	57.7 ± 10.2	48.3 ± 8.1	83.8 ± 8.1	86.9 ± 4.8	86.6 ± 6.1
	VD+P	44.6 ± 5.3	52.4 ± 11.9	40.6 ± 5.7	82.6 ± 5.3	81.3 ± 8.6	79.4 ± 8.9
	GSR+VD+P	49.5 ± 6.4	47.9 ± 8.4	44.7 ± 6.1	82.9 ± 7.2	83.4 ± 11.4	85.9 ± 7.7
$N = 7$	GSR	57.7 ± 8.4	61.6 ± 10.3	57.5 ± 6.0	85.9 ± 7.0	84.9 ± 11.1	93.5 ± 4.3
	VD	40.9 ± 7.5	48.8 ± 12.0	46.1 ± 6.6	88.0 ± 5.8	82.9 ± 11.6	80.9 ± 4.8
	GSR+VD	32.8 ± 7.9	52.7 ± 10.2	46.6 ± 7.5	85.2 ± 6.5	87.5 ± 7.5	82.6 ± 5.7
	GSR+P	42.7 ± 13.3	59.6 ± 8.8	46.1 ± 12.7	80.8 ± 6.7	86.5 ± 7.3	86.1 ± 7.2
	VD+P	43.1 ± 9.5	52.3 ± 14.5	39.4 ± 9.0	79.3 ± 7.7	80.7 ± 7.4	79.9 ± 12.8
	GSR+VD+P	46.9 ± 10.6	49.5 ± 10.9	35.8 ± 9.2	81.6 ± 10.4	80.0 ± 15.3	83.7 ± 9.7
$N = 8$	GSR	60.2 ± 11.2	65.2 ± 8.8	58.9 ± 9.6	83.8 ± 8.5	89.0 ± 6.4	92.2 ± 5.2
	VD	35.3 ± 11.3	43.3 ± 14.9	38.6 ± 5.8	85.9 ± 6.8	80.0 ± 10.1	78.2 ± 9.1
	GSR+VD	33.0 ± 7.9	49.3 ± 16.8	46.3 ± 10.9	83.2 ± 12.0	77.3 ± 11.2	85.2 ± 6.4
	GSR+P	47.3 ± 11.0	58.1 ± 11.1	48.5 ± 11.4	78.5 ± 7.4	82.8 ± 11.0	85.9 ± 8.8
	VD+P	50.2 ± 11.0	42.8 ± 11.8	41.6 ± 9.2	78.1 ± 7.2	83.8 ± 10.3	83.8 ± 10.2
	GSR+VD+P	51.0 ± 12.6	49.0 ± 14.4	37.3 ± 9.6	78.8 ± 9.6	85.5 ± 9.8	80.9 ± 6.8
$N = 9$	GSR	58.7 ± 12.9	-	57.9 ± 9.2	82.2 ± 10.5	-	92.3 ± 5.8
	VD	37.8 ± 12.3	-	40.6 ± 6.6	82.1 ± 8.0	-	77.8 ± 9.3
	GSR+VD	34.5 ± 8.7	-	47.2 ± 9.1	81.8 ± 14.0	-	83.9 ± 5.6
	GSR+P	45.3 ± 11.8	-	47.8 ± 8.9	77.2 ± 8.5	-	84.8 ± 9.2
	VD+P	49.3 ± 12.0	-	40.7 ± 8.7	79.8 ± 9.0	-	81.9 ± 8.9
	GSR+VD+P	46.5 ± 10.2	-	36.8 ± 10.2	75.4 ± 9.5	-	78.9 ± 9.3

[a]: Support size to this question is limited to $N = 8$ as this question was not asked during laps 1, 4.

Table C.12: Hard and soft accuracy metrics ($M\% \pm SD\%$) over 10-fold cross validation for self-reported comfort (Q1, Q3) and perceived safety (Q2) classification using various input configurations and support set sizes (N). Results are shown for different support set sizes ($N = 0$ to $N = 9$) used in fine-tuning the model. SMOTE was applied to address class imbalance. Bold values indicate the best-performing configuration for $N \leq 5$.

Figure C.4(a) displays the hard accuracy performance for Q1 across all 10 folds, for each input signal configuration and support set size. A triangle marks the mean accuracy, and the vertical lines show the standard deviation across folds. This setup allows for assessment of how the number of personalized samples ($N_{support}$) and input signal configuration affect model performance. Results for soft accuracies for Q1 are shown in figure C.4(b), and hard and soft accuracies for Q2 and Q3 are shown in Figure C.5(a), C.5(b), C.6(a) and C.6(b), respectively.



(a) Hard accuracy (Q1)

(b) Soft accuracy (Q1)

Figure C.4: Mean and standard deviation of the accuracies for Q1 for $N = 0$ to $N = 9$ support pairs across all input configurations, evaluated over the 10-fold cross validation. The shaded region and dashed line represent the baseline classifier's mean and standard deviation accuracies.

Figure C.4(a) demonstrates a clear improvement in hard accuracy with participant-specific fine-tuning for all input configurations. Across all configurations, accuracy rises sharply as the number of support pairs (N) rises from 0 to roughly 4-5, confirming the practicality of user-adapted training. Beyond this point, however, the trajectories diverge.

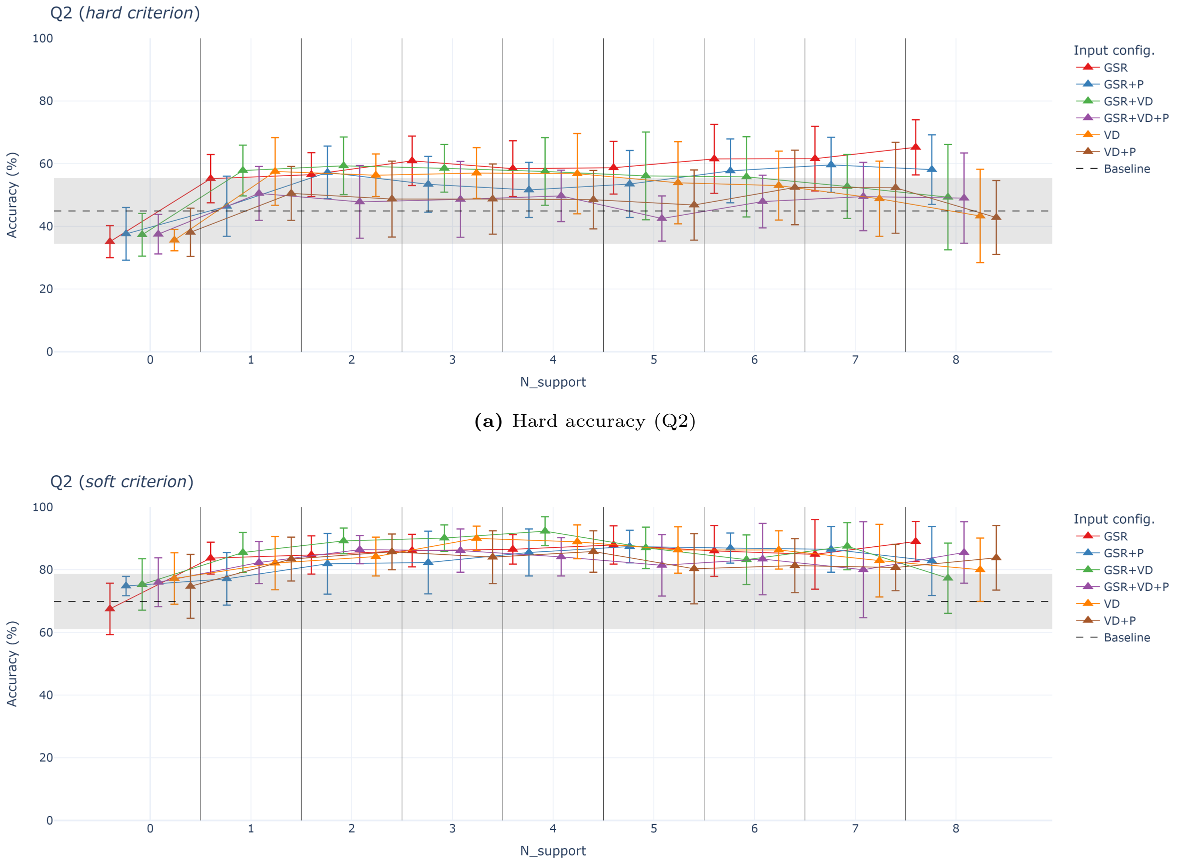
- GSR-only continues to improve and remains marginally above the majority-class baseline.
- Configurations that incorporate vehicle-dynamics or perception features (VD, P) peak around $N = 4, 5$ and then decline, often dropping below the majority-class baseline for $N > 6$.

The most plausible explanation is over-fitting: scenario-specific VD and P signals may lead the model to rely too heavily on scenario-specific patterns present in the fine-tuning data, which do not generalize well to new scenarios. In contrast, the GSR signal appears more participant-specific and less dependent on the scenario itself, allowing the model to generalize more effectively when applied to unseen scenarios. Standard deviations increase substantially for larger N across all configurations, reflecting the small residual test set per participant; with few samples, performance can swing from exceptionally high to very poor, inflating the variance.

Figure C.4(b) shows a similar pattern to the hard accuracy results. Across all input configurations, accuracy improves as N increases between $N = 0$ and $N = 3, 4$, after which most configurations plateau or show a slight decline.

Again, the GSR-only configuration remains relatively stable across all values of N and avoids a performance decline as seen in the multimodal configurations. Similarly, increased standard deviations are observed for high values of N.

Considering both hard and soft accuracy metrics, the GSR-only configuration at $N = 4$ presents a practical optimum for this task. At this support size, the model achieves strong performance with a relatively low standard deviation across cross-validation folds, while still maintaining a sufficiently large query set for reliable evaluation. In summary, while combining input modalities was beneficial in the general models (Table C.8), Figure C.4(a) indicates that the GSR, when calibrated to the individual, offers more gains.



(a) Hard accuracy (Q2)

(b) Soft accuracy (Q2)

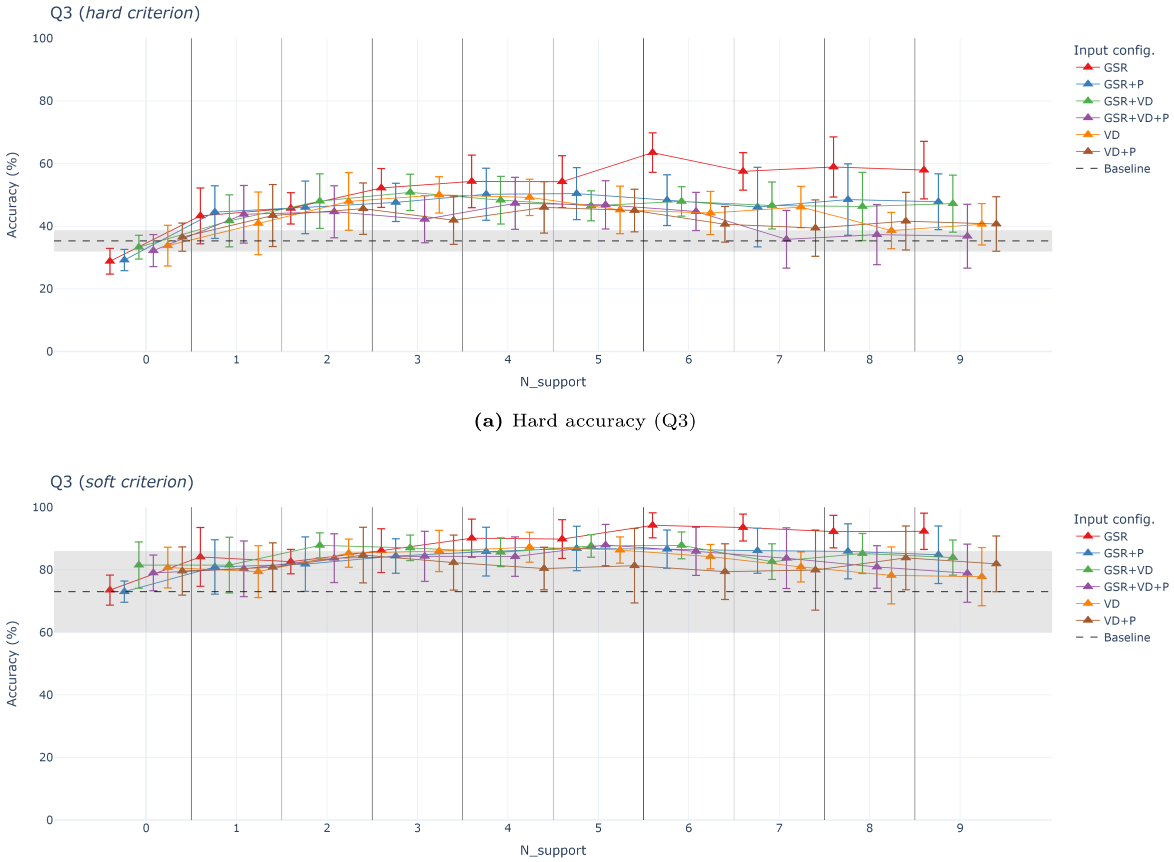
Figure C.5: Mean and standard deviation of the accuracies for Q2 for $N = 0$ to $N = 8$ support pairs across all input configurations, evaluated over the 10-fold cross validation. The shaded region and dashed line represent the baseline classifier's mean and standard deviation accuracies.

Figures C.5(a) and C.5(b) mirror the trends observed for Q1. A single participant-specific support pair ($N = 1$) yields the largest accuracy gain for every configuration, confirming how perceived safety is highly individual. Performance then climbs until roughly $N = 3, 4$, after which performance flattens or declines for most configurations.

Under the soft criterion, configurations that include vehicle dynamics (GSR+VD, VD) briefly surpass GSR-only at $N = 3, 4$ but lose ground as N increases further, suggesting overfitting. GSR-only either continues to improve or remains stable across the entire support range and never drops below the majority baseline. This robustness indicates again that the GSR is less entangled with scenario-specific patterns and therefore generalizes better to unseen data.

The expanding vertical bars at larger N are a direct consequence of the shrinking query set, which limits the reliability of the performance estimates.

Balancing accuracy, variance and evaluation set size, $N = 4$ emerges as the most defensible support size for perceived safety models. Although vehicle dynamics-related configurations outperform the GSR-only in soft accuracy, the GSR-only configuration remains the favorable configuration as it still outperforms other configurations in hard metrics its greater stability.



(a) Hard accuracy (Q3)

(b) Soft accuracy (Q3)

Figure C.6: Mean and standard deviation of the accuracies for Q3 for $N = 0$ to $N = 9$ support pairs across all input configurations, evaluated over the 10-fold cross validation. The shaded region and dashed line represent the baseline classifier's mean and standard deviation accuracies.

Figures C.6(a) and C.6(b) show that predicting overall ride comfort follows the same general patterns observed for Q1 and Q2. Accuracy improves consistently from $N = 0$ to $N = 3, 4$ across all configurations, after which most plateau or decline. The GSR-only configuration remains the only configuration that remains stable beyond $N = 4$, while other configurations tend to drop toward or below the majority-class baseline. This pattern again suggests overfitting driven by scenario-specific features.

As in previous results, the vertical variance bars increase with higher N, reflecting greater variance across cross-validation folds due to the reduced size of the query set.

Taken together, the results in Figure C.6 further support earlier conclusions: A support size of $N = 4$ and the GSR-only configuration remain the most robust and defensible choice for predicting subjective comfort ratings.

C.3.5. Final Model Evaluation

The final model for comfort and safety prediction adopts the following optimal configuration over all three comfort metrics (Q1, Q2, Q3):

- **Input channels**: GSR
- **Class imbalance handling**: SMOTE-only oversampling
- **Fine-tuning**: Support pair set size $N = 4$

With this setup, hard accuracies of 58.1%, 58.4% and 54.3% and soft accuracies of 88.5%, 86.5% and 90.1% for perceived comfort, safety and overall ride comfort, respectively, are achieved.

The following paragraphs evaluate the final model more in-depth across perceived comfort, safety and overall ride comfort.

Perceived Comfort
Table C.13 reports the mean, maximum, minimum and standard deviations of the performance metrics across the 10-fold cross-validation. Each personalized model is evaluated on the test participant it was fine-tuned on; each performance metric of a fold is the average performance over all test participants in that fold.

	Hard				**Soft**			
Metric	**Mean (%)**	**Max. (%)**	**Min. (%)**	**Std. (%)**	**Mean (%)**	**Max. (%)**	**Min. (%)**	**Std. (%)**
Accuracy	58.1	68.8	50.0	5.3	88.5	83.3	96.7	4.1
Precision	34.2	49.8	26.0	6.8	77.1	92.0	64.7	7.2
Recall	42.3	52.9	36.2	5.0	77.3	90.8	68.1	5.8
F1 score	36.0	49.9	28.2	6.2	76.1	91.0	65.7	6.6

Table C.13: Performance metrics over 10-fold cross-validation in the perceived comfort (Q1) classification model based on GSR data (phasic and tonic components) with a support set size $N = 4$.

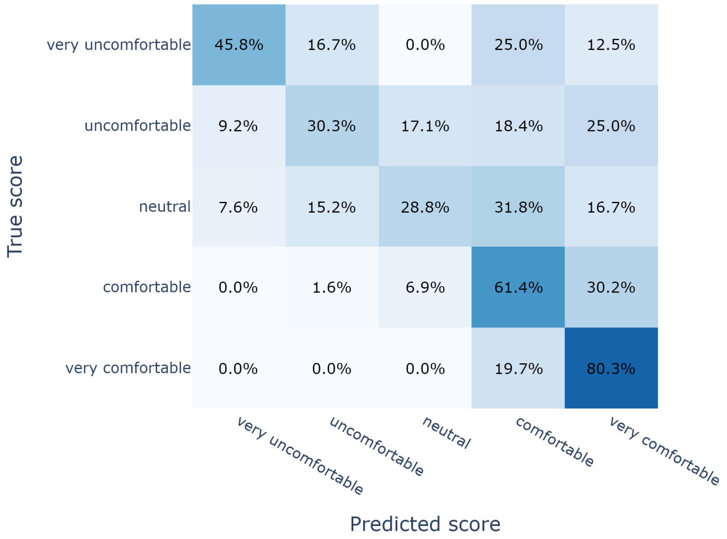
Figures C.7 and C.8 present the aggregated confusion matrix and ROC analysis for perceived comfort (Q1), respectively. The ROC analysis includes class-wise curves (left) and micro- and macro-averaged curves (right), providing a comprehensive view of model performance across and within classes.
The micro-average combines all predictions across classes and reflects the model's global ability to distinguish correct from incorrect predictions, while the macro-average calculates the average performance per class, treating each class equally regardless of its frequency. A close alignment between micro- and macro-average indicates a uniform model performance across classes, while a substantially higher micro-average may signal a bias towards majority classes.

In addition to this, Cohen's κ is reported to quantify the agreement between predicted and true labels, while accounting for the agreement expected by chance. It is defined as:

$$\kappa = \frac{p_0 - p_e}{1 - p_e} \tag{C.3}$$

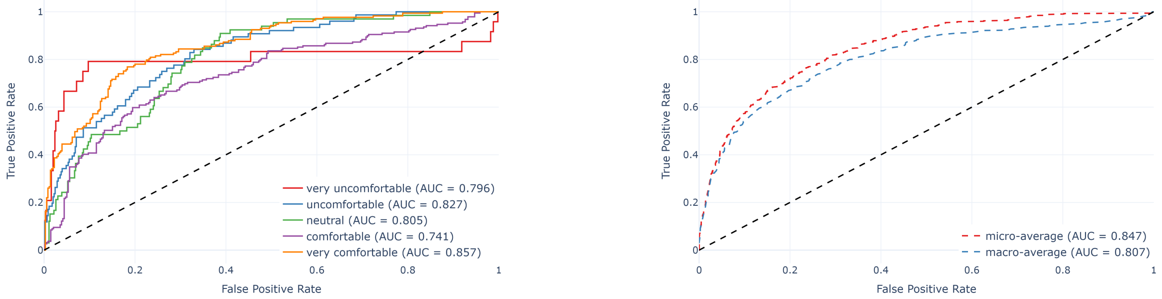
Where p_0 denotes the observed agreement (i.e., raw accuracy) and p_e the expected agreement, which is the probability that the prediction and truth match by chance, given the distribution of labels. Given the strong imbalance in the dataset, accuracy can be misleading, as it may largely reflect correct majority-class predictions rather than meaningful learning. Cohen's κ corrects for this by discounting agreement that could occur purely by chance.
The value of κ ranges from -1, indicating complete disagreement where predictions consistently mismatch the true labels, to 0, which reflects chance-level agreement where the model's accuracy can be attributed entirely to random chance rather than true predictive ability, up to 1, representing perfect agreement where predictions exactly match the true labels across all classes.
According to interpretation guidelines, proposed by Landis and Koch [16], κ values between 0.01 to 0.20 indicate slight agreement, 0.21 to 0.40 fair agreement, 0.41 to 0.60 moderate agreement, 0.61 to 0.80 substantial agreement and 0.81 to 1.00 almost perfect agreement. Higher values reflect increasingly consistent predictions that go beyond chance and thus strong evidence of effective learning. Including Cohen's κ in evaluations ensures performance gains reflect true robust learning rather than exploitation of the data imbalance.

For perceived comfort, Cohen's κ reached 0.409 under hard metric evaluation and 0.643 under soft metric evaluation under this final model configuration. In contrast, the general model with GSR-only input achieved substantially lower values of 0.077 and 0.312, respectively.



True score \ Predicted score	very uncomfortable	uncomfortable	neutral	comfortable	very comfortable
very uncomfortable	45.8%	16.7%	0.0%	25.0%	12.5%
uncomfortable	9.2%	30.3%	17.1%	18.4%	25.0%
neutral	7.6%	15.2%	28.8%	31.8%	16.7%
comfortable	0.0%	1.6%	6.9%	61.4%	30.2%
very comfortable	0.0%	0.0%	0.0%	19.7%	80.3%

Figure C.7: Aggregated confusion matrix over 10-fold cross-validation for the perceived comfort (Q1) classification using GSR data (phasic and tonic components) with a support set size $N = 4$.



(a) Class-wise

(b) Micro-, macro-averaged

Figure C.8: Aggregated Receiver Operating Characteristic (ROC) curve over 10-fold cross-validation for the perceived comfort (Q1) classification using GSR data (phasic and tonic components) with a support set size $N = 4$. Both class-wise ROC curves (left) and micro-, macro-averaged ROC curves (right) have corresponding Area Under Curve (AUC) values indicated in the legend.

Perceived Safety

Table C.14 summarizes the 10-fold cross-validation results for the perceived safety (Q2) model.

	Hard				**Soft**			
Metric	**Mean (%)**	**Max. (%)**	**Min. (%)**	**Std. (%)**	**Mean (%)**	**Max. (%)**	**Min. (%)**	**Std. (%)**
Accuracy	58.4	75.0	46.7	8.9	86.5	93.3	77.7	4.7
Precision	33.4	56.7	21.3	9.8	75.1	85.6	64.9	6.6
Recall	44.7	67.1	34.2	9.8	78.4	89.3	67.9	6.7
F1 score	35.8	60.0	22.3	10.7	75.5	86.5	64.7	6.7

Table C.14: Performance metrics over 10-fold cross-validation in the perceived safety (Q2) classification model based on GSR data (phasic and tonic components) with a support set size $N = 4$.

Figures C.9 and C.10 illustrate the aggregated confusion matrix and ROC curves for the perceived safety (Q2) scores, respectively.

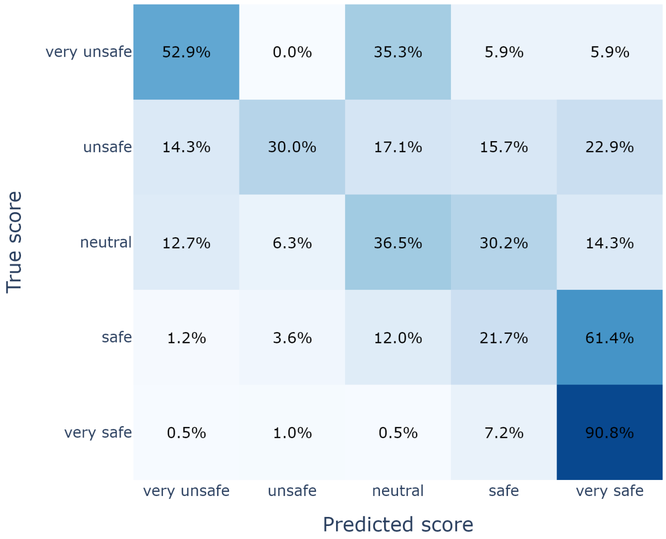
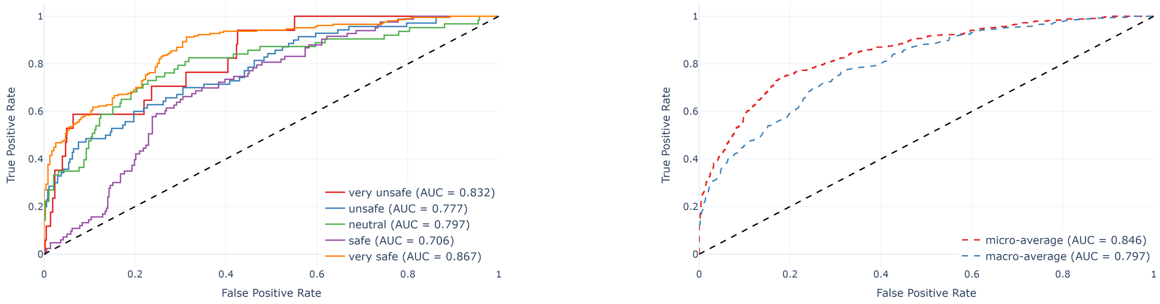


Figure C.9: Aggregated confusion matrix over 10-fold cross-validation for the perceived safety (Q2) classification using GSR data (phasic and tonic components) with a support set size $N = 4$.



(a) Class-wise

(b) Micro-, macro-averaged

Figure C.10: Aggregated Receiver Operating Characteristic (ROC) curve over 10-fold cross-validation for the perceived safety (Q2) classification using GSR data (phasic and tonic components) with a support set size $N = 4$. Both class-wise ROC curves (left) and micro-, macro-averaged ROC curves (right) have corresponding Area Under Curve (AUC) values indicated in the legend.

In perceived safety, the final model yielded a Cohen's κ of 0.376 under hard metrics and 0.652 under soft metrics, markedly surpassing the general model's scores of 0.048 and 0.178.

Overall Ride Comfort

Table C.15 presents the cross-validated performance metrics for the overall ride comfort (Q3) model.

	Hard				**Soft**			
Metric	**Mean (%)**	**Max. (%)**	**Min. (%)**	**Std. (%)**	**Mean (%)**	**Max. (%)**	**Min. (%)**	**Std. (%)**
Accuracy	54.3	65.0	36.1	8.3	90.1	98.3	77.8	6.1
Precision	30.6	38.9	23.7	4.6	84.0	98.9	56.0	11.6
Recall	41.1	54.3	26.0	7.1	83.5	97.7	61.7	10.1
F1 score	33.3	44.0	18.9	6.5	82.5	98.6	57.9	10.9

Table C.15: Performance metrics over 10-fold cross-validation in the overall ride comfort (Q3) classification model based on GSR data (phasic and tonic components) with a support set size $N = 4$.

Figures C.11 and C.12 show the aggregated confusion matrix and ROC curves for the overall ride comfort (Q3) classification task.

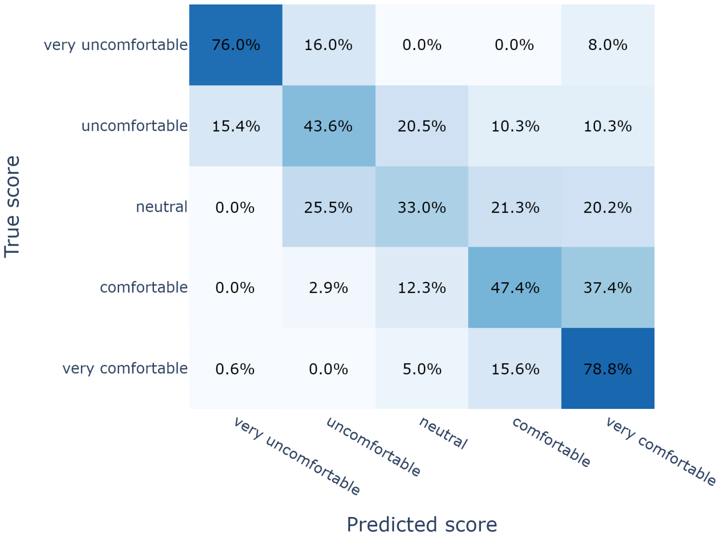
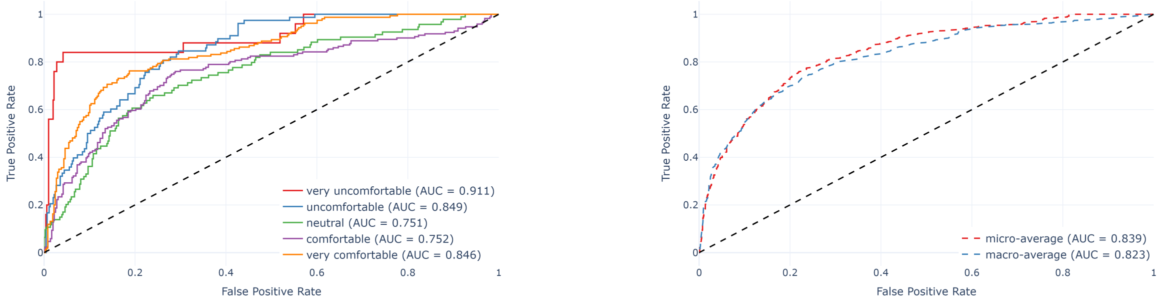


Figure C.11: Aggregated confusion matrix over 10-fold cross-validation for the overall ride comfort (Q3) classification using GSR data (phasic and tonic components) with a support set size $N = 4$.



(a) Class-wise

(b) Micro-, macro-averaged

Figure C.12: Aggregated Receiver Operating Characteristic (ROC) curve over 10-fold cross-validation for the overall ride comfort (Q3) classification using GSR data (phasic and tonic components) with a support set size $N = 4$. Both class-wise ROC curves (left) and micro-, macro-averaged ROC curves (right) have corresponding Area Under Curve (AUC) values indicated in the legend.

For overall ride comfort predictions, the final model attained a Cohen's κ of 0.397 and 0.748 under hard and soft evaluation metrics, respectively. In comparison, the general model reached considerably lower values of 0.126 and 0.371

C.4. Discussion

The focus of this chapter was on predicting passengers' perceived comfort and safety in automated driving. Four subquestions were derived for this purpose and will be answered here:

1. **How accurately can the GSR alone predict objective driving style (calm vs. aggressive)?**

   The GSR demonstrates strong performance in the prediction of objective driving style with an accuracy of 88.61% (± 3.77%) over a 10-fold cross validation, with 87.73% (± 4.75%) precision, 97.70% (± 8.86%) recall, 88.61% (± 3.98%) F1 score and an Area Under the Curve of the ROC of 0.938 (Table C.2, Figure C.3). These results show a strong discriminative capacity of GSR in distinguishing between calm and aggressive driving styles.

The predictive power of the GSR is maximized when it is decomposed and both phasic and tonic components are used as input (Tables C.3, C.4, C.5).

2. **How accurately can the GSR alone predict perceived comfort, perceived safety and overall ride comfort?**

   The GSR demonstrates poor performance in the prediction of subjective comfort ratings with 33.0% (± 5.4%), 41.5% (± 6.0%), 30.8% (± 4.8%) accuracies for Q1, Q2 and Q3, respectively (Table C.7). These results fall short of the majority-class baseline accuracies of 38.7% (± 5.4%), 44.9% (± 10.5%), 35.3% (± 3.4%), indicating that the model fails to provide meaningful predictive value.

3. **To what extent do vehicle dynamics and/or perception data improve GSR-based models for subjective comfort metrics?**

   Multimodal fusion with vehicle dynamics and perception improved performance by only 4-8% in accuracy across all three subjective comfort measures. For Q1 and Q3, models based on objective driving features (vehicle dynamics and perception) achieved the best performance, whereas for Q2, the fusion of GSR with one of the objective driving features achieved the best performance. However, these models only marginally outperformed the majority-class baseline for some input configurations, while in others they failed to surpass it altogether.

   A three-way fusion (GSR, vehicle dynamics and perception) did not show consistent benefits, likely due to the increased model complexity or missing perception entries.

4. **What challenges arise when predicting subjective ratings, and how can these challenges be addressed to improve model performance?**

   During this study, the following challenges arose:

   1. **Class imbalance**: A strongly skewed distribution of questionnaire responses toward positive responses (Figure A.8), mirroring patterns seen in other comfort-related studies [14], [32]. This imbalance biases the model toward "(very) comfortable/safe" responses and impairs its ability to learn patterns associated with "(very) uncomfortable/unsafe".
   2. **Questionable response validity**: Various participants responded feeling "very comfortable/safe" for scenarios driven under the aggressive driving style, raising concerns whether this response was genuine, influenced by social desirability bias, or misunderstanding. Invalid or noisy subjective labels assigned to each time series can mislead the model during training and significantly degrade its predictive accuracy.
   3. **Ambiguous score thresholds**: Without clear reference points, participants may have struggled to discriminate between adjacent comfort levels, leading to inconsistent or coarse ratings that mask fine-grained variations in perceived comfort. This label ambiguity introduces more label noise that weakens the model's ability to learn and degrades performance.
   4. **Inter-participant variability**: Both GSR and subjective comfort scores vary substantially across participants, challenging the model's generalization performances.

   To address each challenge, the following approaches were employed:

   1. A Synthetic Minority Oversampling Technique (SMOTE) approach was applied to the training set to address the class imbalance. Although this did not substantially increase overall accuracy, it improved precision, recall and F1 score, surpassing the majority-class baseline on these metrics and enhancing the model's ability to predict minority classes (Table C.8).
   2. Excluding participants from the data addressed the questionable response validity (Table C.10). However, this led to a decline in model performance, indicating that the loss of training data outweighed the benefits of removing potentially unreliable labels (Table C.11. Consequently, this approach is not further pursued.
   3. To address ambiguous score thresholds, a *near-fit* criterion (soft) was introduced: predictions within one class of the ground-truth were counted as true positive. This soft evaluation metric

accounts for participants' difficulty in making fine-grained comfort judgments and aligns better with the practical goal of broadly distinguishing between comfort and discomfort, rather than capturing subtle variations.

4. The final challenge, high inter-participant variability, was addressed through user-adapted fine-tuning. For each participant, the final projection layer of the model was fine-tuned using N support pairs (i.e., the same scenario under the calm and aggressive driving style) and evaluated on the remaining $10 - N$ (or $8 - N$) query pairs. Using $N = 3$ or 4 yielded the best trade-off between accuracy and stability across cross-validation folds, while preserving sufficient data for test evaluation. This approach improved hard accuracies by $15 - 20\%$ and soft accuracies by $10 - 20\%$ across input configurations (Table C.12, Figures C.4-C.6).

A stratified 10-fold cross-validation was adopted during this study to test for robustness in place of a leave-one-out cross-validation (LOOCV). This design choice was originally chosen to mitigate the risk that a single participant with atypical GSR or comfort dynamics would disproportionately influence the evaluation metrics, thereby providing a more stable estimate of generalization performance. Nevertheless, future work could consider LOOCV to examine its potential impact on the reported results.

A final model configuration is presented, which applies SMOTE to the training set and user-adapted fine-tuning with $N = 4$ as the optimal configuration to address the challenges that arose with this subjective comfort prediction task. Under hard metric evaluation, it yields 58.1%, 58.4% and 54.3% accuracy for perceived comfort, safety and overall ride comfort, respectively. Soft metric evaluation raises these scores to 88.5%, 86.5% and 90.1%, respectively (Tables C.13, C.14, C.15).

Cohen's κ values further support the improvements over the general model, rising from 0.077 to 0.409 for perceived comfort, from 0.048 to 0.376 for perceived safety, and from 0.126 to 0.397 for overall ride comfort under hard evaluation. Under soft evaluation, they increase from 0.312 to 0.643, 0.178 to 0.652, and 0.371 to 0.748, respectively. These substantial gains demonstrate that the user-adapted models effectively learn meaningful patterns in the GSR data, in constrast to the general model whose performance, particularly under hard evaluation, is largely driven by chance.

Confusion matrices across all three comfort metrics (Figures C.7, C.9, C.11) reveal a strong performance at the extremes: the model shows relatively higher accuracies at the "very uncomfortable/unsafe" and "very comfortable/safe" instances, while most misclassifications occur between adjacent responses such as "uncomfortable/unsafe", "neutral" and "comfortable/safe".
Class-wise ROC curves (Figures C.8(a), C.10(a), C.12(a)) confirm this pattern with AUC values substantially higher for the extreme responses, indicating that the models are better at distinguishing these extremes. Together, these results suggest that the configuration is already reliable for a binary comfort-alert application that only triggers when passengers approach extreme discomfort.

These soft metric results exceed prior binary classifiers in accuracy for trust (78.2%, [1]), motion sickness (77%, [27]), comfort (71.9%, [30]) and stress (73%, [36]). In hard accuracy, the model surpasses a 4-class comfort classifier (55.99%, [22]) but falls short of a 4-class motion sickness classifier (86% [31]) and a 10-class comfort classifier that also employs a similar soft metric allowance (92.4% [6]). Notably, these studies relied on multiple physiological input modalities, whereas the proposed configuration solely relies on the GSR signal.

A key limitation of the current approach lies in the way the GSR input is standardized. To preserve the physiological contrast between calm and aggressive driving, the GSR signal was z-standardized on a per-participant. While this scaling method yielded the highest model performance, it inherently requires prior GSR recordings from each participant in both low- and high-arousal conditions. As such, the models are not directly deployable in a plug-and-play fashion for new users. This limitation poses minimal concern for the subjective comfort and safety models, as those already rely on participant-specific fine-tuning. However, it significantly limits the deployability of the driving style classification model, which can be operated independently of prior user data. A workaround could involve recording a short baseline GSR under low-arousal conditions, for example, calm driving or a stationary rest, before deployment. This baseline could then serve as a reference to scale the measured GSR, or to detect deviations that could indicate elevated arousal or anomalous events.
A second limitation lies in the event-based nature of the dataset and modeling approach. In this study,

each scenario was annotated with three comfort-related ratings, collected post-hoc. As a result, the model is restricted to generating three comfort-related predictions per scenario and is limited to offline evaluation. While the driving style classifier could, in principle, be implemented for online use, the experiment involved stopping the Vehicle Under Test after each scenario to collect subjective ratings. These stationary periods would also need to be annotated either as a distinct stationary class or as the existing “calm” class. Moreover, a real-time classification would still require a sliding window of approximately 30 seconds to enable reliable phasic and tonic decomposition of the GSR, as the cvxEDA algorithm is not designed for online use.
Consequently, real-time comfort predictions remain an open challenge, requiring further research into both online-capable signal decomposition and a solution for the scaling limitation.

Chapter B already highlights the importance of expanding the dataset with additional participants and a broader range of driving scenarios. The predictive modeling framework presented in this chapter would also benefit from such expansion. A larger and more diverse dataset would improve model generalizability and open the door to training a more complex deep learning architecture.
As previously mentioned, it is recommended to record a short baseline GSR segment for each participant and research how this is best used to enable and facilitate real-time models that no longer require both low- and high-arousal data for scaling.
Lastly, while collecting continuous comfort ratings throughout a whole lap could in theory be used to develop a real-time comfort classifier, this approach was deemed impractical by Siemens Digital Industries Software in earlier experiments. Continuous feedback introduced confusion and distraction for participants, ultimately compromising the quality of the data. For this reason, the current post-hoc comfort evaluation remains the most feasible option.

The current dataset, despite its event-based structure, could still lend itself to the development of a real-time anomaly detection framework. One possible approach would be to use a sliding window across the full time series of one lap and assign the scenario’s comfort score to a surrounding time interval. In cases where the reported comfort scores are particularly low, these intervals could be labelled as anomalous. An anomaly detection model trained on such segments could then learn to identify deviations from typical physiological patterns that occur during moments of comfort. This approach could allow for a real-time flagging model of discomfort episodes, despite the absence of continuous comfort ratings.

C.5. Conclusion

This chapter has demonstrated that the Galvanic Skin Response holds substantial promise as an objective indicator of passenger comfort in automated driving. A subject-independent model reliably distinguishes the calm driving style from the aggressive driving style configuration of this experiment with an accuracy of 88.61%.

When calibrated onto individual participants, GSR-based models predicted subjective comfort ratings with hard accuracies of 58.1%, 58.4% and 54.3% for perceived comfort, safety and overall ride comfort, respectively. Applying a soft criterion increases these accuracies to 88.5%, 86.5% and 90.1%, respectively. The elevated soft accuracy scores confirm the GSR’s ability to capture broad fluctuations in passenger comfort, even though the strict classification of finer comfort nuances remains challenging.

The findings in this study emphasize the importance of user-adapted calibration. Without it, model performance remains limited and largely driven by chance due to dataset imbalance. Personalized calibration, however, allows the GSR to become a highly informative and responsive signal for passenger comfort assessment.

D

Generative AI Acknowledgment

For this study, Generative AI, primarily Large Language Models (LLMs), were employed for the following purposes:

Light-editing:
To improve both writing quality and efficiency, LLMs were occasionally used to refine sentences or paragraphs or to suggest better wording. This was applied to both the scientific paper, within the boundaries of the IEEE guidelines, and the final report. Typical prompts for this purpose were:

> [sentence/paragraph]
>
> 1. Read this text (optional: I don't like [last part]); can you help me improve readability? Remember, the text is for a scientific paper/report, so maintain an appropriate academic tone. Do not change the structure or content.
>
> 2. Read this text; I don't like this [word] in this sentence. Can you give me 5 alternatives that fit here?

The outcomes were never directly copied and pasted into the paper or report. Instead, selected phrases or words were integrated into the original text to preserve my own writing style in the paper and report and the tone I wanted for the final work.

Debugging:
LMMs were used for debugging purposes if the debugging issues in question were relatively common problems For more specific or package niche issues, however, these models often failed to provide a solution. In those cases, solutions were found by inspecting the source code, looking into forums or searching elsewhere online. A common prompt for debugging was:

> [code snippet]
>
> This code above gives the following error:
>
> [error message]
>
> Why does this code give me this error, and how can I fix it?

Sometimes, these suggestions were applied directly. If they did not give the same error, I always double-checked that the code still produced the intended outcome. Most of the time, however, the suggestions required adaptation to be fit in the specific context and logic of the existing codebase.

Prototyping Plots:
The final use of LLMs was for quickly prototyping plots and figures. This typical prompt looked like:

> I need your help visualizing [subject] in Python using [matplotlib/plotly].
>
> Write me a function that takes [input] and makes a plot that:
> - [list of requirements]

While the generated code significantly improved efficiency, it consistently required manual adjustments to correct minor flaws or to better align with my preferences.

For text refinement, OpenAI's GPT-4o model was primarily used, and Grammarly was employed to catch spelling mistakes, whereas Anthropic's Claude models, integrated in SiemensGPT, supported debugging and plotting.

Generative AI was **never** used as a source for information, interpretation of results, or methodological decision making; it purely served as a practical aid during this study and the writing of this report. Its use significantly streamlined the workflow and improved productivity; for example, generating exploratory plots through Generative AI support often reduced time by a factor of five compared to doing it from scratch manually.

References

[1] Ighoyota Ben. Ajenaghughrure, Sonia Claudia Da Costa Sousa, and David Lamas. “Psychophysiological modelling of trust in technology: Comparative analysis of algorithm ensemble methods”. en. In: *2021 IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*. Herl’any, Slovakia: IEEE, Jan. 2021, pp. 000161–000168. ISBN: 978-1-7281-8053-3. DOI: 10.1109/SAMI50585.2021.9378655. (Visited on 12/05/2024).

[2] Matthias Beggiato, Franziska Hartwich, and Josef Krems. “Physiological correlates of discomfort in automated driving”. en. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 66 (Oct. 2019), pp. 445–458. ISSN: 13698478. DOI: 10.1016/j.trf.2019.09.018. (Visited on 12/05/2024).

[3] Mathias Benedek and Christian Kaernbach. “A continuous measure of phasic electrodermal activity”. en. In: *Journal of Neuroscience Methods* 190.1 (June 2010), pp. 80–91. ISSN: 01650270. DOI: 10.1016/j.jneumeth.2010.04.028. (Visited on 12/10/2024).

[4] Jason Braithwaite J. et al. *A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments*. Tech. rep. Birmingham: Behavioural Brain Sciences Centre, University of Birmingham, UK, 2015.

[5] N. V. Chawla et al. “SMOTE: Synthetic Minority Over-sampling Technique”. In: *Journal of Artificial Intelligence Research* 16 (June 2002). arXiv:1106.1813 [cs], pp. 321–357. ISSN: 1076-9757. DOI: 10.1613/jair.953. (Visited on 05/21/2025).

[6] Maciej Piotr Cieslak. “Towards accurate ride comfort evaluation using biometric measurements and neural networks”. en. PhD thesis. Coventry University, 2019.

[7] Harald Devriendt et al. “A Multimodal Sensor Setup for In Situ Comparison of Driving Dynamics, Physiological Responses and Passenger Comfort in Autonomous Vehicles”. en. In: 2025. DOI: 10.54941/ahfe1005852. (Visited on 03/20/2025).

[8] Nicole Dillen et al. “Keep Calm and Ride Along: Passenger Comfort and Anxiety as Physiological Responses to Autonomous Driving Styles”. en. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Honolulu HI USA: ACM, Apr. 2020, pp. 1–13. ISBN: 978-1-4503-6708-0. DOI: 10.1145/3313831.3376247. (Visited on 12/05/2024).

[9] Euro NCAP. *Crash Avoidance: Frontal Collisions Protocol*. Protocol. version: 0.9. Euro NCAP, Dec. 2024.

[10] Alessandro Gabrielli et al. “A Framework to Study Autonomous Driving User Acceptance in the Wild”. en. In: *Intelligent Autonomous Systems 16*. Ed. by Marcelo H. Ang Jr et al. Vol. 412. Series Title: Lecture Notes in Networks and Systems. Cham: Springer International Publishing, 2022, pp. 123–140. ISBN: 978-3-030-95891-6 978-3-030-95892-3. DOI: 10.1007/978-3-030-95892-3_10. (Visited on 12/05/2024).

[11] Alberto Greco et al. “cvxEDA: A Convex Optimization Approach to Electrodermal Activity Processing”. In: *IEEE Transactions on Biomedical Engineering* 2016 (Apr. 2016), pp. 797–804. DOI: 10.1109/TBME.2015.2474131.

[12] Eléonore H. Henry et al. “Car sickness in real driving conditions: Effect of lateral acceleration and predictability reflected by physiological changes”. en. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 97 (Aug. 2023), pp. 123–139. ISSN: 13698478. DOI: 10.1016/j.trf.2023.06.018. (Visited on 12/17/2024).

[13] Francisco Hernando-Gallego, David Luengo, and Antonio Artes-Rodriguez. “Feature Extraction of Galvanic Skin Responses by Nonnegative Sparse Deconvolution”. en. In: *IEEE Journal of Biomedical and Health Informatics* 22.5 (Sept. 2018), pp. 1385–1394. ISSN: 2168-2194, 2168-2208. DOI: 10.1109/JBHI.2017.2780252. (Visited on 12/27/2024).

[14] Stefanie Horn et al. “User evaluation of comfortable deceleration profiles for highly automated driving: Findings from a test track study”. en. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 105 (Aug. 2024), pp. 206–221. ISSN: 13698478. DOI: 10.1016/j.trf.2024.05.025. (Visited on 06/22/2025).

[15] Tugrul Irmak, Daan M. Pool, and Riender Happee. “Objective and subjective responses to motion sickness: the group and the individual”. en. In: *Experimental Brain Research* 239.2 (Feb. 2021), pp. 515–531. ISSN: 0014-4819, 1432-1106. DOI: 10.1007/s00221-020-05986-6. (Visited on 12/05/2024).

[16] J. R. Landis and G. G. Koch. “The measurement of observer agreement for categorical data”. eng. In: *Biometrics* 33.1 (Mar. 1977), pp. 159–174. ISSN: 0006-341X.

[17] Erika Lutin et al. “Feature Extraction for Stress Detection in Electrodermal Activity:” en. In: *Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies*. Online Streaming, — Select a Country —: SCITEPRESS - Science and Technology Publications, 2021, pp. 177–185. ISBN: 978-989-758-490-9. DOI: 10.5220/0010244601770185. (Visited on 05/27/2025).

[18] Dominique Makowski et al. “NeuroKit2: A Python toolbox for neurophysiological signal processing”. en. In: *Behavior Research Methods* 53.4 (Aug. 2021), pp. 1689–1696. ISSN: 1554-3528. DOI: 10.3758/s13428-020-01516-y. (Visited on 11/21/2024).

[19] Haolan Meng et al. “Study on physiological representation of passenger cognitive comfort: An example with overtaking scenarios”. en. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 102 (Apr. 2024), pp. 241–259. ISSN: 13698478. DOI: 10.1016/j.trf.2024.03.003. (Visited on 03/20/2025).

[20] Drew M. Morris, Jason M. Erno, and June J. Pilcher. “Electrodermal Response and Automation Trust during Simulated Self-Driving Car Use”. en. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 61.1 (Sept. 2017), pp. 1759–1762. ISSN: 1071-1813, 2169-5067. DOI: 10.1177/1541931213601921. (Visited on 12/05/2024).

[21] Dario Niermann, Alexander Trende, and Andreas Luedtke. “Tracking and Evaluation of Human State Detections in Adaptive Autonomous Vehicles”. en. In: *HCI International 2020 - Posters*. Ed. by Constantine Stephanidis and Margherita Antona. Vol. 1224. Series Title: Communications in Computer and Information Science. Cham: Springer International Publishing, 2020, pp. 378–384. ISBN: 978-3-030-50725-1 978-3-030-50726-8. DOI: 10.1007/978-3-030-50726-8_50. (Visited on 12/02/2024).

[22] Shiwei Peng et al. “Comfort of Autonomous Vehicles Incorporating Quantitative Indices for Passenger Feeling”. en. In: *Journal of Shanghai Jiaotong University (Science)* 29.6 (Dec. 2024), pp. 1063–1070. ISSN: 1007-1172, 1995-8188. DOI: 10.1007/s12204-022-2531-5. (Visited on 12/05/2024).

[23] Jaume R. Perello-March et al. “Driver State Monitoring: Manipulating Reliability Expectations in Simulated Automated Driving Scenarios”. In: *IEEE Transactions on Intelligent Transportation Systems* 23.6 (June 2022). Conference Name: IEEE Transactions on Intelligent Transportation Systems, pp. 5187–5197. ISSN: 1558-0016. DOI: 10.1109/TITS.2021.3050518. (Visited on 12/18/2024).

[24] Vishnu Radhakrishnan et al. “Measuring Drivers’ Physiological Response to Different Vehicle Controllers in Highly Automated Driving (HAD): Opportunities for Establishing Real-Time Values of Driver Discomfort”. en. In: *Information* 11.8 (Aug. 2020), p. 390. ISSN: 2078-2489. DOI: 10.3390/info11080390. (Visited on 12/12/2024).

[25] Elena N. Schneider et al. “Electrodermal Responses to Driving Maneuvers in a Motion Sickness Inducing Real-World Driving Scenario”. en. In: *IEEE Transactions on Human-Machine Systems* 52.5 (Oct. 2022), pp. 994–1003. ISSN: 2168-2291, 2168-2305. DOI: 10.1109/THMS.2022.3188924. (Visited on 12/05/2024).

[26] Shili Sheng et al. *A Case Study of Trust on Autonomous Driving*. arXiv:1904.11007 [cs]. July 2019. DOI: 10.48550/arXiv.1904.11007. (Visited on 12/18/2024).

[27] Oluwaseyi Elizabeth Shodipe and Robert S. Allison. "Modelling the relationship between the objective measures of car sickness". en. In: *2023 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*. Regina, SK, Canada: IEEE, Sept. 2023, pp. 570–575. ISBN: 979-8-3503-2397-9. DOI: 10.1109/CCECE58730.2023.10289000. (Visited on 12/05/2024).

[28] Joseph Smyth et al. "Exploring the utility of EDA and skin temperature as individual physiological correlates of motion sickness". en. In: *Applied Ergonomics* 92 (Apr. 2021), p. 103315. ISSN: 00036870. DOI: 10.1016/j.apergo.2020.103315. (Visited on 12/05/2024).

[29] Jork Stapel, Alexandre Gentner, and Riender Happee. "On-road trust and perceived risk in Level 2 automation". en. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 89 (Aug. 2022), pp. 355–370. ISSN: 13698478. DOI: 10.1016/j.trf.2022.07.008. (Visited on 12/17/2024).

[30] Haotian Su and Yunyi Jia. "Study of Human Comfort in Autonomous Vehicles Using Wearable Sensors". en. In: *IEEE Transactions on Intelligent Transportation Systems* 23.8 (Aug. 2022), pp. 11490–11504. ISSN: 1524-9050, 1558-0016. DOI: 10.1109/TITS.2021.3104827. (Visited on 12/05/2024).

[31] Ruichen Tan et al. "Motion Sickness Detection for Intelligent Vehicles: A Wearable-Device-Based Approach". en. In: *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*. Macau, China: IEEE, Oct. 2022, pp. 4355–4362. ISBN: 978-1-6654-6880-0. DOI: 10.1109/ITSC55140.2022.9922392. (Visited on 12/17/2024).

[32] Laurin Vasile et al. "Comfort and Safety in Conditional Automated Driving in Dependence on Personal Driving Behavior". en. In: *IEEE Open Journal of Intelligent Transportation Systems* 4 (2023), pp. 772–784. ISSN: 2687-7813. DOI: 10.1109/OJITS.2023.3323431. (Visited on 04/21/2025).

[33] P.H. Venables and D.A. Mitchell. "The effects of age, sex and time of testing on skin conductance activity". en. In: *Biological Psychology* 43.2 (Apr. 1996), pp. 87–101. ISSN: 03010511. DOI: 10.1016/0301-0511(96)05183-6. (Visited on 04/23/2025).

[34] Xiaodi Xiang et al. "Research on vehicle comfort testing and evaluation based on the characterization of passenger motion sickness degree". en. In: *Third International Conference on Biomedical and Intelligent Systems (IC-BIS 2024)*. Ed. by Zulqarnain Baloch and Pier Paolo Piccaluga. Nanchang, China: SPIE, July 2024, p. 112. ISBN: 978-1-5106-8127-9 978-1-5106-8128-6. DOI: 10.1117/12.3036840. (Visited on 12/16/2024).

[35] Tianxiang Zhan et al. *Time Evidence Fusion Network: Multi-source View in Long-Term Time Series Forecasting*. en. arXiv:2405.06419 [cs]. Sept. 2024. DOI: 10.48550/arXiv.2405.06419. (Visited on 04/25/2025).

[36] Pamela Zontone et al. "Stress Evaluation in Simulated Autonomous and Manual Driving through the Analysis of Skin Potential Response and Electrocardiogram Signals". en. In: *Sensors* 20.9 (Apr. 2020), p. 2494. ISSN: 1424-8220. DOI: 10.3390/s20092494. (Visited on 12/05/2024).