Optimization and Personalization of Assisted Driving Functions based on Implicit Driver Feedback

Zur Erlangung des akademischen Grades eines
DOKTORS DER INGENIEURWISSENSCHAFTEN (Dr.-Ing.)
von der KIT-Fakultät für Elektrotechnik und Informationstechnik des
Karlsruher Instituts für Technologie (KIT)
angenommene

DISSERTATION

von

M.Sc., Schwager, Robin
geboren in: Mühlacker, Deutschland

Tag der mündlichen Prüfung:	05.05.2026
Hauptreferent:	Prof. Dr.-Ing. Sören Hohmann
Korreferent:	Prof. Dr. Klaus Bengler

Preface

This thesis is the result of my work as a research assistant at Dr. Ing. h.c. F. Porsche AG and the Institute of Control Systems (IRS) at the Karlsruhe Institute of Technology (KIT). Without the support of many people, this thesis would not have been possible. First and foremost, I would like to thank Prof. Dr.-Ing. Sören Hohmann for supervising my research and his continued support over the course of the past three years. Furthermore, I would like to thank Prof. Dr. Klaus Bengler for his interest in my work and for the assessment of this thesis. I appreciate your participation in the evaluation committee of my thesis.

Many thanks to all colleagues at Porsche who supported me with my work. Especially, I want to thank Dr. Michael Grimm for supervising my thesis from the company side. Thank you for your continued support, your genuine interest in my work, and our fruitful and interesting discussions on the many aspects of this thesis. Furthermore, I would like to thank Dr. Sebastian Söhner for his supervision of our research campus at Porsche and his efforts in providing us with all the resources required for our research. Moreover, I would like to express my heartfelt thanks to all past and current members of the InnoCampus. Thank you for providing such a unique and lighthearted working atmosphere. Especially, I want to thank my fellow PhD students at Porsche: Stefan Roos, Lukas Ewecker, Tim Brühl, Tin Stribor Sohn, and Tim Eberhardt; and all my students who supported me with my work: Mattis Bodynek, Lena Dalke, Lennart Querfurth, Simon Hartmann, Xin Liu, Matej Švaral, Lukas Schick, and Andrea Anastasio. It was a pleasure to work with all of you. I will always fondly remember my time in the InnoCampus.

Last but not least, I want to thank my friends and family for their unconditional support throughout these three years. Finally, Kaitlyn, I am very grateful for your love and all your continued encouragement. You are wonderful. Thank you for everything.

Mühlacker, September 2025

Letzte Nacht habe ich geträumt,
ich sei ein Pandas DataFrame.

Matej Švaral, a very committed intern

Abstract

When using assisted driving functions, drivers frequently take over the vehicle control to adjust the system's behavior according to their personal preferences. However, voluntary driver interventions are rarely analyzed in related work, although a correlation between intervention frequency and driver dissatisfaction is generally assumed. Thus, this thesis focuses on the analysis of driver intervention behavior during naturalistic driving. In a first test group study, participants manually annotated the reasons behind each intervention while using an in-production Predictive Longitudinal Driving Function (PLDF). The study results show that the majority of interventions are voluntary adjustments based on deviating personal preferences, indicating a high optimization potential. Building on these findings, a framework for the derivation of necessary adjustments to map-based driving functions is proposed and applied to the recorded dataset. The analysis reveals that generalized policy adjustments are rarely applicable due to inconsistent individual driving behavior. Instead, personalized adjustments to the PLDF's speed profile are proposed solely in specific locations, where the drivers consistently intervene. Additionally, to enable the application of the developed algorithms without manual annotation, an automatic classification pipeline is developed using the annotated intervention dataset. Based on these foundations, a prototypical self-learning PLDF is designed that iteratively adjusts its speed profile based on relevant driver interventions in a traded control setting. The applicability of this prototype is evaluated in a driving simulator-based test group study. The study finds both a significant increase in driver satisfaction and a significant reduction in intervention frequency, confirming the applicability of the prototype. Thus, this thesis demonstrates that driver interventions are a valid source for feedback for the personalization of map-based assisted driving functions.

Kurzfassung

Während der Nutzung von assistierten Fahrfunktionen greifen Fahrer häufig in das aktive System ein, um das Systemverhalten an ihre persönlichen Präferenzen anzupassen. Diese freiwilligen Fahrereingriffe wurden in der einschlägigen Forschung bislang jedoch nur selten untersucht, obwohl eine Korrelation zwischen der Eingriffshäufigkeit und der Fahrerunzufriedenheit allgemein angenommen wird. Diese Dissertation befasst sich daher mit der Analyse des Eingriffsverhaltens von Fahrern unter realen Bedingungen. In einer ersten Probandenstudie annotierten die Teilnehmenden unter realen Fahrbedingungen die Gründe für jeden Fahrereingriff während der Nutzung einer serienmäßig verfügbaren prädiktiven longitudinalen Fahrfunktion. Die Ergebnisse der Studie zeigen, dass über die Hälfte aller Eingriffe freiwillige Anpassungen des Fahrverhaltens aufgrund von abweichenden persönlichen Präferenzen darstellen. Dies deutet auf ein erhebliches Optimierungspotenzial hin. Basierend auf diesen Erkenntnissen wird eine Methodik zur Ableitung notwendiger Anpassungen kartenbasierter Fahrfunktionen entwickelt und auf den erhobenen Datensatz angewendet. Die Analyse des Datensatzes verdeutlicht, dass generalisierte Anpassungen der Fahrstrategie aufgrund inkonsistenten individuellen Fahrverhaltens nur eingeschränkt anwendbar sind. Stattdessen werden personalisierte Anpassungen des Geschwindigkeitsprofils der Fahrfunktion nur für spezifische Streckenabschnitte vorgeschlagen, in denen die Fahrer konsistent eingreifen. Um die Anwendung der entwickelten Algorithmen auch ohne manuelle Annotation zu ermöglichen, wird zudem eine automatische Klassifikations-Pipeline entwickelt, die auf dem Datensatz annotierter Fahrereingriffe trainiert wird. Auf diese Erkenntnisse aufbauend wird eine prototypische selbstlernende longitudinale Fahrfunktion entwickelt, welche sich iterativ an die individuellen Fahrerpräferenzen anpasst, indem sie aus deren Eingriffsverhalten lernt. Die Anwendbarkeit des Prototyps wird in einer weiteren Probandenstudie in einem Fahrsimulator evaluiert. Die Ergebnisse der Studie zeigen sowohl eine signifikante Steigerung der Fahrerzufriedenheit als auch eine signifikante Reduktion der Eingriffshäufigkeit und bestätigen damit die generelle Anwendbarkeit des Prototyps. Diese Dissertation zeigt somit, dass Fahrereingriffe eine valide Feedbackquelle für die Personalisierung kartenbasierter Fahrfunktionen darstellen.

Contents

1 Introduction

Advanced Driver Assistance Systems (ADASs) represent a rapidly growing field of research and development in the automotive sector. According to recent reports, approximately 60 % of all new vehicles sold globally are equipped with some form of ADAS technology [Mar]. Driving functions[1], also called *continuous automation systems*, are types of ADASs that actively take over parts of the vehicle control for extended periods of time during driving [Gas+17; Win+15]. In the J3016 standard [Sae], the Society of Automotive Engineers (SAE) International defines six levels of driving automation for driving functions, ranging from Level 0, *no driving automation*, to Level 5, *full driving automation*. Currently, most in-production driving functions are at Levels 1 and 2, while Level 3 systems are gradually being introduced to the market [Yan]. One key distinction is that driving functions of Level 3 and above allow drivers to disengage from the driving task and perform secondary activities, such as reading a newspaper. In contrast, drivers are still required to continuously monitor the driving function's behavior when using Level 1 and 2 driving functions. At these lower levels, the driver must be ready to take over the vehicle control at any time, either when the driving function leaves its Operational Design Domain (ODD) or in order to correct potential errors made by the system [Sae]. As a result, the drivers may notice discrepancies between their preferred driving style and the driving function's behavior. Drivers may then intervene and temporarily take over the vehicle control to adjust its behavior to their preferences. Such voluntary interventions typically occur within the driving function's ODD, i.e., in situations that the function is designed to handle. By intervening in this way, drivers implicitly provide feedback on aspects of the system's behavior with which they are dissatisfied. However, related research rarely investigates the underlying reasons for these driver-initiated takeovers [Ger+21; Mor+20] and rarely leverages the potential feedback contained in them [GJ23], although a correlation between intervention frequency and driver satisfaction is generally assumed [GJ23; Lee+21; MZ21; Wan+13].

The personalization of driving functions is an active research field [HHW20; Yi+20], as it is widely acknowledged that driving styles vary between individual drivers [Chu+23; Sag+15; TBAMG04]. Existing personalization approaches for longitudinal driving functions primarily focus on Adaptive Cruise Control (ACC) driving scenarios, during which the ego vehicle maintains a safe distance to a leading vehicle. Most of these approaches aim to imitate an individual driver's manual driving behavior as closely as possible, based on the assumption that individual drivers generally prefer an automated driving style that mirrors their manual driving style [GJ23; HHW20; Yi+20]. In contrast, other researchers propose a continuous personalization process that incorporates driver

[1] The term *driving function* is defined in Section 2.1.1.

feedback during function usage [AT05; GJ23; HHW20]. Such feedback could potentially be derived from the aforementioned driver interventions. As an example, the field of Interactive Imitation Learning (IIL) builds upon the premise that expert interventions contain valuable feedback, which can be used to iteratively update a policy [Cel+22]. Accordingly, the investigation of driver interventions during naturalistic driving could provide information about potentially necessary optimizations or personalizations of the underlying driving function in order to better comply with the individual drivers' preferences.

This thesis is conducted in cooperation with Dr. Ing. h.c. F. Porsche AG, hereinafter referred to as *Porsche*. Accordingly, the driving function analyzed in this thesis is an in-production Level 1 Predictive Longitudinal Driving Function (PLDF) developed by Porsche. The PLDF takes over the longitudinal vehicle control and predictively generates an appropriate speed profile by taking into account legal speed limits and road topography. In addition, the PLDF includes standard ACC functionality. Thus, the analyzed driving function also incorporates map-based behavior, in contrast to most related research on driving function personalization, which primarily focuses on generalized ACC policies.

1.1 Research Contributions and Thesis Structure

The structure of this thesis is illustrated in Figure 1.1. In Chapter 2, the fundamentals and terminology, as well as the state of research relevant to this thesis are introduced. Furthermore, the resulting research gap is defined and the resulting research objectives are derived.

As outlined above, it is assumed that voluntary driver-initiated takeovers contain implicit feedback on potentially necessary adjustments to the underlying driving function. However, related work provides no in-depth analysis of the underlying reasons for driver interventions during naturalistic driving with longitudinal driving functions. Therefore, the first primary research objective of this thesis is the generation of a dataset containing annotated interventions during naturalistic driving in a test group study. The employed study design and the creation of the intervention dataset are presented in Chapter 3. This dataset is then used to investigate the underlying reasons for driver-initiated takeovers and to assess whether these interventions provide usable feedback for driving function adjustments. To this end, the different types of recorded interventions are discussed and a hierarchical labeling taxonomy for driver interventions is introduced. Furthermore, the dataset is generally analyzed and the questionnaire results are evaluated.

The second primary research objective is the development of a methodology to derive necessary adjustments to the driving function based on the recorded driver interventions in a data-driven manner. This methodology evaluates the consistency of the recorded driver intervention behavior in order to identify suitable driving function adjustments. The analysis primarily examines whether a personalization of the driving function's behavior is required, and whether the necessary changes can be represented by generalized driving policy adjustments, as it is commonly done in related research on driving function personalization. In Chapter 4, this developed framework is introduced and applied to the dataset from Chapter 3. Finally, the resulting recommended adjustment strategies are discussed.

The third primary research objective is the development of an automatic driver intervention classification pipeline. This automatic classification enables the application of the developed analyses and algorithms without the manual labeling process conducted during the first test group study. The complete classification pipeline is described in Chapter 5 after the relevant employed classification algorithms are introduced. Furthermore, necessary data preprocessing steps are highlighted and the resulting classification performance of the evaluated algorithms is analyzed. Based on the classification results, potential ways to further improve the performance are discussed.

Finally, the fourth research objective applies the preceding findings to develop a prototypical self-learning driving function that is iteratively updated based on performed driver interventions. To this end, an algorithm is designed that generates updated target speed profiles from recorded interventions. The developed prototype is subsequently evaluated in a second test group study to demonstrate its ability to adapt to individual driver preferences. Chapter 6 presents the complete self-learning driving function framework and the employed driving simulator-based study design. The results of the simulator study are discussed with a focus on the self-learning driving function's effects on driver satisfaction with the system and the drivers' intervention behavior over the course of the experiments.

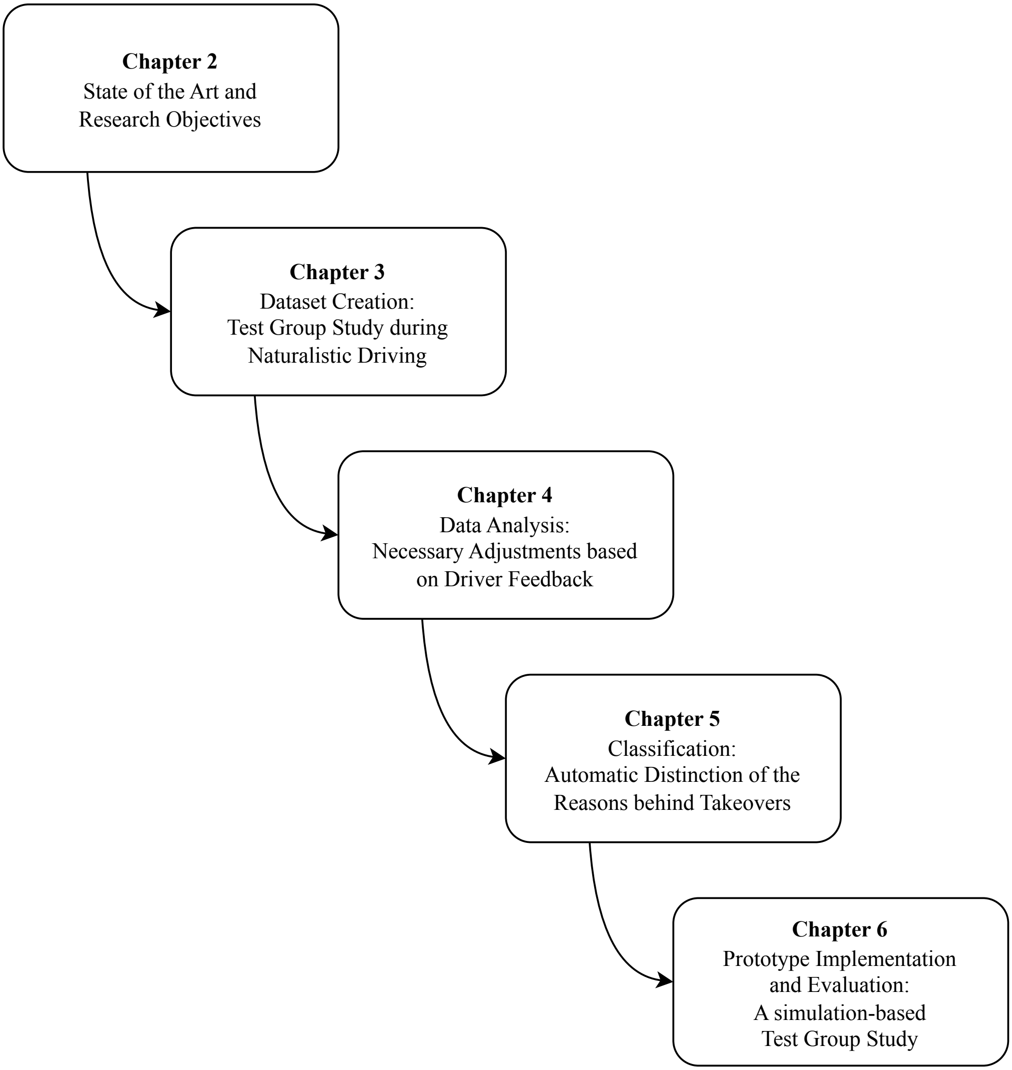


Figure 1.1: The structure of this thesis.

2 State of the Art and Research Objectives

In this chapter, the state of the art relevant to the present thesis is introduced. First, Section 2.1 provides a general introduction to the fundamentals and terminology required throughout this work. Recent research in the field of driver-initiated takeovers and their underlying reasons is highlighted in Section 2.2. In Section 2.3, a literature review on different driving function personalization approaches is presented. Subsequently, Section 2.4 discusses research on the preferred automated driving style of humans. Relevant IIL approaches are examined in Section 2.5. Finally, the reviewed literature is discussed in Section 2.6, resulting research gaps are identified, and the research objectives of this thesis are introduced.

2.1 Fundamentals and Terminology

The fundamentals required for this thesis primarily focus on the different types of ADASs, their classification, and the interactions between driver and automation during their use. Relevant terminology is also introduced throughout this section.

2.1.1 Advanced Driver Assistance Systems

ADASs are an increasingly important field of research and development in the automotive sector. The following subsection aims to explain and classify the many different types of ADASs on the current market.

Definition 2.1 (ADAS)
ADASs include a range of in-vehicle technologies designed to assist the driver in various aspects of the driving task. The main goal of ADASs is to improve driving comfort and safety. They encompass functions that can automate parts of the vehicle control task, provide additional information, warn drivers of potentially hazardous situations, and even automatically take over the vehicle control in emergency situations [Gas+17; Win+15].

Gasser *et al.* [Gas+17] and Winner *et al.* [Win+15] define three main categories of ADASs depending on their principle of operation. First, there are *information and* *warning systems*. These communicate relevant information to the driver or warn them about potentially hazardous situations. Information and warning systems use the human-machine interfaces of the vehicle to visually and acoustically communicate with the driver, e.g., via icons in the instrument cluster, or by playing sound notifications. Common information and warning systems include the traffic sign recognition, lane departure warning, and blind spot monitoring.

The second and most relevant category for this thesis is the field of *continuous automation systems*. These systems support the driver with the active task of driving and are henceforth also called *driving functions*.

Definition 2.2 (Driving Function)
A driving function, also called a continuous automation system, is an ADAS which actively takes over and automates parts of the vehicle control for a prolonged time during driving. In this context, the driving task can be divided into two main control tasks: longitudinal and lateral vehicle control. The longitudinal control describes the vehicle's acceleration control, primarily via the accelerator and brake pedals, while the lateral control describes the vehicle's steering control. Driving functions can be activated and deactivated at will be the driver. Common examples include cruise control, ACC, and lane centering assist [Gas+17; Win+15].

The third category are the *temporarily intervening systems*, which automatically and only temporarily take over the vehicle control during potentially hazardous situations. These systems permanently monitor the driver and their surroundings, similar to the information and warning systems. However, if a potential hazard is detected, the system may automatically take over the vehicle control and try to avoid a potential accident. Examples of such systems are the autonomous emergency brake and the active lane keep assist [Gas+17; Win+15].

2.1.2 Levels of Driving Automation

In the field of continuous automation functions, the SAE International defined six levels of driving automation in their J3016 standard [Sae]. These levels define the degree of automation of driving functions. Relevant aspects include the functionality of the automated system, the driver's responsibilities while the automation is active, and the ODD for which the function was designed. In the following paragraphs, the six levels of driving automation are explained in detail.

Definition 2.3 (ODD)
The ODD describes under which conditions a driving function can be safely operated. This includes geographical, temporal, environmental, or traffic-related limitations. If the driving function's ODD is left, in most cases, the driver is required to take back control of the previously automated vehicle control tasks [Lee+20; Sae].

Level 0 - No Driving Automation: The driver fully controls the longitudinal and lateral vehicle control tasks without any continuous automation systems. Information and warning systems may still be active [Sae].

Level 1 - Driver Assistance: Level 1 of the driving automation levels describes the field of driver assistance. Level 1 driving functions are also called *assisted driving functions*.

Definition 2.4 (Assisted Driving Function)
An assisted driving function is a Level 1 continuous automation system. It takes over either the longitudinal or lateral control of the vehicle in a specified ODD, and the driver still controls the other components of vehicle control. The driver bears full responsibility for the vehicle's behavior and must continuously observe the driving function. If the driving function makes a mistake or leaves its ODD, the driver must take over the vehicle control again [Sae].

Level 2 - Partial Driving Automation: Similar to Level 1 driving automation, but in this case, the driving function takes over both the longitudinal and lateral control of the vehicle in a specified ODD [Sae].

Level 3 - Conditional Driving Automation: The driving function takes over the full driving task in a specified ODD, e.g., during highway driving. While the system remains active and within its ODD, the vehicle manufacturer takes full responsibility for the vehicle's behavior. The driver may perform secondary tasks, e.g., reading a newspaper, and is not required to observe the driving function. The driving function may request driver support, e.g., due to leaving its ODD, and the driver must return to the driving task within a specified takeover time [Sae].

Level 4 - High Driving Automation: Equivalent to Level 3 driving automation, with the key difference that the driver may fully detach from the driving task, e.g., by sleeping. The vehicle must always be capable of reaching a safe state even without interventions by a human driver [Sae].

Level 5 - Full Driving Automation: During level 5 driving, also called *autonomous driving*, the vehicle operates completely autonomously. A driver is never required, and the system completely takes over the driving task in all situations and ODDs [Sae].

2.1.3 Hierarchical Model of the Driving Task

For the analysis of human driving behavior while using ADASs, a model for the abstraction of the driving task is required. In the context of ADASs, Michon's hierarchical model is commonly used to categorize the different aspects of the human driving task [Ger+21; MS20; Sch12; WXC14]. Michon [Mic85] describes that, from a driver's perspective, the driving task can be divided into three hierarchical levels of control: *strategical planning*, *tactical maneuvering*, and *operational control*. The model is depicted in Figure 2.1 and generally follows a top-down approach. However, unforeseen factors during the execution of the lower levels can also cause changes in the higher-level strategies of the driver. This is depicted as the bidirectional information flow between the three levels of the driving task.

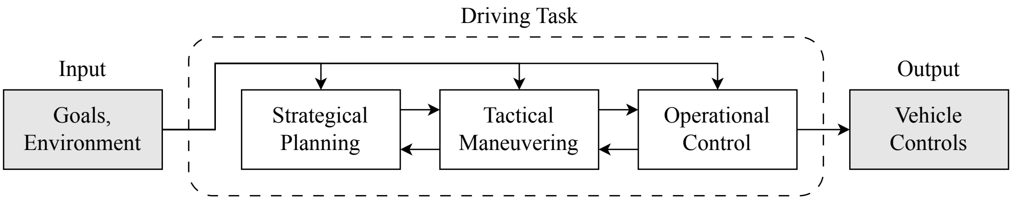


Figure 2.1: Hierarchical model of the driving task (adapted from [Mic85]).

On the strategical planning level, the driver primarily makes decisions regarding the navigation task of the drive. The route choice may be influenced by various factors, such as traffic conditions, fuel consumption, and personal preferences. Decisions at the strategical planning level are typically made over relatively long time periods [MS20; Mic85; WXC14]. The tactical maneuvering level describes how the chosen route is executed in terms of individual maneuvers. This includes, e.g., the decision whether to stay behind a slower vehicle or to overtake it. Decisions at the tactical maneuvering level are typically made within a few seconds [Mic85; WXC14]. Finally, the operational control level defines the execution of maneuvers, including steering, accelerating, and braking control. Decisions on this level generally take less than a second and feel automatic to the driver [Mic85; WXC14].

2.1.4 Human-Machine Cooperation and Transitions of Control

When using driving functions, both the driver and the system can control the vehicle. This combination of driver and system control over the same task can be classified as human-machine cooperation.

Definition 2.5 (Cooperation)
Hoc defines cooperation as follows:

> *Two agents are in a cooperative situation if they meet two minimal conditions.*
>
> 1. *Each one strives toward goals and can interfere with the other on goals, resources, procedures, etc.*
> 2. *Each one tries to manage the interference to facilitate the individual activities and/or the common task when it exists.*
>
> *The symmetric nature of this definition can be only partly satisfied [Hoc01].*

Based on this definition of cooperation, especially the operation of Level 1 and 2 driving functions falls under the term of human-machine cooperation. There, the driver is still considered to be in the loop, needs to continuously monitor the driving function's behavior, and potentially still take over parts of the vehicle control [Mar+20; Sae]. Two main subcategories of human-machine cooperation exist in the state of the art: *shared control* and *traded control*. These can be primarily distinguished by whether only one agent is active at a time or whether both agents are active at the same time [WPA23; Mar+20].

Definition 2.6 (Shared Control)
Under the term shared control*, a human-machine cooperation is defined where both computer and human are working on the same task at the same time [SVB78]. In the context of ADASs, this means that both the driving function and the human driver actively control the same aspects of the driving task [Mar+20].*

An example of a Level 1 shared control lateral driving function is the active lane centering assist, where both driver and driving function assert forces on the steering wheel at the same time [Mar+20].

Definition 2.7 (Traded Control)

Traded control defines a human-machine cooperation where both computer and human are working on the same task but at different times [SVB78]. In the context of ADASs, this means that the driving function and the driver take turns at controlling specific aspects of the vehicle control but never do so at the same time [WPA23; Mar+20].

An example of a Level 1 traded control longitudinal driving function is ACC, where either the driving function or the driver controls the vehicle's acceleration at any time. While the ACC controls the vehicle, the driver primarily monitors the system without interacting with the longitudinal vehicle controls. However, if the driver decides to take over the vehicle control, e.g., by pressing the brake pedal, the ACC automatically relinquishes its control over the vehicle. In this manner, the vehicle control is traded between the driver and the automation [WPA23; Mar+20].

Transitions of Control during Traded Control

As introduced in Chapter 1, this thesis focuses on a longitudinal driving function that falls into the category of traded control. Accordingly, the process of trading the vehicle control between the driver and automation, along with the associated terminology, is discussed in this section. Terminology in the research literature in the field of control transitions is often ambiguous, and several works attempt to provide a unified nomenclature [Lu+16; Mag+22]. Since no universally accepted terminology is available in the state of the art, the terminology adopted in this thesis is chosen to best align with its focus and objectives. This terminology is primarily based on the works of Martens *et al.* [Mar+08], Lu *et al.* [Lu+16], and Maggi *et al.* [Mag+22]. The fundamental building block of this terminology is the so-called *Transition of Control (ToC)*.

Definition 2.8 (ToC)
A ToC describes a change in the automation status of a driving function. More specifically, it describes a change of the agent that partially or fully controls the driving task, i.e., the longitudinal and/or lateral control of the vehicle. The driving task can therefore be shifted from the human to the automation and vice versa [Lu+16; Mag+22; Tof+09].

Multiple types of ToCs exist, and they are generally distinguished based on the answers to the following three questions [Lu+16; Mag+22]:

1. Who initiates the ToC?
2. Who is in control after the ToC?

3. Is the ToC optional or mandatory?

The answers to the first two questions can be either *the driver* or *the system*. Based on these first two questions, four distinct types of ToCs can be defined. The first type, a *driver-initiated transition from driver to automation*, describes a transition of the vehicle control from the driver to the automation that is always voluntary. An example of this ToC type is the simple manual activation of an assisted driving function, such as ACC, by the driver [Lu+16; Mar+08].

The second ToC type is the *driver-initiated transition from automation to driver* [Lu+16; Mar+08]. In the context of this thesis, this ToC is also called a *driver-initiated takeover*.

Definition 2.9 (Driver-initiated Takeover)
A driver-initiated takeover describes the manual driver-initiated transition of the full or partial vehicle control from the driving function to the driver. During such a driver-initiated takeover, the driving function is deactivated by the driver and manual driving is resumed [Lu+16; Mar+08].

A driver-initiated takeover can be based on a voluntary decision of the driver to take over the vehicle control, or it can be mandatory based on the current driving situation. The primary reasons for a mandatory driver-initiated takeover include leaving the driving function's ODD and correcting a mistake made by the driving function [Lu+16; Tof+09].

When a driving function, especially but not limited to Level 3, detects that it leaves its ODD, it may warn the driver to return to the driving task. Such a ToC is called a *system-initiated transition from automation to driver* or, in this thesis, a *system-initiated takeover* [Lu+16; Mag+22].

Definition 2.10 (System-initiated Takeover)
A system-initiated takeover describes the automatic system-initiated transition of the full or partial vehicle control from the driving function to the driver. Such a system-initiated takeover is triggered when the driving function detects that it is about to leave its ODD, has already left its ODD, or has run into a system malfunction, and can therefore no longer perform the previously automated driving task.

Depending on the level of driving automation, the driving function can either immediately hand back the control to the driver by directly deactivating itself, or it can perform a so-called *handover* of the driving task, which gives the driver some time to take over the vehicle control before the driving function deactives itself. As the driving function cannot safely control the vehicle outside of its ODD, a system-initiated takeover can never be voluntary and is always mandatory [Lu+16; Mag+22; MA+20; Tof+09].

The fourth and final type of ToCs is the *system-initiated transition from driver to automation*. Such a ToC occurs when the driver manually controls the vehicle and a temporarily intervening driving function detects a potentially hazardous situation. The driving function then automatically intervenes and takes over control of the vehicle to protect the driver or other traffic participants. Therefore, a system-initiated transition from driver to automation can also never be voluntary and is always mandatory. An example system that utilizes such ToCs is the automated emergency brake, which only activates during near-crash situations and overrides the driver's inputs [Lu+16].

2.1.5 Personalization

The personalization of driving functions is one of the main subjects of this work. Therefore, the concept of personalization and the feedback on which this personalization is based on must be defined. Tuzhilin [Tuz09] defines personalization as follows: "Personalization tailors certain offerings by providers to consumers based on certain knowledge about them, on the context in which these offerings are provided with a certain goal in mind". Adomavicius and Tuzhilin [AT05] further extend this by describing personalization as an iterative process in the form of a feedback loop, consisting of three steps: understanding consumers, delivering personalized offerings, and measuring the personalization impact. These steps were transferred to the automotive context by Hasenjäger, Heckmann, and Wersing [HHW20], where the consumer corresponds to the driver and the offering represents the used driving function. Based on these sources, the definition of personalization in this work is as follows:

Definition 2.11 (Personalization)
Personalization, in the context of this work, is the process of tailoring an ADAS, such as a driving function, to a driver's individual preferences based on certain knowledge about them. The goal of the personalization is to increase driver satisfaction with the system. The process of personalization is an iterative one, consisting of three steps: First, to observe the driver and analyze their driving behavior. Second, to adjust the ADAS based on the gathered knowledge and making it available to the driver. And third, to measure the impact of the chosen personalizations and adjust them again, if necessary. These steps may be executed once, multiple times, or continuously [AT05; HHW20; Tuz09].

Personalization can be achieved in two different ways, either *explicitly* or *implicitly*. The main difference between both personalization types is the Human-Machine-Interface

(HMI) which is used as input channel to gather the driver's preferences. In the context of this work, explicit and implicit personalization are defined as follows:

Definition 2.12 (Explicit Personalization)
Explicit personalization provides the driver with the option to manually and deliberately choose their preferences via a dedicated input channel, such as a settings menu. The used input channel must be one of the HMI input devices which are not required for the primary driving task, e.g., a dedicated settings menu. The driver's intentions are therefore explicitly stated and no deeper interpretation is needed [FP06; HHW20].

Definition 2.13 (Implicit Personalization)
Implicit personalization does not provide the driver with dedicated input channels to select their preferences but instead uses implicit feedback. This implicit feedback is provided indirectly by the driver while executing the primary driving task either fully manually or in cooperation with an active driving function. The used input channels are therefore those HMI input devices which are required for the primary driving task, such as the steering wheel and pedals. The driver's intentions are implicitly provided in the form of their driving behavior, and therefore some form of interpretation is required to derive the driver's preferences [FP06; HHW20].

Implicit personalization has the advantage of allowing for complex driving function adjustments while the driver provides feedback by simply executing the driving task. Disadvantages include the complex interpretation process of the driver's behavior on the system side. A wrong interpretation of the driver's intentions might cause a mismatch between the driver's expectations and the system's behavior, leading to a lower satisfaction and potentially the feeling of control loss [HHW20]. Advantages of explicit personalization include the low interpretation effort on the system side and a higher feeling of control for the driver, since they may explicitly choose each adjusted setting. However, explicitly providing feedback requires the driver's attention and effort, which might discourage them from doing so. Furthermore, the complexity of the possible settings must be limited to avoid overwhelming the driver and to limit the potential complexity of the adjustments [HHW20]. This thesis, as well as most recent research on driving function personalization, focuses on implicit feedback mechanisms.

2.1.6 Driving Style

The main reason for the personalization of driving functions is that individual drivers exhibit different driving styles. These driving styles may differ significantly from person to person [Chu+23; Sag+15; TBAMG04]. Sagberg *et al.* [Sag+15] and Chu *et al.* [Chu+23] conducted literature reviews and summarized common definitions of driving style. Based on their analyses, a driving style is defined as follows in the context of this thesis:

Definition 2.14 (Driving Style)
The driving style describes a generalization of an individual human's driving behavior and habits. This includes both conscious and subconscious decision-making. The driving style may be influenced by the driver's personality and external factors, such as traffic flow, daytime, and road type. It is generally regarded as a relatively stable aspect of the driving behavior, as it is based on habits that develop over time [Chu+23; Sag+15].

In related work, drivers' individual driving behavior is commonly clustered and categorized into different driving styles [Chu+23; Sag+15]. The two main data sources used in these analyses are either questionnaires filled out by the drivers or recorded driving data, e.g., from instrumented vehicles [Chu+23]. The generated categories or clusters may range from simple distinctions between aggressive and defensive drivers [Gel+16; VLZ14] to more complex categorizations, e.g., via the multidimensional driving style inventory [TBAMG04].

Generally, when analyzing the human driving style via recorded driving data, so-called *driving style indicators* are extracted from the driving data and used for further analyses [Sag+15]. These indicators include measures such as the preferred strength of accelerations and decelerations, the general speed choice, the mean Time Headway (THW) and Time to Collision (TTC) to a leading vehicle, and more [BGJ19; CMM02; Gel+16; KMS20; LCB16; MZ21].

2.1.7 Evaluation of Driver Satisfaction

Two of the primary goals of ADASs are the improvement of the driver's safety and comfort. However, many drivers are reluctant to use ADASs due to various reasons, ranging from general distrust toward automation to dissatisfaction with a driving function's specific behavior and driving style [Bel+18; NW23; Oth21]. To address these shortcomings, many studies investigate the trust and acceptance of drivers toward various ADASs [Ade09; BK13; Gao+24; KBB18; Lar+15; Lee+21; MZ21].

Driver trust toward an ADAS is defined as the attitude that the system will support the driver in situations where they are vulnerable [LS04; KBB18]. The driver trust is therefore especially relevant for ADASs that are designed to operate in critical scenarios, such as temporarily intervening systems and driving functions of Level 3 and above. Driver trust is significantly influenced by the user's experience with ADASs, with experienced drivers generally exhibiting higher trust than inexperienced drivers [BK13; Cle+22]. However, the focus of this thesis is a Level 1 PLDF, which is not designed to handle critical driving scenarios. Furthermore, this thesis focuses on the behavior of experienced drivers who are already familiar with ADASs, as discussed in Section 3.1.1. Accordingly, driver trust is not addressed throughout this thesis, but the primary focus is on evaluating the driver satisfaction.

The terms *driver acceptance* and *driver satisfaction* are often used interchangeably in related work. Multiple models exist on the individual acceptance of technology [Dav89; Ajz85; Ven+03], and some of these have also been specifically applied to ADAS usage [Ade09; Rah+17; Rah+18]. Many of these models define acceptance as the behavioral intention to use a system. Depending on the model, this intention of use may depend on the perceived usefulness of the system, the perceived ease of use, the expected performance of the system, the attitude toward the system, and social influences. However, these models rarely provide ready-made questionnaires. Instead, they primarily provide guidance on the topics that questionnaires should address. Consequently, related studies often employ custom-made questionnaires adapted to their specific use case [Ade09; Gao+24; Lar+15; MZ21].

One commonly used questionnaire concerning system acceptance is introduced by van der Laan, Heino, and De Waard [VHD97]. This questionnaire, hereinafter called the *system acceptance scale*, was designed for the assessment of acceptance of advanced transport telematics systems with the goal of being especially simple and concise. The questionnaire consists of nine five-point rating-scale items which consists of two subscales: perceived usefulness and driver satisfaction. The full questionnaire is depicted in Table 2.1. The usefulness subscale consists of the five items with an odd index in the table, whereas the satisfaction subscale consists of the four items with an even index. The usefulness and satisfaction scores are calculated as the mean value of their respective items, whereas the system acceptance score is calculated as the mean across all nine items. In related work, this questionnaire has been commonly used to evaluate the driver acceptance of ADASs [BK13; EBS17; Him+25; MZ21].

Effects of Driver Interventions on Satisfaction

In some related studies, the driver satisfaction is evaluated by analyzing the frequency of driver interventions while using a driving function [GJ23; Lee+21; MZ21; Wan+13]. In these works, it is argued that a driver who is dissatisfied with a driving function's behavior will intervene and adjust the function's behavior accordingly. Therefore, driver dissatisfaction is shown by a high intervention frequency while using a driving function.

Table 2.1: The nine items of the system acceptance scale by van der Laan, Heino, and De Waard [VHD97].

Index	My judgements of the (...) system are... (please tick a box on every line)						
1	useful	_	_	_	_	_	useless
2	pleasant	_	_	_	_	_	unpleasant
3	bad	_	_	_	_	_	good
4	nice	_	_	_	_	_	annoying
5	effective	_	_	_	_	_	superfluous
6	irritating	_	_	_	_	_	likeable
7	assisting	_	_	_	_	_	worthless
8	undesirable	_	_	_	_	_	desirable
9	raising alertness	_	_	_	_	_	sleep-inducing

However, no related study could be found that proves this correlation between driver dissatisfaction and intervention frequency.

2.1.8 Predictive Longitudinal Driving Function

This chapter introduces the driving function that is the focus of the analyses in this thesis. The driving function is a Level 1 PLDF developed by Porsche, which is readily available in production vehicles. The system is officially named *Porsche InnoDrive*. This PLDF was chosen because a subjectively high frequency of driver interventions was identified during expert interviews. Thus, a fully developed driving function that is currently on the market and actively used by customers can be analyzed allowing an investigation of its potential for optimization. The information presented in this chapter is based on the internal Porsche documentation as well as the official owner's manual [Dr.]. The functionality and contained feature set of Porsche InnoDrive is highly similar to the PLDFs of other manufacturers, such as BMW AG's *Assisted Driving Mode* [BMW] or Mercedes-Benz AG's *Active Distance Assist DISTRONIC* [Mer].

The PLDF encompasses state-of-the-art ACC functionality and includes additional features specifically designed for free-driving scenarios. These features include the automatic adoption of the legal speed as the function's set speed and the automatic adaptation of the speed to high-curvature road segments, such as bends, roundabouts, and turns at junctions. The PLDF is designed for operation on well-surfaced country roads and motorways. By utilizing sensor and map data, the PLDF calculates a predictive speed profile for the road ahead, aiming to comply with legal speed limits and to reduce the vehicle speed in road segments where driving at the legal speed would be infeasible, unsafe, or uncomfortable. Additionally, the PLDF ensures a safe distance to a potential leading vehicle by continuously measuring its relative distance and velocity to the ego vehicle. If the leading vehicle is detected to be slower than the ego vehicle, the PLDF decelerates accordingly. This capability is referred to as *ACC functionality* in this thesis, as it aligns with the core focus of traditional ACC systems [PM08; PF10; Var+15; Win+15]. Generally, the PLDF's chosen longitudinal speed follows the formula shown in Equation 2.1:

$$v_{\mathrm{PLDF}} = \min(v_{\mathrm{set}}, v_{\mathrm{curvature}}, v_{\mathrm{ACC}}). \tag{2.1}$$

v_{set} describes the set speed of the PLDF, which mirrors the legal speed if not changed by the driver. $v_{\mathrm{curvature}}$ describes the maximum speed at which a curve, turn, or roundabout can be comfortably traversed, and v_{ACC} describes the speed which ensures a safe distance to the leading vehicle. The speed driven by the PLDF, v_{PLDF}, is then chosen as the minimum of those three velocities.

The PLDF's designated ODD encompasses free-driving scenarios and driving scenarios with a leading vehicle. In these situations, the PLDF is designed to take over the full longitudinal vehicle control. However, all reactions to other traffic besides a leading vehicle fall outside of the function's ODD, and the driver is required to constantly observe the vehicle's surroundings. As soon as the vehicle leaves the PLDF's ODD, the driver is required to take over the vehicle control and handle said situation. Examples for situations that fall outside of the PLDF's ODD are slowing down at red traffic lights, giving way to crossing traffic in junctions and roundabouts, as well as maneuvers like overtaking, changing lanes, and merging.

The PLDF is controlled via a dedicated ADAS control stalk mounted on the steering column. This control stalk can be used to activate and deactivate the PLDF, as well as to adjust the PLDF's set speed and the desired distance to the leading vehicle. The following subsections describe how drivers can interact with the PLDF in detail.

Manual Set Speed Adjustments

The PLDF calculates a predictive speed profile for the upcoming road within a defined look-ahead distance and always plans to comply with the legal speed limit. The so-called *set speed* of the PLDF refers to the desired speed at which the vehicle is intended to drive on a straight road segment. This set speed is generally set automatically to the detected legal speed limit. The legal speed limit is obtained from both the map data, using Global Positioning System (GPS)-based localization, and a camera-based Traffic Sign Detection (TSD) system. In case of discrepancies between the map data and TSD information, the TSD is prioritized.

The driver may manually adjust the set speed using the ADAS control stalk. The resulting deviation of the set speed from the legal speed is referred to as a *set speed offset*. The offset remains active only until a new legal speed limit is reached, at which point it is discarded and the set speed is set to the new legal speed. Accordingly, the driver can modify the set speed of the PLDF on straight roads only temporarily. In this thesis, the process of adjusting the set speed of the PLDF is called a *set speed adjustment*.

Definition 2.15 (Set Speed Adjustment)
The term set speed adjustment *refers to the manual modification of the set speed of the PLDF or similar longitudinal driving functions. In the context of the PLDF, the driver may either voluntarily adjust the current set speed due to deviating personal preferences, or they may serve to correct an incorrectly chosen set speed of the PLDF, e.g., due to an erroneous TSD or incorrect map data.*

Driver Interventions via the Gas and Brake Pedal

The PLDF's operation principle falls under the categorization of traded control systems. When the PLDF is active, it takes over the full longitudinal vehicle control. Therefore, as long as the PLDF stays within its defined ODD, no longitudinal driver inputs are required. However, the driver may always decide to take over the longitudinal vehicle control via a driver-initiated takeover. This deactivates the driving function and the driver is in full control of the longitudinal vehicle control again. The vehicle control is therefore traded but never shared.

The status of the PLDF describes whether the function is currently active or whether the driver controls the vehicle. The underlying state machine of the PLDF status is depicted in Figure 2.2. When the driver activates the PLDF via the ADAS control stalk, the PLDF status is set to *active*. The PLDF can then be deactivated by using the control stalk again or by pressing the brake pedal. Deactivating the function using the control stalk is also called a *cancel*. When deactivated, the PLDF status changes to *standby* and the driver controls the vehicle longitudinally again. To return to the active status, the driver must manually resume the PLDF's execution via the control stalk. When the driver presses the gas pedal stronger than the PLDF intends to accelerate, the PLDF status changes to *passive* and the vehicle accelerates accordingly. As soon as the driver releases the gas pedal again, the PLDF status automatically changes from passive to active. The passive status therefore refers only to a temporary deactivation of the PLDF as long as the driver presses the gas pedal. Besides using the gas and brake pedal, the driver can also interact with the PLDF by increasing or decreasing its set speed via the control stalk. However, this does not affect PLDF's current status.

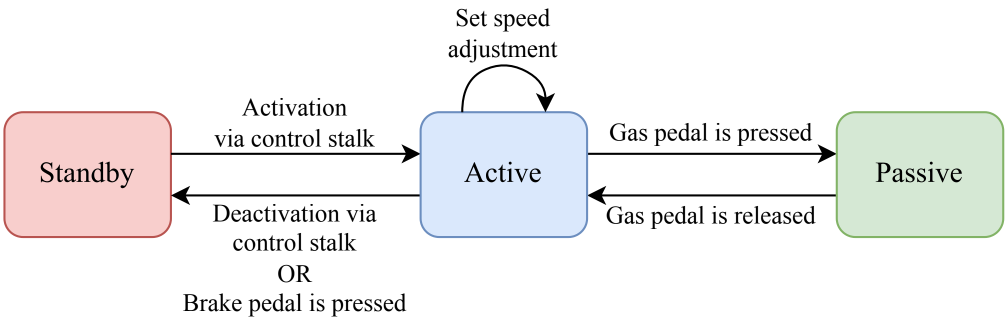


Figure 2.2: Simplified schematic of the PLDF's state machine. Depending on the driver action, the PLDF switches between the three possible states.

As explained, the driver may interact with the PLDF in two primary ways. First, by adjusting the system's set speed, which does not influence the PLDF status. Second, by pressing the gas or brake pedals, which automatically deactivates the PLDF until the driver returns the vehicle control to the system. In the context of traded control systems, such pedal interventions are therefore considered to be driver-initiated takeovers.

Definition 2.16 (Pedal Intervention)
In the context of this work, a pedal intervention *is defined as a short-term driver-initiated takeover of vehicle control via either the gas or the brake pedal. Pedal interventions occur if the driving function is active and the driver briefly intervenes to adjust the function's behavior in a specific situation, after which the function is reactivated. Pedal interventions can be either voluntary or mandatory. Voluntary pedal interventions are based on deviating personal preferences by the driver in certain situations. Mandatory pedal interventions are required either to handle situations outside of the PLDF's ODD, or to correct potential function errors. In the context of this work, the manual deactivation of the PLDF via the control stalk is also considered a pedal intervention, since it has the same effect as briefly pressing the brake pedal.*

Both ways of interacting with the PLDF are then summarized under the general term of a *driver intervention*.

Definition 2.17 (Driver Intervention)
In the context of this work, a driver intervention is a short-term interaction by the driver with an active driving function with the goal to adjust the function's behavior in a specific situation. This categorization includes both pedal interventions and set speed adjustments.

2.2 Research on Driver-Initiated Takeovers

In the field of ToCs, both system-initiated and driver-initiated takeovers are the topic of recent research. Although Gershon *et al.* [Ger+21] and Morando *et al.* [Mor+20] argue that driver-initiated takeovers are more common during real-world driving, most research focuses on system-initiated takeovers [BBT24; Ger+21; Mor+20]. The research topic of system-initiated takeovers is particularly relevant for driving functions of Level 3 and above, as drivers may fully disengage from the vehicle control and perform secondary tasks [MA+20; Sae]. Related research therefore focuses on the driver's readiness for a takeover [BRK17], the design of system-initiated takeover requests [MA+20], and the driver's takeover performance, e.g., the time needed to take over the vehicle control again [ES17; PGR23; Ryd+23].

However, driver-initiated takeovers are rarely analyzed compared to system-initiated takeovers [Ger+21; Mor+20]. Recent research on driver-initiated takeovers focuses on topics such as their impact on vehicle dynamics and traffic flow [PM08; PF10; Var+15], the influence of the driver trust on the likelihood of intervening in potentially critical scenarios [BBT24; OF+21], and the classification of different types of driver-initiated takeovers [Ger+21; PM08; PF10; Yan+23]. This section highlights research that specifically examines the classification of driver-initiated takeovers and the underlying reasons behind them.

2.2.1 Types of Driver-Initiated Takeovers during Adaptive Cruise Control

In 2006, the Dutch Road Authority Rijkswaterstaat conducted a field operational test in the Netherlands to evaluate the effects of driving with ACC and lane departure warning systems [ABH07]. During the field operational test, 19 participants drove vehicles equipped with these systems over a period of five months, while recording the driving data via data loggers. The results showed that ACC was most frequently used on highways, where it was active for more than 40 % of the total driving time. On secondary roads, ACC was used 22 % of the time, whereas in urban areas, it was only used $4\,\%$ of the time. The main findings of the study indicate that the used ADASs improved traffic safety, reduced fuel consumption, and increased driver comfort.

The dataset recorded in this field operational test was analyzed in multiple other works. Pauwelussen and Minderhoud [PM08] and Pauwelussen and Feenstra [PF10] used the dataset to analyze the drivers' behavior while using ACC. They especially analyzed the drivers' motives to manually take over the vehicle control. The authors extracted segments of driving data around each ToC in the dataset where the vehicle speed was within a range of $70\,\mathrm{km/h}$ to $130\,\mathrm{km/h}$. By analyzing the extracted segments, they found that approximately $80\,\%$ of driver-initiated takeovers were initiated with the gas pedal and only $20\,\%$ were conducted with the brake pedal. The authors then deduced four main reasons for driver-initiated takeovers by analyzing multiple features from the driving data, such as the THW and the relative speed to the leading vehicle.

First, in a *speed adaptation prior to a lane change*, drivers may either overtake a leading vehicle or merge onto another lane. Second, in a *speed adaptation to avoid overtaking by ACC on the wrong lane*, drivers may deactivate the ACC to avoid illegally overtaking another vehicle. Third, *overruling due to offensive or defensive behavior* is described as accelerating or braking in response to another vehicle attempting to merge in front of the ego vehicle. In such cases, the driver may seek to create either a sufficient or an insufficient gap for the other vehicle to merge into. Finally, *reaching the system support boundaries in a safety-critical situation* describes a situation in which the ACC's maximum acceleration or deceleration capacity are insufficient to avoid a potentially hazardous situation, prompting the driver to intervene, e.g., in case of a sudden and harsh deceleration of the leading vehicle [PM08; PF10].

2.2.2 Types of Driver-Initiated Takeovers during Level 2 Driving

Gershon *et al.* [Ger+21] and Yang *et al.* [Yan+23] both analyzed and categorized different types of driver-initiated takeovers during naturalistic driving with level 2 driving functions. As their database, they used the Massachusetts Institute of Technology (MIT) Advanced Vehicle Technology (AVT) study dataset [Fri+19], which is a publicly available dataset containing naturalistic real-world driving data of human drivers using SAE Level 1 and 2 driving functions. Their findings and the used dataset are described in the following section.

Massachusetts Institute of Technology Advanced Vehicle Technology Study Dataset

The MIT AVT study [Fri+19] aims to collect naturalistic driving data using in-production ADASs from multiple different vehicle manufacturers. The dataset is publicly available and, at the time of writing this thesis, the study is still ongoing and the recorded dataset continues to grow. At that time its official paper was released in 2019, the MIT AVT dataset contained recorded drives from 122 different drivers across 29 different vehicles over a time period of 37 months. The aggregated dataset thus featured over $823{,}000\,\mathrm{km}$ of naturalistic driving, spanning over the entire continental United States of America. The used vehicle fleet consisted of 23 Tesla Model S and Model X vehicles, two Volvo S90 Series, two Range Rover Evoque, and two Cadillac CT6. Each vehicle was equipped with custom data logging devices that recorded the data of three to six inward- and outward-facing cameras, GPS data, Inertial Measurement Unit (IMU) data, and raw vehicle bus data. However, access to the vehicle bus data is partially restricted due to unavailable information on how to decode many of the signals.

Classification of Driver-Initiated Takeovers based on the Hierarchical Model of the Driving Task

Gershon *et al.* [Ger+21] used the MIT AVT dataset to analyze the ToCs during naturalistic driving with Cadillac Super Cruise (SC). Cadillac SC is a hands-free level 2 driving function designated for use on compatible limited-access multi-lane highways only. The driving function encompasses ACC functionality to ensure a safe distance to the leading vehicle and performs active lane centering. In order to activate SC, ACC must first be active and then the driver may switch to SC via an additional button press. SC will perform automatic lane changes to overtake slower leading vehicles, if possible. However, the driver may also manually change lanes, which temporarily deactivates SC and reverts the function to ACC. All situations outside of the above-explained highway driving fall outside of the function's ODD [Cad].

The authors used the driving data of 14 participants over the course of one month each. During that time, the participants drove a cumulated distance of over $35{,}000\,\mathrm{km}$ and used the ADASs with which the vehicles were equipped on a voluntary basis. In the dataset, SC was used on approximately $40\,\%$ of the driven distance on highways. The authors used the automation status, which can be read from the vehicle bus signals, to automatically detect driver- and system-initiated takeovers in the recorded data. They first analyzed the general number and types of ToCs and then focused on the analysis of full driving function disengagements from level 2 to manual driving.

In total, the authors found 2610 driver- or system-initiated takeovers, out of which $84\,\%$ were driver-initiated and $16\,\%$ were system-initiated. The authors attribute the driver-initiated takeovers to risk mitigation, execution of functions beyond the ODD, and to deviating driver preferences in how the driving task should be executed. Figure 2.3 depicts the number of found driver-initiated takeovers between the three automation levels. As can be seen, $44\,\%$ of driver-initiated takeovers caused a reversion from SC to ACC, mainly due to manual lane changes and gas pedal interventions, and only $17\,\%$ of driver-initiated takeovers fully disengaged SC directly back to Level 0 via a brake pedal press or a manual button press. The remaining $39\,\%$ of interventions occurred while only ACC was active.

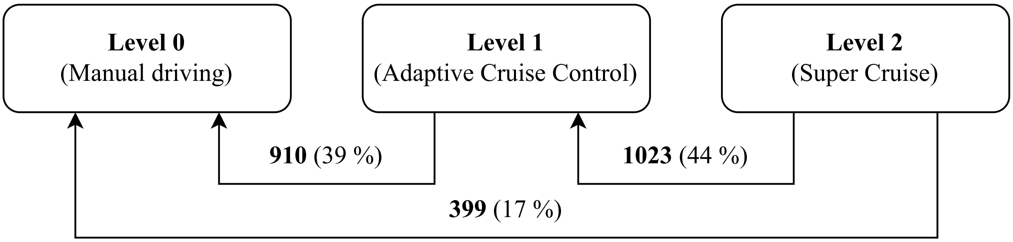


Figure 2.3: Number of driver-initiated takeovers during Cadillac SC usage between the three possible levels of automation (adapted from [Ger+21]).

The authors then chose the 399 full automation disengagements from level 2 to Level 0 for a deeper analysis. The authors used the hierarchical model of the driving task by Michon [Mic85] to categorize each full automation disengagement as either a *strategic ToC*, *maneuver ToC*, or *control ToC*, depending on whether the disengagement involved long-term planning, tactical maneuvers, or immediate vehicle control actions. Using this taxonomy, two trained annotators separately labeled each of the 399 full disengagements accordingly. The following distribution was found for the different ToC types: strategic ToCs: $31.6\,\%$, maneuver ToCs: $11.8\,\%$, and control ToCs: $56.4\,\%$. Most interventions are of the control type, where the driver briefly reacts to an environmental input, while maneuver ToCs are relatively rare. The authors state that many maneuver ToCs, especially overtaking and merging interventions, only reverted the automation back to Level 1, as they were mostly executed via the gas pedal or manual steering. Thus, these interventions were not annotated.

Hierarchical Clustering of Driver-Initiated Takeovers

Yang *et al.* [Yan+23] also utilized the MIT AVT dataset to analyze and categorize ToCs in the naturalistic driving data of eight Cadillac SC and eight Tesla Auto Pilot (AP) drivers. Tesla AP is a Level 2 driving function with an ODD comparable to that of Cadillac SC.

The authors generated a dataset for their analyses by automatically extracting driver-initiated takeovers using the ADAS status from the vehicle bus data. For SC, 214 takeovers were extracted, and for AP, 525 takeovers were identified. The extracted dataset primarily focuses on highway driving. For each takeover, features regarding the vehicle kinematics and contextual information were extracted within a fixed time frame before and after each intervention. The vehicle kinematics features include information such as the vehicle speed, acceleration, and steering wheel angle. The contextual features consist of information such as road type, relative distance to the leading vehicle, distance to the next highway exit, and upcoming congestion. Only the vehicle kinematics features were then used to perform agglomerative hierarchical clustering, which identified 5 different types of takeovers. The contextual features were subsequently used to analyze the found clusters in greater detail.

With 467 instances, the so-called *normal takeovers* are the most common type. According to the authors, these driver-initiated takeovers are characterized by small changes in vehicle speed and steering control. The second largest cluster is composed of 171 *braking takeovers*. These takeovers feature high decelerations, commonly before highway exits or congestion. The 54 *accelerating takeovers* feature high accelerations often in the context of a slower leading vehicle and lane changes. Overtaking maneuvers therefore likely fall into this category. Lastly, 32 so-called *evasive maneuver takeovers* were found with strong decelerations and high steering. These interventions likely contribute to hazardous situations. Additionally, for AP only, 14 *right-swerve takeovers* were observed. These interventions are thought to be exclusive to AP as it can also be operated outside of highways. The authors state that the normal and braking takeovers primarily occurred during high-speed highway driving, while accelerating, evasive maneuver, and right-swerve takeovers were more common during slow car-following scenarios.

2.2.3 Summary and Limitations

Only a few studies could be found which analyze the different types of driver-initiated takeovers with Level 1 and 2 driving functions during naturalistic driving [Ger+21; PM08; PF10; Yan+23]. These studies primarily focus on multi-lane highway driving. During the analyzed scenarios, driver-initiated takeovers were found to be an integral and common aspect of naturalistic driving function usage. They were also found to be significantly more common than system-initiated takeovers during level 2 driving [Ger+21].

For ACC driving, most interventions were found to fall outside of the driving function's ODD. The identified reasons for driver initiated takeovers were mainly focused on manual lane changes, merging, and safety-critical scenarios [PM08; PF10]. For the categorization of driver-initiated takeovers during level 2 driving, two different approaches were highlighted. Gershon *et al.* [Ger+21] manually annotated full function disengagements and assigned them to one of three categories based on the hierarchical model of the driving task by Michon [Mic85]. As potential reasons for the takeovers, risk mitigation, the execution of functionality beyond the system's ODD, and deviating driver preferences were named. However, these takeover reasons were not directly linked to the derived categories [Ger+21]. Yang *et al.* [Yan+23] also categorized driver-initiated takeovers during level 2 driving. They conducted a hierarchical clustering of vehicle dynamics features extracted from the interventions. Their analysis identified five categories of driver-initiated takeovers, which are primarily distinguished by velocity and steering changes during the takeovers [Yan+23].

The highlighted works of research therefore provide a first coarse top-down analysis of the different types of driver-initiated takeovers. However, this also represents their main limitation. The conducted classifications are solely based on post-drive analyses of the driver-initiated takeovers. However, no ground truth labels from the drivers are provided regarding their reasons for intervening in the respective scenarios. This missing granular ground truth data limits the depth of the conducted analyses to coarse top-down approaches. No granular analysis of the drivers' reasons behind each takeover could be found.

2.3 Driving Function Personalization Approaches

It is generally accepted that driving styles differ across individual drivers [Sag+15; TBAMG04]. Therefore, many works focus on the personalization of driving functions in order to increase the driver satisfaction and comfort with the system [HHW20; PC19; Yi+20]. In this section, recent research on the personalization of driving functions is highlighted. As introduced in Section 2.1.5, ADAS personalization can be distinguished into explicit and implicit approaches. Most of the recent research focuses on implicit personalization approaches, as they allow for more complex adjustments by automatically deriving the driver's preferences from their driving behavior. These implicit personalization approaches can be found for almost all types of ADASs. This includes longitudinal and lateral driving functions, temporarily intervening systems, and information and warning systems [HHW20; PC19; Yi+20]. For this thesis, however, the research focused on longitudinal driving functions is especially relevant and will be highlighted in this section.

Most research in the field of longitudinal driving function personalization focuses solely on ACC driving with an active leading vehicle [BSDP08; BGJ19; CMM02; Che+17; Gel+16; GJ23; HHW20; KMS20; LCB16; Wan+13; Yi+20]. Only few authors investigate the personalization of additional functionality. For example, Kuderer, Gulati, and Burgard [KGB15] additionally focus on personalized lane changes in their system, while Ramyar *et al.* [Ram+17] investigate lane changes and free-driving behavior while their ACC is active.

2.3.1 General Personalization Approach

As described in Section 2.1.5, Hasenjäger, Heckmann, and Wersing [HHW20] define the general steps of the implicit personalization process as follows: First, the driving behavior is observed and recorded. Second, a model is built from the human driving behavior. And third, the model's performance is validated. If necessary, this process may be conducted iteratively to continuously adjust and optimize the driving function according to the driver's wishes. These three steps resemble the general personalization process described by Adomavicius and Tuzhilin [AT05].

Implicit personalization approaches are generally data-driven, i.e., they require some form of database to derive the driver's preferences from. As explained in Section 2.1.5, the input channels used in implicit approaches are the HMI input devices which are also required for the primary driving task, such as the steering wheel and the pedals.

Therefore, recordings of manual driving data are commonly used as a database for the implicit derivation of driver preferences [HHW20; Yi+20]. This data can either be recorded in instrumented vehicles during naturalistic driving [BSDP08; CMM02; KMS20; KGB15; LCB16; Ram+17; Wan+13], or it can be recorded in a human-in-the-loop simulator setting [BGJ19; Che+17; GJ23].

2.3.2 Driver Models

Based on the recorded manual driving data, so-called *driver models* are created which model the human driving behavior. These driver models are often used as part of the vehicle controller of the personalized driving function [HHW20; WXC14]. Wang, Xi, and Chen [WXC14] describe the different types of driver models in recent research and summarize them into three categories: parametric, non-parametric, and semi-parametric models.

Parametric models use mathematical formulas to describe the driving behavior. The parameters in the formulas are derived from the recorded manual driving data of individual drivers. Often, the parameters of the models are fitted via, e.g., a least squares regression of the model's outputs and the recorded manual driving data [BSDP08; BGJ19; Wan+13]. Other authors prefer to derive physical model parameters by analyzing driving style indicators, such as preferred accelerations, THW, and TTC [KMS20]. Furthermore, group-based approaches exist that aim to personalize the driving function by assigning drivers into clustered groups based on their driving style. The driver model's parameters are then taken from the predefined groups into which the individual drivers fall [CMM02; Gel+16]. Non-parametric models are black boxes which commonly use machine learning approaches to model driver behavior. Examples of the used algorithms include Reinforcement Learning (RL)-based Q-Learning [Che+17], hidden Markov models [LCB16], and random forests [Ram+17]. And finally, semi-parametric models combine both parametric and non-parametric models as subsystems into one combined model.

2.3.3 Evaluation of the Model Performance

The final step of driving function personalization is the evaluation of the created model's performance. According to the previously established definition of personalization, this should be done by gathering feedback from the driver while they use the personalized driving function. However, in most related work, the driving behavior of the personalized model is compared only to the recorded manual driving data. This approach is based on the assumption that drivers generally prefer a driving style which is similar to their own [HHW20]. This assumption is discussed in more detail in Section 2.4.

The most common approach to evaluate the personalized model's performance is the comparison of the model's generated outputs to the recorded manual driving data. To generate the personalized model's output, the recorded driving data is often replayed in an offline simulation. There, the leading vehicle is replayed exactly as it was recorded during the manual driving sessions. Then, the personalized model is run offline and its simulated driving behavior can be compared to the manual driving. Most commonly, the mean squared error of the ego vehicle's spacing gap, velocity, or acceleration is calculated between the recorded manual driving and the personalized model's output [BSDP08; CMM02; Che+17; KMS20; Ram+17]. Other authors claim that driving style indicators are better suited for the comparison [BGJ19; LCB16]. In most cases, it is argued that the found error values are low, indicating a good model performance. However, the calculated metrics are often not compared to other models and therefore it cannot be stated if the error is truly low.

Only a few authors compare their found metrics to other models. Chen *et al.* [Che+17] claim that they simulated a traditional ACC system in addition to their personalized ACC and compared the model performances to the recorded manual driving data. They report a significantly lower mean squared error of the ego velocity and the gap spacing, indicating a better imitation of the manual driving style of their model compared to the traditional ACC. Another example is provided by Lefèvre, Carvalho, and Borrelli [LCB16], who use the manual driving data of five different drivers in their experiments. For each driver, they trained a personalized model on the driver's individual driving data, and an average model on the data of the four remaining drivers. They then showed that their personalized model performs better than the average model at imitating the driving style of the individual drivers.

The research of Wang *et al.* [Wan+13] is the only work which was found that evaluates the personalized driving function in a real vehicle during naturalistic driving on highways. The authors personalized an ACC based on recorded manual driving data from a first field test. Their personalized driving function is then successfully used in a second field test in real traffic scenarios where it is compared to the drivers' manual driving behavior and a standard ACC. The authors state that the drivers filled out a questionnaire after the field test, however they do not share the questionnaire results and they do not state how many drivers participated in the second field test. Their analyses show that the personalized ACC's driving style resembles the driving style of an example driver more closely than the standard ACC. They also show that their personalized ACC reduces the frequency of brake pedal interventions compared to the standard ACC and therefore argue that their system has increased acceptability.

Another work which uses questionnaires to evaluate the driver's satisfaction is presented by de Gelder *et al.* [Gel+16]. There, the authors do not directly assume that the manual driving style of drivers also reflects their preferred automated driving style with ACC. Instead, they chose to conduct a test group study with 20 participants using a freely parameterizable ACC. They performed multiple gap closing maneuvers with different parameterizations and let the drivers evaluate how well they liked each maneuver with a questionnaire. Finally, the drivers also performed manual gap closing maneuvers to demonstrate their preferred manual driving style. Based on the recorded manual drives and the evaluations of the ACC settings from the questionnaires, the authors created driving style clusters and trained an automatic classifier on the data. With this classifier, they are able to derive the preferred ACC parameters from a driver's recorded manual driving. However, the resulting personalized driving function is not tested in a test group study again afterward.

2.3.4 Continuous Driving Function Adjustment

The ideal personalization process is described as a continuous cyclical process, where after the evaluation of the model performance, another adjustment cycle may begin to further optimize the created driver model [AT05; HHW20]. However, most related work only focuses on a one-shot adjustment of the driving function based solely on recordings of manual driving. Only few works on ACC personalization also mention the need for continuous updates [BSDP08; Che+17; GJ23; HHW20; LCB16; Wan+13].

Both Wang *et al.* [Wan+13] and Bifulco, Simonelli, and Di Pace [BSDP08] describe that inconsistent driving behavior is a natural limitation of personalized driving functions. Therefore, both introduce a personalized ACC which is able to learn online while the driver is manually controlling the vehicle. In the case of Wang *et al.* [Wan+13], the driving function is continuously updated while the driver controls the vehicle, and as soon as the driver hands over control to the ACC, the currently learned parameters are used. On the other hand, Bifulco, Simonelli, and Di Pace [BSDP08] propose to recalibrate the driving function only on demand. Their personalized ACC therefore features a learning mode which can be manually activated. While the learning mode is active, the driver may demonstrate their preferred driving style, which is then imitated when the function is activated afterward. The authors of both works state multiple reasons why such a recalibration might be needed, such as changing environmental and road conditions, or a driver's preferences which change over time. The inconsistency of human demonstrations during manual driving is also stated as a limitation of personalized driving functions in other work [KGB15].

Another reason why continuous updates of the personalized driving function might be needed is described by Lefèvre, Carvalho, and Borrelli [LCB16]. At the end of their paper, they briefly mention that they conducted first tests of their personalized driving function in a real vehicle. There, they realized that especially drivers with an aggressive manual driving style would often prefer a more defensive automated driving style instead. Therefore, they state that continuously updating their personalized ACC, while the driver already uses it, might be beneficial instead of only training the model based on recordings of manual driving. Similar results are also found by de Gelder *et al.* [Gel+16]. In their analysis, which compared manual driving styles to preferred ACC parameterizations, they found that most drivers preferred a slightly more defensive automated driving style.

The work of Guo and Jia [GJ23] is the only study found that continuously and iteratively updates the driving function based on feedback provided by the driver while using it. This feedback is provided in the form of driver interventions. The authors argue that drivers express their dissatisfaction with the driving function by taking over the vehicle control to adjust function's behavior. Therefore, an Inverse Model Predictive Control (IMPC) approach is proposed that adjusts its parameters based on recorded driver interventions during ACC car following. Initially, the IMPC is configured to mimic the behavior of an already established generic driver model. The driver is then placed in a driving simulator to experience the driving function's behavior, where they may intervene and adjust the vehicle's velocity if they feel uncomfortable. The interventions are recorded and the IMPC's parameters are updated after each drive based on the intervention data. The driver then uses the adjusted driving function again, iteratively updating it via their interventions. The approach was tested with an aggressive and a conservative driver in a driving simulator. For both drivers, the intervention frequency decreased to 0 % after two iterations, indicating a successful personalization.

2.3.5 Summary and Limitations

Many different personalization approaches for longitudinal driving functions were highlighted in this section, and they generally follow the personalization process described in Section 2.1.5. All found personalized longitudinal driving functions primarily focus on ACC driving with an active leading vehicle. The goal of the created driver models is generally to imitate individual drivers' prerecorded manual driving data as closely as possible.

The highlighted personalization approaches feature multiple limitations. The first limitation is their reliance on the assumption that drivers prefer their manual driving style also as their automated driving style. However, some authors mention that this statement might not hold true [Gel+16; GJ23; LCB16]. Another limitation mentioned by some authors is the inconsistent driving behavior of human drivers, which can be difficult to learn from when using generalized policies [BSDP08; KGB15; Wan+13]. The final limitation is the limited evaluation of the personalized driving functions. Generally, the goal of driving function personalization is to increase the driver's satisfaction with the system. However, the proposed personalized driving functions are rarely driven and evaluated by the study participants after the personalization process again. Due to this lack of driver feedback, the personalized driving functions are also rarely continuously updated, although the need for such updates is mentioned by several authors [BSDP08; Che+17; GJ23; HHW20; LCB16; Wan+13]. Only one approach could be found that features such continuous updates based on driver feedback [GJ23]. There, driver interventions are used as feedback while the driving function is currently used. These interventions are then iteratively used to update the driving function until the driver is satisfied with its behavior.

2.4 Preferred Automated Driving Style

As explained in Section 2.3, most research on personalized driving functions is based on the assumption that drivers prefer their manual driving style also as their automated driving style. However, some authors mention a discrepancy between the manual driving style and the preferred automated driving style [Gel+16; GJ23; LCB16]. Therefore, this section highlights recent research on preferred automated driving styles in order to analyze whether drivers truly prefer their manual driving style also as their automated driving style, or whether a more defensive driving style is generally preferred.

In related work, many test group studies can be found which analyze the participants' manual driving styles and then compare them to their preferred automated driving styles. Most commonly, driving simulators are used in these studies [Bas+17; Gri+16; Lee+21; MZ21; Pen+24]. However, a few authors also conducted field tests in real vehicles [Gel+16; Vas+23; Yus+16]. The results of these test group studies are mixed. Many studies find that drivers generally prefer a more defensive driving style than their own manual driving style [Bas+17; Pen+24; Yus+16], while other works argue that drivers prefer an automated driving style close to their own [Gel+16; Gri+16; MZ21; Vas+23]. However, some of the studies which find that drivers prefer an automated driving style close to their own also argue that there is still a slight tendency toward more defensive driving styles [Gel+16; Gri+16], and that very aggressive automated driving is generally disliked by all drivers [Gri+16]. The analyzed studies are briefly discussed in the following paragraphs.

2.4.1 Test Group Studies in Driving Simulators

Most of the reviewed test group studies are conducted in driving simulators [Bas+17; Gri+16; Lee+21; MZ21; Pen+24]. Two similar studies were conducted by Basu *et al.* [Bas+17] and Griesche, Dotzauer, and Käthner [Gri+16], with the former focusing on diverse driving scenarios on an extensive simulation map, and the latter solely focusing on overtaking maneuvers of a slower leading vehicle on a two-lane highway. Both studies were split into two sessions. In the first session, the manual driving behavior of each participant was recorded in the relevant scenarios. In a second session on a different day, drivers experienced four prerecorded driving styles, replayed as automated driving styles of a driving function, one of which corresponded to their own behavior. The drivers were then asked to evaluate which of the experienced automated driving styles they prefer the most. Basu *et al.* [Bas+17] additionally gave the participants the option to evaluate how well the experienced driving styles resembled their own manual driving. They found that participants generally preferred a significantly more defensive automated driving style than their own manual driving. Interestingly, they also found that most participants preferred the driving style which they thought was their own. However, there was little correlation to their actual own driving style. Griesche, Dotzauer, and Käthner [Gri+16], on the other hand, found that for the analyzed overtaking maneuvers, drivers generally preferred a driving style which was similar to their own. The individual drivers rated their manual driving style not always as the best, but positive on average. However, the authors also mention a slight tendency toward more defensive driving styles and state that too aggressive and dangerous driving styles were disliked by all participants.

Ma and Zhang [MZ21] and Peng *et al.* [Pen+24] both did not record and analyze the manual driving behavior of their participants, but instead used questionnaires to classify their manual driving style as either aggressive or defensive drivers. They then let each participant experience a predefined aggressive and defensive driving style in their simulation environment in various driving scenarios. In the study by Ma and Zhang [MZ21], participants evaluated their trust and acceptance of each automated driving style via multiple questionnaires. Participants could additionally manually take over the vehicle control if they wished to. Their analyses show that participants rated the driving function with their respective driving style significantly better in both trust and acceptance. They also found that the driver-initiated takeover frequency is significantly lower if their manual driving style and the automated driving style align. Peng *et al.* [Pen+24], on the other hand, found that the defensive automated driving style is more comfortable for all drivers. They additionally let drivers evaluate the perceived naturalness of each driving style. Interestingly, they found that participants rated the driving style which was more similar to their own manual driving style to be more natural, although the more defensive driving style was generally preferred in terms of driving comfort.

Lee *et al.* [Lee+21] focused their analysis on the driver trust toward the driving function in intersection scenarios. In their study, the participants first drove manually through the different intersection types and then they experienced a driving function with a conservative, a moderate, and an aggressive automated driving style. Based on driving style indicators, they computed the similarity of the manual driving to the automated driving style. The driver trust was measured based on how often they perform pedal interventions. This approach is based on the assumption that drivers who do not trust the driving function commonly intervene in the active system. Their analysis then shows that the used driving style had no effect on the general intervention frequency. However, with the conservative driving style, drivers more commonly pressed the gas pedal, while the aggressive driving style prompted more brake pedal interventions. The authors treat both intervention types as the same for their trust analysis.

2.4.2 Test Group Studies in Real Vehicles

Three test group studies were found which conduct field tests to evaluate which automated driving style is preferred compared to the participants' manual driving styles [Gel+16; Vas+23; Yus+16]. The test group study conducted by de Gelder *et al.* [Gel+16] was already explained in Section 2.3. To summarize, their study focuses on ACC gap closing scenarios, where they first let each study participant demonstrate their manual driving behavior and then they let the drivers experience and evaluate different ACC parameterizations. For both parts of the study, the participants are seated in the driver seat of a real vehicle. Their analysis found that most drivers generally prefer an ACC gap closing driving style close to their own. However, there is a trend toward slightly more defensive automated driving styles.

Vasile *et al.* [Vas+23] and Yusof *et al.* [Yus+16] both also conducted test group studies in real vehicles. However, in both studies, the driver is not seated in the driver seat when experiencing the automated driving style. In the test group study by Vasile *et al.* [Vas+23], a Level 3 driving function is tested on German highways. Participants first demonstrated their manual driving style on the highway while sitting in the driver seat. Based on driving style indicators extracted from the recorded data, each driver's manual driving style was derived. Then, the participants were seated in the passenger seat, where they experienced a defensive driving function while a safety driver was seated in the driver seat. The perceived driving comfort was subsequently evaluated via questionnaires. The results show that drivers with defensive manual driving styles generally rated the driving comfort as high with the defensive driving function, while aggressive drivers gave low comfort ratings.

Yusof *et al.* [Yus+16] conducted a test group study in a real vehicle with a *Wizard of Oz* setup, i.e., the car is driven manually by a driver but the participant is told that the vehicle drives autonomously. The manual driving styles of the participants were assessed via a questionnaire based on which they were categorized as either aggressive or defensive drivers. Then, the participant was seated in the back seat of the vehicle and multiple different scenarios were driven, each with an aggressive, a defensive, and a very defensive driving style. The drivers evaluated the experienced driving styles using a questionnaire. The results show that all drivers, no matter their manual driving style, preferred the defensive automated driving style over the very defensive and aggressive driving styles.

In both cases, it is unclear how well applicable the results of the test group studies are to naturalistic driving, since the driver is not seated in the driver seat. However, it is generally accepted that the driving experience may significantly differ depending on where a human is seated in the vehicle [Har+22; IMW20; Itt+23].

2.4.3 Summary and Limitations

As was explained in this section, related work is not conclusive on whether drivers prefer their own manual driving style also as their automated driving style. However, most studies find at least a tendency toward more defensive automated driving styles [Bas+17; Gel+16; Gri+16; Pen+24; Yus+16].

The main limitations of the proposed studies are the rare use of field tests with the participants in the driver seat, and the often limited number of analyzed scenarios. Most highlighted studies were conducted either in driving simulators [Bas+17; Gri+16;

Lee+21; MZ21; Pen+24], or if the participant was seated within a real vehicle, they were rarely seated in the driver seat [Gel+16] but in the passenger seat [Vas+23] or the back seat [Yus+16] instead. While many studies focus on diverse driving situations [Bas+17; MZ21; Vas+23; Pen+24; Yus+16], others only analyze one specific driving situation, such as gap closing scenarios [Gel+16], overtaking maneuvers [Gri+16], and intersection scenarios [Lee+21].

2.5 Interactive Imitation Learning Approaches

Imitation Learning (IL) falls into the field of machine learning and focuses on learning a task from human demonstrations. Its main ideas are based on the human teaching process, where an expert teacher demonstrates how a task is solved and a learner tries to copy the expert's behavior. The approach of directly imitating the manual demonstrations of humans is called *Behavioral Cloning (BC)*, however, many different approaches exist in the field of IL. Especially interesting for this thesis are IIL approaches. These approaches focus on receiving human feedback intermittently while the learner is currently executing the to-be-learned task, allowing for an online improvement of the robot's behavior. For this thesis, the main relevant type of IIL approaches are based on absolute feedback in the state-action transition space. Here, the human supervisor provides demonstrations of the optimal behavior in the current situation during the execution of the task [Cel+22]. This way of providing feedback therefore resembles how drivers intervene during the use of driving functions.

The following sections explain relevant approaches which use absolute corrections in the state-action transition space. Since these approaches are closely related to BC and build upon it, BC is additionally explained, even though it is not technically an IIL approach [Cel+22].

2.5.1 Behavioral Cloning

BC [BS95] is considered the most traditional IL approach. This approach uses a prerecorded dataset of expert demonstrations as its database. The contained state-action pairs are then learned by training, e.g., a neural network, using supervised learning methods. The agent therefore learns to imitate the expert's behavior in the situations covered in the training dataset. The primary steps of BC are shown in Algorithm 1. As BC learns from offline data instead of human feedback, it is not considered an IIL approach. However, many IIL approaches build upon the foundations of BC [BS95; Cel+22; Spe+22].

Algorithm 1 Behavioral Cloning [BS95; Cel+22]

Collect expert demonstrations: Record expert performing the to-be-learned task.
Construct a dataset of the recorded state-action pairs.
Train a policy on the dataset using supervised learning.
return Trained policy.

BC has multiple limitations, including its need for large amounts of training data due to its proneness to the so-called *covariate shift* [AN04; Cel+22; Osa+18; Spe+22]. Covariate shift is an effect that occurs when the learner reaches states which were never visited by the expert in the provided training dataset. When reaching these previously unknown states, the learner does not have any data on how to behave because it was not trained to handle said situation, potentially leading to catastrophic mistakes. For example, when training a policy to perform active lane centering, the expert demonstrations would only contain ideal trajectories of driving within the lane markings. However, if the trained policy makes a small mistake and crosses over one of the lane markings, it has no information on how to salvage the situation by steering back onto the road, since this never happened in the training data [Cel+22; Osa+18].

2.5.2 Dataset Aggregation

Dataset Aggregation (DAgger) [RGB11] is an IIL approach that aims at addressing covariate shift by querying expert feedback in the form of corrections. It uses an iterative approach where the trained agent generates a trajectory based on its current policy. Each visited state at every time step is then corrected by either a human annotator or another algorithm. The provided demonstrations are added to the training dataset, and the agent is trained again. This process is repeated iteratively until the policy performs sufficiently well [Cel+22; RGB11; Spe+22]. Algorithm 2 depicts the primary steps conducted during DAgger.

Algorithm 2 Dataset Aggregation [Cel+22; RGB11]

```
Train initial policy on initial dataset using BC.
while Policy not good enough do
    Execute the current policy.
    Collect states visited by the current policy.
    Query an expert to provide actions on each visited state.
    Update dataset with collected state-action pairs.
    Retrain policy on updated dataset using supervised learning.
return Final policy.
```

DAgger therefore mitigates covariate shift by allowing the trained policy to make mistakes and then providing corrective demonstrations on how to fix these mistakes. However, DAgger also introduces high labeling efforts, since each visited state needs to be annotated by the supervisor. DAgger was not specifically intended for human users, but rather another expert policy could be used to provide expert annotations, such as a model-based controller or a planner system [Cel+22; Spe+22]. For human annotators, the labeling process is described as tedious, not user-friendly, and prone to mislabeling [Cel+22; Las+17]. Due to these limitations, further research on more human-centric approaches was conducted with the goal to reduce the required labeling efforts [Cel+22].

2.5.3 Approaches based on Human-Gated Interventions

As explained in the previous paragraph, the main limitation of DAgger is its tedious annotation process, which needs feedback on every state visited by the trained policy. Therefore, research was conducted on how to reduce the required labeling effort, resulting in approaches where expert corrections are only provided if they are really needed. These approaches can be distinguished into robot-gated and human-gated intervention approaches. Robot-gated approaches rely on an estimation by the trained agent of whether it requires feedback on its current actions. Since the robot decides whether an intervention is required, these approaches are called *robot-gated*. On the other hand, human-gated approaches allow the supervising expert to decide themselves when to intervene and take over control of the agent [Cel+22]. Human-gated interventions are therefore especially relevant for this thesis, since they resemble the intervention behavior of human drivers while using driving functions.

The general approach followed by most related work on human-gated interventions starts with the training of an initial policy, commonly using BC. Then, the policy is executed under the supervision of a human expert, who takes over control of the system by intervening if necessary. The recorded interventions are then stored in a training dataset and used to update the policy iteratively [Bi+20; Chi+22; Goe+19; Kel+19; Man+20; Spe+22]. Various tasks are performed in these works, including driving a vehicle [Bi+20; Kel+19; Spe+22], manipulation tasks using a robot arm [Chi+22; Man+20], and landing a quadcopter [Goe+19]. Figure 2.4 visualizes the general workflow of IIL approaches based on human-gated interventions.

The main difference between human-gated IIL algorithms is how the intervention-based training process is designed. For example, Kelly *et al.* [Kel+19] add the recorded intervention data to the initial demonstration dataset used for the BC and then retrain the policy on the combined dataset. In contrast, Goecks *et al.* [Goe+19] update their policy solely based on the intervention data. Mandlekar *et al.* [Man+20] argue that the states where the supervisor did not intervene should also be used as training data, in addition to the intervention data, since no interventions indicate good policy behavior. Two particularly interesting approaches are introduced by Bi *et al.* [Bi+20] and Spencer *et al.* [Spe+22], which both contain measures to compensate for the delays caused by

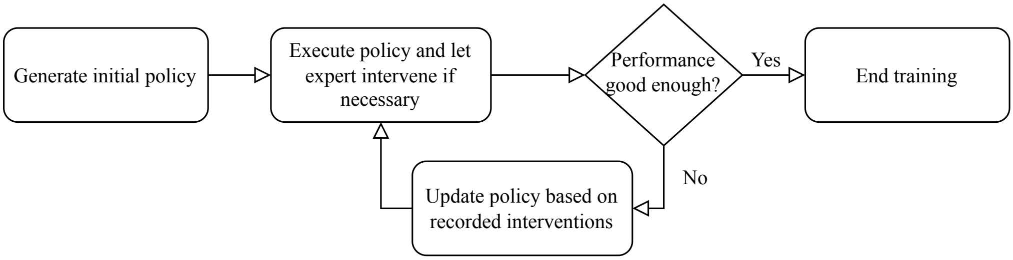


Figure 2.4: The general training workflow of IIL approaches based on human-gated interventions.

human reaction times when intervening. Spencer *et al.* [Spe+22] additionally introduce an approach to learn from all parts of the driven trajectory by leveraging the implicit information provided by the intervention timing. These approaches are explained in the following sections.

Compensation of Human Reaction Times via Interpolation

Bi *et al.* [Bi+20] introduce an approach which aims at compensating for the human reaction time of the supervisor when intervening. They argue that, when unwanted behavior occurs, the supervisor's intervention is delayed due to their reaction time. Therefore, the intervention already occurs in a bad state, and it takes some time to recover the agent to a desired state. In their proposed approach, the state-action pairs around the start of an intervention are removed and interpolated in order to create a smooth intervention trajectory instead. The smoothed trajectory is then used to update the trained policy. The authors use lane keeping as an example task to introduce their idea, which is depicted in Figure 2.5.

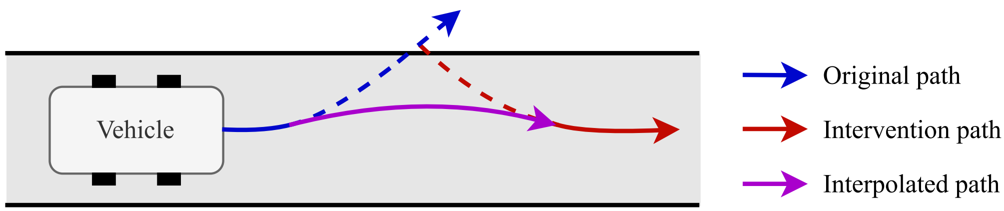


Figure 2.5: Trajectory interpolation around the start of interventions during lane keeping (adapted from [Bi+20]). Dashed lines indicate the parts of the original trajectory which are overwritten by the interpolation.

As can be seen, the agent makes a mistake by steering the vehicle off the road, indicated by the blue arrow. The human operator then intervenes as depicted by the red arrow, steering the vehicle back to the lane center. Due to the delay in the human reaction, the combined trajectory in red and blue does not represent desirable behavior, since the agent steers off the road and then drives back. Therefore, the authors remove segments of the trajectories around the start of the intervention and interpolate the trajectory to generate a more desirable path, here depicted by the purple arrow [Bi+20].

Expert Intervention Learning

Spencer *et al.* [Spe+22] introduce an approach named *Expert Intervention Learning (EIL)*, which aims at learning from the timing of interventions in addition to copying the demonstrated actions. Based on the intervention timings, parts of the driven trajectory are separated into good and bad state-action pairs. The authors state that the timing, and therefore also the absence of an intervention, contains implicit feedback which should be leveraged. Their approach is therefore related to Mandlekar *et al.* [Man+20], who state that a trajectory without interventions is implicitly labeled as good by the operator. EIL focuses on the assignment of three potential categories to each state of the driven trajectory: so-called *good states*, *bad states*, and *intervention states*. Their approach is explained with the example of driving a remote controlled car in a safe corridor, similar to the lane keeping task used above. Figure 2.6 depicts how the three categories are assigned to different parts of an example trajectory [Spe+22].

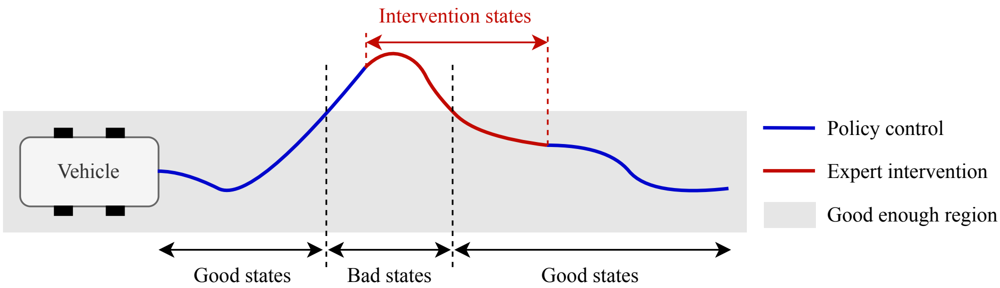


Figure 2.6: Flagging of trajectory segments as good states, bad states, and intervention states (adapted from [Spe+22]).

The work of Spencer *et al.* [Spe+22] is based on the idea that a so-called *good enough region* exists, where no intervention is executed by the operator as long as the agent stays within that region. This aspect of human behavior to accept small deviations from the perceived optimal behavior is also described by other authors [Hoc01; WXC14], where it is described as *human complacency*.

The good enough region is depicted in Figure 2.6 by the gray background. As can be seen, all states where the vehicle stays within this region are labeled as good states. Conversely, all states where the vehicle drives outside of this region are labeled as bad states. In the beginning, the trained policy controls the vehicle, depicted by the blue path. However, after some time the vehicle leaves the good enough region and, shortly after, the human operator intervenes by steering the vehicle back. This is depicted by the red segment of the driven path. The operator then hands back control to the policy, which proceeds to drive within the good enough region. All visited states are labeled as either good states or bad states, while some states are additionally labeled as intervention states [Spe+22].

While intervention states are objectively detectable, the good enough region is deduced from the intervention timing. Spencer *et al.* [Spe+22] assume that the state in which the operator takes over the vehicle control is already a bad state, mainly due to the reaction time needed for the human to take over the vehicle control. Therefore, a time threshold is used to assign states shortly before the start of the intervention as bad states. Conversely, only the first part of the intervention trajectory contains bad states, since the operator directly steers the vehicle back into the good enough region. Therefore, another time threshold is used to assign states shortly after the start of the intervention as bad states as well. The rest of the intervention contains desirable state-action pairs, since they demonstrate how to recover the vehicle close to the borders of the good enough region. After the good, bad, and intervention states have been identified, they are aggregated into buckets which are then used to update the policy. This is done by positively reinforcing good and intervention states, while negatively reinforcing bad states [Spe+22].

Although the authors report good results of their experiments, they also note limitations in their approach. One of the main limitations is the goal of only reaching a "good enough" performance. The authors state that if one can settle for good enough, their approach can be used to quickly learn a policy with decent performance. However, if an optimal policy should be learned, the expert must be queried everywhere, i.e., a DAgger approach would be required instead. Another limitation, which is stated by the authors, is the reliance on a consistent expert supervisor. If the supervisor behaves inconsistently, the learner's performance quickly reaches a cap beyond which it does not further improve [Spe+22].

2.5.4 Summary and Limitations

In this chapter, an introduction to IL and IIL approaches was given, with a focus on the work of Bi *et al.* [Bi+20] and Spencer *et al.* [Spe+22]. The main quality of the highlighted IIL approaches is the possibility to learn from human interventions while the trained policy is currently being executed. This allows for a natural and intuitive teaching process which is more data efficient and robust than previous approaches such as BC and DAgger [Cel+22]. Potential preprocessing steps for intervention data are explained in the works of Bi *et al.* [Bi+20] and Spencer *et al.* [Spe+22]. While Bi *et al.* [Bi+20] mainly focus on smoothing intervention trajectories to compensate for human reaction times, Spencer *et al.* [Spe+22] segment the trajectories around interventions into good states, which should be reinforced, and bad states, which should be avoided.

One limitation of the introduced algorithms is their reliance on consistent human behavior. The proposed algorithms aim at learning generalized policies for given tasks, but their quality suffers if the human demonstrations, on which they build upon, are inconsistent [Cel+22; Spe+22]. This limitation is shared with the driving function personalization approaches highlighted in Section 2.3 [BSDP08; KGB15; Wan+13]. Another limitation is that IL approaches generally require large amounts of data and commonly feature tedious annotation processes, such as DAgger. While IIL approaches aim at compensating this high need for data, they commonly settle for only a good enough performance [Cel+22; Spe+22].

2.6 Discussion and Research Objectives

In the preceding sections, research in the field of driver-initiated takeovers and the personalization of longitudinal driving functions was highlighted. Additionally, related research was highlighted that focuses on the topics of IIL approaches and on which automated driving style human drivers prefer compared to their manual driving style. In each section, the limitations of current research in the field were discussed separately. In this section, the findings are combined and discussed. Based on the research gaps of the current state of the art, the research objectives of this thesis are formulated.

2.6.1 Investigation of Driver Interventions during Naturalistic Driving

As was highlighted in Section 2.2, only few works exist that investigate the naturalistic takeover behavior of human drivers during real-world driving. However, the conducted analyses follow top-down approaches and only categorize the driver interventions on a coarse level [Ger+21; PM08; PF10; Yan+23]. The absence of an in-depth analysis of the underlying reasons for driver-initiated takeovers during naturalistic driving can likely be attributed to the absence of ground-truth annotations by the drivers. The found works solely utilize post-drive analyses of the recorded interventions to deduce potential categories and reasons for driver-initiated takeovers. However, a dataset featuring driver-initiated takeovers with ground-truth annotations by the drivers themselves could not be found in the current state of the art. Therefore, the recording of such a dataset in a test group study and the in-depth analysis of the human intervention behavior is the first research objective of this work.

Further limitations of research in this field are the almost exclusive focus on multi-lane highway driving scenarios [Ger+21; PM08; PF10; Yan+23]. Similarly, related work in the field of driving function optimization almost exclusively focuses on ACC car following scenarios with only limited analyses of other driving scenarios such as free driving [KGB15; Ram+17]. However, Level 1 and 2 driving functions may also be used on other road types and in more diverse driving scenarios. Therefore, the recorded dataset should focus on varied driving scenarios on different road types, as they are naturally encountered during driving, including both car following scenarios and free driving.

Finally, another assumption that is commonly made in related work is the correlation between driver-initiated takeovers and dissatisfaction with the driving function [GJ23; Lee+21; MZ21; Wan+13]. However, no related work could be found that proves this assumption in a test group study. Therefore, the conducted test group study should also investigate the effects of driver-initiated takeovers on the satisfaction of the drivers.

To summarize, the resulting research objectives are as follows:

- Investigate naturalistic human intervention behavior with an assisted longitudinal driving function by recording a dataset in varied driving scenarios.
- Investigate the reasons behind driver interventions based on ground-truth annotations by the drivers.
- Investigate the effects of driver interventions on the driver satisfaction.

2.6.2 Optimization Potential Analysis based on Driver Interventions

In Section 2.3, research on longitudinal driving function personalization approaches was discussed. These works primarily focus on ACC car following scenarios. Almost all highlighted approaches follow the workflow of first recording manual driving data and then adjusting a driving function's behavior to mimic this manual driving style as closely as possible. These approaches are based on the assumption that drivers generally prefer their manual driving style also as their automated driving style. However, as discussed in Section 2.4, the results of the studies conducted on this topic are inconclusive. While some studies claim that the assumption is correct and drivers truly prefer their manual driving style as their automated driving style as well [MZ21; Vas+23], most studies find tendencies toward more defensive automated driving styles [Bas+17; Gel+16; Gri+16; Pen+24; Yus+16]. Mimicking the driver's manual driving behavior might therefore not be the ideal goal of driving function personalization approaches.

Generally, the main goal of driving function personalization approaches should be to improve the driver's satisfaction with the driving function, as it is stated in Definition 2.11. Therefore, the satisfaction with the personalized driving function must be measured while the function is used. As mentioned in Section 2.6.1, a correlation between driver interventions and the driver's dissatisfaction is assumed. If this correlation is confirmed, driver interventions could be used to deduce optimization potential of the driving function, as they highlight aspects of the driving function with which drivers are dissatisfied. Similarly, IIL approaches are based on the idea that driver interventions provide valuable feedback that can be used to optimize the system. Therefore, one research objective is to analyze driver interventions in the newly recorded dataset to determine if adjustments to the driving function are required and, if so, how those adjustments should be conducted.

Both the investigated driving function personalization approaches and the highlighted IL approaches aim at learning generalized policies. However, it is commonly stated that human behavior may be inconsistent, which negatively affects the performance of the trained algorithms [BSDP08; Cel+22; KGB15; Spe+22; Wan+13]. Therefore, the analysis of the intervention behavior in the dataset should also focus on whether a generalized policy can correctly represent the human driving behavior.

Based on the identified research gaps, the resulting research objectives are as follow:

- Develop a bottom-up, data-driven method to derive necessary adjustments to the driving function based on recorded driver interventions.
- Investigate whether these adjustments could be represented by a generalized driving policy and whether a personalization of the adjustments is required.

2.6.3 Development and Evaluation of a Prototypical Self-Learning Personalized Driving Function

As explained in Section 2.6.2, most driving function personalization approaches conduct a one-shot personalization of the driving function with the goal of mimicking the driver's manual driving behavior as closely as possible. However, the effect of their personalization on the driver's satisfaction is rarely analyzed afterward. In Definition 2.11, the ideal personalization approach is described as a cyclical process that should continuously investigate the driver's satisfaction while using the driving function and adjust it accordingly. An example of such a continuous update process can be seen in recent IIL approaches. These approaches are based on the idea that expert interventions contain feedback that can be used to optimize the observed system. Only one driving function personalization approach could be found in related work, that also utilizes these driver interventions as feedback to update their driver model [GJ23]. This approach follows the same general workflow as IIL approaches, which seems promising for the continuous personalization of driving functions.

One goal of this thesis is therefore to develop such a continuous personalization approach for the PLDF utilizing the implicit feedback contained in driver interventions. Based on the results of the data-driven analyses conducted for the previous research objectives, the existing personalization and IIL approaches may be re-evaluated to determine their applicability to the analyzed PLDF. The developed driving function prototype should then be tested in another test group study to validate its design.

Accordingly, these resulting research objectives are investigated:

- Develop a prototypical PLDF that learns from driver interventions iteratively in a cooperative traded control setting.

- Test the applicability of the proposed prototype and its effects on the driver satisfaction in a test group study.

2.6.4 Development of an Automatic Classification Framework for Driver Interventions

The research objectives outlined above conduct analyses based on the manually annotated dataset obtained from the test group study described in Section 2.6.1. The investigations in Section 2.6.1 and Section 2.6.2 focus on examining driver intervention behavior over an extended period of time, which requires the manually annotated reasons behind each driver intervention. However, the dependence on manual annotations strongly limits the scalability of the developed approaches, particularly for the analysis of large-scale datasets, such as vehicle fleet data. Furthermore, the research described in Section 2.6.3 aims to develop a prototypical driving function capable of directly learning from the observed driver interventions during driving. However, in order to use driver interventions for such a prototype during naturalistic driving, driver interventions need to be manually annotated, which strongly limits the applicability in a customer-oriented setting. Consequently, the final research objective of this thesis is the development of a framework for the automatic classification of driver interventions, utilizing the recorded and labeled dataset from Section 2.6.1.

As discussed in Section 2.2, some related studies investigate human intervention behavior, but their approaches to automatically derive the reasons behind each driver intervention are limited to coarse top-down approaches by their underlying datasets, which lack granular ground-truth labels provided by the drivers themselves [Ger+21; PM08; PF10; Yan+23]. Therefore, no classification framework for the granular reasons behind driver interventions could be found in the state of the art.

To summarize, the resulting research objectives are as follows:

- Develop a method for the automatic classification of driver interventions based on the recorded dataset from Section 2.6.1 as a proof of concept.
- Evaluate the classification framework's performance on the recorded dataset.

3 Dataset Creation: Test Group Study during Naturalistic Driving

This chapter focuses on the design and results of a test group study with the goal of investigating human intervention behavior during naturalistic driving with an assisted driving function. This study addresses the research objectives stated in Section 2.6.1. Parts of the data analyses conducted in the context of this test group study were the topic of one bachelor thesis [Dal23] and one master thesis [Liu24], which were supervised within the scope of this doctoral thesis, and resulted in a publication [Sch+24a]. The bachelor thesis, master thesis, and publication include only a subset of the analyses and results presented in this chapter.

The primary objective of the test group study is the creation of an annotated dataset which contains driver interventions during naturalistic driving and the drivers' intentions behind each intervention. Based on the recorded dataset, a taxonomy of the reasons behind driver interventions is created. Additionally, it is investigated whether an effect of driver interventions on the drivers' satisfaction with the system can be found. The test group study conducted in this chapter is called the *dataset creation test group study* and the resulting dataset is hereinafter called the *driver intervention dataset* to distinguish them from the test group study conducted in Chapter 6.

This chapter is structured as follows: In Section 3.1, the design of the test group study is explained, including the study setup, procedure, questionnaire design, and more. Section 3.2 then addresses the creation of the annotated dataset based on the raw data recorded in the study. In Section 3.3, the results of the study are shown, including an analysis of the created dataset and the questionnaire results. Finally, in Section 3.4, the findings are summarized and the resulting research contributions are highlighted.

3.1 Study Design

This section focuses on the design of the dataset creation test group study. It includes the whole design process starting with the definition of the study's scope up until its execution.

3.1.1 Study Scope

The scope of the test group study is mainly defined by the research objectives stated in Section 2.6.1. These focus on the generation of a dataset featuring naturalistic driving data including ground truth annotations by the drivers about the reason behind each intervention. Another research focus is to investigate the effect of the driver's intervention behavior on their satisfaction with the driving function. Additional requirements on the study are defined by the company cooperation and during preliminary tests.

As this thesis was conducted in cooperation with Porsche, the first requirement from the company side is that the chosen driving function selected for this test group study is the PLDF, described in Section 2.1.8. Consequently, the study was conducted using instrumented Porsche vehicles equipped with the PLDF. This choice of driving function should not restrict the generalizability of the study results to other PLDFs, as comparable systems from other manufacturers provide highly similar feature sets [BMW; Mer]. Due to restrictions imposed by the company, only Porsche employees who received a specific driving training are allowed to drive with these vehicles. Therefore, it was decided that all participants must be part of the driver assistance systems department, in which this thesis was also supervised. However, employees who took part in the development of the PLDF were excluded from participating. The choice of participants is covered in Section 3.1.5.

Another requirement is defined by the research objectives stated in Section 2.6.2. As stated there, the consistency of driver behavior should be analyzed in recurring scenarios. Therefore, it was decided that each participant must drive on the chosen route multiple times, to allow for the comparison of their behavior across multiple drive-throughs.

During the initial design stages of the test group study, preliminary studies were conducted with five individual drivers. From the experience gained during these preliminary studies, some requirements were additionally derived. One of the goals of this test group study is to record naturalistic driving data in varied driving scenarios. However, the preliminary tests have shown that drivers who are unfamiliar with the PLDF tend to drive more cautiously in the beginning. Drivers unfamiliar with the system also tend to intervene in situations where no intervention would have been required, due to uncertainty regarding the system's ODD. These observations are consistent with the findings of other studies that report different driver behavior depending on previous experience with an ADAS [BK13; Cle+22; Hoc+18]. Another observation made during the preliminary tests is that participants drive more cautiously on unknown routes compared to familiar ones. Therefore, one requirement of the study is that the participants are both familiar with the PLDF and the driven route.

Based on the above stated research objectives and requirements, the following study scope is derived:

- Record driver interventions during naturalistic driving with the PLDF in varied driving scenarios.
- Record the driving behavior of each participant over the course of multiple drives on the same route.
- Annotate each driver intervention with the driver's reason to intervene.
- Investigate the driver's satisfaction with the PLDF using a questionnaire.
- Study participants must be familiar with the PLDF and the driven route.

The investigation of the driver satisfaction is based on the following hypothesis:

Hypothesis 3.1 (Effect of Driver Interventions on Satisfaction)
Drivers who frequently intervene and take over the vehicle control are less satisfied with the PLDF compared to drivers who intervene less frequently.

3.1.2 Chosen Route and Study Duration

As outlined in the study scope, one objective is to record naturalistic driving data on routes that are familiar to the driver and repeatedly traversed. Based on this requirement, the daily commutes of each participant were selected. These routes are familiar to the drivers and are naturally driven twice per day, once in each direction. The office of the Porsche driver assistance systems department is located in Mönsheim, southwestern Germany, and is surrounded primarily by rural country roads and small villages. Furthermore, a nearby highway ramp provides access to major city centers such as Stuttgart and Karlsruhe. Thus, the commuter routes around this office location offer diverse driving scenarios, as required by the study scope.

Although recording multiple consecutive commutes in a single session would be straightforward, it would likely influence the participants' driving behavior, as they would be required to drive the same route repeatedly for several hours. Therefore, it was decided to record the everyday commutes of the participants over the course of one week. This duration was chosen as a compromise between recording a sufficient number of drives and limiting the study duration.

The study was initially planned for three months with two instrumented vehicles. However, due to technical issues with one of the vehicles and the data logging devices, the total study duration extended to six months.

3.1.3 Study Setup

As explained in the study scope, the objective is to record three types of data:

1. Driving data with the PLDF that includes the driver interventions.
2. The ground truth annotations by the drivers containing the reason behind each driver intervention.
3. The driver's satisfaction with the system after experiencing it.

For the recording of the driving data, two instrumented vehicles were used: One Porsche Cayenne from the year 2023 and one Porsche 911 from the year 2022. However, due to technical problems with the Porsche 911, the vehicle could not be used in the second half of the study. Therefore, most of the drives in the final dataset were recorded with the Porsche Cayenne. However, the choice of vehicle is not expected to influence the study results, as the feature set of the PLDF is identical across both vehicles.

Both vehicles were equipped with the standard in-production PLDF and a vehicle bus data logging device, which is set up to record all FlexRay bus signals. On the FlexRay, all relevant signals for the PLDF and the following analyses are sent. Because it was decided to record the everyday commutes of the participants, no co-driver was seated in the vehicle to manage the annotation of the driver interventions. Instead, a smartphone with a voice recording app was attached to each vehicle's windshield in a phone mount. Using the smartphone, drivers annotated their interventions by briefly pressing on the smartphone screen to start the recording, stating their reason for the intervention, and then pressing on the screen again to stop the recording. There are two main reasons why voice recordings were chosen for the annotation of the interventions. First, voice recordings are intuitive and minimize the distraction during driving compared to, e.g., a complex touch screen setup with predefined categories. Second, the goal of this study is to record and analyze the detailed reasons behind each driver intervention. Therefore, it was decided to not limit the possible labels beforehand by using predefined categories. Finally, a questionnaire was used to evaluate the driver satisfaction with the system. This questionnaire is explained in the next section.

3.1.4 Questionnaire

The translated questionnaire used in this study is depicted in Appendix A.1. The first page inquires about demographic data such as age, general driving experience, and experience with ADASs. Due to data protection reasons and the low number of participants, the gender of participants was not recorded. The second page contains a custom questionnaire assessing the driver satisfaction with the function. It includes five items using five-point Likert scales [Lik32] ranging from *strongly agree* to *strongly disagree*. The driver satisfaction was assessed using a single item titled *I was satisfied with the driving function in general*, while the other items inquire about the satisfaction with the frequency of driver interventions and how much drivers liked the PLDF outside of situations they intervened in. The questionnaire used in this study was kept relatively simple as the primary objective was to record a dataset of annotated driver interventions at the time this study was planned. The third page of this study's questionnaire contains text boxes asking about the participant's intervention behavior. Finally, a large text box provides space for free-form feedback on the PLDF.

3.1.5 Participants

The participants of this study were recruited internally within the driver assistance systems department of Porsche. As stated before, employees who took part in the development of the PLDF were excluded from participating. The participants were required to have experience with the PLDF, and to ensure diversity of the driving scenarios in the dataset, their commute should not predominantly consist of highway driving.

In total, 26 participants took part in the study. However, due to technical problems with the data logging devices and the test vehicles, the data of four participants could not be used. Additionally, data from five participants were excluded from the dataset due to poor annotation quality or noncompliance with the study instructions. Therefore, a total of 17 participants are contained in the final study dataset. The age distribution of the valid participants is shown in Table 3.1.

Table 3.1: Age distribution of the dataset creation test group study participants.

	$20 < \text{Age} \leq 30$	$30 < \text{Age} \leq 40$	$40 < \text{Age} \leq 50$	$50 < \text{Age} \leq 60$
Number of participants	7	6	4	0

3.1.6 Study Procedure

The study procedure for a single participant is outlined in this section. At the beginning, participants completed the first page of the questionnaire, without being shown the subsequent pages, before receiving an explanation of the study task. They were instructed to use the study vehicle for one week on their daily commutes, during which they should use the PLDF as much as possible by avoiding prolonged periods of manual driving. They were also instructed to take over vehicle control whenever they deemed it necessary. After each takeover, participants were asked to return control to the PLDF and briefly annotate the reason for their intervention using the mounted smartphone. Each voice annotation should contain the following information:

- Driver input,
- Situation,
- Reason,
- Desired behavior.

The *driver input* describes the action taken by the driver, e.g. pressing the brake pedal or adjusting the PLDF's set speed. Examples for the *situation* include curves, turns, straight roads, etc. The *reason* describes why the driver intervened, with the two main categories being either that the situation fell outside the PLDF's ODD or that the driver intervened due to a deviating personal preference. Finally, the *desired behavior* describes how the PLDF should have behaved instead, e.g., driving faster, slower, or accelerating earlier. This information is only required if the reason for the intervention was a deviating personal preference. Drivers were informed that they did not need to follow a rigid annotation style and could instead use free text to describe each intervention, provided it contained the relevant information explained above. They were also instructed to always drive the same route from work to home without additional stops, ensuring that drives on the same route could be better compared in Chapter 4. Additional drives beyond the commutes could be annotated on a voluntary basis.

After receiving the instructions, a brief test drive was conducted with each participant to verify whether the instructions were understood. The participant additionally received a written manual that elaborates the instructions in detail. After all open questions were answered, the vehicle was handed over to the participant. One week later, the participant handed back the vehicle and filled out the second and third pages of the questionnaire.

3.2 Dataset Creation

The data recorded during the study consists of raw FlexRay vehicle bus data and voice annotation files created by the drivers. Prior to any analyses on the dataset, this raw data was preprocessed and the voice annotations were translated into machine-readable labels. The preprocessing steps conducted on the dataset can be summarized as follows:

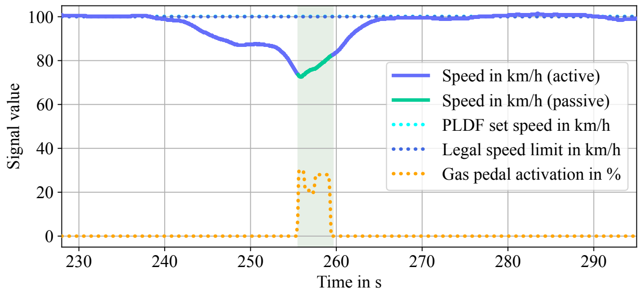
1. Scale the raw bus signal values to their corresponding physical values.
2. Remove all drives from the dataset which are not annotated or do not contain the participant's commuter route. Due to the high manual annotation effort, it was decided to focus solely on the commuter routes, even if the participants also annotated different drives.

3. Check the quality of each participant's voice annotations and driving data. If the quality of their voice annotations was too poor to derive the required annotation labels, or if the participant did not comply with the study instructions, their data was removed from the dataset. An example for noncompliance is not using the PLDF for extended periods of time.

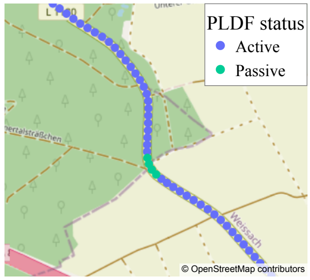
3.2.1 Labeling Strategy

After the dataset was preprocessed and filtered, the intervention data was labeled using the recorded voice annotations. The goal of the labeling process is to assign a label to each intervention in the dataset. For this manual labeling task, a custom annotation tool was developed. The tool iterates over every single driver intervention and voice annotation in the dataset and allows the annotator to assign labels to each intervention based on the information from the voice recordings. The file creation timestamp of the voice recordings is used to synchronize them with the time series data. As explained in Section 2.1.8, there are two main types of interventions in the dataset: pedal interventions and set speed adjustments. Pedal interventions are automatically detected via the status signal of the PLDF, and set speed adjustments are automatically detected if the PLDF's set speed deviates from the current legal speed limit. The labeling tool then visualizes the time series data around each detected intervention in a signal-over-time plot and the corresponding GPS traces in a map plot. The timestamp of each voice annotation is additionally visualized. Based on the visualized data and the voice recording's content, which is shown in the tool via speech-to-text, the annotator may label the intervention using five different fields. Four of these fields were already introduced in Section 3.1.6 as the necessary content of each voice annotation. Additionally, a fifth field named *system* is introduced that describes which functionality is overwritten by the driver intervention. The possible options are either *PLDF* to describe free-driving scenarios or *ACC* to describe interventions in the context of car following scenarios. Therefore, the resulting label fields are *system*, *driver input*, *situation*, *reason*, and *desired behavior*.

A simplified signal-over-time plot and its corresponding GPS traces of an example intervention are depicted in Figure 3.1. As can be seen in the GPS traces in Figure 3.1b, the intervention takes place on a high-curvature country road. The legal speed and the corresponding set speed are continuously at $100\,\mathrm{km/h}$, but to accommodate for the high road curvature, the PLDF reduces the velocity in the first curve to approximately $80\,\mathrm{km/h}$ and in the second curve to approximately $70\,\mathrm{km/h}$. However, although the PLDF behaves as intended and does not objectively make a mistake, the driver decides to intervene by pressing the gas pedal, thereby increasing the vehicle speed in this curve above the PLDF's intended speed. In Figure 3.1a, this can be seen by the vehicle speed signal turning from blue, which indicates an active PLDF, to green, which indicates a passive PLDF. After the driver releases the gas pedal, the PLDF automatically reactivates and the vehicle accelerates to the set speed again. The driver provided a voice annotation



(a) The driver performs a voluntary gas pedal intervention to increase the vehicle speed in a curve above the PLDF's intended speed.



(b) The corresponding GPS traces of the gas pedal intervention.

Figure 3.1: Signal-over-time plot of an example gas pedal intervention and the corresponding GPS traces. The vehicle speed signal changes its color based on the current PLDF status. Interventions are additionally highlighted by a green background. The GPS traces are colored based on the current PLDF status.

at around 270 s, stating that they would have preferred a higher velocity in the second curve and therefore intervened by pressing the gas pedal. In the annotation tool, this intervention was then translated as follows:

- System: *PLDF*
- Driver input: *Gas*
- Situation: *Curve*
- Reason: *Personal preference*
- Desired behavior: *Higher speed*

As explained above, the possible annotation labels were not predefined before the study. Instead, the potential options for each annotation field were continuously expanded during the annotation process. Whenever a new annotation was translated which could not be described with the existing labels, new options were added. This bottom-up labeling strategy ensures that all potential types of interventions are thoroughly described. The final list of labels for each field and some additional explanations of the labels are shown in Appendix A.2. Using these labels, almost all interventions could be described in detail. Only few interventions had such exceedingly rare reasons or situations, that the label *other* was assigned to one or multiple fields. If a field was irrelevant or not applicable, the label *n/a* was assigned. Interventions with missing or incomplete voice annotations were still annotated if the reasons behind them were obvious from the vehicle bus and GPS data. However, cases with unclear driver intentions had to be labeled with the reason *unknown*. Using this labeling strategy, all interventions in the dataset were annotated. Solely the final deactivation of the PLDF before reaching the navigation goal was excluded from the annotation, as the PLDF was not reactivated afterward.

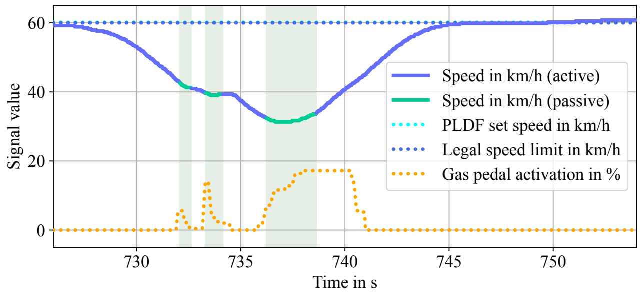
3.2.2 Merging of Interventions

In Figure 3.2, two examples are shown for voluntary gas pedal interventions in roundabouts. In both cases, the drivers aimed to increase the vehicle speed above the PLDF's planned velocity due to a deviating personal preference. As can be seen, the driver presses the gas pedal three separate times in Figure 3.2a, which causes three separate pedal interventions. These pedal interventions are automatically detected via the PLDF status signal and accordingly, all three interventions require an annotation. In this case, they received the same labels, except for their situation label. The first two interventions were annotated with the situation *Roundabout - Approach*, while the third intervention was annotated with the situation *Roundabout - Driving through*. In Figure 3.2c, a similar scenario is depicted. However, this time, the driver performs only one continuous gas pedal intervention. The intervention encompasses the whole roundabout including approaching, driving through, and leaving it. During the annotation process, it was decided that interventions such as these should be labeled with the situation *Roundabout - Driving through*. However, it is often ambiguous which segment of the roundabout an intervention belongs to.

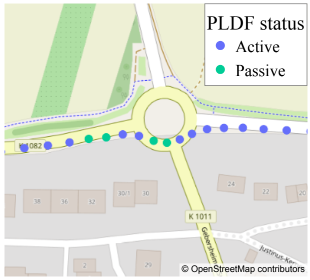
Another limitation of the described annotation approach is that each pedal intervention is counted separately, which affects subsequent statistical analyses. Comparing Figure 3.2a and Figure 3.2c, in both cases the driver decided that they preferred to drive faster through a roundabout. However, when comparing the number of performed interventions, the first driver would have performed three interventions where the second driver only performed one intervention. The only difference is how the drivers executed each intervention. Therefore, it was decided that interventions with the same driver intention that occur in quick succession, such as the example in Figure 3.2a, should be grouped together and counted as one single intervention for the statistical analyses. As a result of this merging of interventions, and due to their inherent ambiguity, the additional situation information, such as *approach* and *driving through* is dropped. Instead, both driver interventions in Figure 3.2 are annotated with the general situation *Roundabout*. The resulting annotation of both interventions in Figure 3.2 is as follows:

- System: *PLDF*
- Driver input: *Gas*
- Situation: *Roundabout*
- Reason: *Personal preference*
- Desired behavior: *Higher speed*

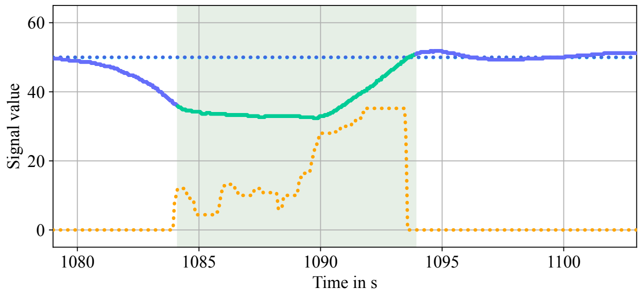
The original granular situation labels with *approach*, *entering*, *driving through* and *leaving* are still contained in the annotated dataset. However, for the following analyses, the general situation labels are used instead.



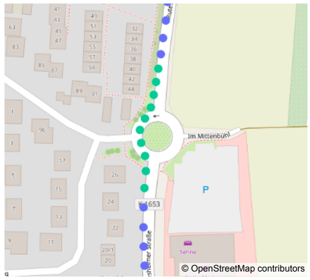
(a) The driver presses the gas pedal three separate times while driving through a roundabout.



(b) The corresponding GPS traces of the three consecutive gas pedal interventions.



(c) The driver performs a prolonged gas pedal intervention while driving through a roundabout.



(d) The corresponding GPS traces of the single prolonged gas pedal intervention.

Figure 3.2: Signal-over-time plots of gas pedal interventions in roundabouts and the corresponding GPS traces. In both cases, the drivers intervened voluntarily to increase the vehicle speed above the PLDF's intended speed in and around the roundabout.

3.3 Results

In this section, the results of the first test group study are described. This includes a general overview of the recorded dataset, an in-depth analysis of the found types of driver interventions, and an analysis of the questionnaire results.

3.3.1 Dataset Overview

In Table 3.2, an overview of the recorded dataset is provided. As can be seen, the dataset features a total of 3347 driver interventions, out of which 2438 are pedal interventions and 909 are set speed adjustments. This number is relatively high, as it resembles an average intervention rate of 0.64 interventions per minute, which shows that both pedal interventions and set speed adjustments are frequently performed by the drivers.

These results align with related research, highlighted in Section 2.2, that also reports a relatively high frequency of driver interventions while using Level 1 and 2 driving functions.

Table 3.2: Overview of the annotated driver intervention dataset.

Total number of participants	17
Total number of drives	158
Total duration	92.8 h
Total distance	4334 km
Total number of interventions	3347
Number of pedal interventions	2438
Number of set speed adjustments	909
Average number of drives per participant	9.29
Average duration per drive	35.6 min
Average distance per drive	27.0 km
Average number of interventions per drive	21.0
Average number of interventions per minute	0.64

The dataset contains a total of 158 drives which results in an average number of 9.29 drives per participant. This number can be explained by the fact that drivers either used the vehicle from Monday evening to Friday morning, which results in eight commutes, or from Monday evening to the following Monday morning, which results in ten commutes. Participants could voluntarily decide whether to use the vehicle only until Friday or until the following Monday. Therefore, each participant provided either four or five commutes in each direction. Two drivers additionally provided a voluntary sixth annotated commute on the weekend.

Out of the 4334 km driven, 14.3 % were driven on highways, while 25.9 % were driven inside towns and cities. The remaining 59.8 % are comprised of country roads and other street categories. 30.92 % of the time, the ACC was active due to the presence of a slower leading vehicle, whereas the remaining 69.08 % of the dataset consists of free-driving scenarios without preceding slower vehicles.

3.3.2 Types of Interventions

During the annotation process, 164 unique label combinations were assigned to the interventions. These granular labels were then analyzed and grouped together to create a bottom-up hierarchical label taxonomy consisting of four Annotation Abstraction Levels (AALs). The first level, Annotation Abstraction Level One (AAL1), describes the original label combinations, consisting of the five label fields used during annotation. These granular original labels feature a high redundancy and are therefore grouped together based on their underlying driver intention, e.g., driving faster in a roundabout due to a deviating personal preference, or braking at a red traffic light. The resulting second level, Annotation Abstraction Level Two (AAL2), consists of 55 labels. These driver intentions are then further grouped into 14 clusters based on the general situations in which they occur, resulting in the third level, Annotation Abstraction Level Three (AAL3). The fourth and final level, Annotation Abstraction Level Four (AAL4), comprises four labels which distinguish the main reasons why a driver may intervene. These are a deviating personal preference, mandatory interventions in situations which fall outside of the PLDF's ODD, interventions due to incorrect information from the sensing state or map, and other interventions. Table 3.3 summarizes the structure of the AALs used in this study.

Table 3.3: The general content and number of distinct labels in each of the four hierarchical AALs.

Level	Content	Number of labels
AAL1	Raw annotations	164
AAL2	Driver intentions	55
AAL3	General situations	14
AAL4	Main intervention reasons	4

Table 3.4 depicts the third and fourth levels of the label taxonomy. Additionally, the absolute and relative numbers of interventions are reported. A limitation of these measures is that the data of participants who drove longer distances and intervened more frequently have a disproportionate influence on the relative label distribution compared to drivers who had shorter commutes. To mitigate this imbalance, the relative distribution of labels was normalized across all participants and is provided as well.

As illustrated by the data, interventions attributed to deviating personal preferences are by far the most prevalent, accounting for a total of 1917 instances and comprising $53.40\,\%$ of the normalized distribution. This indicates a high optimization potential of the PLDF, as these types of interventions fall within its ODD. Therefore, by adjusting the PLDF to the drivers' individual preferences, the number of performed interventions could

Table 3.4: AAL3 and AAL4 of the hierarchical label taxonomy. For each label of AAL3, the absolute number of interventions, as well as the relative and normalized distributions are included.

AAL4	AAL3	Absolute	Relative	Normalized
Personal preference	Speed adjustment on straight roads	968	28.92 %	25.18 %
	Speed adjustment in high-curvature segments	458	13.68 %	14.38 %
	Adjustment of acceleration timings onto speed limits	366	10.94 %	10.06 %
	Adjustment of acceleration strength	79	2.36 %	2.27 %
	Adjustment of ACC distance	46	1.37 %	1.51 %
Outside of ODD	Right-of-way interactions	340	10.16 %	14.34 %
	Traffic light interactions	208	6.21 %	7.55 %
	Multi lane interactions	122	3.65 %	2.49 %
	Reactivate after stop	112	3.35 %	2.59 %
	Other traffic interactions	16	0.48 %	0.49 %
Incorrect input data	Incorrect ACC sensing state	212	6.33 %	7.16 %
	Incorrect TSD or map information	189	5.65 %	5.53 %
Other	Missing annotation	105	3.14 %	2.54 %
	Unintentional intervention	78	2.33 %	2.41 %
	Other	48	1.43 %	1.50 %

potentially be halved. The interventions based on personal preferences mainly focus on the adjustment of the vehicle speed and acceleration timings in various scenarios. 798 total interventions and 27.45 % of the normalized distribution are attributed to situations which fall outside of the PLDF's ODD. These are the only interventions which are necessary by design of the PLDF and mainly comprise interactions with other traffic participants, such as yielding in right-of-way interactions, or reacting to traffic lights. The third main reason for driver interventions is the correction of wrong input data of the PLDF, which makes up 401 interventions and 12.69 % of the normalized distribution. The PLDF's behavior is mainly based on information from the TSD, the map, and the ACC target vehicle detection. If any of these inputs is erroneous, the driver is required to correct the PLDF's behavior accordingly. These categories describe the three main reasons for driver interventions. However, some interventions do not fall into these three categories, such as unintentional or not annotated interventions. These interventions are therefore attributed to the *other* category and make up 6.45 % of the normalized distribution, i.e., 231 interventions in total. For more detailed information on the lower abstraction levels, Appendix A.3 features tables presenting the AAL2 labels and their distribution in the dataset. The AAL1 labels are not shown in this thesis, since all relevant information is contained in AAL2 and the AAL1 labels only introduce more noise and redundancy. AAL1 only represents the output of the annotation process, which the higher abstraction levels are built upon.

Used Driver Input Channels

In Table 3.5, the distribution of the used driver inputs to initiate interventions are shown. As can be seen, the interventions based on deviating personal preferences predominantly consist of gas pedal interventions and set speed increases. Apparently, most drivers would prefer a more dynamic driving style while using the PLDF.

Table 3.5: Distribution of the used driver inputs to initiate an intervention, divided into the four AAL4 labels. The distribution in the full dataset is additionally depicted.

Main reason to intervene	Gas pedal	Brake pedal	Cancel	Set speed increase	Set speed decrease
Personal preference	48.62 %	7.51 %	1.41 %	37.87 %	4.59 %
Outside of ODD	27.82 %	71.93 %	0.25 %	0.00 %	0.00 %
Incorrect input data	55.11 %	30.17 %	2.00 %	11.97 %	0.75 %
Other	35.50 %	32.90 %	12.55 %	9.52 %	9.52 %
Full dataset	43.53 %	27.34 %	1.97 %	23.78 %	3.38 %

Conversely, drivers primarily used the brake pedal for interventions in situations outside of the ODD, as these interventions are predominantly comprised of traffic interactions that involve yielding to other traffic participants, as shown in Table 3.4. In these cases, the driver is required to press the brake pedal in order to yield, if necessary, because the PLDF is designed to enter traffic situations without anticipating potential traffic. However, there are still 27.82 % of gas pedal interventions outside of the ODD. These mostly contain multi-lane traffic interactions, such as overtaking and merging, and the mandatory reactivation of the PLDF after a prolonged stop. Due to the design of the PLDF, all situations that fall outside of its ODD must be handled via pedal interventions. Accordingly, 0 % of interventions outside of the ODD were conducted via set speed adjustments.

Both input channels of the PLDF, the ACC sensing state and the speed limit information, may contain incorrect information. An incorrect ACC sensing state can be divided into two main categories. Either a leading vehicle is not detected and the driver is required to manually press the brake pedal, or a wrong vehicle is detected as the leading vehicle, causing the ACC to unnecessarily decelerate, prompting a gas pedal intervention by the driver. In both cases, a pedal intervention is required. In the case of an incorrect speed limit, both pedal interventions and set speed adjustments may be used to adjust the vehicle speed accordingly. Therefore, interventions based on incorrect input data are predominantly carried out using the pedals, while set speed adjustment are employed only occasionally.

Finally, the remaining *other* interventions are primarily comprised of unintentional und not annotated driver interventions. Therefore, the driver input channels are roughly balanced. Interestingly, the option to deactivate the PLDF with the cancel is seldomly used outside of the *other* category. In total, the cancel functionality was only used 66 times in the whole dataset, out of which 22 interventions fall in the *other* category. Due to its rarity and its functional similarity to briefly pressing the brake pedal, it was decided to generally count cancel interventions as a pedal interventions in the following analyses.

Driver Interventions due to Deviating Personal Preferences

For this thesis, the interventions due to a deviating personal preference are especially important, as they indicate optimization potential within the PLDF's ODD. Interventions outside of the ODD cannot be used to personalize the driving function because they represent situations that the PLDF is not designed to handle. Instead, the development of a new driving function with an extended ODD would be required, which lies outside of the scope of this thesis. Interventions based on incorrect input data also represent an optimization potential for the PLDF, as incorrect perception and map data cause a significant number of driver interventions. However, the development of perception functions and precise map data are other areas of research that fall outside of the scope of this thesis. Thus, only interventions due to deviating personal preferences are further analyzed.

As can be seen in Table 3.4, the majority of interventions due to personal preferences fall into three main categories of AAL3: *speed adjustment on straight roads*, *speed adjustment in high-curvature segments*, and *adjustment of the acceleration timing onto speed limits*. These three categories account for $49.62\,\%$ of the normalized distribution of all interventions in the dataset. The remaining two categories, *adjustment of the acceleration strength* and *adjustment of ACC distance* only make up $3.78\,\%$ of the normalized distribution. In the following paragraphs, these five AAL3 categories are discussed in detail. For a more granular view, Table A.3 in Appendix A.3 depicts the AAL2 labels out of which the five main AAL3 categories consist. Additionally, example plots for the most common and relevant AAL2 intervention types are provided in Appendix A.4.

The highest number of interventions in AAL3 are *speed adjustments on straight roads*, which account for $25.18\,\%$ of the normalized distribution. These consist of set speed increases and decreases as well as gas and brake pedal interventions on straight roads. Set speed increases make up the majority of this category with $17.74\,\%$, followed by accelerations using the gas pedal with $3.79\,\%$. Only $3.65\,\%$ of interventions aim to reduce the vehicle velocity using set speed decreases or brakes. This shows a tendency toward higher speeds, as well as a preference to adjust the vehicle speed using set speed adjustments on straight roads rather than by utilizing the pedals. While both intervention types may occur separate from each other, drivers often combine the interventions by first pressing the gas to accelerate and then setting the set speed to the preferred value.

Interventions focusing on the *speed adjustment in high-curvature segments* are the second most common, accounting for $14.38\,\%$ of the normalized distribution. These consist of gas and brake pedal interventions in situations where a high road curvature requires a lower vehicle speed than the current set speed. Using the curvature information from the map, the PLDF calculates adequate speed profiles in curves, turns, and roundabouts. However, drivers frequently adjust the vehicle speed in these situations via pedal interventions. Roundabouts are the most common situation in which the vehicle speed was adjusted, accounting for $6.21\,\%$ of the normalized distribution. The interventions in roundabouts are almost exclusively gas pedal interventions with only few brake pedal interventions. Speed adjustments in left and right turns account for $5.26\,\%$ of the normalized distribution. The ratio of gas and brake pedal interventions is mostly balanced with no clear preference for higher or lower speeds in turns. And finally, gas and brake pedal intervention in curves account for $2.58\,\%$ of the normalized distribution. Approximately two thirds of these interventions were conducted using the gas pedal and one third using the brakes.

$10.06\,\%$ of the normalized distribution are made up of *adjustments of the acceleration timing onto speed limits*. All three contained AAL2 labels, *earlier acceleration onto higher speed limit*, *later deceleration onto lower speed limit* and *set speed increase before higher speed limit*, focus on the adjustment of the acceleration and deceleration timings onto upcoming speed limits. The PLDF is designed to always abide by the legal speed limit. Therefore, the PLDF accelerates only after a higher speed limit sign is passed, staying at the previous set speed until the new sign is reached. In order to accelerate earlier, drivers may use the gas pedal to increase the vehicle speed shortly before the higher speed limit sign, resulting in an *earlier acceleration onto higher speed limit* intervention. The same result can be achieved by increasing the set speed shortly before the higher speed limit, resulting in a *set speed increase before higher speed limit* intervention. However, the gas pedal is predominantly used in these situations and only few set speed increases were recorded. Conversely, when approaching a lower legal speed limit sign, the PLDF predictively decelerates in order to reach the lower speed limit sign at exactly the corresponding speed. Drivers may choose to press the gas pedal or briefly deactivate the PLDF in order to interrupt this deceleration process and keep a higher vehicle speed for longer. This intervention type is named *later deceleration onto lower speed limit*. Earlier acceleration interventions are more common with 6.61 % of the normalized distribution, compared to later decelerations with 3.46 %.

Turning to the less frequent AAL3 intervention types, *adjustments of the acceleration strength* account for only 2.27 % of the normalized distribution. In these interventions, the driver presses the gas pedal to accelerate more strongly in situations where the PLDF already accelerates. This may either happen after a complete vehicle stop, or while accelerating to the current set speed following a lower speed scenario, such as a lower legal speed limit. Crucially, the driver does not aim to initiate the acceleration earlier, nor to exceed the set speed, but only to accelerate stronger to the current set speed.

Interestingly, the least common AAL3 intervention type based on personal preferences is the *adjustment of ACC distance*, accounting for only 1.51 % of the normalized distribution. There, the most common AAL2 intervention is *ACC stronger acceleration*. This intervention typically occurs in stop-and-go traffic, when the leading vehicle accelerates and the PLDF is perceived as accelerating too slowly, prompting the driver to press the gas pedal, closing the gap. The second most common AAL2 intervention type is *ACC lower distance*. It also features a gas pedal intervention to close the gap to the leading vehicle, but this time during vehicle-following scenarios in flowing traffic. The relatively low number of found ACC-focused interventions is noteworthy, given that most related research focuses exclusively on the personalization of ACC target following behavior. There are multiple factors to consider when trying to determine why this might be the case. First, the majority of related research on driving function personalization has specifically selected ACC target following as the primary optimization objective. As a result, their datasets focus exclusively on vehicle-following scenarios. In contrast, only 30.92 % of the used dataset in this study features vehicle-following scenarios, and only 14.3 % of the dataset was recorded during highway driving. Second, in the present study, the drivers were free to select their preferred ACC gap setting, and adjustments to this setting were not counted as driver interventions. If the drivers could not have chosen the gap setting freely, it is likely that the number of driver interventions to adjust the gap to the leading vehicle would have been higher.

In conclusion, it is important to note that the intervention distributions presented in this section are influenced by the frequency of the corresponding driving situations in the dataset. For example, there are generally less roundabouts in a typical commute compared to straight road segments. Consequently, the intervention distribution is inherently skewed toward more common scenarios, such as straight road segments. This discrepancy is addressed in Chapter 4 by analyzing the intervention frequencies separately for each specific location type.

Frequency of Personal Preference-based Driver Interventions over the Course of the Study

During the annotation process of the dataset, it was observed that drivers intervened less frequently based on personal preferences during their first commute, compared to subsequent drives. This pattern is likely attributable to learning effects over the course of the study. The observation is statistically evaluated by calculating the number of personal preference-based interventions per minute for each drive of each participant. Subsequently, the intervention frequency of each participant in each drive iteration is compared to the mean intervention frequency across all remaining drives of the respective participant. The differences in intervention frequencies are calculated for each individual driver and a Shapiro-Wilk test was applied to assess their normality. Due to the varying number of drives per participant, only the first eight drives of each participant are used for the evaluation. The conducted Shapiro-Wilk test finds that all evaluated score differences are normally distributed, as the resulting p-values exceed a threshold of $5\,\%$. Accordingly, a series of Student's t-tests for paired samples is conducted to determine whether the differences in intervention frequency are statistically significant. Only for the first drive of each participant, a significantly lower intervention frequency is observed ($t(17) = 2.578$, $p = 0.020$). No significant differences are found for later drive iterations. These findings support the initial observation that participants are less likely to intervene during their first drive compared to subsequent drives.

3.3.3 Questionnaire Results

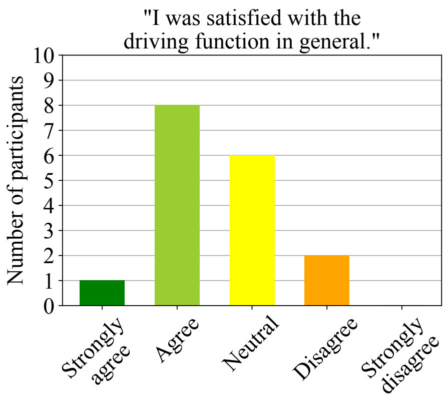
In this section, the questionnaire results are discussed. This includes the quantitative results of the questionnaire's second page, which inquires about the driver's satisfaction with the PLDF, and the qualitative results of the third page, which contains text fields for free-form feedback. The questionnaire used in this study is shown in Appendix A.1. The first questionnaire page is used to collect demographic data and information regarding the participants' experience with ADASs. The age distribution within the dataset is presented in Section 3.1.5, while the remaining information on the first page was not used for further analyses. The primary reason for inquiring about the participants' driving and ADAS experience was to verify that they met the requirement of being sufficiently familiar with the PLDF.

Driver Satisfaction Questionnaire Analysis

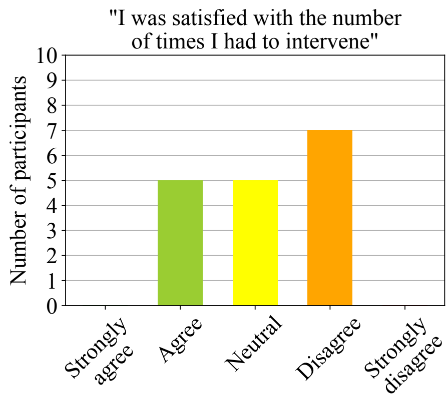
The results of the first four questions of the driver satisfaction questionnaire are illustrated as bar plots in Figure 3.3. Since a Likert scale [Lik32] was used in the questionnaire, the responses are coded from 1.0 for *strongly disagree* to 5.0 for *strongly agree* in the quantitative analysis. As illustrated, drivers are relatively satisfied with the driving function in general, reaching an average score of 3.47 in Figure 3.3a. Most participants responded to the statement that they were generally satisfied with the PLDF with either *agree* or *neutral*. Only one participant strongly agreed, and two participants disagreed with the statement. In Figure 3.3b, the participants gave mixed answers whether they were satisfied with the number of required interventions, reaching an average score of only 2.88. Conversely, drivers unanimously agreed in Figure 3.3c that a reduction of the number of interventions would increase their satisfaction with the system. Finally, in Figure 3.3d, most drivers declared that they were satisfied with the PLDF outside of interventions they intervened in with an average score of 3.94. Only one driver disagreed with this statement and three drivers responded neutrally.

These results show that subjectively drivers would prefer a lower intervention frequency while using the PLDF, and that most drivers are generally satisfied with the PLDF outside of the situations they intervened in. This indicates that a general optimization, or specifically a personalization, of the PLDF based on the feedback contained in driver interventions could be a feasible goal. To evaluate Hypothesis 3.1, a correlation analysis is conducted between the frequency of driver interventions and the general driver satisfaction values taken from Figure 3.3a. Since the satisfaction ratings are ordinal, Spearman's rank correlation coefficient was selected to access the correlation between the satisfaction ratings and the continuous intervention frequency values. Due to the low sample size of 17 participants, Kendall's Tau was additionally calculated, as it is more robust and reliable when dealing with small datasets. Table 3.6 depicts the results of the correlation analysis.

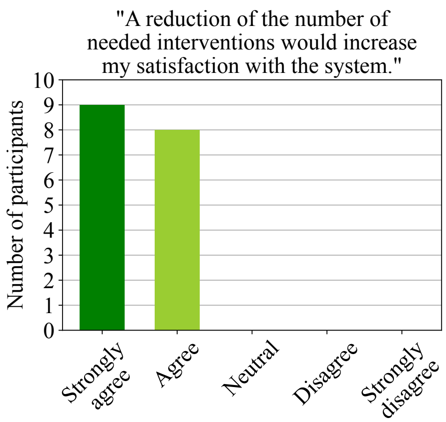
The correlation of the general driver satisfaction with four features was calculated. The *total number of interventions per minute* includes all interventions in the dataset. The other three features contain the intervention frequencies of the different AAL4 labels excluding *other*. A correlation is assumed to be statistically significant if the p-value is smaller than $5\,\%$. The results show a moderate but significant negative correlation between the general driver satisfaction and both the total intervention frequency and the frequency of interventions due to deviating personal preferences. For interventions outside of the ODD and interventions due to incorrect input data, no statistically significant correlation with the reported driver satisfaction values could be found. The found negative correlation of the intervention frequency with the driver satisfaction values confirms Hypothesis 3.1. Drivers who frequently intervened tended to report lower levels of satisfaction with the PLDF, compared to drivers who intervened less often. Moreover, the data indicate that specifically the frequency of interventions based on deviating personal preferences have a significant effect on the driver satisfaction



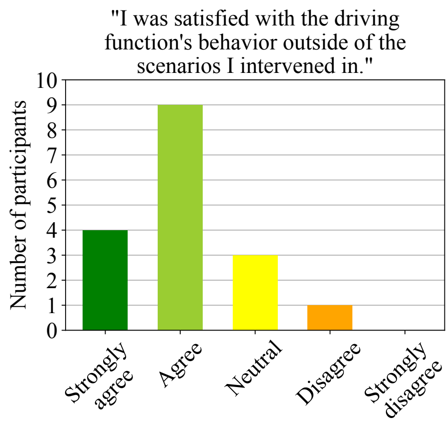
(a) General satisfaction.



(b) Satisfaction with the number of interventions.



(c) Satisfaction with a reduction of interventions.



(d) Satisfaction outside of interventions.

Figure 3.3: Results of the second page questionnaire regarding the driver satisfaction with the PLDF.

compared to other intervention types. One possible explanation for this finding could be that drivers know about the ODD restrictions and are more forgiving about required interventions in these situations. However, if they feel the need to commonly intervene in situations the PLDF was designed to handle, the driver satisfaction with the system is negatively affected. In conclusion, it must be stated that the statistical analyses conducted in this section are based on a rather small sample size of 17 participants. Therefore, Hypothesis 3.1 is evaluated again with more participants in the second test group study of this thesis, covered in Chapter 6.

Table 3.6: Spearman and Kendall correlation coefficients and p-values of the driver satisfaction and the intervention frequencies of the different AAL4 labels. Significant p-values are marked in bold text.

Feature	Spearman ρ	Spearman p-value	Kendall τ	Kendall p-value
Total number of interventions per minute	-0.602	**0.011**	-0.465	**0.019**
Number of personal preference interventions per minute	-0.593	**0.012**	-0.447	**0.025**
Number of interventions outside ODD per minute	-0.315	0.219	-0.268	0.177
Number of incorrect input data interventions per minute	-0.005	0.984	-0.018	0.928

Free-Form Feedback Analysis

On the third page of the used questionnaire, participants had the possibility to provide free-form feedback about the PLDF. In their feedback, drivers commonly described the driving situations in which they were dissatisfied with the driving function and therefore intervened. The described situations generally align with the analysis results of the annotated intervention types in the dataset. Drivers also commonly described the system behavior as too rigid and machine-like, combining too defensive and too aggressive elements. The PLDF always drives at exactly the legal speed limit outside of high-curvature situations, which is commonly described as too defensive, and many drivers would have preferred a slightly higher speed in many situations. However, the strength of accelerations and decelerations onto upcoming lower and higher speed limits are described as quite strong, while the acceleration and deceleration timings are described as rather defensive. Some drivers also mentioned that they miss coasting behavior, e.g., when approaching a lower speed limit sign. Instead, the PLDF is designed to always use the brakes in these situations. To summarize, many drivers would have preferred a smoother acceleration and deceleration profile instead. Regarding the interactions with surrounding traffic, many drivers described that the PLDF's driving style often did not reflect the surrounding traffic flow. Commonly, the PLDF exhibited behavior that made the drivers feel like a traffic obstacle. Especially the following scenario was commonly described by participants: The PLDF decelerates too strongly and early onto an upcoming lower speed limit, while the surrounding traffic flows much more smoothly and decelerates more slowly onto the lower speed limit. Still, many drivers see the benefits of the function, especially outside of urban scenarios. However, due to its limited ODD, the PLDF is described as not suited for high traffic density urban driving.

On the third page, drivers could also describe the situations in which they were dissatisfied with the PLDF's behavior but chose not to intervene. Most drivers answered that they generally adjusted behavior in which they were dissatisfied with the PLDF. Some drivers reported that they did not correct small or brief deviations from their desired behavior. Specifically noteworthy are the answers of the four drivers who reported a low satisfaction with the PLDF outside of scenarios they intervened in. One of these drivers reported that the acceleration and deceleration profile was still too rigid in most situations, as described above. Another driver with an especially long commute reported becoming fatigued by the high frequency of required interventions and chose to accept the PLDF's behavior after a certain time. Similarly, a third participant reported that intervening and correcting the system sometimes felt futile, as the PLDF would continue to drive in the same manner afterward.

The gathered feedback shows that the drivers would generally use the system during their commutes, mostly outside of urban scenarios. However, they feel that many open points should still be addressed. The feedback provided by the driver interventions can be useful to derive possible adaptations to the PLDF, especially because drivers describe that they are mostly satisfied with the PLDF outside of the situations they intervened in. Only four drivers did not agree with this statement, stating that they sometimes were dissatisfied even outside of these situations. Two of these participants expressed fatigue with the frequent need to intervene, noting that their interventions do not lead to long-term changes in the PLDF's behavior. However, if the PLDF was adjusted and potentially personalized based on these driver interventions, this issue could be mitigated.

3.4 Summary and Contributions

In this chapter, a test group study was conducted with 17 participants where each driver intervention was manually annotated by the drivers during naturalistic driving. Based on these ground truth annotations, a four-level hierarchical label taxonomy was introduced, which describes the reasons behind driver interventions in a bottom-up manner. The results show that drivers frequently intervened while using the PLDF, and that $53.40\,\%$ of these interventions are based on deviating personal preferences of the drivers in situations which fall within the PLDF's ODD. This indicates a high optimization potential of the driving function, since the drivers commonly intervened and adjusted the system behavior in situations which the PLDF was designed to manage. By intervening, the drivers provided feedback about their preferred function behavior in these situations. The most common AAL3 intervention types due to personal preferences were *speed adjustment on straight roads*, *speed adjustment in high-curvature segments*, and *adjustment of the acceleration timing onto speed limits*. These three categories account for almost $50\,\%$ of the normalized distribution of all interventions in the dataset.

Notably, *adjustments of ACC distance* make up less than 2 % of interventions, although most related research on longitudinal driving function personalization focuses solely on ACC driving. The analysis of the questionnaire results shows that drivers would subjectively prefer a reduction of the needed driver interventions while using the PLDF. This was further strengthened by the results of a correlation analysis between the reported general driver satisfaction and the intervention frequency of each driver. The results indicate that both the general intervention frequency, and especially the frequency of interventions based on personal preferences, are negatively correlated to the reported driver satisfaction. Therefore, drivers who intervene more often are less satisfied with the driving function than drivers who intervene less often. Most drivers also stated that they were generally satisfied with the PLDF's behavior outside of scenarios they intervened in. This further indicates that reducing the intervention frequency of drivers while using the PLDF could increase the driver satisfaction with the system. However, it must be stated that the results of this study are based on a small sample size of only 17 participants. Therefore, further research is needed to fully prove this correlation and investigate the effects of potential function adjustments.

To summarize, the research contributions of this chapter are as follows:

- An analysis of the ground truth reasons behind driver interventions during naturalistic driving with the PLDF.
- The introduction and application of a four-level hierarchical label taxonomy for driver interventions based on the recorded dataset.
- A correlation analysis showing a significant negative correlation between the driver intervention frequency and the reported satisfaction with the PLDF.

4 Data Analysis: Necessary Adjustments based on Driver Feedback

This chapter focuses on the derivation of necessary adjustments to the PLDF based on the results of the test group study in Chapter 3. The goal of these adjustments is the reduction of the intervention frequency based on deviating personal preferences. Therefore, this chapter builds upon the research objectives defined in Section 2.6.2. Based on the created intervention dataset, a data-driven method is developed to derive necessary adjustments to the PLDF. The developed methodology focuses on whether a general adjustment of the PLDF is required for all drivers, or if personalized changes are necessary. It is also investigated whether the necessary adjustments can be represented by a generalized driving policy. The developed methodology and the results of its application to the recorded dataset were published in a research paper [Sch+24b].

This chapter is structured as follows: In Section 4.1, the underlying methodology of the developed framework for the derivation of necessary adjustments is explained. In Section 4.2, this framework is then applied to the recorded dataset from Chapter 3, and the results are presented. Finally, in Section 4.3, this chapter is summarized and the resulting research contributions are highlighted.

4.1 Methodology

In this section, the developed methodology for the derivation of necessary adjustments to the PLDF is introduced. The methodology builds upon the results from Chapter 3, which state that drivers intervene in situations where they are dissatisfied with the PLDF. Accordingly, adjustments to the PLDF are primarily necessary in the specific situations where drivers consistently intervene. As mentioned in Section 3.3.2, the presented intervention distributions contain general trends, but they are skewed toward the most common situations in the dataset. Therefore, the intervention frequencies must be put into the context of their corresponding commute by calculating the individual Intervention Rate (IR) in each specific situation for each driver separately. Based on these IRs, necessary adjustments to the PLDF can be derived.

The methodology is structured as follows: First, the intervention types relevant to the adjustment are defined. Second, potential approaches for adapting the PLDF's behavior are discussed. Third, the aforementioned IRs are calculated by filtering and aggregating all situations in the dataset where a relevant intervention was possible. Finally, an algorithm is presented for the derivation of necessary adjustments to the PLDF based on the calculated IRs.

4.1.1 Definition of Relevant Intervention Types

In Chapter 3, the driver interventions are classified based on their underlying driver intention and the situations in which they occur. The interventions based on deviating personal preferences by the drivers within the PLDF's ODD are chosen as the input for the function adjustment, because these voluntary interventions show the drivers' preferences in situations that the PLDF was designed to manage without driver interventions. The situations in which these personal preference-based interventions occur can be split into free-driving scenarios without relevant traffic interactions, and traffic-related interventions especially involving a potential leading vehicle. The interventions involving the PLDF's behavior in traffic-related scenarios are the AAL3 *adjustment of ACC distance* interventions and the AAL2 *stronger acceleration after stop* interventions. These intervention types make up less than 3 % of the normalized distribution. In contrast, the remaining interventions based on personal preferences in free-driving scenarios account for around 50 % of driver interventions in the dataset. These interventions generally occur only in specific locations, such as straight roads, roundabouts, and speed limit changes. Due to their prevalence in the dataset, the function adjustment in this chapter focuses on these location-based interventions. This includes the three AAL3 labels *speed adjustment on straight roads*, *speed adjustment in high-curvature segments*, and *adjustment of the acceleration timing onto speed limits*. Only the location-based AAL2 intervention type *stronger acceleration to current set speed* is excluded, since it is exceedingly rare with only 33 instances in the dataset, even though its corresponding driving situation is highly common. The resulting six relevant locations for the function adjustment, *Straight Road (SR)*, *Speed Limit Increase (SLI)*, *Speed Limit Decrease (SLD)*, *Roundabout (R)*, *Curve (C)*, and *Turn (T)*, are depicted in Table 4.1 with their corresponding AAL2 labels. In these locations, the drivers may either increase or decrease the vehicle speed using the gas pedal, the brake pedal, or the set speed. As can be seen, the AAL3 label *speed adjustment on straight roads* is directly used as the SR location. However, the other AAL3 labels were split into their specific locations. *Speed adjustment in high-curvature segments* was split into R, C, and T, and *adjustment of the acceleration timing onto speed limits* was split into SLI and SLD. The location abbreviations R, C, and T are primarily used in tables and figures. However, for better readability, they will not be abbreviated in the main text.

4.1.2 Potential Driving Function Adjustment Approaches

In Figure 4.1, the general information processing pipeline of the PLDF is illustrated. The output of the PLDF, highlighted in green, is the calculated speed profile, which serves as the focus of the adjustments discussed in this chapter. As inputs, the PLDF receives information from both the perception stack and map data. The map data includes

Table 4.1: The chosen intervention locations for the driving function adjustment and their corresponding AAL2 labels.

Location	Abbreviation	AAL2 Labels
Straight Road	SR	Set speed increase on straight road
		Set speed decrease on straight road
		Straight road acceleration
		Straight road deceleration
Speed Limit Increase	SLI	Earlier acceleration onto higher speed limit
		Set speed increase before higher speed limit
Speed Limit Decrease	SLD	Later deceleration onto lower speed limit
Roundabout	R	Roundabout higher speed
		Roundabout lower speed
Curve	C	Curve higher speed
		Curve lower speed
Turn	T	Right turn higher speed
		Right turn lower speed
		Left turn higher speed
		Left turn lower speed

information such as speed limits, road curvature, and intersection data, and may be enriched with additional data if necessary. The perception algorithms used mainly include the TSD and the detection of potential leading vehicles. While adjustments of the perception algorithms are theoretically feasible, they fall within a different area of research and are therefore beyond the scope of this thesis.

The input information is then processed in the PLDF to generate an appropriate speed profile. The PLDF follows a general design that defines its ODD and underlying algorithms. The behavior of these algorithms is defined by a set of parameters, which are specified during the development process. The PLDF's general design should not be

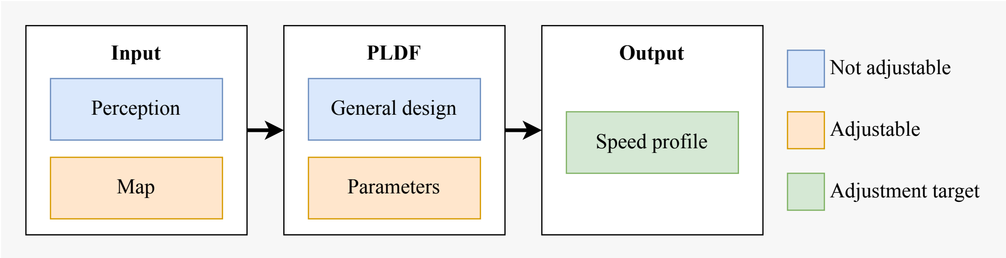


Figure 4.1: General PLDF information processing pipeline. The goal of this chapter is the adjustment of the resulting speed profile according to the driver's preferences, marked in green. Adjustable and not-adjustable components are highlighted in orange and blue, respectively.

changed either, as the objective of this thesis is not to develop a new driving function, but instead to optimize and, if necessary, personalize the existing driving function. Instead, as discussed in Section 2.3, driving functions are commonly personalized via parameter adjustments, provided that their parameterization is accessible. By adjusting these parameters, the general driving policy is altered globally, i.e., in all related situations. For example, a driver who generally prefers higher curve speeds would benefit from an adjustment of the PLDF's maximum lateral acceleration parameter in curves. By changing this parameter, the PLDF's driving behavior in all curves can be adjusted. However, if a driver is generally satisfied with the velocities in curves but only intervenes in a few specific locations, then a general parameter adjustment would not be feasible for them. Instead, the map in these specific locations should be enriched with additional information. This way, the PLDF's behavior can be adjusted only in specific locations instead of globally in all similar locations. It must be noted that such map enrichments are only feasible for location-based intervention types. Traffic-related interventions are not location-specific and therefore only an adjustment of the general driving policy would be feasible in these cases. However, in this chapter, the focus lies solely on location-based intervention types, as explained in Section 4.1.1.

To summarize, without changing the general design of the PLDF or its perception stack, parameter adjustments and location-based map adjustments are the main ways in which the behavior of the PLDF may be changed. This defines the first dimension of possible driving function adjustments: the *location dependency*. There, either *global* parameter adjustments or *local* map adjustments are possible. The second dimension of possible adjustments is the *individualization*. As explained in Section 2.3, it is generally accepted in the related work that individual driving styles differ and that personalization is needed to adjust driving functions to the preferences of individual drivers. However, personalization is only needed if the individual drivers' preferences truly differ from each other. For example, if every driver in the recorded dataset would prefer to accelerate earlier onto upcoming higher speed limits, then no individualization would be necessary for this intervention type. Instead, all drivers would benefit from the same adjustment. This type of adjustment is then called a *collective* adjustment. However, if some drivers prefer an earlier acceleration onto upcoming higher speed limits, and other drivers do not, then a personalization is necessary, resulting in an *individual* adjustment. By combining the two dimensions for driving function adjustments, location dependency and individualization, the matrix depicted in Figure 4.2 is created. The resulting driving function adjustment strategies are defined as follows:

<table>
<tr><th colspan="2" rowspan="2"></th><th colspan="2">Location Dependency</th></tr>
<tr><th>Global</th><th>Local</th></tr>
<tr><td rowspan="2">Individu-
alization</td><td>Collective</td><td>Global Collective (GC)
Adjust general function parameters for everyone.</td><td>Local Collective (LC)
Enrich map information for everyone.</td></tr>
<tr><td>Individual</td><td>Global Individual (GI)
Adjust general function parameters for individual drivers.</td><td>Local Individual (LI)
Enrich map information for individual drivers.</td></tr>
</table>

Figure 4.2: The four potential approaches of adjusting a map-based driving function's generated speed profile.

- *Global Collective (GC)* describes a general adjustment of the PLDF's parameters for all drivers. GC optimization is generally the goal of function application during development, since it aims to optimize the driving function's behavior for all drivers in all situations. A GC adjustment is applicable if, e.g., all drivers generally prefer higher velocities in all curves.
- *Global Individual (GI)* describes an adjustment of the PLDF's parameters for an individual driver. All related driving function personalization approaches analyzed in Section 2.3 conduct a GI adjustment. A GI adjustment is applicable if, e.g., only individual drivers generally prefer higher curve speeds in all curves.
- *Local Collective (LC)* describes an adjustment needed only at a specific map location for all drivers. A LC adjustment is applicable if, e.g., all drivers prefer a higher speed only in a few specific curves. This could also indicate wrong or missing information in the map at the corresponding location, e.g., a wrong curvature representation, since all drivers unanimously intervene there.
- *Local Individual (LI)* describes an adjustment needed for an individual driver at a specific map location. A LI adjustment is applicable if, e.g., a specific driver prefers a higher speed in only a few specific curves. However, since not all other drivers also intervene in these locations, the adjustment is only applicable for this individual driver.

Based on the intervention behavior of the individual drivers, the adequate adjustment strategy for each location type can be derived. This means that the proposed framework must be applied to each location type separately. For example, a driver who always increases the set speed by $3\,\mathrm{km/h}$ on straight roads needs a GI adjustment only for straight roads. If they also consistently intervene in, e.g., two out of six specific curves in their commute by pressing the brake, they additionally need a LI adjustment in these curves.

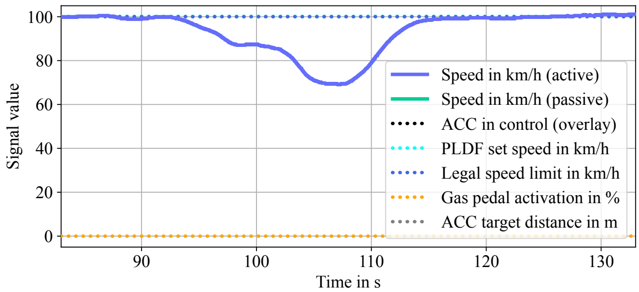
4.1.3 Calculation of Intervention Rates

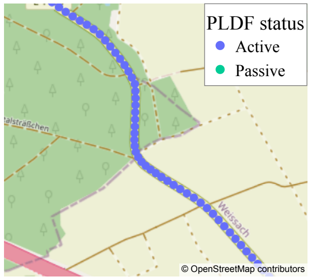
In order to derive the adequate adjustment strategy for each location type, the local and global IRs of each individual driver must be calculated. An IR describes the frequency with which drivers intervene in specific locations when it was possible to do so, and they can be used to judge the consistency of driver behavior. The IRs are always calculated separately for speed increases and speed decreases in each location. In order to calculate the specific IRs, first the dataset must be preprocessed. This preprocessing has the goal of finding all drive-throughs of the chosen location type and assigning them with one of the following labels:

- An intervention was possible and the driver did not intervene.
- An intervention was possible and the driver intervened by increasing the vehicle speed.
- An intervention was possible and the driver intervened by decreasing the vehicle speed.
- An intervention was not possible.

Based on these intervention possibility labels, the IRs are calculated. First, a specific location type is chosen for the analysis. Then, the dataset is filtered to extract all drive-throughs of the selected location type. Custom filters were developed to identify the relevant situations by analyzing both the PLDF's bus signals and map information. For example, roundabouts and turns can be directly read from the map data, speed limit changes are detected via the TSD signal, and sharp curves are detected by calculating the PLDF's maximum curve speed from the road curvature and comparing it to the current set speed. After the relevant locations are extracted from the dataset, it is evaluated whether it was possible for the driver to conduct the corresponding intervention during each drive-through. This evaluation mainly focuses on whether surrounding traffic, such as slower leading vehicles, influenced the driver. Only free-driving scenarios, which are not influenced by surrounding traffic, are used for the calculation of the IRs.

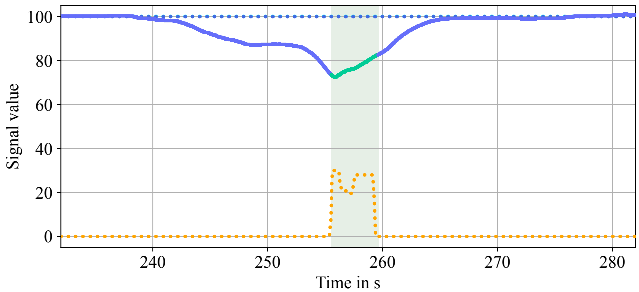
Figure 4.3 contains examples for three of the possible categories each drive-through can be assigned to. In Figure 4.3a and Figure 4.3b, a free drive on a high-curvature road segment is depicted. In this situation, no surrounding traffic limited the driver and they could have intervened either by pressing the gas pedal or the brake pedal. However, in this case, they chose not to do so. In Figure 4.3c and Figure 4.3d, the same location

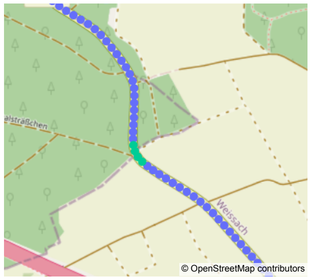




(a) A free-driving scenario through a curve without a driver intervention. The PLDF decelerates due to a high-curvature road segment. In this situation, a gas or brake pedal intervention would have been possible, but the driver did not choose to do so.

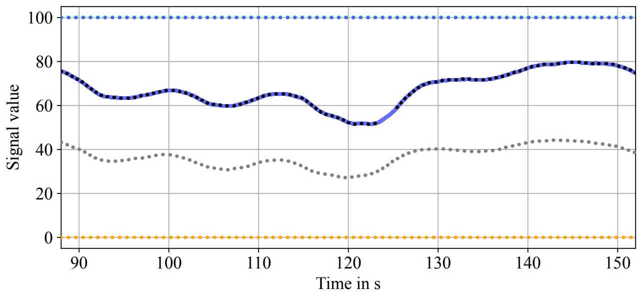
(b) The corresponding GPS traces of the curve drive-through without an intervention.

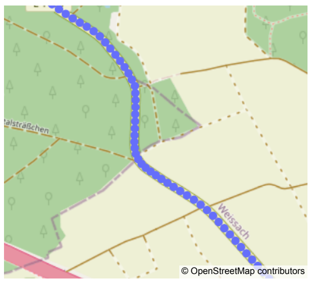




(c) A free-driving scenario through a curve with a driver intervention. In this case, an intervention was again possible and the driver chose to intervene by pressing the gas pedal.

(d) The corresponding GPS traces of the curve drive-through with a gas pedal intervention.





(e) A vehicle-following scenario in the same high-curvature road segment. In this case, no gas pedal intervention was possible since the driver followed a slower leading vehicle. Pressing the gas pedal in this situation could potentially lead to a hazardous situation.

(f) The corresponding GPS traces of the curve drive-through with a slower leading vehicle.

Figure 4.3: Example plots depicting the three intervention possibility categories in the same curve location.

is depicted during another free drive without a leading vehicle. This time, the driver chose to intervene by pressing the gas pedal. Finally, Figure 4.3e and Figure 4.3f depict a drive-through in the same location with a significantly slower leading vehicle. As illustrated by the black dotted overlay above the vehicle speed signal, the ACC limits the PLDF's velocity in order to keep a safe distance to the leading vehicle. Accordingly, the driver was not able to intervene in this situation by pressing the gas, even if they would have wanted to do so. By pressing the gas in this situation, the driver could have possibly caused a hazardous situation. Conversely, it is not conclusive whether the driver would have wanted to use the brake to decelerate in this curve, since the leading vehicle already slowed them down. Therefore, situations involving slower leading vehicles are not used for the calculation of the IRs.

After data filtering, the drive-throughs of each specific location type are clustered based on their GPS locations and the vehicle heading. First, the Euclidean distances between all found drive-throughs, where an intervention was possible, are calculated. Drive-throughs with a Euclidean distance below a specific threshold are then grouped together. The used thresholds were determined empirically through qualitative analysis of the generated clusters. For all location types except straight roads, a threshold of $100\,\mathrm{m}$ was used. However, different interventions on the same straight road may occur across longer distances. Therefore, a higher threshold of $1000\,\mathrm{m}$ is applied, in addition to a postprocessing step that analyzes the underlying map data and separates found clusters that occur on different road segments. The found clusters are then split based on their drive-through direction using Density-Based Spatial Clustering of Applications with Noise (DBSCAN). In locations on continuous roads, the clusters are simply split into the two opposite directions of the road, one cluster for each direction. In locations where multiple roads intersect, such as roundabouts and turns, different clusters are generated based on the direction the location is entered and left. This way, the same roundabout or intersection may contain multiple separate clusters, depending on where the drivers entered and left it. Finally, numerical identifiers are assigned to the resulting unique location clusters. Via this preprocessing, a dataset is generated that contains all drive-throughs of a specific location type where an intervention was possible. Both a driver identifier and a location identifier are assigned to each of these drive-throughs. Using this dataset, four different types of speed increase IRs and speed decrease IRs can be calculated for each location type, based on the four function adjustment strategies introduced in Section 4.1.2:

- $\mathrm{IR}^{+}_{\mathrm{GC}}$ and $\mathrm{IR}^{-}_{\mathrm{GC}}$ of all drivers in all locations.
- $\mathrm{IR}^{+}_{\mathrm{GI}}$ and $\mathrm{IR}^{-}_{\mathrm{GI}}$ of each individual driver in all locations.
- $\mathrm{IR}^{+}_{\mathrm{LC}}$ and $\mathrm{IR}^{-}_{\mathrm{LC}}$ of all drivers in each specific location.
- $\mathrm{IR}^{+}_{\mathrm{LI}}$ and $\mathrm{IR}^{-}_{\mathrm{LI}}$ of each individual driver in each specific location.

Each IR is calculated as the ratio between the number of drive-throughs where a driver intervened by increasing or decreasing the vehicle speed and the total number of drive-throughs where such an intervention would have been possible. The speed increase IRs are denoted as IR^{+}, while speed decrease IRs are denoted as IR^{-}. Let

$\mathcal{D} = \{0, 1, \ldots, N_{\mathcal{D}} - 1\}$ be the set of driver indices, where $N_{\mathcal{D}}$ denotes the total number of drivers with $N_{\mathcal{D}} = 17$. Let $\mathcal{L} = \{0, 1, \ldots, N_{\mathcal{L}} - 1\}$ be the set of location indices, where $N_{\mathcal{L}}$ denotes the total number of distinct location clusters per location type. Throughout, $j \in \mathcal{D}$ denotes a driver index and $k \in \mathcal{L}$ a location index. Let $c^{+}_{j,k}$ represent the number of drive-throughs with a speed increase intervention by driver j at location k. Let $c^{-}_{j,k}$ represent the number of drive-throughs with a speed decrease intervention by driver j at location k. Let $c^{\text{total}}_{j,k}$ represent the total number of drive-throughs by driver j at location k where an intervention was possible.

$\mathrm{IR}^{+}_{\mathrm{GC}}$ and $\mathrm{IR}^{-}_{\mathrm{GC}}$ are calculated by aggregating the respective interventions on all drive-throughs across all drivers and all locations:

$$\mathrm{IR}^{+}_{\mathrm{GC}} = \frac{\sum_{j\in\mathcal{D}} \sum_{k\in\mathcal{L}} c^{+}_{j,k}}{\sum_{j\in\mathcal{D}} \sum_{k\in\mathcal{L}} c^{\text{total}}_{j,k}} \tag{4.1}$$

$$\mathrm{IR}^{-}_{\mathrm{GC}} = \frac{\sum_{j\in\mathcal{D}} \sum_{k\in\mathcal{L}} c^{-}_{j,k}}{\sum_{j\in\mathcal{D}} \sum_{k\in\mathcal{L}} c^{\text{total}}_{j,k}} \tag{4.2}$$

$\mathrm{IR}^{+}_{\mathrm{GI}}$ and $\mathrm{IR}^{-}_{\mathrm{GI}}$ are computed for each driver individually, aggregating across all locations:

$$\mathrm{IR}^{+}_{\mathrm{GI}}(j) = \frac{\sum_{k\in\mathcal{L}} c^{+}_{j,k}}{\sum_{k\in\mathcal{L}} c^{\text{total}}_{j,k}} \quad \text{for each } j \in \mathcal{D} \tag{4.3}$$

$$\mathrm{IR}^{-}_{\mathrm{GI}}(j) = \frac{\sum_{k\in\mathcal{L}} c^{-}_{j,k}}{\sum_{k\in\mathcal{L}} c^{\text{total}}_{j,k}} \quad \text{for each } j \in \mathcal{D} \tag{4.4}$$

$\mathrm{IR}^{+}_{\mathrm{LC}}$ and $\mathrm{IR}^{-}_{\mathrm{LC}}$ are computed for each specific location, aggregating across all drivers:

$$\mathrm{IR}^{+}_{\mathrm{LC}}(k) = \frac{\sum_{j\in\mathcal{D}} c^{+}_{j,k}}{\sum_{j\in\mathcal{D}} c^{\text{total}}_{j,k}} \quad \text{for each } k \in \mathcal{L} \tag{4.5}$$

$$\mathrm{IR}^{-}_{\mathrm{LC}}(k) = \frac{\sum_{j\in\mathcal{D}} c^{-}_{j,k}}{\sum_{j\in\mathcal{D}} c^{\text{total}}_{j,k}} \quad \text{for each } k \in \mathcal{L} \tag{4.6}$$

$\mathrm{IR}^{+}_{\mathrm{LI}}$ and $\mathrm{IR}^{-}_{\mathrm{LI}}$ are calculated for each driver at each specific location:

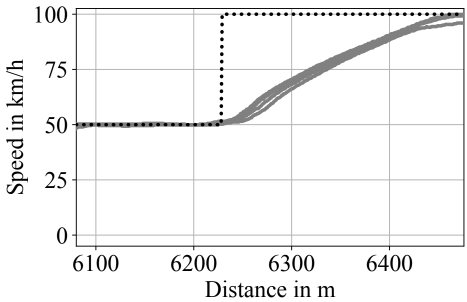
$$\mathrm{IR}^{+}_{\mathrm{LI}}(j,k) = \frac{c^{+}_{j,k}}{c^{\text{total}}_{j,k}} \quad \text{for each } j \in \mathcal{D},\ k \in \mathcal{L} \tag{4.7}$$

$$\mathrm{IR}^{-}_{\mathrm{LI}}(j,k) = \frac{c^{-}_{j,k}}{c^{\text{total}}_{j,k}} \quad \text{for each } j \in \mathcal{D},\ k \in \mathcal{L} \tag{4.8}$$

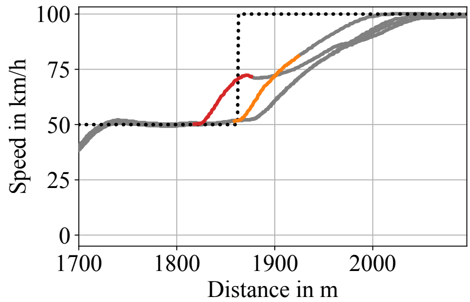
Examples for LI driver behavior with different resulting IRs are illustrated in Figure 4.4. There, multiple drive-throughs of individual drivers in three different SLI locations are depicted. Each location is traversed four separate times. In Figure 4.4a, the participant drives through the same SLI location four times and never intervenes. Therefore, their $\mathrm{IR}^{+}_{\mathrm{LI}}$ and $\mathrm{IR}^{-}_{\mathrm{LI}}$ in this location equal $0\,\%$, i.e., no adjustment to the driving function's behavior is necessary here. In Figure 4.4b, another SLI location is depicted. There, the driver intervenes in two out of four drive-throughs via the gas pedal, resulting in a $\mathrm{IR}^{+}_{\mathrm{LI}}$ of $50\,\%$ and a $\mathrm{IR}^{-}_{\mathrm{LI}}$ of $0\,\%$. In this case, the driver behavior is inconsistent and it is unclear whether they would prefer an adjustment of the PLDF's behavior here. Finally, in Figure 4.4c, a driver intervenes four out of four times by pressing the gas pedal while driving through the SLI location, resulting in a $\mathrm{IR}^{+}_{\mathrm{LI}}$ of $100\,\%$ and a $\mathrm{IR}^{-}_{\mathrm{LI}}$ of $0\,\%$. Accordingly, the PLDF's behavior should be adjusted in this location for this specific driver.

4.1.4 Algorithm for the Derivation of Necessary Adjustments

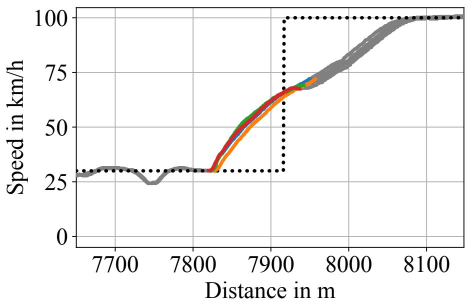
Based on the calculated IRs, necessary adjustments of the PLDF are derived using the algorithm illustrated in Figure 4.5. The algorithm uses IR thresholds to define the required consistency of driver intervention behavior for an adjustment to be considered applicable or not applicable. Relatively high values, close to $100\,\%$, should be chosen for the upper thresholds $\tau^{\mathrm{high}}_{\mathrm{GC}}$, $\tau^{\mathrm{high}}_{\mathrm{GI}}$, $\tau^{\mathrm{high}}_{\mathrm{LC}}$, and $\tau^{\mathrm{high}}_{\mathrm{LI}}$. These upper thresholds define the minimum IR necessary for an adjustment to be applicable. Conversely, relatively low values, close to $0\,\%$, should be chosen for the lower thresholds $\tau^{\mathrm{low}}_{\mathrm{GC}}$, $\tau^{\mathrm{low}}_{\mathrm{GI}}$, $\tau^{\mathrm{low}}_{\mathrm{LC}}$, and $\tau^{\mathrm{low}}_{\mathrm{LI}}$. These lower thresholds define the maximum IR until no adjustment of the PLDF's behavior is necessary in the respective situation. However, no objectively correct threshold exists and the chosen thresholds should reflect the available sample size in the dataset. For example, relatively high sample sizes are available for the driver-focused cases GC and GI. Therefore, stricter thresholds, i.e., close to the ideal values of $0\,\%$ and $100\,\%$, should be chosen there. In contrast, the location-focused cases, LC and LI, feature significantly lower sample sizes due to the underlying study design. In the case of LI, the sample sizes mostly range from one to five valid drive-throughs per location and driver. Consequently, less strict thresholds must be chosen in these cases, i.e., higher values for $\tau^{\mathrm{low}}_{\mathrm{LI}}$ above the ideal value of $0\,\%$, and lower values for $\tau^{\mathrm{high}}_{\mathrm{LI}}$ below the ideal value of $100\,\%$. Another factor that should be accounted for is that drivers intervene less frequently in their first commute compared to subsequent drives, as was shown in Section 3.3.2. Additionally, related work describes a so-called *complacency component* of human behavior, i.e., drivers do not intervene if the difference between the desired behavior and the actual behavior is too small [Hoc01; Spe+22; WXC14]. Both of these effects may cause the driver to intervene less frequently, although they would prefer an adjustment of the PLDF's behavior in the respective situation. Hence, the upper thresholds $\tau^{\mathrm{high}}_{\mathrm{GC}}$, $\tau^{\mathrm{high}}_{\mathrm{GI}}$, $\tau^{\mathrm{high}}_{\mathrm{LC}}$, and $\tau^{\mathrm{high}}_{\mathrm{LI}}$ may deviate more strongly from their ideal value of $100\,\%$ compared to the deviation of the lower thresholds $\tau^{\mathrm{low}}_{\mathrm{GC}}$, $\tau^{\mathrm{low}}_{\mathrm{GI}}$, $\tau^{\mathrm{low}}_{\mathrm{LC}}$, and $\tau^{\mathrm{low}}_{\mathrm{LI}}$



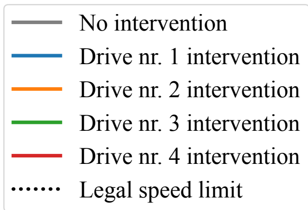
(a) A SLI location with a IR^{+}_{LI} of 0 % and a IR^{-}_{LI} of 0 %. The participant drove through this specific location four times and never intervened.



(b) A SLI location with a IR^{+}_{LI} of 50 % and a IR^{-}_{LI} of 0 %. The participant intervened in two out of four drive-throughs by pressing the gas pedal in this specific location.



(c) A SLI location with a IR^{+}_{LI} of 100 % and a IR^{-}_{LI} of 0 %. The participant drove through this specific location four times and intervened in all four drive-throughs by pressing the gas pedal.



(d) Legend of the speed-over-distance plots.

Figure 4.4: Example plots of three SLI locations with different LI IRs. Each location plot contains the driving data of one individual driver. The speed signal of each drive-through is depicted by a gray line when there was no intervention. Segments of the vehicle speed where the driver intervened are colored. The legend of all three location plots is depicted in Figure 4.4d.

from their ideal value of 0 %. The allowed amount of deviation from the ideal thresholds should be negatively proportional to the available sample sizes, since single outliers may have a stronger effect on the resulting IRs when lower sample sizes are used, compared to cases with high sample sizes. This adjusted threshold choice is reflected during the application of the algorithm in Section 4.2.

First, IR^{+}_{GC} and IR^{-}_{GC} of the chosen location type are calculated, as explained in the previous section. If both IR^{+}_{GC} and IR^{-}_{GC} are lower than τ^{low}_{GC}, drivers in the dataset rarely or never conducted personal preference-based interventions in this location type. Therefore, no adjustment of the PLDF's behavior is necessary. Conversely, if either IR^{+}_{GC} or IR^{-}_{GC} is higher than the threshold τ^{high}_{GC}, a GC adjustment should be conducted for this location type, since almost all drivers consistently intervened similarly in all

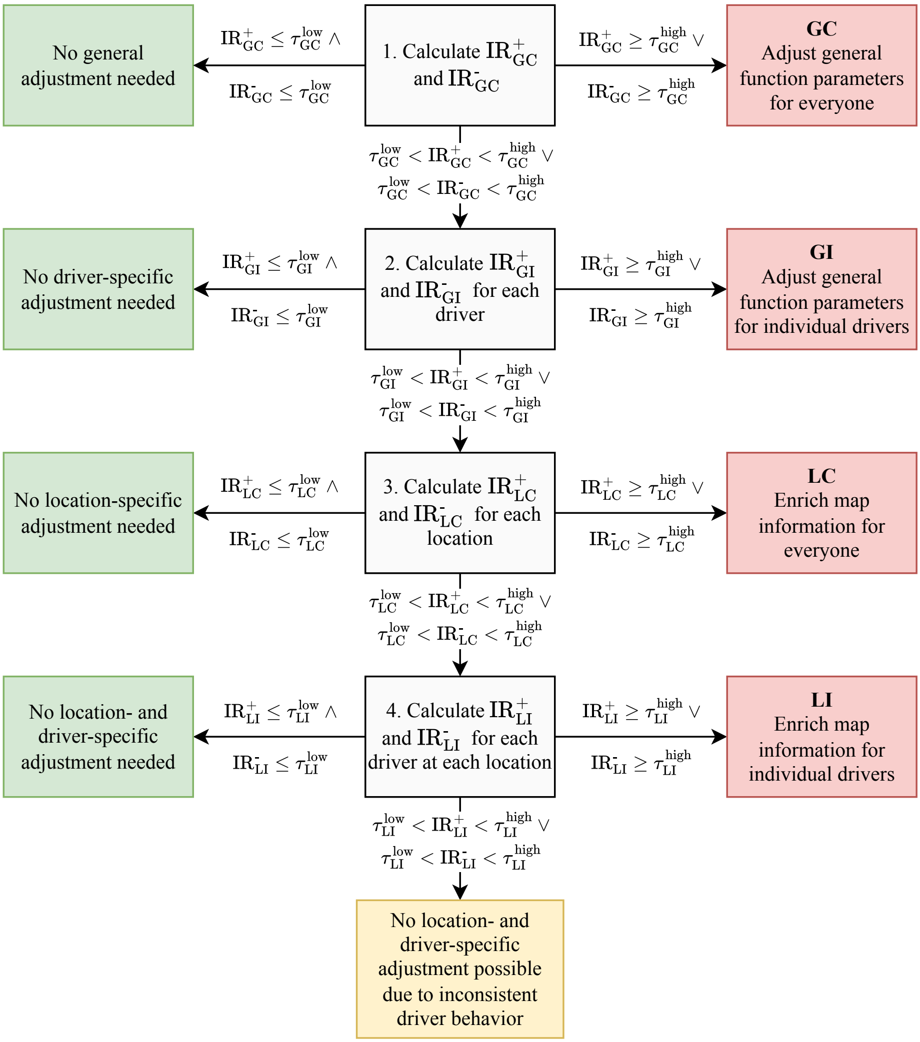


Figure 4.5: Algorithm for the derivation of necessary adjustments to the PLDF based on the calculated IRs. The algorithm is applied separately to each location type. Depending on the chosen thresholds, both expressions on the right-hand side and the bottom of a gray block may be satisfied. In this case, the right-hand side path takes precedence. The formulas in the figure are presented as shown to limit visual complexity.

corresponding drive-throughs in the dataset. If IR^{+}_{GC} reaches τ^{high}_{GC}, the PLDF's velocity in this location type should generally be increased, whereas a speed decrease is applicable if IR^{-}_{GC} reaches τ^{high}_{GC}. However, if the value of IR^{+}_{GC} or IR^{-}_{GC} falls between both thresholds, then no GC adjustment is applicable and the possibility of a GI adjustment is evaluated next.

IR^{+}_{GI} and IR^{-}_{GI} are calculated for each driver separately. If one of both IRs of individual drivers reaches τ^{high}_{GI}, a GI adjustment is applicable for them. Conversely, if their IRs do not exceed τ^{low}_{GI}, no adjustment to the PLDF's behavior is necessary for these drivers, since they almost never intervene in the respective location. For drivers who reach IRs between the upper and lower threshold, no GI adjustment is applicable because they intervene too inconsistently on a global scale. Therefore, the applicability of location-focused adjustments is evaluated for these remaining drivers.

Next, IR^{+}_{LC} and IR^{-}_{LC} are calculated for all corresponding locations in the dataset separately. Again, if the local IR of any location reaches the upper threshold τ^{high}_{LC}, a LC adjustment is applicable, since all drivers in the dataset unanimously and consistently intervene in this specific location. If neither IR^{+}_{LC} nor IR^{-}_{LC} exceed τ^{low}_{LC}, no adjustment is needed in the respective location. However, if drivers inconsistently intervene in some locations, the possibility of a LI adjustment is evaluated. Depending on the underlying dataset, the number of locations with drive-throughs by multiple participants may be limited. If the sample size of drivers at a given location is insufficient, the evaluation of a LC adjustment may be skipped in order to avoid drawing unreliable or invalid conclusions.

Finally, if neither a GI adjustment is applicable for a given driver, nor the LC adjustment sufficiently addresses all relevant locations, a potential LI adjustment is evaluated. The respective IRs, IR^{+}_{LI} and IR^{-}_{LI}, are calculated for each driver at each location separately. The PLDF's speed profile is then adjusted accordingly in locations where the corresponding participant consistently intervened with an IR greater than τ^{high}_{LI}. Conversely, locations with IRs lower than τ^{low}_{LI} do not require any adjustments, since the driver did not or only rarely intervene there. Finally, locations with IRs between both thresholds feature inconsistent driver behavior and it is not conclusive whether the participant preferred an adjustment in this location. In these locations with inconsistent individual driver behavior, this approach reaches its limits and deeper scene understanding would be required to derive the reasons behind the drivers' inconsistent behavior.

4.2 Results

In this section, the proposed framework for the derivation of necessary adjustments is applied to the driver intervention dataset recorded in Chapter 3. First, the dataset is filtered, and the relevant drive-throughs of the defined location types are locally accumulated. Subsequently, the IRs for the four function adjustment strategies, GC, GI, LC, and LI, are computed. The results are presented and interpreted in the following sections.

The complete results of the GI and LI analyses include tables containing extensive data. To enhance readability, these results were summarized and illustrated using representative examples. The complete results are presented in Appendix B.1.

4.2.1 Global Collective: Analysis of General Intervention Rates of All Drivers

First, the applicability of a GC adjustment is evaluated. This type of adjustment is applicable if all drivers consistently intervene in all drive-throughs of specific location types. Thus, the GC IRs, IR^+_{GC} and IR^-_{GC}, were calculated across all drivers and all locations of each location type. The results are depicted in Table 4.2. There, the higher speed IR, IR^+_{GC}, and the lower speed IR, IR^-_{GC}, are shown, as well as the total number of intervention opportunities in the dataset of the corresponding location types. As illustrated, no location type reaches an IR of close to $100\,\%$. Therefore, no GC adjustment of the PLDF's behavior in these locations should be conducted. Drivers most commonly increased the vehicle speed in roundabouts, with a IR^+_{GC} of $66.3\,\%$. However, they also decreased the vehicle speed in $1.4\,\%$ of roundabouts and did not intervene in $32.3\,\%$ of roundabouts, although an intervention would have been possible. All other calculated IRs are generally lower and range from $0\,\%$ to $12.3\,\%$ for IR^-_{GC} and from $15.9\,\%$ to $34.2\,\%$ for IR^+_{GC}. IR^+_{GC} is generally higher than IR^-_{GC} across all location types. This result aligns with the findings in Section 3.3.2, where it was shown that drivers tend to increase the vehicle speed substantially more frequently than they decrease it when intervening based on personal preferences. SLI and SLD are the only location types where drivers never decreased the vehicle speed. Notably, turns are the only locations where the lower speed IR almost reaches the same level as the higher speed IR, with a IR^+_{GC} of $17.8\,\%$ and a IR^-_{GC} of $12.3\,\%$. The number of intervention opportunities varies widely, with more than 2000 SR locations but only 362 roundabouts where an intervention was possible. Since no location type reaches a IR^+_{GC} or IR^-_{GC} of close to $100\,\%$, no GC adjustment should be applied. However, the data shows that drivers commonly intervene, and for none of the six location types, a combined higher and lower speed IR of $0\,\%$ is reported. Therefore, a more granular adjustment approach based on individual drivers is evaluated next.

Table 4.2: IR^+_{GC}, IR^-_{GC}, and the total number of intervention opportunities for each location type.

	Location type					
	SR	**SLI**	**SLD**	**R**	**C**	**T**
IR^+_{GC}	$34.2\,\%$	$27.0\,\%$	$19.7\,\%$	$66.3\,\%$	$15.9\,\%$	$17.8\,\%$
IR^-_{GC}	$3.2\,\%$	$0.0\,\%$	$0.0\,\%$	$1.4\,\%$	$3.0\,\%$	$12.3\,\%$
Number of intervention opportunities	2463	1028	956	362	807	528

4.2.2 Global Individual: Analysis of Individual Intervention Rates

Since no GC adjustment is applicable for the PLDF, this section focuses on the evaluation of a potential GI adjustment based on the drivers' intervention behavior. A GI adjustment aims to personalize the PLDF's general driving policy if individual drivers consistently intervene in all drive-throughs of specific location types. Therefore, IR^{+}_{GI} and IR^{-}_{GI} were calculated for each driver individually in all locations of each type. The complete results are presented in Table B.1 in Appendix B.1, depicting both IR^{+}_{GI} and IR^{-}_{GI}, as well as the respective number of intervention opportunities. IR^{+}_{GI} and IR^{-}_{GI} of four selected example drivers are additionally illustrated in Figure 4.6.

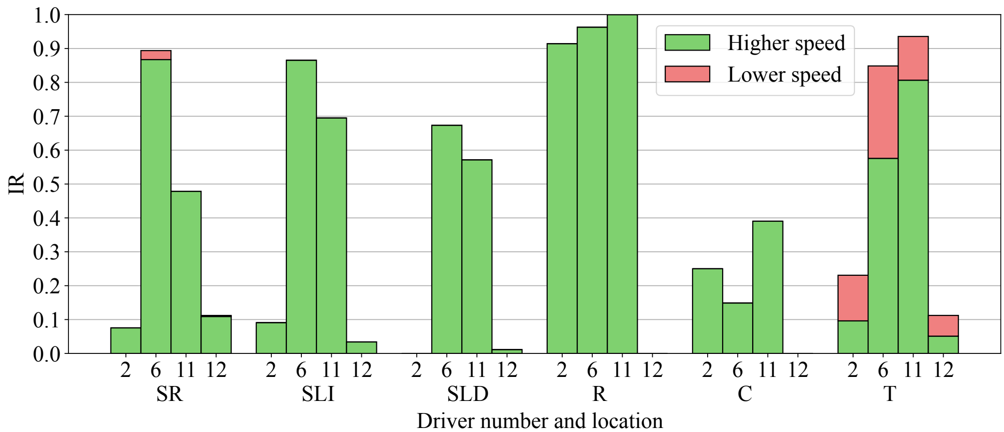


Figure 4.6: IR^{+}_{GI} and IR^{-}_{GI} of four selected example drivers in each location type.

As illustrated by the data, the individual IRs among drivers differ considerably. Some drivers almost never intervene, such as driver number twelve, while other drivers reach high IRs in almost all situations, such as driver number six. It can also be seen that some drivers intervene frequently only in specific location types, but not in others. For example, driver number two rarely intervenes in SR, SLI, and SLD locations, but reaches a IR^{+}_{GI} of 91.4 % in roundabouts. Only one driver reaches a IR^{+}_{GI} of exactly 100 % for one location type: driver number eleven in roundabouts. However, some other drivers also reach generally high IRs.

In Section 4.1.4, it was defined that a GI adjustment is applicable for drivers whose IR^{+}_{GI} or IR^{-}_{GI} reaches the τ^{high}_{GI} threshold, which should be set close to 100 %. In the following analysis, the results for threshold values of 80 %, 90 %, and 100 % are investigated. Table 4.3 depicts the number of drivers who reached these values for τ^{high}_{GI} in the different location types. If a threshold τ^{high}_{GI} of 90 % is chosen, a GI adjustment would be applicable for four drivers in roundabout locations. Only these four drivers intervene consistently enough in this location type. With a threshold of 80 %, a GI adjustment would be applicable for a seven drivers in the roundabouts and for five drivers also in other location types. However, it is debatable whether a threshold of 80 % is appropriate for this evaluation. For example, driver number 11 reaches both a IR^{+}_{GI} of 80.6 % and a IR^{-}_{GI} of 12.9 % in such situations. As a result, this driver would receive a GI adjustment toward higher turn speeds when a τ_{GI}^{high} of 80 % is applied, despite also frequently intervening by pressing the brake pedal in the same context. Accordingly, a threshold of 90 % to 100 % for τ_{GI}^{high} is recommended to ensure that adjustments are made only when the driver preferences are clear.

Table 4.3: Number of drivers with either IR^{+}_{GI} or IR^{-}_{GI} above different τ_{GI}^{high} thresholds. These drivers intervene so frequently in the respective location types, that a GI adjustment of the PLDF's behavior is applicable.

τ_{GI}^{high}	Number of drivers per location type					
	SR	**SLI**	**SLD**	**R**	**C**	**T**
80 %	2	1	1	7	0	1
90 %	0	0	0	4	0	0
100 %	0	0	0	1	0	0

Conversely, no adjustment to the PLDF's behavior is necessary for drivers who almost never intervene in specific locations. Therefore, three different values are tested for the lower IR threshold τ_{GI}^{low}: 0 %, 2.5 %, and 5 %. The number of drivers who reach IRs at or below these thresholds is depicted in Table 4.4. As can be seen, some drivers almost never intervene in specific location types. Especially in SLD and curve locations, consistently low IRs are commonly found. Accordingly, these drivers do not require any adjustments to the PLDF's behavior in these situations.

Table 4.4: Number of drivers with both IR^{+}_{GI} and IR^{-}_{GI} below different τ_{GI}^{low} thresholds. These drivers intervene so rarely in the respective location types, that no adjustment of the PLDF's behavior is necessary.

τ_{GI}^{low}	Number of drivers per location type					
	SR	**SLI**	**SLD**	**R**	**C**	**T**
5 %	1	3	8	3	7	0
2.5 %	1	1	7	3	7	0
0 %	0	0	4	3	6	0

To evaluate the general applicability of a personalization of the PLDF's behavior, the statistical significance of the individual differences in the drivers' intervention frequencies is tested. For each location type, IR^{+}_{GI} and IR^{-}_{GI} were considered, except for SLI and SLD, where IR^{-}_{GI} was excluded due to its constant value of $0\,\%$ across all drivers. First, for each location type, the distribution of intervention frequencies across drivers was evaluated for normality using Shapiro-Wilk tests with a p-value threshold of $5\,\%$. Since the Shapiro-Wilk test results indicate that not all features are normal distributed, a Friedman test was chosen for the subsequent analysis, comparing $k = 10$ conditions over $n = 17$ subjects. The results indicate a statistically significant difference in the individual intervention behavior ($\chi^2(9) = 60.11$, $p = 1.28 \cdot 10^{-9}$). The associated effect size (Kendall's $W = 0.393$) suggests a moderate effect. Therefore, these results suggest that a personalization of the PLDF's behavior may be appropriate.

To summarize, the intervention behavior of individual drivers differs significantly, and a GI adjustment is proposed for some drivers especially in roundabout locations, depending on the chosen $\tau^{\text{high}}_{\text{GI}}$. Conversely, some drivers rarely intervene in specific location types and do not require any adjustment to the PLDF's driving policy, depending on the chosen $\tau^{\text{low}}_{\text{GI}}$. However, most drivers reach IR values between both thresholds, even for the most lenient values chosen. These drivers behave too inconsistently on a global scale to adjust the PLDF's general driving policy accordingly. Therefore, the proposed location-focused adjustments are evaluated in the next sections.

It must be noted that the available sample size for roundabout locations is relatively small for some drivers in comparison to other location types. As a result, a substantial portion of the roundabout drive-throughs may originate from the same few specific roundabouts encountered repeatedly during the participants' commutes. Due to this limited sample size, it remains uncertain whether this localized behavior can also be generalized to routes beyond the participants' commutes in the dataset. Accordingly, a GI adjustment should only be considered if a sufficient number of data points is available to support the findings. Ideally, these data points should also include driving data from a broader set of locations beyond the participants' commutes.

4.2.3 Local Collective: Analysis of Collective Location-based Intervention Rates

In this section, the applicability of a LC adjustment of specific locations in the map for all drivers is evaluated. A LC adjustment is applicable if all drivers intervene in the same manner in almost every drive-through of a specific map location. The map information for this location can then be enriched with additional information from the drivers' behavior, or potentially incorrect map information may be identified and fixed. Hence, IR^{+}_{LC} and IR^{-}_{LC} are calculated for each unique location in the dataset. If either IR^{+}_{LC} or IR^{-}_{LC} of a specific location reaches $\tau^{\text{high}}_{\text{LC}}$, a LC adjustment should be conducted in this location. Another requirement for a LC adjustment is that the underlying dataset contains driving data of a representative number of drivers in the respective locations, since a LC adjustment affects the collective map data of the whole vehicle fleet.

However, the underlying dataset in this thesis does not provide the necessary sample size of drivers in the specific locations. Due to the study design, participants predominantly drove on different routes, resulting in relatively few overlapping road segments. Table 4.5 depicts the distribution of unique map locations that were driven through by different numbers of individual drivers in the dataset. As illustrated, 76.71 % of unique locations in the dataset were driven through by only one driver and only 1.86 % were driven through by at least five drivers, which corresponds to a total number of 31 unique locations. Out of these 31 locations, only one reaches an IR of more than 80 %: One roundabout location has driving data of five different drivers and a $\mathrm{IR}^{+}_{\mathrm{LC}}$ of 93.75 %. However, driving data of only five participants is not representative, and no LC adjustment should be conducted based on such a small sample size. Instead, analyses should be conducted based on extensive vehicle fleet data. Using the automatic classification of driver interventions proposed in Chapter 5, the driver interventions in customer fleet data could be classified and used for large-scale analyses. With such an extensive dataset, a potential LC adjustment may be evaluated again. However, with the underlying dataset of this study, no recommendation for or against a LC adjustment of specific map locations can be made.

Table 4.5: The distribution of individual drivers who drove through a unique map location for each location type.

Unique drivers	**Distribution of unique locations**						
	SR	**SLI**	**SLD**	**R**	**C**	**T**	**Total**
1	79.2 %	76.68 %	72.47 %	67.82 %	76.45 %	78.36 %	76.71 %
2	14.18 %	14.62 %	17.81 %	18.39 %	18.53 %	16.37 %	15.91 %
3	4.01 %	4.35 %	5.67 %	11.49 %	4.25 %	1.75 %	4.50 %
4	1.08 %	1.98 %	1.62 %	0.00 %	0.00 %	0.58 %	1.02 %
5+	1.54 %	2.38 %	2.42 %	2.30 %	0.78 %	2.92 %	1.86 %
Total number of unique locations	649	253	247	87	259	171	1666

4.2.4 Local Individual: Analysis of Individual Location-based Intervention Rates

As explained in the previous sections, no general policy adjustments of the PLDF's behavior are applicable for most drivers in most location types. The underlying dataset also does not allow for the recommendation of any LC map adjustments. Therefore, the applicability of LI adjustments to specific locations for each individual driver is evaluated in this section. A LI adjustment is applicable if an individual driver consistently intervenes in a specific map location, i.e., their $\mathrm{IR}^{+}_{\mathrm{LI}}$ or $\mathrm{IR}^{-}_{\mathrm{LI}}$ reaches $\tau^{\mathrm{high}}_{\mathrm{LI}}$. Conversely, if an individual driver almost never intervenes in a specific location, i.e., neither $\mathrm{IR}^{+}_{\mathrm{LI}}$ nor $\mathrm{IR}^{-}_{\mathrm{LI}}$ exceeds $\tau^{\mathrm{low}}_{\mathrm{LI}}$, no adjustment in this specific location is required. However, if drivers show inconsistent intervention behavior, i.e., either $\mathrm{IR}^{+}_{\mathrm{LI}}$ or $\mathrm{IR}^{-}_{\mathrm{LI}}$ falls between both thresholds $\tau^{\mathrm{low}}_{\mathrm{LI}}$ and $\tau^{\mathrm{high}}_{\mathrm{LI}}$, then no adjustment may be made, since it is not conclusive whether the driver would prefer an adjustment of the PLDF's behavior in this location or not. Accordingly, each specific location for each individual driver can be classified into three possible categories:

- *LI adjustment applicable*,
- *inconsistent driver behavior*,
- *no adjustment neccessary*.

Due to the study design, most unique locations feature four to five drive-throughs of individual participants. This number is additionally reduced by the necessary filtering of drive-throughs where no intervention was possible, resulting in an average of 2.61 drive-throughs per unique location and individual driver. The primary objective of the following analysis is to evaluate the local consistency of individual driving behavior across multiple drive-throughs. Accordingly, a minimum of two valid drive-throughs per location is required, although larger sample sizes are generally preferred. In this thesis, a threshold of at least three valid drive-throughs per location is selected as a compromise between limited sample sizes and the reliability of the evaluation. The distribution of valid drive-throughs per unique location in the remaining dataset is depicted in Table 4.6. As illustrated, $51.50\,\%$ of unique locations feature three or more valid drive-throughs per individual driver.

The thresholds used for the evaluation of the LI adjustment must be selected with consideration to the limited sample sizes. As described in Section 4.1.4, a higher deviation of the upper threshold $\tau^{\mathrm{high}}_{\mathrm{LI}}$ from its ideal value of $100\,\%$ should be allowed than for the lower threshold $\tau^{\mathrm{low}}_{\mathrm{LI}}$ from its ideal value of $0\,\%$. This is motivated by two main factors. First, drivers were observed to intervene less frequently during their first commute compared to subsequent drives, likely due to learning effects, as discussed in Section 3.3.2. Second, related research has shown that drivers often do not intervene in situations with only small deviations to the desired behavior due to complacency, although they would generally prefer a function adjustment [Hoc01; Spe+22; WXC14].

Table 4.6: Distribution of valid drive-throughs per unique location and individual driver for each location type.

Valid drive-throughs	**Distribution of unique locations and drivers**						
	SR	**SLI**	**SLD**	**R**	**C**	**T**	**Total**
1	26.93 %	24.09 %	29.86 %	24.06 %	38.53 %	35.86 %	29.42 %
2	17.26 %	15.13 %	20.55 %	22.56 %	20.88 %	24.89 %	19.08 %
3	20.25 %	24.65 %	19.45 %	24.81 %	16.18 %	23.63 %	20.82 %
4	21.06 %	24.37 %	20.27 %	16.54 %	19.12 %	12.24 %	19.99 %
5	11.28 %	9.52 %	7.95 %	9.77 %	3.82 %	2.95 %	8.43 %
6+	3.23 %	2.24 %	1.91 %	2.26 %	1.46 %	0.42 %	2.26 %
Total number of unique locations and drivers	869	357	365	133	340	237	2301

Accordingly, the drivers' internal threshold for initiating an intervention may generally be higher than that for not intervening. Therefore, an upper threshold $\tau_{\mathrm{LI}}^{\mathrm{high}}$ of $66.\overline{6}\,\%$ is employed, which corresponds to the fraction $\frac{2}{3}$, while the lower threshold $\tau_{\mathrm{LI}}^{\mathrm{low}}$ is set to $25\,\%$, which corresponds to the fraction $\frac{1}{4}$. The selected thresholds are chosen with consideration to the minimum sample size of three drive-throughs per unique location. For example, if a driver does not intervene during the first drive-through but intervenes during the subsequent two drive-throughs, the resulting IR of $66.\overline{6}\,\%$ should be counted as consistently intervened. However, if they intervened only once in three valid drive-throughs, the resulting IR of $33.\overline{3}\,\%$ should be counted as inconsistent driver behavior. The resulting possible intervention frequencies for three, four, and five drive-throughs are illustrated in Figure 4.7. The applicability of each combination for a LI adjustment is highlighted via the color coding. However, future adjustments of the PLDF should be based on extensive customer data, and thus should employ larger sample sizes and stricter IR thresholds.

To evaluate the applicability of a LI adjustment for each individual driver in each unique location, the dataset is filtered and aggregated as described in the previous sections. The resulting dataset for the LI evaluation contains 1185 unique locations where an individual driver had at least three valid drive-throughs with the possibility to intervene. $\mathrm{IR}_{\mathrm{LI}}^{+}$ and $\mathrm{IR}_{\mathrm{LI}}^{-}$ are then calculated and compared to the upper and lower

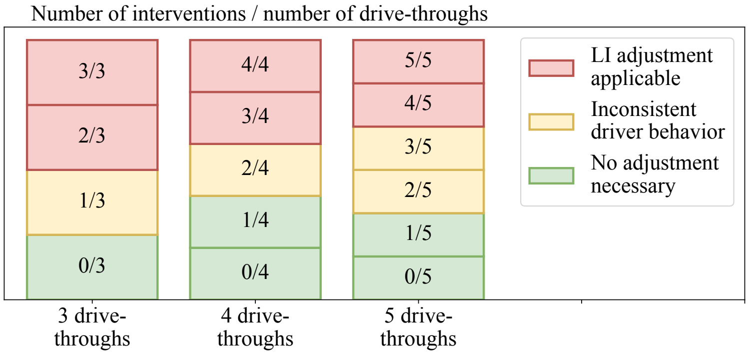


Figure 4.7: Potential combinations of intervention counts and drive-throughs, and the resulting applicability of LI adjustments for three, four, and five drive-throughs. Cases in which a LI adjustment is applicable, since the IR reaches the $\tau_{\mathrm{LI}}^{\mathrm{high}}$ threshold of $66.\overline{6}\,\%$, are marked in red. Cases in which no LI adjustment is applicable, since the IR does not exceed the $\tau_{\mathrm{LI}}^{\mathrm{low}}$ threshold of $25\,\%$, are marked in green. Cases in which the IR lies between both thresholds are marked in yellow.

thresholds, $\tau_{\mathrm{LI}}^{\mathrm{high}}$ and $\tau_{\mathrm{LI}}^{\mathrm{low}}$. Using the defined thresholds, each pair of unique locations and individual drivers is classified into one of three categories mentioned above. The distributions of these three categories are aggregated across all drivers and depicted in Table 4.7. As illustrated by the data, the driver behavior tends to be locally consistent, as only $11.56\,\%$ of all valid locations feature inconsistent driver behavior. In contrast, drivers consistently intervened in $29.20\,\%$ of locations, whereas they intervened so rarely that no adjustment is necessary in $59.24\,\%$ of locations. This distribution is largely consistent across all location types, with the exception of roundabouts, where $70.42\,\%$ of valid locations are applicable for a LI adjustment. Roundabouts are also the only location where the number of applicable LI adjustments and the number of inconsistent cases is higher than the number of locations where no adjustment is needed. In all other location types, the distribution of locations where no adjustment is necessary consistently exceeds $50\,\%$, while locations with inconsistent driver behavior are the least frequent at approximately $10\,\%$.

Locations with inconsistent driver behavior, even at the LI level, represent the primary limitation of this adjustment strategy. In such cases, it is unclear why drivers only intervene occasionally. Therefore, the evaluation of the applicability of LI adjustments in these locations is not possible. To address this uncertainty, larger drive-through sample sizes are required to observe the driver behavior over an extended period of time. If their behavior stabilizes with additional data, the applicability of a LI adjustment may be reevaluated. Conversely, if the inconsistency persists, the driver behavior may be influenced by external factors such as mood or weather conditions, which are beyond the scope of this thesis.

The presented results focus on the number of unique locations with at least three valid drive-throughs where an adjustment for individual drivers may be applicable.

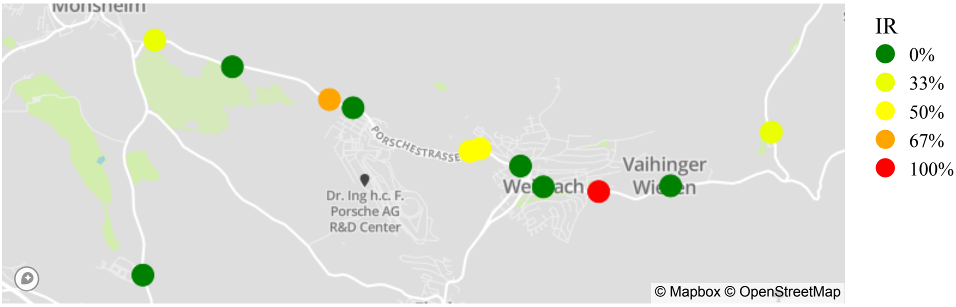
Table 4.7: Distribution of LI adjustment applicability categories in each location type aggregated across all drivers.

Category	Distribution of categories in unique locations						
	SR	SLI	SLD	R	C	T	Total
LI adjustment applicable	33.40 %	23.04 %	18.78 %	70.42 %	18.12 %	26.88 %	29.20 %
Inconsistent driver behavior	11.75 %	11.98 %	8.84 %	16.90 %	10.87 %	11.83 %	11.56 %
No adjustment necessary	54.85 %	64.98 %	72.38 %	12.68 %	71.01 %	61.29 %	59.24 %
Total number of locations	485	217	181	71	138	93	1185

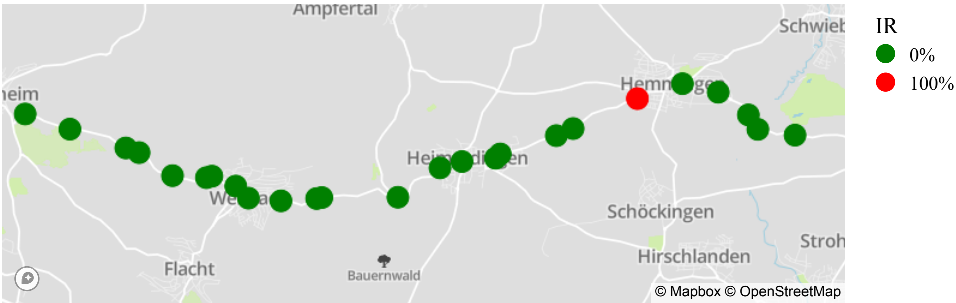
When these results are set into the context of the number of interventions conducted in these locations, it is found that a total of 1500 personal preference-based interventions occurred in these locations. 84.13 % of these interventions occurred in locations where either $\mathrm{IR}^{+}_{\mathrm{LI}}$ or $\mathrm{IR}^{-}_{\mathrm{LI}}$ reached the upper threshold $\tau^{\mathrm{high}}_{\mathrm{LI}}$, whereas only 15.87 % of interventions were associated with inconsistent driver behavior. Consequently, if LI adjustments were implemented in the respective high-consistency locations such that drivers no longer intervene, the number of personal preference-based interventions could potentially be reduced by up to 84.13 % in these locations.

As Table 4.7 only presents the data aggregated across all drivers, the complete data for each individual participant is provided in Table B.2 to Table B.5 in Appendix B.1. Additionally, the IRs of two example participants in SLI locations are visualized using map plots in Figure 4.8. Figure 4.8a depicts the data of a participant with relatively inconsistent intervention behavior. In the illustrated segment of their commute, the participant has six locations with an IR of 0 %, two locations where the IR reaches $\tau^{\mathrm{high}}_{\mathrm{LI}}$, and four locations where they behaved inconsistently with IRs of $33.\overline{3}$ % and 50 %. In contrast, the participant in Figure 4.8b shows a more consistent intervention behavior. They never intervened in any SLI location during their commute, except in one location where they reached a consistent $\mathrm{IR}^{+}_{\mathrm{LI}}$ of 100 %. Accordingly, for this driver, an adjustment of the PLDF's SLI behavior is only required at the single location where they consistently intervened.

To summarize, in the majority of locations, a LI adjustment is applicable, as most drivers intervene relatively consistently on a location-specific level. The 11.56 % of locations with inconsistent driver behavior represent the primary limitation of the LI approach.



(a) The LI IRs of driver number six in SLI locations on a selected map segment of their commute.



(b) The LI IRs of driver number nine in SLI locations on a selected map segment of their commute.

Figure 4.8: Map plots depicting the LI IRs of two example drivers for SLI locations. The unique locations are highlighted using colored markers. The markers are colored based on the maximum of IR^{+}_{LI} and IR^{-}_{LI} of the individual drivers in the corresponding locations.

In these cases, drivers intervene irregularly, and no reliable adjustment to the PLDF's behavior can be derived. Nonetheless, the objective of the PLDF's adjustment is to reduce the number of personal preference-based interventions, and among the four proposed strategies, the LI adjustment is the most promising. As explained above, 84.13 % of personal preference-based interventions occur in locations with consistent individual driver interventions, which could be mitigated through a successful LI adjustment.

4.3 Summary and Contributions

This chapter focuses on the derivation of necessary adjustments to the PLDF's driving behavior based on driver feedback in the form of personal preference-based interventions. The test group study analysis in Chapter 3 found that most personal preference-based driver interventions occur during free driving scenarios in certain location types. Therefore, potential adjustment strategies for map-based driving functions are discussed in the methodology, and a framework for the derivation of necessary adjustments based on driver interventions is presented. This framework revolves around the evaluation of the applicability of four adjustment strategies, GC, GI, LC, and LI. By calculating collective and individual IRs in specific locations and across all locations in the dataset, the applicability of each strategy is evaluated. The results show that no general adjustment of the PLDF's driving policy for all drivers, i.e., a GC adjustment, is applicable, since the participants' individual intervention behavior differs significantly. However, it is also found that personalized driving policy adjustments, i.e., GI adjustments, are only feasible for some drivers in specific location types, since most drivers behave rather inconsistently on a global policy level. This result is noteworthy, since all related state-of-the-art personalization approaches highlighted in Section 2.3 focus on such generalized individual policy adjustments. Instead, LI adjustments are found to be the most promising approach, since drivers behave most consistently on a location-specific level. Out of the analyzed locations, inconsistent intervention behavior was only observed in $11.56\,\%$ of locations. Moreover, $84.13\,\%$ of the analyzed personal preference-based interventions are performed consistently in specific locations. Therefore, this number also represents the potential reduction of personal preference-based interventions if a LI adjustment was conducted in the respective locations. Accordingly, an adjustment of the map data for specific locations and individual drivers is proposed for most drivers in most location types.

The primary limitation of the proposed framework is the limited sample sizes in the underlying dataset used for the evaluation. Especially for the evaluation of location-specific adjustments, the drive-through sample sizes are low, and relatively lenient thresholds must be used accordingly. Due to the study design that focuses on individual commutes instead of predefined routes, the dataset does not feature locations with a representative number of individual drivers for an evaluation of potential LC adjustments. Accordingly, extensive fleet data is necessary to evaluate the LC and LI adjustment strategies in more depth. When the framework is applied to the real vehicle fleet, significantly higher sample sizes and stricter thresholds should be employed before an adjustment of the PLDF's behavior is executed. Additionally, data from a broader range of routes should be included in the dataset to more accurately assess the applicability of potential GI adjustments beyond the participants' typical commutes. Another limitation of the proposed framework is that it focuses solely on what kind of adjustments are needed and where they are needed. However, it is not evaluated how these adjustments should be conducted specifically. For example, no specific parameters are derived for a GI adjustment, and no location-specific speed profiles are generated for LI adjustments. Therefore, Chapter 6 focuses on this derivation of new target speed profiles based on the performed driver interventions.

To summarize, the research contributions of this chapter are as follows:

- The introduction and application of a framework for the derivation of necessary adjustments to map-based driving functions utilizing personal preference-based driver interventions.

- An in-depth analysis of the general, location-specific, and individual driving behavior in specific location types in the underlying dataset with the following results:
  - No general driving policy adjustment for all drivers in the dataset is applicable, since the individual drivers' behavior differs significantly. Therefore, a personalization is recommended.
  - General driving policy adjustments are applicable only for some individual drivers in specific location types. Therefore, no generalized driving policy may represent the individual driver behavior in most cases.
  - Location-specific adjustments for individual drivers show the most promising results for the reduction of personal preference-based driver interventions. Therefore, this approach is further pursued.

5 Classification: Automatic Distinction of the Reasons behind Takeovers

The preceding chapters focused on the analysis of human intervention behavior and the derivation of necessary adjustments to the PLDF based on the dataset recorded in Chapter 3. The creation of this dataset first required manual voice annotations by the drivers for each driver intervention, and then the voice annotations and bus signal data were manually processed and combined into the resulting driver intervention dataset. This manual annotation process considerably limits the applicability of the developed analyses outside of the dataset used in this thesis. Similarly, the subsequent Chapter 6 focuses on the development of a prototypical driving function that iteratively learns from the driver interventions during driving. However, the necessity of manual annotation strongly limits the applicability of such a self-learning driving function in an in-production vehicle environment. Therefore, the objective of this chapter is the development of a pipeline for the automatic classification of driver interventions. Such an automatic classification pipeline facilitates the development of a customer function without requiring any manual labeling efforts. Additionally, already existing datasets containing large-scale vehicle data, such as fleet data, may be labeled using the developed classification pipeline. This chapter therefore covers the research objectives stated in Section 2.6.4. The development and implementation of the classification framework were conducted over the course of a master thesis [Sch24], which was supervised within the scope of this doctoral thesis. The corresponding research paper has been accepted for publication and is currently in press [Sch+26b]. The master thesis and the publication include only a subset of the methodology and results presented in this chapter.

This chapter explains the development and evaluation of an automatic driver intervention classification pipeline and is structured as follows: First, the fundamentals of Multivariate Time Series Classification (MTSC) are explained in Section 5.1, and candidate MTSC models are identified. The methodology behind the developed classification pipeline is explained in Section 5.2. Section 5.3 highlights the results of the trained classification models on the driver intervention dataset. Additionally, strengths and weaknesses of the best-performing model are analyzed, and improvement potentials are derived. Finally, Section 5.4 summarizes this chapter and highlights its research contributions.

5.1 Fundamentals of Multivariate Time Series Classification

The driver intervention dataset created in Chapter 3 consists of various signals from different Electronic Control Units (ECUs) and sensors in the vehicles. Each of the recorded signals contains a series of values that are indexed and ordered by timestamps. Therefore, each signal's data may be described as a *time series*, according to Definition 5.1. Furthermore, the combination of related time series signals in the dataset may be called a *Multivariate Time Series (MTS)*, according to Definition 5.2.

Definition 5.1 (Time Series)
Hamilton defines a time series as follows:

> *A time series is a collection of observations indexed by the date of each observation. Usually we have collected data beginning at some particular date [...] and ending at another [...] [Ham94].*

A singular time series is also commonly called a Univariate Time Series (UTS) [Ham94; Rei93].

Definition 5.2 (MTS)
Reinsel defines a MTS as "a vector of [...] time series variables" and describes that "such multivariate processes arise when several related time series processes are observed simultaneously over time, instead of observing just a single series as is the case in UTS analysis" [Rei93].

The MTS data in the underlying dataset contains segments of assisted driving and segments of driver interventions, which are accordingly labeled by the drivers. As the objective of this chapter is the automatic classification of these labels based on the MTS, this task falls into the field of MTSC, according to Definition 5.3.

Definition 5.3 (MTSC)
MTSC aims to build predictive models that assign discrete labels to segments of MTS data [Dha+20; Rui+21]. Ruiz et al. describe MTSC as "a form of machine learning where the features of the input vectors are real valued and ordered", i.e., MTS data [Rui+21].

Time series classification has been studied extensively, and there are a multitude of different established models to choose from. While most research focuses on Univariate Time Series Classification (UTSC) approaches, the popularity of MTSC approaches has been rising recently [Rui+21]. Both Dhariyal *et al.* [Dha+20] and Ruiz *et al.* [Rui+21] review multiple recent MTSC algorithms and compare their performances on different benchmark tasks in the University of East Anglia (UEA) MTSC archive [Bag+18]. The five best-performing MTSC algorithms on the UEA MTSC benchmarks are listed below [Dha+20; Rui+21]:

- InceptionTime [Faw+20a],
- Long Short Term Memory Fully Convolutional Network (LSTM-FCN) [Kar+17; Kar+19],
- RandOm Convolutional KErnel Transform (ROCKET) [DPW20],
- HIerarchical VotE COllective of Transformation-based Ensemble (HIVE-COTE) [LTB16; Bag+20],
- Word ExtrAction for time SEries cLassification plus Multivariate Unsupervised Symbols and dErivatives (WEASEL+MUSE) [SL18].

While all five algorithms show excellent results on the UEA MTSC benchmarks, the training times and required computational resources differ significantly between the models. Especially for HIVE-COTE and WEASEL+MUSE, Dhariyal *et al.* [Dha+20] and Ruiz *et al.* [Rui+21] report an exceptionally high demand for computational resources and long training times for both models. During preliminary tests of the proposed algorithms, this was also confirmed for the driver intervention dataset used in this thesis. Therefore, HIVE-COTE and WEASEL+MUSE are excluded, and InceptionTime, LSTM-FCN, and ROCKET are selected as candidate models for the following evaluations. Another well-established approach, Dynamic Time Warping (DTW) [SC78], was long considered the "gold standard" [Rui+21] for UTSC and MTSC and is commonly used during benchmarks as a baseline. However, the above-mentioned algorithms significantly outperform DTW on the used benchmarks [Dha+20; Rui+21]. Therefore, DTW was also excluded from the potential candidate algorithms.

The following paragraphs introduce the three chosen algorithms InceptionTime, LSTM-FCN, and ROCKET in more depth. As DTW, HIVE-COTE, and WEASEL+MUSE were not chosen for the subsequent development of the automatic driver intervention classification pipeline, they are briefly explained in Section C.1 of Appendix C.

5.1.1 InceptionTime

InceptionTime [Faw+20a] is a deep Convolutional Neural Network (CNN)-based classifier, inspired by the Inception-v4 architecture [Sze+17], which is traditionally used in computer vision. The network consists of two sequential residual blocks, followed by a global average pooling layer, and a fully connected softmax layer that outputs the classification probabilities. Each residual block's input is transferred to the next block's input via a shortcut linear connection, which mitigates problems inherent to deep neural networks, such as the vanishing gradient problem [Faw+20a; Sze+17]. The residual blocks consist of three Inception modules each. Figure 5.1 depicts the structure of such an Inception module. As illustrated, multiple convolutional and pooling filters are applied to the input data. The first step is the so-called *bottleneck* layer. This layer applies m filters with a length and stride of 1 to all M dimensions of the input MTS data, thus reducing its dimensionality to $m < M$. The transformed m time series are then fed into n convolutional kernels of different lengths. Each of these kernels generates one transformed time series for each input time series. The used convolutions allow the network to identify potentially relevant patterns in the data, while the dimensionality reduction is used to ensure better generalization and lower model complexity. Parallel to the bottleneck and convolutional layers, a sliding MaxPooling window is applied to all input dimensions separately, followed by a bottleneck convolution over all dimensions, reducing them to one. This MaxPooling and convolutional filter is used to deal with small irregularities in the data. The resulting time series of the parallel paths are then concatenated to form the $m \cdot n + 1$ dimensional output of the inception module. These operations are repeated for each individual Inception module in the proposed framework [Faw+20a].

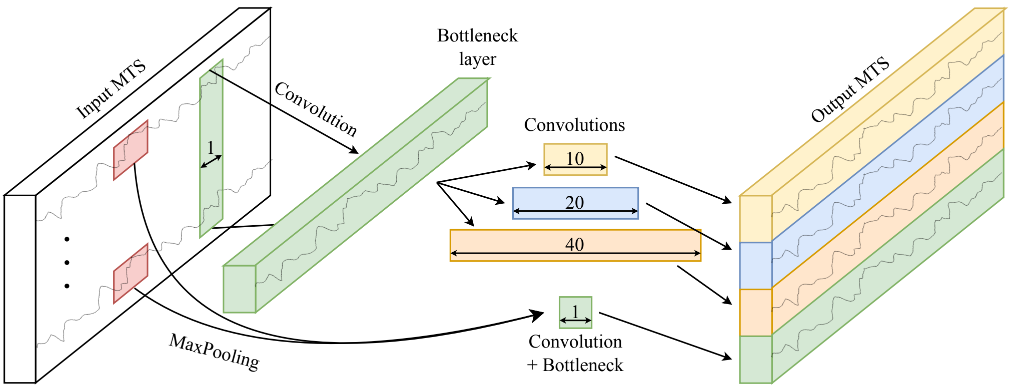


Figure 5.1: Architecture of an Inception module used in the InceptionTime MTSC classification model (adapted from Fawaz *et al.* [Faw+20a]).

5.1.2 Long Short Term Memory Fully Convolutional Network

LSTM-FCN [Kar+17; Kar+19] employs two parallel processing paths that combine the structure of Long Short Term Memory (LSTM) models with Fully Convolutional Networks (FCNs). The general architecture of LSTM-FCN is depicted in Figure 5.2. On one path, the input data is fed into a FCN which consists of three convolutional layers with Batch Normalization (BN) and Rectified Linear Unit (ReLU) activation functions. So-called *squeeze-and-excitation* blocks are additionally used after the first two convolutional layers each. This block is described as a learned self-attention mechanism on the output feature maps of prior layers. According to the authors, this block captures the inter-correlations between different variables in the MTS. After the FCN, a global pooling layer is used to fix the model's output dimensions. Parallel to the FCN path, a LSTM module with dropout is used that learns the long-term temporal connections in the variables. Before the input data is fed into the LSTM, its time and variable dimensions are switched in a so-called *dimension shuffle* block. Therefore, instead of receiving each timestep of the MTS step by step, the LSTM receives all timesteps of one time series variable in the MTS at once, and sequentially receives the following variables one by one. The results of both paths are then concatenated, and the final predictions are calculated via a SoftMax layer [Kar+17; Kar+19].

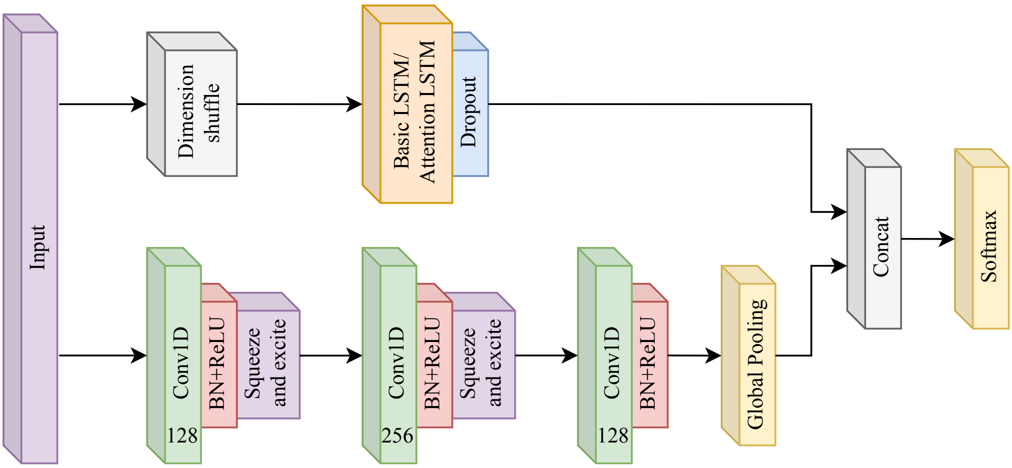


Figure 5.2: General architecture of LSTM-FCN (adapted from Karim *et al.* [Kar+19]).

5.1.3 RandOm Convolutional KErnel Transform

As the name implies, ROCKET [DPW20] also uses convolutional kernels for classification, similar to InceptionTime and LSTM-FCN. However, the weights of its convolutional kernels are not trained. Instead ROCKET consists of several thousand randomly initialized one-dimensional convolutional kernels. These kernels' parameters, i.e., length, weights, dilation, padding, and bias, are fully randomized and fixed, i.e., they are not changed during training. This design choice builds upon the theory that a high number of random convolutional kernels may have an advantage compared to trained kernels, especially on small datasets where it may be difficult to train a generalized model. These one-dimensional convolutional kernels are applied to a random subset of input channels of the MTS data, with the transformed features being computed as a convolution across these dimensions while sliding over the time axis. Subsequently, two features are extracted from each kernel's output data: the maximum value and the proportion of positive values. These output features are then concatenated and used to train a simple linear classifier [DPW20]. The architecture of ROCKET is illustrated in Figure 5.3. Notably, both Dhariyal *et al.* [Dha+20] and Ruiz *et al.* [Rui+21] report that ROCKET achieves outstanding results on the UEA MTSC benchmarks, while its training process is at least ten times faster compared to other evaluated models.

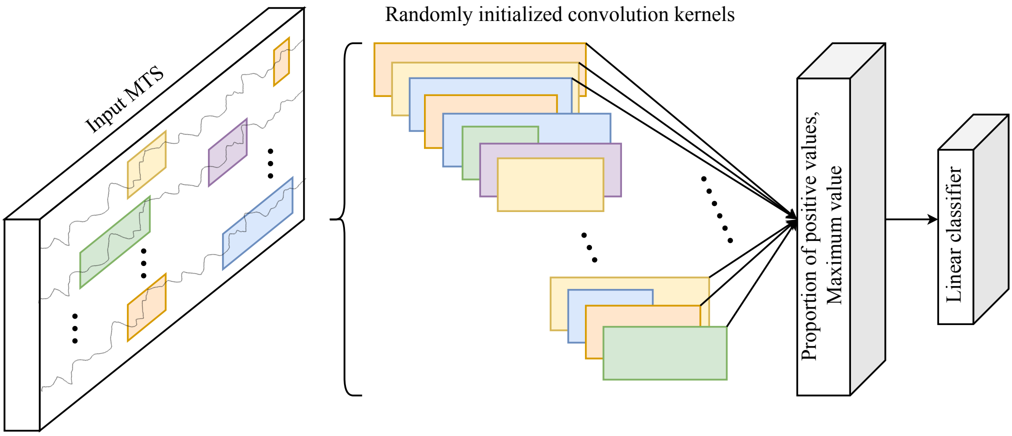


Figure 5.3: ROCKET architecture (adapted from Salehinejad *et al.* [Sal+22]).

5.2 Methodology

This section covers the methodology of the developed automatic classification pipeline for driver interventions. It goes into detail about the specific objectives of the classification, necessary preprocessing of the underlying dataset, and the final pipeline for model training and evaluation.

5.2.1 Classification Objectives

The general objective of the classification task is to automatically classify the reasons behind each driver intervention, i.e., the labels assigned during annotation, based on the respective recorded bus signals. This classification task is split into three possible granularities, which are introduced as follows:

- Full classification,
- Binary classification,

- Semi-binary classification.

First, in the *full classification* task, the specific reason behind each driver intervention is identified. The used classes generally align with the AAL2 labels introduced in Section 3.3.2. However, some modifications to the AAL2 labels, which are covered in Section 5.2.2, were necessary to suit the classification task. A high performance on the full classification task is the ideal result for the developed classification pipeline, as it poses the most granular and therefore complex task. As this full classification task might be difficult due to the high number of classes and the inherent imbalance of the dataset, the *binary classification* task is introduced next. In a binary classification, the intervention labels are grouped based on whether they are relevant or irrelevant for a function adjustment. Generally, any intervention that falls outside of the PLDF's ODD may be considered irrelevant, while all interventions that are based on deviating personal preferences may be considered relevant interventions. However, the definition of relevant and irrelevant classes may depend on the focus of the planned function adjustment. A high performance on the binary classification task is the minimum requirement for the developed classification pipeline and its primary objective, as it enables the distinction between interventions that should be used for the function adjustment and interventions that should not be used. Finally, the *semi-binary classification* task focuses on distinguishing all relevant classes while still grouping the irrelevant classes together without further distinction. This classification task is proposed since the distinction of all irrelevant classes in the dataset does not provide any benefit for a function adjustment that is solely based on the relevant intervention types. The three granularity levels of the classification tasks are illustrated in Figure 5.4. This approach is chosen to systematically evaluate the trade-off between classification detail and model performance. It establishes binary classification as minimum viability benchmark while still aiming for the ideal, more informative full classification. In theory, the binary or semi-binary classification task should be less complex than a full classification, while still providing important information for a function adjustment.

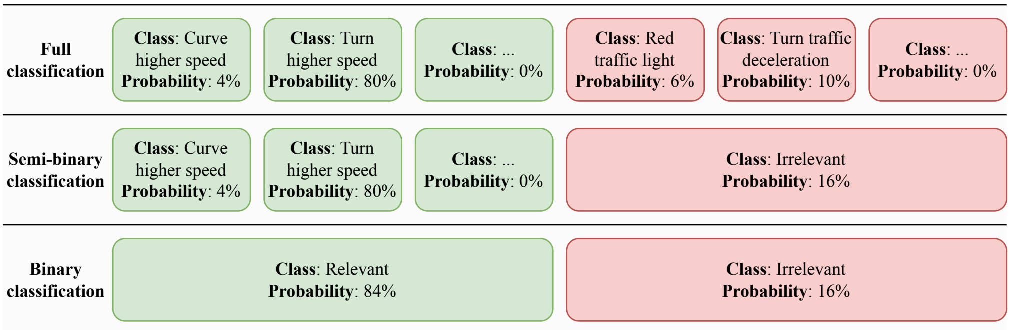


Figure 5.4: The three different classification tasks. Each rounded rectangle represents a class with a class name and an example predicted probability. Relevant classes are marked in green, while irrelevant classes are highlighted in red. Due to space limitations, the remaining classes in the high-granularity tasks are abbreviated using triple dots.

Three versions of the classification dataset are created using the three different label strategies, and the evaluated classification models are trained on each dataset separately. The performance of a model trained on a more complex classification task, such as the full classification, can still be evaluated on simpler classification tasks by calculating so-called *pseudo-binary* and *pseudo-semi-binary* scores. To do so, the ground truth and predicted classes are aggregated based on their binary labels during model performance evaluation. For example, a model trained on the full classification task assigns a probability to each possible class for a given intervention, and the class with the highest probability is chosen as the predicted class. By comparing these predicted labels with the corresponding ground-truth classes, the full classification performance can be calculated using conventional metrics such as precision, recall, and F1-score. For the pseudo-binary evaluation, the predicted probabilities of all relevant and irrelevant classes are aggregated separately, and the higher aggregated probability determines the predicted binary label. This binary prediction is then compared with the intervention's ground-truth binary label, and conventional performance metrics are computed to obtain the pseudo-binary scores. The pseudo-semi-binary scores are calculated analogously, with the difference that the probabilities of irrelevant classes are aggregated, while the relevant class probabilities are retained without modification. These pseudo-scores enable a direct comparison of models trained on datasets with different levels of granularity.

5.2.2 Classification Labels

The used classes for the classification generally follow the established AAL2 logic, i.e., they encompass the granular reasons behind each driver intervention. However, some modifications are necessary to adjust the AAL2 classes for the classification:

1. Assignment of binary relevance,
2. Exclusion of interventions without annotations,
3. Exclusion of unintended interventions,
4. Exclusion of set speed adjustments,
5. Merging of highly similar classes.

First, it is decided which classes are relevant for the underlying classification task, and binary labels are assigned to each AAL2 label accordingly. As defined in Section 4.1.1, the focus of the PLDF adjustments lies on free driving scenarios without involvement of an active ACC due to potential leading vehicles. Thus, all personal preference-based intervention types in free driving scenarios are considered relevant, while personal preference-based interventions focused on an adjustment of the ACC behavior are considered irrelevant. All intervention types that fall outside the PLDF's ODD are also considered irrelevant. Interventions due to incorrect input data are inherently divided into corrections of the ACC sensing state and corrections of incorrect legal speed information. As the adjustment of the PLDF's sensing state is beyond the scope of this thesis, the ACC-focused corrections are considered irrelevant. Based on the results of Chapter 4, a LI adjustment is proposed, i.e., local enrichments of the map information for individual drivers. Therefore, not only subjectively desired changes to the PLDF's speed on straight roads can be learned, but also corrections of objectively incorrect speed limits, if the driver intervenes accordingly. Thus, the AAL2 labels based on incorrect TSD or map information are considered relevant for the binary classification. Finally, the AAL4 label *other* is primarily composed of interventions with missing annotations, unintentional interventions, and highly uncommon intervention types which are not applicable as feedback for a function adjustment. Interventions with missing annotations and unintentional interventions are excluded from the dataset, as they provide no usable information for the classification model to train on. The remaining *other* labels are then considered irrelevant.

Next, all set speed adjustments are excluded from the dataset. As discussed in Section 3.3.2, $0\,\%$ of interventions outside of the PLDF's ODD are conducted via a set speed adjustment. They solely occur in personal preference-based interventions, during the correction of incorrect legal speed information, and in unintentional interventions. As unintentional interventions are removed from the dataset, all remaining set speed adjustments can be considered relevant. Accordingly, no classification is required in their case, since all set speed interventions should be used as feedback for the function adjustment, if performed consistently. Therefore, all set speed interventions are removed from the dataset, and the trained models solely focus on the classification of pedal interventions.

Finally, the remaining dataset is inherently imbalanced and features a high number of 48 different classes, out of which the least common 24 classes make up only $13.27\,\%$ of the dataset. Thus, highly similar intervention types are merged together in order to reduce the number of classes. Two strategies for the merging of classes were employed. First, classes were manually selected which inherently contain feedback on the same desired adjustment of the PLDF's behavior. This includes, e.g., the merging of interventions in turns. Instead of separating left and right turns, the combined location *turn* is used instead. Additionally, personal preference-based interventions on straight roads are merged with interventions that aim at correcting wrongly detected legal speed limits. During annotation, there was no indication in the recorded bus signals whether an intervention was conducted due to a deviating personal preference or an objectively incorrect legal speed. Instead, solely the voice annotation content was used to derive the underlying reason for the intervention. However, as explained above, both types of interventions can be used as input for a LI adjustment with the same effect of adjusting the local speed profile for individual drivers. Accordingly, the respective intervention types are merged, except for the class *acceleration due to incorrect lower speed limit in map*, which is a special case that can be accurately detected by trained annotators without corresponding voice annotations, as explained in Section A.3. As is also explained in Section A.3, multiple classes exist where the correct label was highly ambiguous during annotation. These classes include, e.g., *stronger acceleration after stop*, *ACC stronger acceleration*, and *ACC lower distance*. To identify potential candidates for the merging, preliminary model trainings on the full dataset were conducted to create initial confusion matrices. Similar classes exhibit high inter-class confusion, as the trained models struggle to distinguish between them. Then, hierarchical agglomerative clustering is used to create class similarity dendrograms based on which candidate classes for the merging can be identified. Similar approaches are used in related work for the creation of labeling hierarchies from highly granular classes [CO19; GSC02]. The candidates are then qualitatively evaluated and applicable combinations are merged based on the requirements that only small irrelevant classes are merged, and that the inherent reason and situation for the interventions strongly overlap. Finally, the names of the resulting classes are slightly shortened compared to their AAL2 counterparts due to space limitations in plots and tables. The final 29 classification labels are depicted in Table C.4 and Table C.5 in Appendix C.2. Appendix C.2 also contains detailed information about all excluded, renamed, and merged AAL2 labels.

5.2.3 Dataset Preprocessing

In order to create an intervention dataset that is feasible for the training of MTSC models, the original driver intervention dataset must be preprocessed. This section outlines all necessary preprocessing steps.

Extraction of Interventions

The original dataset consists of full-length drive recordings separated into individual commutes. For the classification task, only segments with driver interventions are relevant. These are extracted from the dataset and saved as single instances with their corresponding class label. Then, the unused intervention types defined in Section 5.2.2 are excluded from the dataset. During the initial annotation process in Section 3.2.2, consecutive driver interventions with the same underlying driver intention in the same situation were grouped together as one intervention. However, the goal of the classification is to create a model that can automatically classify each pedal intervention in previously unseen data. While the pedal interventions can be automatically detected via the PLDF status signal, no information is available on which pedal interventions should be merged before they are classified. Accordingly, the intervention merging process in the original dataset is reverted for the classification, so that each singular pedal intervention is classified separately. For example, in Figure 3.2a, three consecutive gas pedal interventions in a roundabout were grouped together as one intervention for the original dataset. However, for the classification, each pedal intervention is included separately with the same classification label of *roundabout higher speed*. The resulting dataset contains 3195 instances of annotated pedal interventions. As aforementioned, the classes in the dataset and their respective numbers of instances are depicted in table Table C.4 and Table C.5 in Appendix C.2.

The most promising state-of-the-art MTSC models for this classification task were introduced in Section 5.1: InceptionTime, LSTM-FCN, and ROCKET. For LSTM-FCN and ROCKET, it is explicitly stated that they require uniform, fixed-length MTS sequences as input data [DPW20; Dha+20; Kar+17; Kar+19]. For InceptionTime, no such information is stated in its publication [Faw+20a]. However, its official implementation on GitHub requires a fixed input shape for the MTS sequences [Faw+20b]. Therefore, a uniform sequence length $T_{sequence}$ must be defined for all interventions in the dataset. If an intervention is shorter than $T_{sequence}$, zero padding is used to fill the missing signal values, while an intervention that is longer than $T_{sequence}$ must be cut accordingly by removing excess values. Figure 5.5 depicts the number of interventions in the dataset with different durations. As illustrated, most interventions are relatively short, with over $90\,\%$ featuring a duration of less than $20\,\mathrm{s}$. Conversely, only a total of 26 interventions are longer than $80\,\mathrm{s}$. All 26 of these interventions belong to irrelevant intervention types, primarily the *stopover* and *red traffic light* classes. Therefore, their extended durations can be attributed to prolonged standing times.

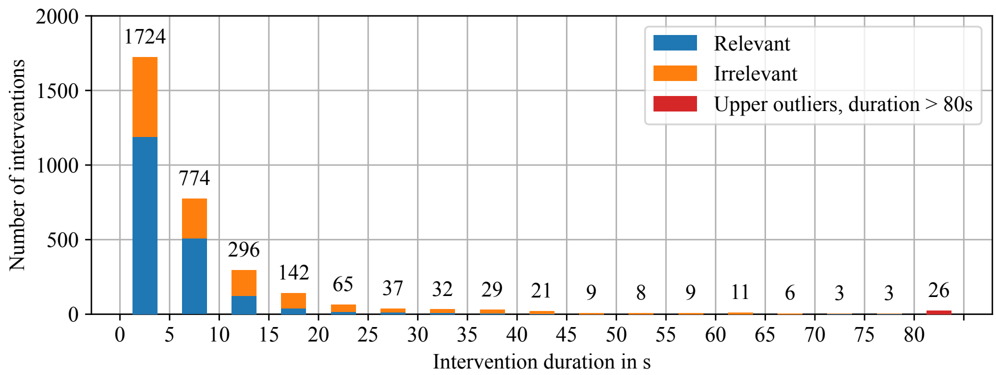


Figure 5.5: Histogram depicting the number of interventions with different durations in the dataset.

When choosing a low value for $T_{sequence}$, longer durations are cut off, potentially lowering the performance of trained models on these longer intervention types. Conversely, if a $T_{sequence}$ that is too long is chosen, the performance on shorter intervention types might suffer, since the trained model could face problems focusing only on the relevant part of the input sequence. Ultimately, the correct choice of $T_{sequence}$ depends on the chosen MTSC model, and how well it can handle either cut-off or zero-padded interventions. Therefore, hyperparameter tuning is performed to determine the optimal $T_{sequence}$ value for each model. The following candidate values are evaluated during hyperparameter tuning: $20\,\mathrm{s}$, covering $91.9\,\%$ of interventions, $40\,\mathrm{s}$, covering $97.0\,\%$, $60\,\mathrm{s}$, covering $98.5\,\%$, and $80\,\mathrm{s}$, covering $99.2\,\%$ of interventions in the dataset. The hyperparameter tuning is covered more in Section 5.2.4, and the full results are depicted in Table C.8. The optimal values for T_{sequence} vary considerably across the three evaluated models. While ROCKET performs best with a low $T_{sequence}$ of $20\,\text{s}$, InceptionTime and LSTM-FCN prefer higher values of $60\,\text{s}$ and $80\,\text{s}$, respectively. This discrepancy can likely be attributed to differences in the underlying model architectures. Generally, interventions using longer $T_{sequence}$ values capture more information, due to not being cut off, but they also contain more zero-padded values for shorter interventions. Therefore, deep neural network-based architectures, such as InceptionTime and LSTM-FCN, might be better suited to handle longer sequence cutouts, as their inherent architecture allows them to focus on the relevant center segment of each sequence, while ignoring irrelevant zero paddings. In contrast, ROCKET relies on randomly initialized convolutional kernels applied across the entire MTS sequence, followed by a rule-based feature extraction and a simple linear classifier. As a result, the added noise of the zero-padding might negatively influence ROCKET's performance on longer cutouts.

As explained above, pedal interventions are automatically detected and extracted from the continuous driving data via the PLDF status signal. However, the surrounding context information around each intervention might also contain important information for the classification. Thus, signal data within a specific T_{cutout} before and after each intervention is additionally included in each extracted intervention sequence. The best-performing value for T_{cutout} is again evaluated via hyperparameter tuning. Values of $0\,\text{s}$, $2.5\,\text{s}$, $5\,\text{s}$, $7.5\,\text{s}$, $10\,\text{s}$, $12.5\,\text{s}$, and $15\,\text{s}$ are tested across all model types. As depicted in Table C.8, LSTM-FCN reaches the highest performance with a T_{cutout} of $7.5\,\text{s}$, while both InceptionTime and ROCKET perform best using a value of $10\,\text{s}$. The preferred values only deviate slightly across the models, independent of the model structure. These results show that the inclusion of context information surrounding the interventions is beneficial, as all tested models perform worse for a T_{cutout} of $0\,\text{s}$. For better visualization, Figure 5.6 depicts the concept of $T_{sequence}$ and T_{cutout} around an intervention.

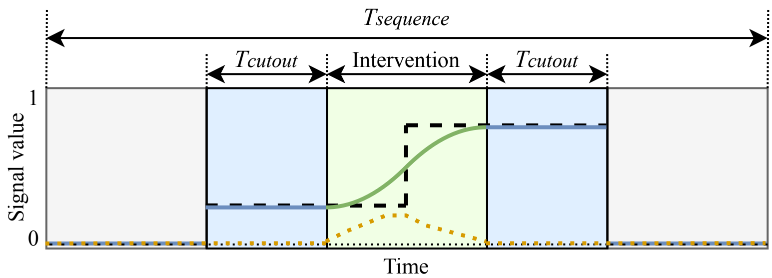


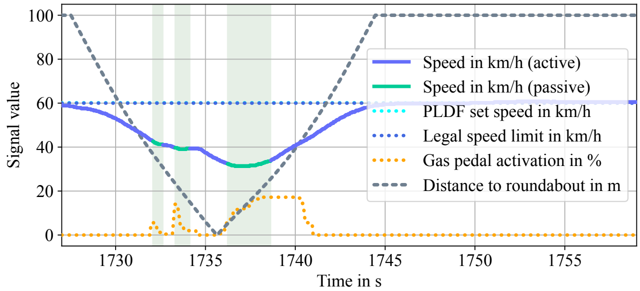
Figure 5.6: Visualization of $T_{sequence}$ and T_{cutout} around an intervention. The pedal intervention is highlighted in green, while T_{cutout} is applied before and after the intervention, highlighted in blue. The surrounding space in gray is filled with zeros for all signals until the sequence duration reaches $T_{sequence}$. If $T_{sequence}$ is shorter than the intervention including T_{cutout}, the sequence is cropped instead of using zero-padding.

Signal Preprocessing

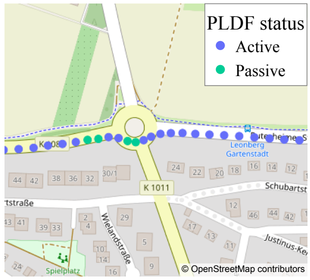
The used bus signals for the classification are mostly composed of signals directly related to the PLDF and the ACC, sensor and map data, and driver inputs, such as the gas and brake pedal activation. Most of these signals only require minimal preprocessing in the form of normalization between the values of zero and one. However, particularly the information from the map data and the PLDF status signal require further preprocessing steps. The PLDF status signal is a categorical variable that represents the three possible statuses *active*, *passive*, and *standby*. For the classification, this status signal is represented via three separate one-hot encoded signals that may either be zero or one, depending on the current PLDF status. Throughout this chapter, these three signals are regarded as one *PLDF status* signal for simplicity. Some signals from the map data also require further preprocessing, specifically the position of three map features: roundabouts, traffic lights, and turns. The location of these features is read from the map data and is initially represented by an active boolean flag at the timestep it is traversed. For example, the roundabout signal equals *true*, represented by a value of one, for one timestep during which a roundabout is entered. Otherwise, the signal equals *false*, represented by a value of zero. These binary indicators are preprocessed to instead represent the distance to such a map feature, capped at a certain maximum threshold. This threshold is called *maximum map feature distance*, and its optimal value is again evaluated via hyperparameter tuning. Out of the tested values of $50\,\mathrm{m}$, $100\,\mathrm{m}$, $150\,\mathrm{m}$, and $200\,\mathrm{m}$, all three models reached the highest performance with a *maximum map feature distance* of $100\,\mathrm{m}$. The performance using the original binary representations was also tested, but all models performed considerably worse compared to the models trained with distance-based map features. Presumably, the explicit representation of distances to upcoming and past map features can be more effectively interpreted by the employed MTSC models than the original binary representation. Furthermore, the original binary signals are only active for a single or a few time steps when passing the respective location. In contrast, the distance-based representations affect a larger number of time steps, resulting in a stronger indication of close-by locations for the MTSC models. Finally, the distance-based preprocessing also allows the models to receive information about upcoming or past map features that lie outside of the currently provided cutout.

Figure 5.7 depicts two examples for the distance-based representation of map features. As can be seen, the distance signals are capped at a *maximum map feature distance* of $100\,\mathrm{m}$ if the respective map feature is farther away. When approaching the map feature, the signal value decreases accordingly, until reaching zero for a brief moment at the exact location of the map feature. After the map feature is passed, the signal value increases again until reaching its maximum value. Due to the original feature representation in the map data, only the beginning point of each map feature is used as its location for the distance calculation. For example, the roundabout in Figure 5.7a is represented by the location where the driver enters it.

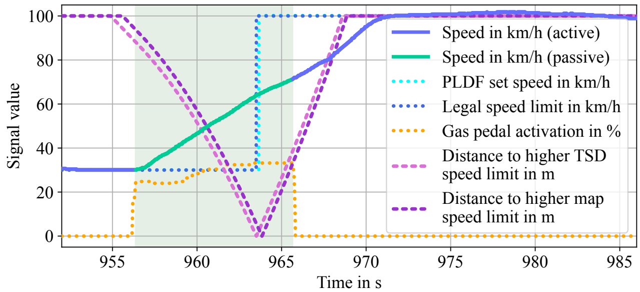
Two signals are contained in the dataset that represent the current speed limit: the TSD



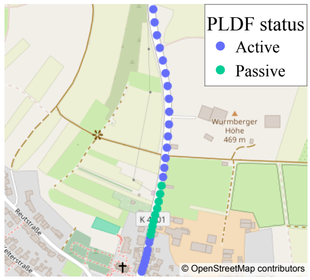
(a) Three consecutive *roundabout higher speed* interventions with the *distance to roundabout* signal.



(b) The corresponding GPS traces of the *roundabout higher speed* intervention.



(c) A *speed limit earlier acceleration* intervention with the *distance to higher TSD speed limit* and *distance to higher map speed limit* signals.



(d) The corresponding GPS traces of the *speed limit earlier acceleration* intervention.

Figure 5.7: Signal-over-time plots and the corresponding GPS traces of example interventions with distance-based map feature signals. Interventions are highlighted with a green background. Each plot depicts only the relevant distance signal for the respective situation. The map distance signals are capped at a *maximum map feature distance* of 100 m.

speed limit and the speed limit from the map data. Both of these signals are normalized and contained in the dataset, since some classes specifically revolve around deviations between map and TSD speed limits, e.g., *slower map speed acceleration*. There, the PLDF starts to decelerate due to a lower upcoming speed limit in the map data that does not exist in reality. Accordingly, the driver intervenes by pressing the gas pedal to override the incorrect speed limit information. In this case, the deviation between map and TSD-based legal speed limits causes a driver intervention. Since many intervention types are related to the speed limit information, additional experiments were conducted to test the effect of providing the distance-based representations also for speed limit changes, resulting in a notable increase in model performance. Therefore, the distance-based representations are also added for changes in both the map and TSD-based legal speed signals. An example of this preprocessing is shown in Figure 5.7c, where the *distance to higher TSD speed limit* and *distance to higher map speed limit* signals are presented during a *speed limit earlier acceleration* intervention. As can be seen, the locations of the map and

TSD-based legal speed limits are slightly shifted due to inaccuracies in the map data. The final list of used signals in the classification dataset is provided in Table C.6 and Table C.7 of Appendix C.3, including a short description of each signal.

5.2.4 Model Training and Evaluation Pipeline

As defined in Section 5.1, the three state-of-the-art MTSC models InceptionTime, LSTM-FCN, and ROCKET are selected as the most suitable candidates for this classification problem. Accordingly, each model is trained on the three different classification tasks introduced in Section 5.2.1: binary, semi-binary, and full classification. During training, a five-fold cross-validation is employed, which corresponds to an 80:20 split between the used training and testing dataset. Therefore, each model is trained five times on different training splits, and each intervention in the dataset is used in the test split once. This ensures more robust results that are only weakly affected by the chosen test splits. Stratified cross-validation additionally ensures that the distribution of classes remains equal across all folds, which is especially important due to the high class imbalance in the underlying dataset. The class imbalance is additionally addressed via a weighted loss used during training, which assigns higher losses to underrepresented classes. For the evaluation of the model performance, recall, precision, and primarily the F1-score are calculated for each model across all five folds. To compare the performance of models that were trained on differently labeled versions of the dataset, the pseudo-binary and pseudo-semi-binary scores are calculated, if applicable. Early stopping is used during training, i.e., the training is stopped if the loss on the test split did not improve over the last ten epochs. When early stopping is triggered, the training process ends and the best-performing model is saved.

The hyperparameters used during training are tuned for each model separately. The three preprocessing hyperparameters, $T_{sequence}$, T_{cutout}, and the *maximum map feature distance*, are relevant for all three model types. Each model additionally contains model-specific hyperparameters, which are also tuned separately. Most model-specific parameters affect either the model structure, e.g., the number and size of convolutional kernels, or the training process, e.g., the batch size and learning rate. The full results of the hyperparameter tuning are depicted in Table C.8 of Appendix C.4.

5.3 Results

In the following section, the trained models' performances are evaluated after hyperparameter tuning and training. First, the reached F1-scores of the trained models are compared in Section 5.3.1. Then, the classification performance is further analyzed by investigating class confusion in Section 5.3.2 and conducting a Permutation Feature Importance (PFI) analysis in Section 5.3.3. Finally, in Section 5.3.4, limitations and potential improvements of the trained classification models are discussed.

5.3.1 Model Performance Comparison

The three MTSC models, InceptionTime, LSTM-FCN, and ROCKET, are trained independently on the three differently labeled variants of the dataset. For each variant, the respective binary, semi-binary, or full classification scores are calculated, in addition to pseudo-scores for the less granular classification tasks. The resulting F1-scores are reported in Table 5.1.

Table 5.1: F1-scores for InceptionTime, LSTM-FCN, and ROCKET trained on the three different dataset variants. Best scores on each classification task are highlighted for each model via bold text.

		Classification task evaluated on:		
		(Pseudo-) Binary	(Pseudo-) Semi-binary	Full
InceptionTime trained on	Binary:	89.7 %	-	-
	Semi-binary:	90.8 %	84.0 %	-
	Full:	**91.6 %**	**85.1 %**	**75.2 %**
LSTM-FCN trained on	Binary:	89.5 %	-	-
	Semi-binary:	89.5 %	**82.2 %**	-
	Full:	**89.8 %**	82.0 %	**71.0 %**
ROCKET trained on	Binary:	91.7 %	-	-
	Semi-binary:	91.7 %	**86.6 %**	-
	Full:	**92.1 %**	85.7 %	**77.7 %**

As illustrated, the highest F1-scores on the binary classification task range from 89.8 % to 92.1 %, whereas the full classification results range from 71.0 % to 77.7 %. The model performance on the semi-binary task falls between the binary classification and the full classification task, with F1-scores ranging from 82.2 % to 86.6 %. These findings align with the expectation that less granular classification tasks are inherently simpler and therefore yield higher performance scores. ROCKET achieves the highest performance across all three classification tasks, followed closely by InceptionTime, whose results are only marginally lower. In contrast, LSTM-FCN exhibits the lowest performance among the three models, falling behind by several percentage points on each task. The results reveal that the highest performance on a given classification task is frequently obtained by models trained on a different, more granular task. For example, the best binary classification scores are consistently produced by models trained on the full classification task. These findings suggest that the inclusion of more detailed class information during training may enhance a model's ability to discriminate between different driver intervention patterns. In the case of the semi-binary classification, this trend is observed only for InceptionTime. In contrast, LSTM-FCN and ROCKET achieve their highest semi-binary performance when trained explicitly on that task. Since ROCKET achieves the highest classification performance across all three tasks, it is selected as the reference model for the analyses presented in the following sections. Specifically, the ROCKET model trained on the full classification task is chosen, as it achieves the best performance on both the pseudo-binary and the full classification task.

Comparison to Baseline Binary Classifiers

To set the binary classification model performances into context, they are compared to baseline classifiers that aim to reach a high classification score via simple strategies. First, a baseline classifier is evaluated that always chooses the majority class. In the underlying binary classification dataset, 1906 interventions are labeled as relevant, while 1289 interventions are labeled as irrelevant. Accordingly, the proposed baseline classifier achieves an F1-score of $44.6\,\%$, which is far below the scores achieved by the trained models. The second potential baseline classifier leverages the distribution of gas and brake pedal interventions in the dataset. As examined in Section 3.3.2, most personal preference-based interventions correspond to speed increases, whereas interventions outside the PLDF's ODD are predominantly performed via the brake pedal. Accordingly, 1715 out of 2289 gas pedal interventions in the classification dataset are relevant, while 715 out of 906 brake pedal interventions are irrelevant. In this distribution, a clear trend toward relevant gas pedal interventions and irrelevant brake pedal interventions is visible. Therefore, a baseline classifier that predicts all gas pedal interventions as relevant and all brake pedal interventions as irrelevant achieves an F1-score of $75.1\,\%$ on the binary classification task. This relatively high score reflects the substantial bias present in the dataset. However, the trained models achieve considerably higher F1-scores, indicating that their classification performance extends beyond a simple analysis of the driver input type. Furthermore, such well-performing baseline classifiers are only available for the relatively simple binary classification task. Due to the high number of granular relevant and irrelevant classes, baseline classifiers perform considerably worse on the semi-binary and full classification tasks.

5.3.2 Class Confusion Analysis

In this section, the performance of the ROCKET model trained on the full classification task is analyzed in more depth using confusion matrices. The complete class confusion matrix for the full classification task is presented in Figure 5.8. There, the model's performance on each individual class is illustrated. In this matrix, each column corresponds to a ground-truth class, while each row indicates the frequency with which instances were predicted as specific classes, normalized over the rows. Accordingly, the diagonal entries of the matrix represent the recall of each ground-truth class, whereas values outside the diagonal represent the frequency of incorrect predictions. For improved visual clarity, the relevant and irrelevant classes are grouped together and separated by horizontal and vertical black bars. In an ideal scenario, each class in the full classification task would be predicted correctly, resulting in a perfect diagonal containing only recall values of $100\,\%$. However, the classification is not perfect, and two distinct types of class confusion can be observed in the matrix: intra-binary and inter-binary confusion. Intra-binary confusion refers to misclassifications that occur between classes sharing the same binary label. In the confusion matrix, these are represented by the upper-left and lower-right regions within the black bars. For instance, $15\,\%$ of *speed limit stronger acceleration* interventions are predicted by the model as *roundabout higher speed* interventions, since they commonly occur after roundabouts. While such a prediction is incorrect in the context of the full classification task, it would still be considered correct in the pseudo-binary classification, as both classes share the same binary label. Consequently, intra-binary confusion is generally less critical than inter-binary confusion if distinguishing relevant and irrelevant classes is an objective. Inter-binary confusion refers to the incorrect prediction of a class instance as another class with a different binary label. It is represented by the upper-right and lower-left regions of the confusion matrix. For instance, *straight road lower speed* is a relevant class, yet $22\,\%$ of its instances are predicted as *other traffic deceleration* interventions, which are irrelevant. Consequently, inter-binary confusion represents both an incorrect full classification and an incorrect binary classification.

To provide a clearer representation of the pseudo-binary confusion, the frequencies with which each ground-truth class is predicted as either relevant or irrelevant are aggregated and illustrated in Figure 5.9, alongside the number of instances per class. Figure 5.9a presents the relevant classes, while Figure 5.9b displays the irrelevant ground-truth classes. The achieved pseudo-binary recalls are considerably higher for most classes compared to the performance on the full classification. For example, *slower map speed acceleration* reaches a full-classification recall of only $67\,\%$. In contrast, its pseudo-binary recall as a relevant class reaches $100\,\%$, as it is only confused with other relevant classes. The pseudo-binary classification matrix highlights several strengths and weaknesses of the trained model. A notable trend can be observed where acceleration interventions are more frequently classified as relevant, whereas deceleration interventions are more often classified as irrelevant. This observation is consistent with the analyses presented in Section 3.3.2, where it was determined that most relevant interventions are performed using the gas pedal, whereas brake pedal usage is more common among irrelevant classes. In Figure 5.9a, the model achieves recalls exceeding $90\,\%$ for all eight relevant gas pedal intervention types. In contrast, the performance is notably worse for the five relevant brake pedal intervention classes, with recalls ranging from $50\,\%$ to $75\,\%$. A similar trend is evident for the irrelevant ground-truth classes in Figure 5.9b, where deceleration classes generally outperform acceleration classes. Nevertheless, the model

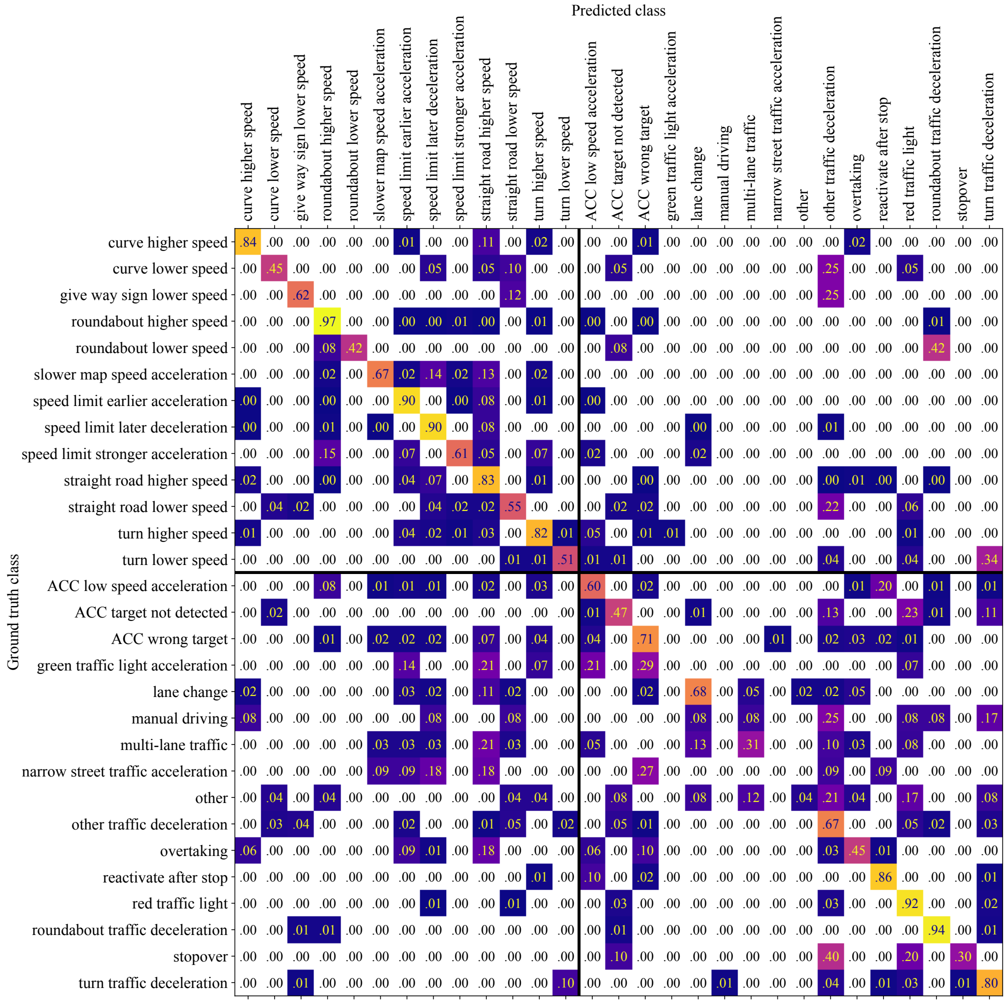
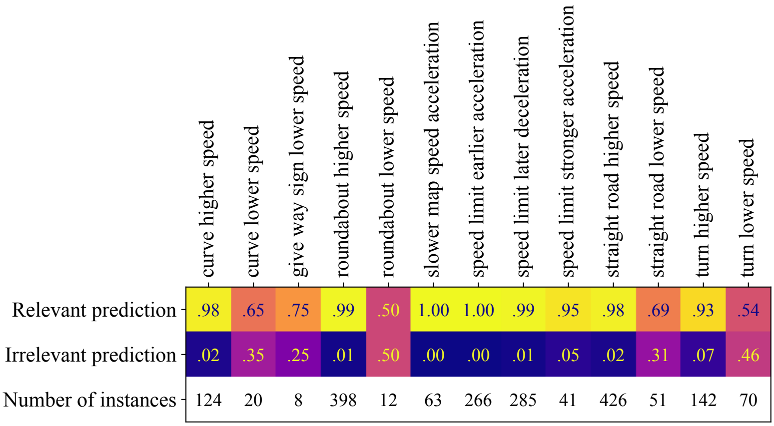


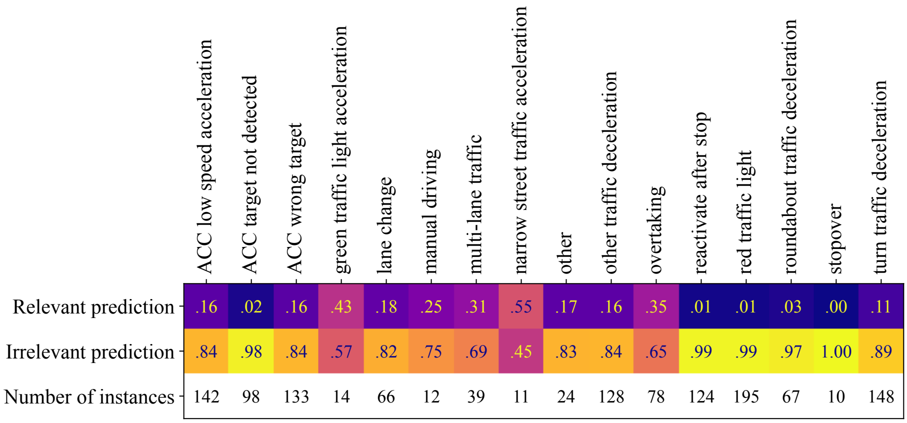
Figure 5.8: Complete class confusion matrix of the trained ROCKET model. Each row represents a ground-truth class, whereas columns represent the predicted classes. The presented confusion values are normalized over each row and rounded to two decimal digits. Relevant and irrelevant classes are separated by horizontal and vertical black bars. Confusion of exactly 0 % is colored with a white background to lower the matrix's visual complexity.

still achieves a strong performance on certain irrelevant acceleration classes, such as *ACC low speed acceleration*, *ACC wrong target*, and *reactivate after stop*. Presumably, the model may be able to distinguish these classes from the relevant classes based on the presence of an active ACC and the low-speed scenarios in which these interventions typically occur.

One reason for the poor pseudo-binary performance on certain classes is the high



(a) Aggregated pseudo-binary confusion matrix of relevant classes. The *relevant prediction* row represents the pseudo-binary recall of each relevant class, whereas the *irrelevant prediction* row represents the proportion of incorrect binary predictions.



(b) Aggregated pseudo-binary confusion matrix of irrelevant classes. The *relevant prediction* row represents the proportion of incorrect binary predictions, whereas the *irrelevant prediction* row represents the pseudo-binary recall of each irrelevant class.

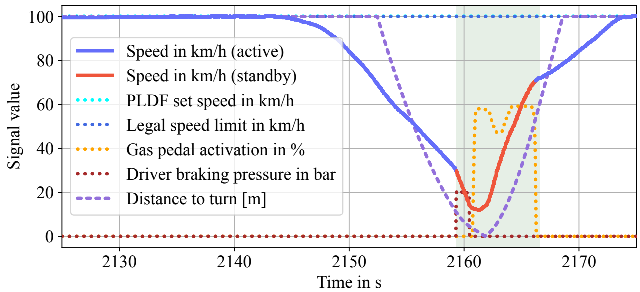
Figure 5.9: Aggregated pseudo-binary confusion matrix of relevant and irrelevant classes. For each ground-truth class, the distribution of relevant and irrelevant predicted classes is depicted. Additionally, the number of instances per class is shown for reference.

inter-binary confusion between interventions occurring in the same location but with different driver intentions. Confusion between gas and brake pedal interventions due to personal preferences and traffic interactions is particularly prevalent. For instance, 51 % of *turn lower speed* interventions are correctly classified, whereas 34 % are misclassified as *turn traffic deceleration*. Similar patterns are observed for other location-specific classes, such as *roundabout lower speed* and *roundabout traffic deceleration*. In these cases, the model fails to consistently determine whether an intervention was triggered due to personal preference or due to traffic interactions. This highlights a key limitation of the dataset: the absence of signals that capture the current traffic situation. For example, when approaching a turn, human drivers perceive the surrounding traffic and intuitively assess whether entering the turn is currently possible or not. A signal that encapsulates such contextual information would likely improve the model's performance substantially. In its current state, only the ACC-related information about the leading vehicle is available in the dataset. Consequently, the model has virtually no information regarding other surrounding or crossing traffic in situations where the driver might be required to yield to other traffic participants. Accordingly, the developed classification model struggles in these potentially traffic-related situations. The inclusion of signals describing the surrounding traffic situation could therefore significantly improve the model's performance in these cases. Such signals could potentially be derived from preprocessed object lists generated by outward-facing sensors, including radars, lidars, and cameras. Two examples of such virtually indistinguishable cases are shown in Figure 5.10. Figure 5.10a depicts a traffic-related deceleration onto an upcoming turn, i.e. a *turn traffic deceleration* intervention. However, the trained model misclassifies it as a personal preference-based intervention and predicts the *turn lower speed* class instead. Conversely, Figure 5.10c shows a ground-truth *turn lower speed* intervention, which the trained model incorrectly assumes to be traffic-related, assigning the *turn traffic deceleration* class. In both cases, it is virtually impossible to distinguish the true reason for the intervention without additional contextual information about traffic in the upcoming road segments.

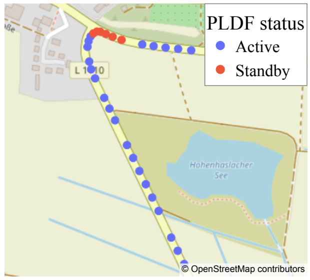
Another factor contributing to the model's varying performance on different classes is the strong imbalance in the dataset. The number of instances per class ranges from as few as eight for *give way sign lower speed* to as many as 426 for *straight road higher speed*. When analyzing the class-specific model performance, a trend is visible that classes with fewer instances generally achieve lower performance compared to classes with higher representation in the dataset. To quantify this relationship, a Spearman rank correlation was computed between each class's full classification F1-score and its number of instances in the dataset. This method was chosen because a Shapiro-Wilk test with a p-value threshold of $5\,\%$ indicated that the distributions of both the F1-scores and the instance numbers deviate from normality. The analysis revealed a strong positive correlation between the two variables ($\rho = 0.789$, $n = 29$, $p = 3.64 \cdot 10^{-7}$). These findings suggest that mitigating the underlying dataset imbalance, e.g., by recording and labeling additional data of currently underrepresented classes, could further improve the model's overall performance.

5.3.3 Permutation Feature Importance (PFI)

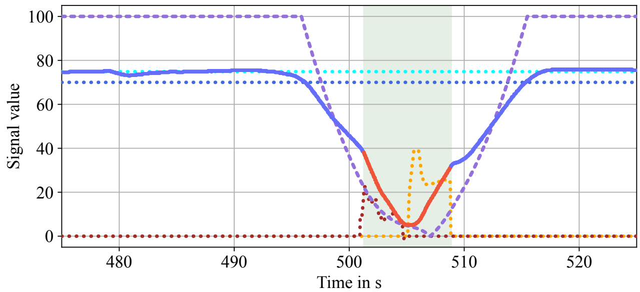
A PFI analysis is conducted on the input signals of the ROCKET model trained on the full classification task to assess the importance of individual signals for the classification performance. During this procedure, the values of each signal in the dataset



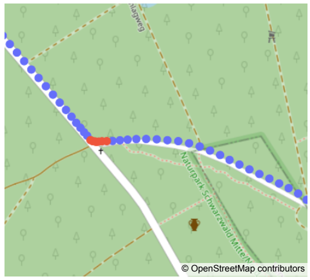
(a) A *turn traffic deceleration* intervention that is predicted as *turn lower speed* intervention by the trained model.



(b) The corresponding GPS traces of the *turn traffic deceleration* intervention.



(c) A *turn lower speed* intervention that is predicted as *turn traffic deceleration* intervention by the trained model.



(d) The corresponding GPS traces of the *turn lower speed* intervention.

Figure 5.10: Signal-over-time plots and the corresponding GPS traces of two brake pedal interventions in turns. In both cases, the trained model misclassifies the reason behind the driver intervention due to missing contextual information about the surrounding traffic. Interventions are highlighted with a green background.

are individually replaced with random noise. Subsequently, the five models obtained from the five-fold cross-validation are evaluated on their respective test splits with the randomized signals, and the resulting full-classification F1-scores are compared to the baseline performance on the unmodified dataset. To ensure robust results, this process is repeated 100 times for each signal, using a different randomization seed in each iteration. The differences in performance are then averaged over the 100 iterations. During preprocessing, the PLDF status signal was decomposed into three one-hot encoded binary signals, representing the statuses *active*, *passive*, and *standby*. Therefore, these three signals are randomized simultaneously for the PFI analysis of the *PLDF status* signal. The results of the PFI are presented in Figure 5.11, where the decrease in F1-score on the full classification task is plotted for each input signal.

As illustrated by the results, signals representing the distances to map features have a substantial influence on the performance of the trained model. In particular, the

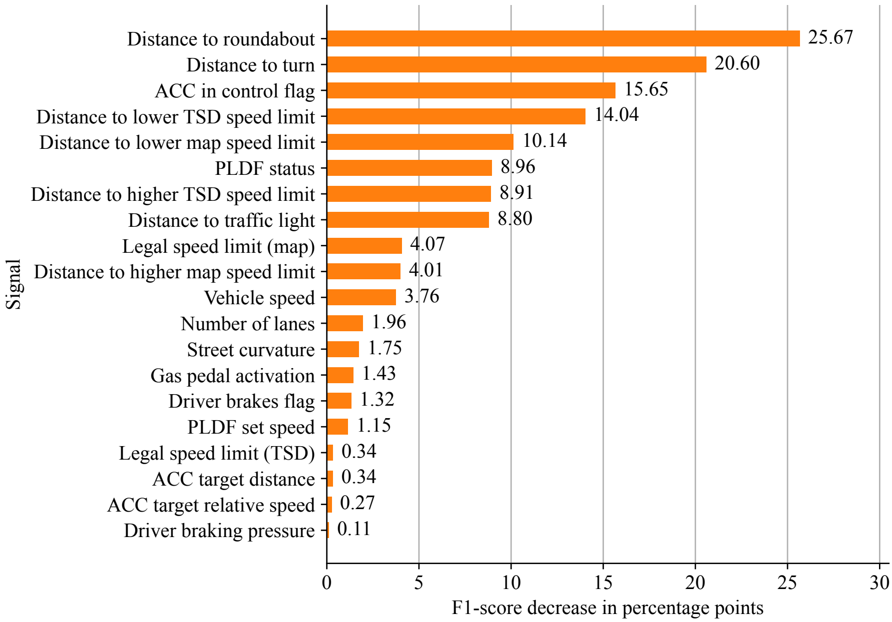


Figure 5.11: PFI results of ROCKET on the full classification task. The decrease in F1-score is depicted for each signal with randomized values.

highest decreases in the F1-score of over 20 % are caused by the randomization of the *distance to roundabout* and *distance to turn* signals. Notably, signals describing the distance to upcoming speed limit changes in both the TSD and the map data also have a considerable impact on model performance, whereas the absolute values of the legal speed limits are less influential. This suggests that information regarding the change of a speed limit is more critical for classifying driver interventions than the absolute speed limit values themselves. The absolute speed limit value from the map data still has a modest effect on the performance of 4.07 %, compared to the negligible impact of the TSD speed limit of 0.34 %. One possible explanation is that the map-based speed limit has a greater impact on the PLDF's behavior, as it predictively reacts to upcoming map-based speed limits, even if they are not detected by the TSD. In addition to the map-related signals, both the *ACC in control flag* and the *PLDF status* signals exhibit a high impact on the model's performance. The relatively high importance of the *ACC in control flag* can likely be attributed to the high number of ACC-related interventions in the dataset, which can be distinguished from free-driving interventions using this flag. Signals with almost no effect on the model's performance include the *ACC target distance* and *ACC target relative speed* signals. This indicates that only the information whether the ACC is currently active is of importance, whereas the exact distances and relative speeds of the ACC target vehicle are less important. The signals describing the activation of the gas and brake pedals are also of comparatively low importance. The reason for this is likely that they are correlated with other, more significant signals, such as the *PLDF status* and the *vehicle speed*, while not contributing additional relevant information for the classification. In summary, the most influential signals are those describing the PLDF's current operational status and the location of the intervention on the map. Additionally, it is assumed that a signal containing information about the surrounding traffic, as proposed in Section 5.3.2, would also achieve a high feature importance, given the currently high class confusion between traffic-related and non-traffic-related interventions. However, further research is required to validate this assumption.

5.3.4 Limitations and Improvement Potential

This section describes the current limitations of the developed classification pipeline, primarily including the limited dataset size and imbalance, the inclusion of traffic-related signals, and the arbitration of the classification results over multiple drive-throughs.

Dataset Size and Imbalance

As described in the previous sections, the underlying dataset exhibits an inherent class imbalance, as certain driver interventions occur considerably more frequently than others. The number of instances per individual class ranges from single-digit or low double-digit counts for rare classes to several hundred for the more common classes. Although a weighted loss function was employed during training to mitigate the impact of this imbalance, a significant correlation between the F1-score and the number of instances per class was identified in Section 5.3.2. Consequently, further research should investigate methods to reduce the effects of this imbalance, e.g., through data augmentation techniques. However, the most effective approach would be to increase the number of instances for underrepresented classes in the dataset. This could be achieved through a large-scale data collection campaign, encompassing the recording and annotation of driving data from a greater number of drivers and across a wider variety of driving scenarios. Such an effort would also address another limitation of the current dataset: its overall limited size. While certain intervention classes contain a relatively high number of instances, the model's ability to generalize beyond the data of the 17 participants in the test group study must still be evaluated.

Inclusion of Traffic-related Signals

As described in Section 5.3.2, a high degree of inter-binary confusion between traffic-related and personal preference-based intervention classes was found in the analyzed confusion matrix. Incorporating signals that represent the surrounding traffic situation could help reduce this confusion and thereby further improve the model performance.

The primary open question is how to preprocess this traffic information in a way so that it can be effectively utilized by the employed MTSC model. For instance, directly feeding the raw object lists from the perception stack into the MTSC model is not feasible due to their complex data structure. Instead, a preprocessing step is required that transforms the relevant information for the classification into time series data. Likely, a dedicated perception module would be necessary to analyze the sensor data and extract the respective information for different traffic situations individually. For example, in turns, roundabouts, and narrow street segments, the information on whether the driver is required to yield to other traffic participants would be required. For traffic light-related interventions, a simple signal representing the current status of the closest traffic light would likely be sufficient. Finally, for multi-lane-related interventions, dedicated detectors for cut-in and merging maneuvers of both the ego vehicle and surrounding traffic would be required.

Arbitration over Multiple Drive-Throughs

The primary objective of the developed classification pipeline is to identify driver interventions which are relevant for an optimization of the PLDF's behavior according to driver preferences. The results of Chapter 4 show that LI adjustments are the most applicable for a personal preference-based adjustment of the PLDF's behavior, as drivers behave relatively consistently over multiple drive-throughs of the same location. Such a LI adjustment is only applicable if the individual driver consistently intervened due to deviating personal preferences in a specific location. Accordingly, the classification results of specific interventions may also be arbitrated over multiple drive-throughs. Since consistent local intervention behavior is necessary for an adjustment of the driving function, especially false-positive classification results can be filtered out over multiple drive-throughs. For example, unintended intervention types were filtered out of the dataset during preprocessing but still occur during naturalistic driving. Accordingly, if a driver intervenes by mistake and the classification pipeline would incorrectly classify this intervention as relevant, it is unlikely that they would consistently perform the same unintended intervention again during subsequent drive-throughs of the same location. Therefore, such a false-positive classification would not necessarily result in an adjustment of the PLDF if not performed consistently.

5.4 Summary and Contributions

In this chapter, an automatic classification pipeline for driver interventions was developed, implemented, and evaluated on the driver intervention dataset created in Chapter 3. The necessary preprocessing steps of the original dataset are explained to create the instance-based dataset required for the classification task. Then, multiple state-of-the-art models are trained on the dataset and their performance on three classification tasks of varying complexity is compared. The best-performing model is ROCKET with an F1-score of 77.7 % on the full classification task and an F1-score of 92.1 % on the pseudo-binary classification task. Notably, it is found that the models trained on the full classification labels consistently outperform the models solely trained on the binary labels. This indicates that the trained models benefit from the granular class information in the full classification labels, compared to simple binary labels. The classification performance was then further evaluated via class confusion and PFI analyses, finding multiple limitations of the underlying dataset used for training. Primarily, these include a strong class imbalance and high inter-binary class confusion due to missing information about surrounding traffic, especially in potential give-way scenarios. Accordingly, the limitations of the current pipeline are discussed, and potential measures are proposed to address them. The two primary measures are as follows: First, expanding the current dataset with additional manually labeled driving data, particularly to increase the representation of underrepresented classes and enhance dataset diversity. Second, incorporating more traffic-related input signals, which could be generated by a dedicated perception module. Nevertheless, the proof of concept is generally successful, and the results demonstrate that, in particular, binary classification of driver interventions is feasible. Future work toward an implementation in customer vehicles should therefore prioritize the proposed measures to both expand the dataset and enrich the input signals in order to further improve model performance.

To summarize, the research contributions of this chapter are as follows:

- It is demonstrated that the automatic classification of driver interventions based on bus signal data is feasible, with the most promising results achieved on the binary classification task.
- The required preprocessing steps to generate a driver intervention dataset suitable for MTSC models are highlighted.
- The performance of multiple state-of-the-art MTSC models is evaluated on the dataset, comparing three classification tasks of varying complexity.
- Potential optimizations of the classification pipeline are proposed, based on detailed analyses of the classification performance.

6 Prototype Implementation and Evaluation: A Simulation-based Test Group Study

The previous chapters focused on analyzing necessary adjustments to the PLDF based on the recorded driver intervention dataset. Building upon these analyses, personalized adjustments to the PLDF's speed profile are proposed for specific locations during free-driving scenarios. This approach differs from related work in the field of driving function personalization, which typically adjusts the general driving policy and focuses primarily on ACC-related use cases. In contrast, this chapter presents the development and implementation of a prototypical driving function adjustment framework that enables the PLDF to learn from driver interventions and personalize the speed profile in locations where drivers consistently intervene. The implementation of such a self-learning driving function is enabled by the automatic classification of driver interventions described in Chapter 5, which eliminates the need for extensive manual labeling. The applicability of the proposed self-learning driving function and its effects on driver satisfaction are evaluated in a driving simulation-based test group study. Accordingly, this chapter addresses the research objectives outlined in Section 2.6.3. Both the developed prototypical driving function and the conducted test group study are covered in a research paper that is currently under review for publication [Sch+26a].

The content of this chapter is split into two main segments. First, the methodology of the proposed self-learning driving function is covered in Section 6.1. Second, the test group study that evaluates the applicability of the developed prototype is covered in Section 6.2 and Section 6.3. Section 6.2 focuses on the design of the test group study, while Section 6.3 presents its results. Finally, Section 6.4 summarizes this chapter and highlights its research contributions.

6.1 Methodology

In this section, the methodology of the developed self-learning driving function prototype is introduced. In Section 6.1.1, the general framework of the self-learning driving function is explained, while Section 6.1.2 provides more detail about the developed Speed Profile Adjustment Algorithm (SPAA).

6.1.1 Self-Learning Driving Function Adjustment Framework

The general personalization process in Definition 2.11 specifies that driving function personalization should be conducted as an iterative process consisting of three main steps:

1. Observing the driving behavior,
2. Adjusting the driving function based on driver feedback,
3. Measuring the impact of the applied personalization and repeating the process, if necessary.

This iterative structure is also reflected in the general workflow of IIL approaches, as illustrated in Figure 2.4. In these approaches, feedback is provided in the form of expert interventions during policy execution, and the effectiveness of an adjustment is evaluated by measuring the intervention frequency. Accordingly, a policy is considered satisfactory once expert interventions are no longer required during its execution.

Building upon the findings of the preceding chapters, both the general personalization process and the IIL workflow are adapted to the objective of location-specific, personalized adjustments of the PLDF based on personal preference-based driver interventions. The resulting framework is depicted in Figure 6.1:

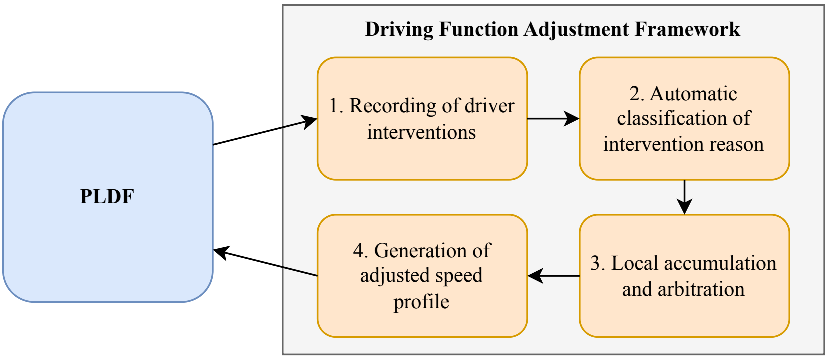


Figure 6.1: General framework of the iterative driving function adjustment based on driver interventions.

As illustrated, the framework consists of four main components: First, the PLDF is executed during naturalistic driving, while driver interventions are recorded as described in Chapter 3. Second, the reason for each intervention is automatically classified using the pipeline introduced in Chapter 5, and irrelevant interventions are removed. Third, relevant interventions are locally accumulated and arbitrated, as proposed in Chapter 4, to determine whether the driver consistently intervenes in the respective locations. Finally, if a LI adjustment is applicable based on the recorded long-term driving behavior, the SPAA is used to generate an adjusted speed profile for the respective location from the driver intervention, as explained in the following Section 6.1.2. This personalized speed profile is then stored for the individual driver, e.g., in a dedicated map layer. When the same location is traversed again, the PLDF applies the adjusted target speed profile instead of its original one, and the driver may intervene again, starting the next iteration of the adjustment process.

6.1.2 Speed Profile Adjustment Algorithm

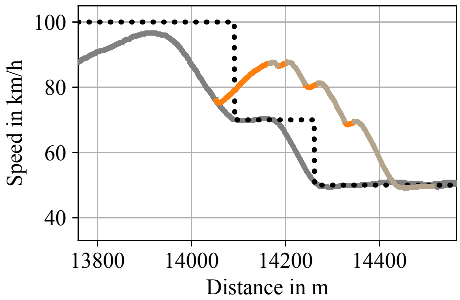
As outlined in the previous section, an algorithm is required that generates a new speed profile from conducted driver interventions in specific locations. For the generation of such a profile, two conditions must be satisfied: first, the intervention must be classified as relevant, and second, the driver must have intervened consistently during previous traversals of the respective location. As investigated in Chapter 5, the best-performing classification task is the binary distinction between relevant and irrelevant intervention types. Consequently, the developed SPAA is designed to be applicable to all relevant intervention types without requiring a distinction between their granular intervention reasons. Two different sub-algorithms are implemented to address the two primary driver input channels: set speed interventions and pedal interventions. In the case of set speed adjustments, the algorithm is relatively simple, as the chosen set speed can be adopted with minor modifications. However, for pedal interventions, a more complex solution is required.

Pedal Intervention Speed Profile Adjustment Algorithm

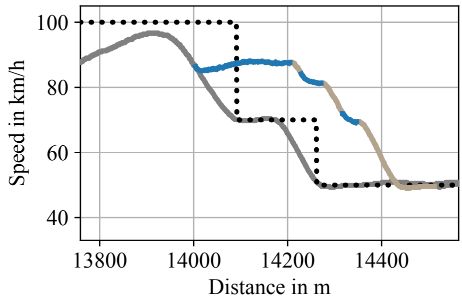
Similar to state-of-the-art IIL approaches, the objective of the pedal intervention SPAA is to derive desired behavior from recorded expert interventions. However, IIL approaches focus on general policy updates, whereas the approach presented in this thesis focuses on speed profile adjustments in specific locations only. Nevertheless, several concepts from IIL approaches can also be applied to this location-specific adjustment. As discussed in Section 2.5.3, multiple approaches exist that aim to compensate for suboptimal expert intervention behavior, most notably by addressing delays caused by human reaction times and undesirable intervention behavior [Bi+20; Spe+22]. For example, Spencer *et al.* [Spe+22] segment a vehicle trajectory containing an expert intervention into so-called *good states*, which should be reinforced in the policy, and *bad states*, which should be avoided. The authors explicitly label trajectory segments surrounding the driver intervention as bad states, since especially the vehicle states preceding the intervention describe the undesirable vehicle behavior that caused the driver to intervene. After a certain delay caused by human reaction times, the driver requires some time to steer the vehicle back into the so-called *good enough region*, after which the trajectory is is flagged as containing good states again. By defining these good and bad states, Spencer *et al.* demonstrate that not all parts of the intervention reflect desirable behavior. Similarly, Bi *et al.* [Bi+20] propose to overwrite trajectory segments around the beginning of the expert intervention by using interpolation. Thus, human reaction delays are compensated and a more desirable trajectory is generated, which is used to update the trained policy. Accordingly, the performed driver intervention should not be directly used as the new target speed profile, but instead effects such as human delays and undesirable intervention behavior need to be compensated. Furthermore, an effect that was observed during the annotation of the dataset in Chapter 3 is the tendency of drivers to overreact when intervening, pressing the gas or brake pedal considerably stronger than necessary out of dissatisfaction with the driving function. Such overreactions also need to be compensated by the developed SPAA.

The only definite information available in each driver intervention is its location and the direction of the desired speed profile adjustment, i.e., either a speed increase or decrease. Consequently, no objective and quantitative metric exists to evaluate the applicability of a new speed profile, apart from assessments by human drivers in a real vehicle or a human-in-the-loop driving simulator. Thus, qualitative analyses of the generated speed profiles were used during the development of the SPAA instead. The development of the SPAA was conducted as follows: First, an intervention catalog was defined containing several representative pedal interventions of all relevant types. Then, an initial version of the SPAA was designed and applied to all interventions in this catalog. The applicability of the generated speed profiles was assessed visually, and the SPAA was updated accordingly. This iterative process was repeated until the SPAA generated applicable speed profiles for all interventions in the catalog. The final parameterization of the SPAA was also determined during this process. Subsequently, the developed SPAA was tested and evaluated by real human drivers in the test group study covered in Section 6.2 and Section 6.3.

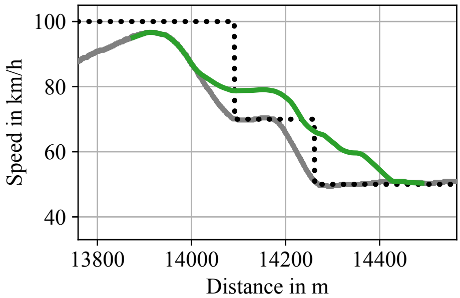
The final design of the pedal intervention SPAA is illustrated in Figure 6.2 using an example intervention. The original PLDF speed profile and the corresponding driver intervention are illustrated in Figure 6.2a. The depicted driving situation contains two consecutive speed limit decreases, indicated by the black dotted line. The PLDF's baseline speed is represented by the gray solid line. As can be seen, the PLDF abides by the speed limit and decelerates predictively onto the upcoming lower speed limits. However, the driver intervenes via the gas pedal four consecutive times, as shown by the orange solid line. Whenever the driver relinquishes the gas pedal, the PLDF takes back the vehicle control and starts to recover the vehicle speed back to its intended baseline, as illustrated in the light-brown-colored solid line. In the corresponding voice annotation, the driver stated that they would have preferred a later and smoother deceleration, maintaining a higher speed for a longer duration. The illustrated driver intervention highlights two key challenges that limit the direct usability of interventions as new target speed profiles: delays and overreactions. Although the driver stated a preference for a later and smoother deceleration, their first intervention occurs only shortly before the PLDF's baseline speed already reaches the first reduced speed limit. The subsequent gas pedal intervention is disproportionately strong, likely overshooting the driver's desired speed at the respective location. Thereafter, the PLDF resumes control and decelerates the vehicle back toward its baseline speed. However, the driver intervenes multiple consecutive times by pressing the gas pedal, creating a relatively jerky speed profile until the PLDF reaches its baseline speed again. This example illustrates why raw driver interventions cannot be directly adopted as new target profiles and underscores the need for an algorithm that derives a suitable adjusted speed profile from the original PLDF baseline speed and the driver intervention.



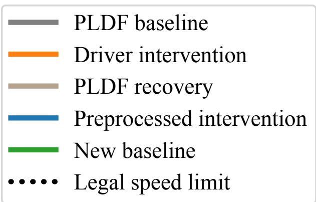
(a) The PLDF's original baseline speed and the performed example driver interventions.



(b) Preprocessing of the driver interventions: Intervention segments are stretched in negative distance direction, and a linearly decreasing offset is applied in order to realign the stretched interventions with the surrounding speed profile.



(c) The preprocessed driver interventions and the PLDF baseline speed are merged and smoothed to create the new baseline speed.



(d) Legend of the speed-over-distance plots.

Figure 6.2: The three main steps of the developed SPAA, illustrated using an example intervention during two consecutive speed limit decreases. Each subplot shows the same driver intervention at the same location, but at different stages of the pedal intervention SPAA process. The combined legend for the three speed-over-distance plots is provided in Figure 6.2d.

Since there is no objectively correct way to derive a new speed profile directly from a driver intervention, the adjusted profile must be approximated from the recorded intervention and the baseline speed. The central concept of the SPAA is to slightly shift the PLDF baseline speed profile toward the driver intervention, while compensating for delayed and disproportionate intervention behavior through appropriate preprocessing of the driver intervention. For the calculation of the new adjusted speed profile, two time series are required, the baseline speed profile over the driven distance d, denoted by $v_{\text{baseline}}(d)$, and the driver speed profile, denoted by $v_{\text{driver}}(d)$. $v_{\text{baseline}}(d)$ and $v_{\text{driver}}(d)$ generally overlap as long as the driver does not intervene. When the driver intervenes, $v_{\text{driver}}(d)$ starts to deviate from $v_{\text{baseline}}(d)$ until the intervention ends and the PLDF recovered the vehicle speed to the baseline speed. The first steps of the SPAA aim to preprocess each driver intervention separately to create the preprocessed driver speed profile $v_{\text{prepro}}(d)$. This preprocessed profile is then merged with $v_{\text{baseline}}(d)$ to create the adjusted speed profile $v_{\text{final}}(d)$. The five main steps of the SPAA are explained in more depth in the following paragraphs.

Step 1: Stretching of intervention distances. In the first step of the algorithm, each segment of $v_{\text{driver}}(d)$ containing a pedal intervention is stretched in the negative distance direction, i.e., to the left hand side of Figure 6.2a, by shifting it toward smaller distance values. In this way, the bad states contained in the preceding trajectory segments that lead to the driver intervention are also overwritten by the SPAA, and human reaction delays are compensated.

Let an intervention segment of the driver speed profile be represented by the distances $(d_i)_{i=0}^{n}$ and the corresponding velocities $(v_i)_{i=0}^{n}$, where n is the final index in the segment. To stretch the interventions, the so-called *stretch factor* α is introduced. Based on the qualitative analyses of the SPAA results, a value of $\alpha = 0.5$ was chosen. The stretched intervention's distance values $(d'_i)_{i=0}^{n}$ are then calculated using the original distance values $(d_i)_{i=0}^{n}$, the stretch factor α, and the maximum distance d_n of the driver intervention:

$$d'_i = d_i - \alpha \cdot (d_n - d_i), \qquad i = 0, \ldots, n. \tag{6.1}$$

The spatial length of the stretched intervention is then defined as

$$L_{\text{S}} = d'_n - d'_0. \tag{6.2}$$

Stretching the distances of the interventions also means overwriting preceding segments of the speed profile which may contain safety-relevant driving behavior, e.g., decelerations in high-curvature road segments. For this reason, the stretch factor is limited by two conditions in order to avoid potentially hazardous driving behavior: First, for exceedingly long interventions, the maximum stretch factor is limited such that the first distance value d_0 is moved at most a distance corresponding to 3 seconds of driving. Second, the stretch factor is reduced near high-curvature road segments depending on the magnitude of local speed variations in the portion of the trajectory to be overwritten. Further details on the computation of the stretch factor are provided in Appendix D.1.

Step 2: Applying an offset to the velocities. To avoid discontinuities at the start of the intervention, a linearly decreasing offset is applied to the intervention velocities $(v_i)_{i=0}^{n}$. This ensures that the first speed value v_0 at the stretched distance d'_0 aligns with the original driver speed $v_{\text{driver}}(d'_0)$. Let

$$\Delta v = v_{\text{driver}}(d'_0) - v_0 \tag{6.3}$$

denote the velocity difference at the beginning of the stretched segment. The offset-corrected velocities $(v'_i)_{i=0}^{n}$ are then given by

$$v'_i = v_i + \Delta v \cdot \left(1 - \frac{d'_i - d'_0}{L_{\text{S}}}\right), \qquad i = 0, \ldots, n. \tag{6.4}$$

Step 3: Constructing the preprocessed driver speed profile. After all intervention segments have been preprocessed, they are incorporated into the original driver speed profile $v_{\text{driver}}(d)$ to obtain the preprocessed driver speed profile $v_{\text{prepro}}(d)$. This is achieved by replacing the velocity values of $v_{\text{driver}}(d)$ at the adjusted distances d'_i with the corresponding preprocessed values v'_i. The resulting preprocessed profile $v_{\text{prepro}}(d)$ is illustrated in Figure 6.2b by the blue and light-brown solid lines. The blue segments represent the preprocessed interventions stretched to the left and smoothly aligned with the surrounding speed profile.

Step 4: Merging the preprocessed and the baseline speed profiles. In regions where the preprocessed driver profile $v_{\text{prepro}}(d)$ deviates from the baseline $v_{\text{baseline}}(d)$, the two profiles are merged by computing their arithmetic mean:

$$v_{\text{mean}}(d) = \frac{v_{\text{baseline}}(d) + v_{\text{prepro}}(d)}{2}. \tag{6.5}$$

This averaging step compensates for disproportionately strong driver interventions while preserving the general shape of the baseline profile. The resulting profile $v_{\text{mean}}(d)$ therefore remains close to the PLDF's baseline speed, but is shifted toward the driver's preferences.

Step 5: Smoothing. Finally, the merged profile $v_{\text{mean}}(d)$ is smoothed using a second-order Savitzky-Golay filter to obtain the final adjusted speed profile $v_{\text{final}}(d)$. The adjusted speed profile generated from the driver intervention is illustrated by the green solid line in Figure 6.2c. As can be seen, the resulting speed profile is closer to the original baseline speed than the raw driver intervention, yet it still captures the driver's intention to decelerate later and more gradually onto the upcoming lower speed limits. Moreover, the human reaction delays are compensated, as the adjusted speed profile begins to deviate from the PLDF baseline speed slightly earlier than the unprocessed driver intervention.

The developed SPAA is designed for iterative use: the individual driver repeatedly drives along the same route, and each time a relevant intervention occurs, the SPAA updates the speed profile accordingly. On the subsequent drive, the newly adjusted speed profile generated by the SPAA is used as the PLDF's new baseline speed. If the driver is still dissatisfied with the new speed profile, they may intervene again, thereby starting the next iteration of the SPAA. This process continues iteratively until the driver is satisfied with the speed profile and stops intervening.

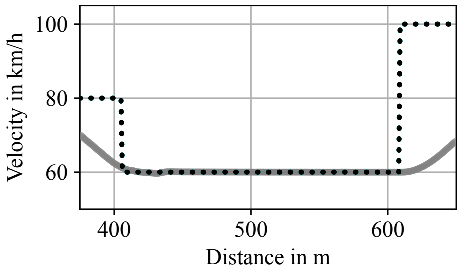
Set Speed Intervention Speed Profile Adjustment Algorithm

As previously discussed, the SPAA for set speed interventions is relatively simple and primarily revolves around adopting the chosen set speed by the driver with only minimal preprocessing. While pedal interventions involve short-term modifications to the PLDF's behavior in specific locations, set speed interventions influence the PLDF's behavior across the entire current road segment. Consequently, set speed interventions involve more long-term decision-making and typically do not contain overreactions by drivers. Accordingly, the set speed adjustment is adopted as demonstrated by the driver with only two necessary processing steps.

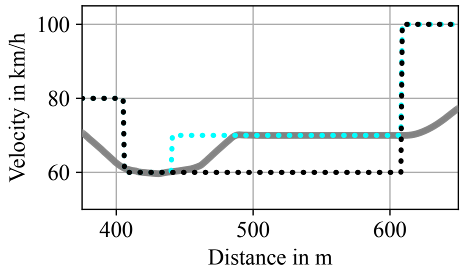
First, set speed interventions can only be conducted in increments of $1\,\mathrm{km/h}$ or $10\,\mathrm{km/h}$. For example, an adjustment of $5\,\mathrm{km/h}$ requires five consecutive activations of the ADAS control stalk, each increasing the set speed by $1\,\mathrm{km/h}$. Therefore, the first preprocessing step consists of aggregating set speed adjustments that were conducted in short succession. For example, in this case the five individual $1\,\mathrm{km/h}$ interventions are combined into a single $5\,\mathrm{km/h}$ intervention. In this thesis, a threshold of $2\,\mathrm{s}$ is applied for the merging of consecutive set speed interventions, i.e., two set speed interventions are combined if the time interval between them is less than $2\,\mathrm{s}$.

Second, set speed adjustments that are performed shortly after entering a new legal speed segment are applied to the entire segment retroactively. During the qualitative analysis of the intervention dataset from Chapter 3 and the preliminary tests of the test group study presented in this chapter, it was observed that drivers frequently set their preferred set speed shortly after passing a new legal speed limit. Therefore, set speed adjustments executed within $4\,\mathrm{s}$ of entering a new legal speed segment are applied retroactively to the entire segment. An example of this retroactive set speed processing is depicted in Figure 6.3.

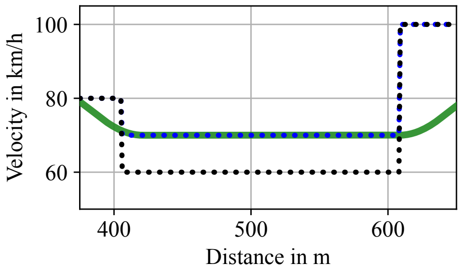
The original set speed in this example location, its corresponding baseline speed profile, and the legal speed limit are illustrated in Figure 6.3a. As can be seen, the original set speed and the legal speed limit overlap, since no set speed intervention was performed. In Figure 6.3b, an example set speed intervention in this road segment is depicted, where the driver increases the set speed from $60\,\mathrm{km/h}$ to $70\,\mathrm{km/h}$ shortly after passing the



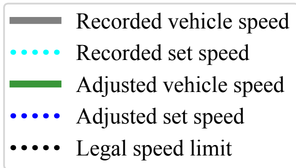
(a) The PLDF’s original baseline speed without any performed set speed changes.



(b) An example set speed intervention, changing the set speed in the current segment from 60 km/h to 70 km/h shortly after passing the lower speed limit sign.



(c) The retroactively adjusted set speed and its corresponding adjusted new baseline speed.



(d) Legend of the speed-over-distance plots.

Figure 6.3: An example for the retroactive adjustment of set speed interventions performed shortly after entering a new legal speed limit segment. Each subplot depicts the same location but at different stages of the set speed intervention SPAA process. The combined legend for the three speed-over-distance plots is provided in Figure 6.3d.

speed limit sign. The proposed processing of the set speed is based on the assumption that the driver would have preferred the chosen set speed for the entire segment, rather than temporarily adhering to the legal speed at the beginning of the segment. The corresponding retroactively adjusted set speed and its associated adjusted speed profile are presented in Figure 6.3c. This new baseline speed is then used during the next drive-through of the respective location by this individual driver.

Finally, a third set speed processing step is proposed, however, it was not implemented in the test group study presented later in this chapter. As stated above, set speed adjustments generally do not involve overreactions by drivers due to dissatisfaction with the PLDF. Therefore, no additional processing steps are required to mitigate disproportionate intervention behavior, as is the case with pedal interventions. Nonetheless, set speed adjustments may still be delayed, and the adjusted speed profile may still contain some bad states preceding the set speed adjustment. These bad states could be overwritten by generally applying the adjusted set speed slightly earlier, compared to the adjustment shown in Figure 6.3 that is only applied if the set speed adjustment was conducted shortly after a legal speed change. Despite the possible benefits, two main reasons led to the decision not to implement this third processing step: First, preliminary driving simulator tests did not indicate that such an adjustment was critically necessary. Second, its specific implementation conflicted with the design of the PLDF mockup employed during the study, resulting in disproportionate implementation efforts. For these reasons, the third processing step was not incorporated into the study.

6.2 Study Design

In this section, the design of the second test group study conducted over the course of this thesis is explained. For the discrimination of the two test group studies, the study in this chapter is called the *prototype test group study* whereas the test group study in Chapter 3 is called the *dataset creation test group study*.

6.2.1 Scope

The scope of the prototype test group study is to evaluate the applicability of the developed prototypical self-learning driving function described in Section 6.1. In particular, the self-learning driving function is compared to the static base implementation of the PLDF that was employed in the dataset creation test group study. The results presented in Chapter 3 showed that drivers were moderately satisfied with the base PLDF on average, and that a high intervention frequency was observed. Furthermore, the intervention frequency was found to correlate significantly with the driver dissatisfaction, suggesting that a reduction of the intervention frequency, especially within the PLDF's ODD, is a valid objective for improving driver satisfaction. Based on these findings, two hypotheses are evaluated in the prototype test group study:

Hypothesis 6.1 (Increase of Driver Satisfaction with the Proposed System)
The driver satisfaction is increased by the proposed self-learning driving function compared to the original baseline PLDF.

Hypothesis 6.2 (Decrease of Intervention Frequency with the Proposed System)
The frequency of driver interventions is reduced by the proposed self-learning driving function compared to the original baseline PLDF.

Furthermore, Hypothesis 3.1 from Section 3.1.1 is revisited. Although the first test group study identified a significant correlation between driver dissatisfaction and intervention frequency, the test group study was based on a relatively small sample size of only 17 participants. For this reason, the prototype test group study reevaluates this hypothesis based on a larger number of participants.

To minimize environmental factors that could influence the study results, such as weather and traffic, the study is conducted in a human-in-the-loop driving simulator. Furthermore, the automatic classification of driver interventions and the accumulation of driver behavior over multiple drive-throughs, as proposed in the general self-learning driving function framework in Section 6.1.1, are omitted for the purposes of this study. The driving scenarios in the simulation are designed so that the PLDF always remains within its ODD, as no other traffic participants are present and no perception or map errors occur that would require driver corrections. Accordingly, the study is limited to free-driving scenarios in which all driver interventions can be assumed to be relevant, rendering the classification step obsolete. In addition, the local accumulation and arbitration of driver interventions are omitted due to time restrictions. Because the study duration per participant is limited, performed driver interventions are directly used to update the baseline speed in the SPAA without prior accumulation and arbitration. Thus, the study focuses in particular on evaluating the applied SPAA and addressing the overarching question of whether updating a driving function based on driver interventions can indeed increase driver satisfaction.

6.2.2 Participants

The recruitment of study participants was organized via an internal Porsche employee volunteer portal. There, Porsche employees across all departments have the opportunity to sign up for internal test group studies. As a requirement for participation, the subjects had to identify themselves as regular users of ADASs. This requirement was chosen to ensure that participants were already familiar with the operation and working principles of ADASs, consistent with the requirements of the dataset creation test group study defined in Section 3.1.1. Due to confidentiality reasons, only Porsche employees were allowed to participate in the test group study.

In total, 46 participants took part in the study, of which 43 completed the procedure with valid results. Two participants did not finish the experiments due to feeling unwell inside the driving simulator, and for one participant, technical issues with the simulator occurred, making their data invalid. The gender and age distribution of the participants is shown in Table 6.1.

Table 6.1: Gender and age distribution of the prototype test group study participants.

Gender	$20 < \text{Age} \leq 30$	$30 < \text{Age} \leq 40$	$40 < \text{Age} \leq 50$	$50 < \text{Age} \leq 60$
Female	8	4	0	2
Male	9	7	6	7

6.2.3 Simulation Environment

The driving simulator used in this study is depicted in Figure 6.4. As illustrated, it is composed of a hexapod with a movable platform onto which a vehicle cockpit is mounted. The hexapod is used to simulate longitudinal and lateral accelerations as well as rotational movements, while vibrations are generated depending on the simulated vehicle speed. In addition, road and vehicle noises are reproduced via a sound system corresponding to the simulated speed. Surrounding the hexapod, high-resolution Light-Emitting Diode (LED) walls are mounted, which display the simulation environment. In contrast to Figure 6.4, a realistic, closed, and fully-equipped vehicle cockpit was mounted on the motion platform during the study. Accordingly, the vehicle interior was not rendered on the LED walls, but only the simulation environment. Due to confidentiality reasons, no photograph of the final simulator setup can be shown in this thesis.

As discussed in the dataset creation test group study in Section 3.1.1, drivers behave differently on familiar routes compared to previously unknown ones. Consequently, the route used in the study should ideally be well known to the participants. However, in this simulation-based study, no high-quality simulation map of a route familiar to all participants is available. Therefore, an internal fictional simulation map is employed, and the study design is adjusted to include multiple introductory drives on the simulation track to allow participants to familiarize themselves with the route. A key requirement for the chosen route is that it contains a sufficient number of locations where relevant intervention types can be conducted. These include straight road segments, speed limit increases and decreases, curves, turns, and roundabouts. However, one limitation of the used hexapod is its maximum rotation angle, which is insufficient to accurately simulate the traversal of turns or roundabouts. In such cases, the hexapod rotates only up to its maximum angle and then remains stationary, while the surrounding LED walls continue to render the rotating environment. This may cause motion sickness for some participants, and therefore turns and roundabouts were excluded from the simulation track. Taking these requirements and limitations into account, the rural road segment shown in Figure 6.5 was selected for the study. As illustrated, the route features frequently changing legal speed limits, mostly $80\,\mathrm{km/h}$ and $100\,\mathrm{km/h}$, and several curves that must be traversed at speeds below the corresponding legal speeds.

Figure 6.4: Porsche human-in-the-loop driving simulator used in the prototype test group study.

The map further incorporates a small village and a short bridge over a highway, with speed limits of $50\,\mathrm{km/h}$ and $60\,\mathrm{km/h}$, respectively. In total, the track is $4.5\,\mathrm{km}$ long, and a complete drive-through takes approximately $3\,\mathrm{min}$ and $30\,\mathrm{s}$.

For the test group study, the PLDF was reimplemented as a Python mockup that communicates with the driving simulator via User Datagram Protocol (UDP). This mockup replicates both the in-production PLDF's driving behavior as well as its user interface. The resulting baseline speed profile, obtained when the PLDF mockup is activated on the simulation track, is shown in Figure 6.6. As illustrated, the PLDF mockup always abides by the legal speed limits and decelerates in high-curvature road segments. The maximum curve speeds, calculated using the original parameterization of the in-production PLDF, are indicated as black bars in the plot.

The objective of the present study is to investigate real-world driving behavior within a simulation-based environment. Accordingly, a realistic driving simulation is required to ensure that the results can be meaningfully transferred to real-world driving. Therefore, the applicability of the employed simulation environment and study design to real-world conditions is discussed in Appendix D.2.

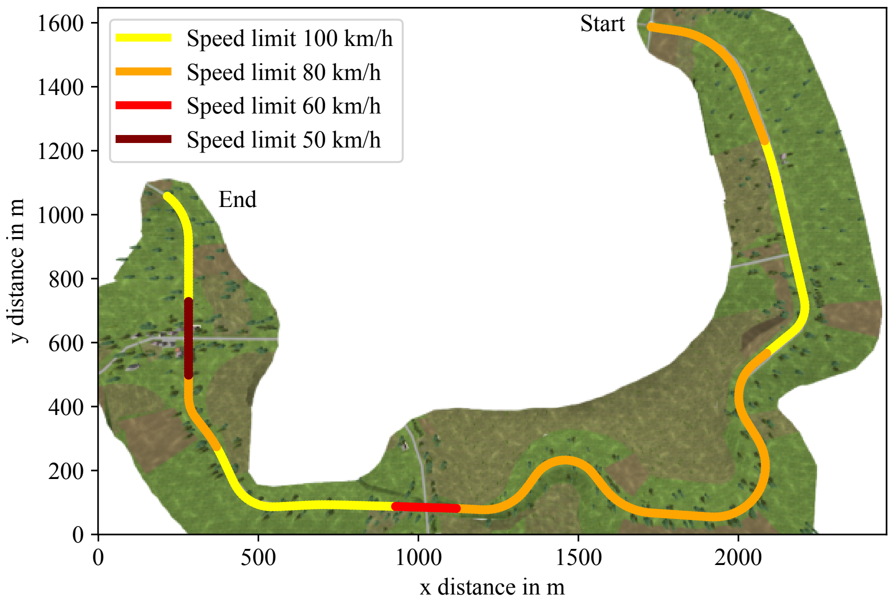


Figure 6.5: Simulation map segment used for the prototype test group study with highlighted speed limits.

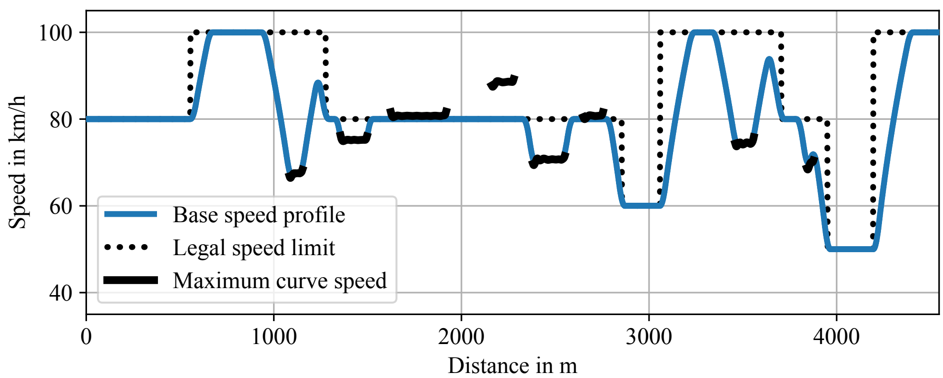


Figure 6.6: Speed-over-distance plot of the PLDF mockup's baseline speed profile on the simulation track without driver interventions.

6.2.4 Study Procedure

The test group study is designed as a blind study, in which each participant tests and evaluates both the baseline PLDF and the self-learning PLDF without being informed about the differences between the two systems. The detailed structure and timeline of a participant session are presented in Table 6.2.

Table 6.2: Structure and timeline of a participant session in the prototype test group study. The stated durations are approximations and may vary for each participant.

Category	Content	Duration
Conversation	Greeting	3 min
	Briefing	10 min
Introductory drives	Test drive, familiarize with simulator	4 min
	Test drive, familiarize with simulation track and PLDF baseline speed	4 min
System A experiments	Test drive with interventions	4 min
	Test drive with interventions	4 min
	Test drive without interventions	4 min
Questionnaire	System A questionnaire	5 min
System B experiments	Test drive with interventions	4 min
	Test drive with interventions	4 min
	Test drive without interventions	4 min
Questionnaire	System B questionnaire	5 min
Conversation	Free-form feedback	5 min
	De-briefing	5 min

Before driving in the simulator, participants are briefed on the content of the study using the written briefing presented in Appendix D.3. They are informed that the focus of the study is to evaluate two versions of the same PLDF, referred to as *system A* and *system B*. System A represents the static baseline PLDF, while system B denotes the self-learning version of the PLDF. Participants are not informed about the differences between the two systems. They are only told that the evaluated driving functions are adaptive to the driver behavior in some way. This information is provided because preliminary tests showed that drivers may become confused by the changing driving behavior of system B if they are not informed about some form of function adjustments. The briefing additionally includes a summary of the study procedure and a description of the used baseline driving function, covering its operation, functionality, and ODD. For the purposes of the study, an additional steering assist is permanently active in the driving simulator which performs active lane centering. This steering assist is intended to support participants on the previously unknown track and reduce their cognitive load, allowing them to focus on the longitudinal vehicle control. The steering assist is implemented to support both hands-on and hands-off driving, depending on the driver preference. However, it is explicitly stated that only the longitudinal control is subject of the evaluation in the questionnaires. Finally, an introduction to the driving simulator and safety instructions are provided. During the study, the participant and the study operator may communicate via a two-way radio system.

After completing the briefing, participants are seated in the simulator and the experiment starts with two introductory drives. During the first drive, participants are instructed to familiarize themselves with the simulator, including manual driving and exploring all possible ways to interact with the driving function. In the second introductory drive, participants are instructed to activate the PLDF, in this case system A, and complete a full drive-through of the simulation track without intervening, in order to familiarize themselves with both the simulation track and the PLDF baseline speed profile. Following the introductory drives, the system A and B tests are conducted, each of which consists of three test drives. During the first two test drives, drivers are instructed to drive with the PLDF as they would in real-world conditions, and that they are free to intervene as they see fit using both the pedals and set speed changes. In the third test drive, the participants are instructed to activate the PLDF without intervening, in order to experience the final speed profile of the driving function. For system A, the baseline speed profile remains unchanged across all drives, mimicking the behavior of the in-production PLDF, which does not adapt its profile to driver preferences. Upon completion of the test drives, participants exit the simulator and evaluate the experienced driving function using the questionnaire described in Section 6.2.5.

System B is then pretrained by applying the SPAA on the interventions recorded during the second drive with system A. Only the second drive with interventions is used for the pretraining, as the analyses in Section 3.3.2 indicate that drivers typically intervene less frequently during their first drive with the PLDF compared to subsequent drives. Accordingly, the second drive resembles the long-term intervention behavior of drivers more closely than the first drive. Thus, the first drive with system B already features an updated PLDF baseline speed. After each drive with system B, the SPAA is applied, further adjusting the baseline speed according to the performed driver interventions. During the third drive, the driver experiences the final baseline speed of system B without performing any interventions. After completing the drives, the participants exit the simulator again and complete the system B questionnaire, which is identical to the system A questionnaire. In a final discussion, participants may provide free-form feedback, which is recorded by the study operator. Subsequently, the operator debriefs the participant, explaining the purpose and content of the study. In total, a 1 h time slot was planned for each participant with a 30 min buffer between sessions. This session duration was recommended by the driving simulator experts at Porsche, as the concentration of the participants tends to decrease over time while the likelihood of motion sickness increases. In practice, the effective duration often exceeded the planned time slot, with an average session length of approximately 1 h and 15 min. This study procedure generally adheres to current best-practice guidelines for simulation-based test group studies [Hoc+18].

One limitation of the study design is that system B requires pretraining on the intervention data from system A, which necessitates a fixed experiment order, i.e., system A before system B. Only with a significantly longer experiment duration, a randomization of the experiment order would have been possible. However, due to participant concentration and motion sickness constraints, the study duration was limited to 1 h. The fixed experiment order introduces the possibility of learning effects, whereby participants might evaluate system B more favorably than system A due to being more comfortable with the task over time. However, the influence of learning effects might be limited in the context of this study, since the driving task is relatively simple and participants already complete five drives with system A before evaluating it, while the system B tests add only three additional drives. Furthermore, the fixed experiment order more realistically reflects the potential customer experience with the self-learning driving function, where the customer first performs multiple drives with the baseline system. Only after a certain number of drives with the baseline PLDF, the speed profile adjustments are conducted based on the recorded intervention behavior, after which the customer experiences the updated speed profile. Thus, it is essentially impossible for a customer to experience system B, trained on their own interventions, without first experiencing system A.

6.2.5 Questionnaire

The translated questionnaires used during the prototype test group study are provided in Appendix D.4. The first page of the questionnaire inquires about the participant's demographic information as well as driving and ADAS experience. This demographic data is filled out by the participants after reading the briefing and before entering the driving simulator. For the evaluation of system A and system B, identical questionnaires are used, consisting of two parts each. First, the system acceptance scale by van der Laan, Heino, and De Waard [VHD97] is used to assess the participant's acceptance of the PLDF, including its satisfaction and usefulness subscales. Second, a custom questionnaire specifically addresses the participant's satisfaction with the PLDF's speed profile in relevant situations. These include the speed choice on straight roads and in curves, the acceleration timing when passing higher and lower speed limit signs, and the overall strength of accelerations and decelerations.

6.3 Results

In this section, the results of the prototype test group study are highlighted. This includes the quantitative evaluation of the three hypotheses stated in Section 6.2.1, as well as qualitative analyses of the generated speed profiles and the recorded free-form feedback.

6.3.1 Driver Satisfaction Analysis

Multiple human factors measures can be derived from the two questionnaires used to evaluate both system A and system B. The first questionnaire is the system acceptance scale [VHD97], which provides a measure of the driver satisfaction with the system, as well as its perceived usefulness. These measures are called the *general satisfaction* and *general usefulness* in the context of this thesis, and can be calculated as the mean value of their respective items in the questionnaire, as explained in Section 2.1.7. Additionally, the *general acceptance* can be calculated as the mean of all nine items in the questionnaire. The subsequent custom questionnaire inquires specifically about the participants' satisfaction with the PLDF's speed profile. Its second item, *I was satisfied with the number of times I had to intervene*, directly describes the *intervention frequency satisfaction*. Finally, by calculating the mean of the questionnaire's last six items inquiring about the satisfaction in specific situations, the *speed profile satisfaction* is calculated. For each of these five scores, the following measures are derived:

- $N_{A>B}$ is the number of participants who rated system A better than system B.
- $N_{A=B}$ is the number of participants who gave equal ratings for system A and system B.
- $N_{A<B}$ is the number of participants who rated system A worse than system B.
- $\bar{A}$ is the mean score of system A across all participants.
- $\bar{B}$ is the mean score of system B across all participants.

For the scores, 5.0 represents the best possible score, while 1.0 is the worst possible score. The resulting values are depicted in Table 6.3.

As can be seen, participants generally assigned higher scores to system B compared to system A. However, some participants still favored system A in some regards, or assigned equal scores for system A and B. The three measures derived from the system acceptance scale, *general satisfaction*, *general usefulness*, and *general acceptance*, show relatively similar results, with a mean score increase of approximately 0.5 for system B compared to system A. In contrast, the custom questionnaire reveals stronger score increases of more than 1.0 for system B. Notably, 40 out of 43 participants rated the speed profile of system B better than system A, while only 34 participants reported a higher general satisfaction with system B. Thus, some participants prefer system B's speed profile, while rating the general system worse than, or equal to, system A. This difference can be explained by the questionnaire design, as the system acceptance scale inquires about the entire system experience, whereas the custom questionnaire focuses specifically on the satisfaction with the PLDF's speed profile. Accordingly, certain

Table 6.3: The questionnaire results of the prototype test group study. Depicted are the number of participants who rated system A better, worse, or equal to system B, and the mean scores for both systems. The best possible score for a system is represented by 5.0, whereas 1.0 describes the worst possible score.

Measure	$N_{A>B}$	$N_{A=B}$	$N_{A<B}$	$\bar{A}$	$\bar{B}$	$\bar{B}-\bar{A}$
General satisfaction	6	3	34	3.69	4.27	0.58
General usefulness	5	3	35	3.71	4.16	0.45
General acceptance	4	2	37	3.70	4.21	0.51
Intervention frequency satisfaction	3	11	29	3.14	4.23	1.09
Speed profile satisfaction	3	0	40	3.06	4.14	1.08

aspects of system B are negatively perceived by some participants, despite the general preference for system B's speed profile. To investigate these aspects, the free-form feedback is analyzed in Section 6.3.5.

Another notable trend in the data is that a relatively high number of eleven participants did not report an increase in their intervention frequency satisfaction, despite reporting a higher satisfaction with the speed profile. This effect may be attributed to the study design, which includes only two drives with interventions per system. Accordingly, if a participant was not fully satisfied with system B's speed profile during the first and second drive, they would continue to intervene in the subsequent drives. Therefore, their intervention frequency remains high for system B, even though they might report a higher satisfaction with the resulting speed profile. To correctly analyze the intervention frequency satisfaction, longer experiments with a higher number of iterations would be required, capturing the participants' long-term intervention behavior. If the SPAA works as intended, drivers would stop intervening after a certain number of iterations, due to being satisfied with system B's speed profile. However, these long-term effects could not be analyzed in this study and should be the topic of future research.

The score differences between system A and B are evaluated using statistical significance tests. A Shapiro-Wilk test is first applied to assess the normality of the score differences. Normality is assumed if the resulting p-value exceeds a threshold of $5\,\%$. For normally distributed data, a paired-samples Student's t-test is used, while the Wilcoxon signed-rank test is applied otherwise. For each of the five evaluated measures derived from the questionnaires, the applied statistical test and its results are reported in Table 6.4. As illustrated, the score differences are statistically significant across all five measures. Thus, Hypothesis 6.1 can be confirmed: The self-learning driving function significantly increases the driver satisfaction with the system in general, but especially with the driving function's speed profile, compared to the original PLDF.

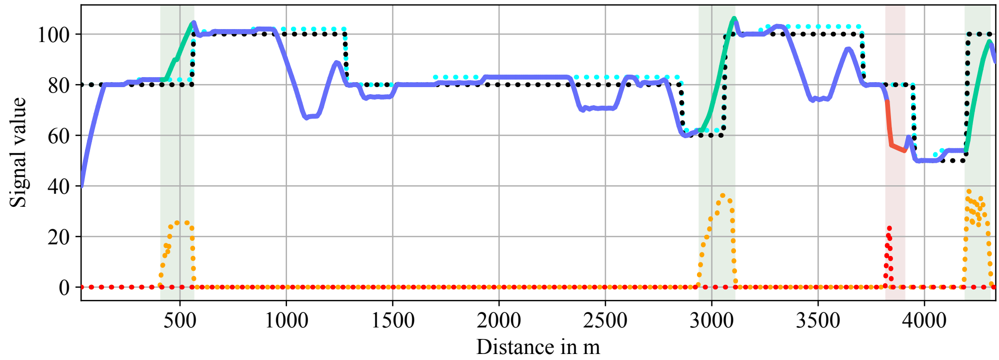
Table 6.4: Statistical significance test results for the five human factors measures derived from the questionnaires. Depending on the normality of the underlying data, either a Student's t-test for paired samples or a Wilcoxon signed-rank test is applied.

Measure	Test	Statistic	p-value
General satisfaction	Paired t-test	$t(42) = 4.60$	$p = 3.85 \cdot 10^{-5}$
General usefulness	Wilcoxon signed-rank	$W = 95.0$	$p = 2.14 \cdot 10^{-5}$
General acceptance	Wilcoxon signed-rank	$W = 103.5$	$p = 2.25 \cdot 10^{-5}$
Intervention frequency satisfaction	Wilcoxon signed-rank	$W = 25.5$	$p = 5.31 \cdot 10^{-6}$
Speed profile satisfaction	Wilcoxon signed-rank	$W = 45.0$	$p = 3.83 \cdot 10^{-9}$

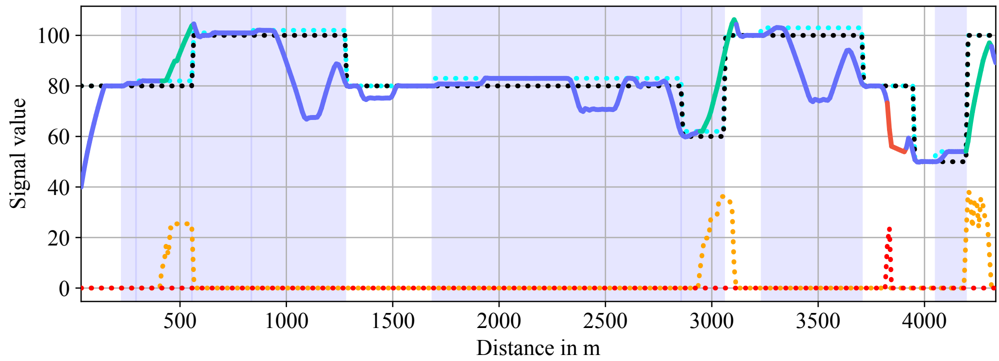
6.3.2 Intervention Rate Analysis

Hypothesis 6.2 states that the frequency of driver interventions should be reduced by system B compared to system A. Therefore, IRs are calculated, similar to the process explained in Section 4.1. In the context of the prototype test group study, an IR defines the relative time of a drive where the participant intervened. These IRs are calculated separately for pedal and set speed interventions, resulting in the *pedal IR* and *set speed IR*, respectively. The *combined IR* is defined as the relative time per drive where any intervention was active. An example drive with system A and highlighted interventions is illustrated in Figure 6.7. As can be seen, the driver performs two gas pedal interventions and one brake pedal intervention in Figure 6.7a. The resulting pedal IR amounts to $11.27\,\%$. The participant also adjusts the set speed in multiple segments of the drive, resulting in a set speed IR of $63.32\,\%$ in Figure 6.7b. The resulting combined IR with $67.96\,\%$ is only slightly higher than the set speed IR in this case, as there are multiple segments where both a pedal and a set speed intervention are currently active, and the combined IR only describes the relative time during which any type of intervention was conducted. For system B drives, the set speed IR is only calculated for set speed changes conducted in the current iteration, since set speed changes from previous iterations are automatically adopted and incorporated into the next baseline speed profile.

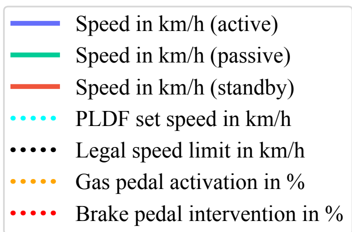
The three different IRs are calculated for each drive with interventions across all 43



(a) Signal-over-time plot of an example drive with system A. Gas and brake pedal interventions are highlighted with a green and red background respectively. The resulting pedal IR amounts to 11.27 %.



(b) Signal-over-time plot of an example drive with system A. Set speed interventions are highlighted with a blue background. The resulting set speed IR amounts to 63.32 %.



(c) Legend of the signal-over-time plots.

Figure 6.7: Signal-over-time plots of the same example drive with system A, highlighting pedal and set speed interventions separately. The combined IR of the depicted example drive amounts to 67.96 %. The legend of both signal-over-time plots is provided in Figure 6.7c.

participants, and the resulting mean values for each iteration are illustrated in Figure 6.8. The first two intervention drives are conducted with system A, whereas the following two drives are conducted with system B. The SPAA is applied after the second iteration with system A and after the first iteration with system B. As illustrated by the data, all three IRs are reduced substantially following the first application of the SPAA. The mean IRs with system A are denoted as $\bar{\mathrm{IR}}_{\mathrm{A}}$, while the mean IRs with system B are denoted as $\bar{\mathrm{IR}}_{\mathrm{B}}$. The combined IR drops from an $\bar{\mathrm{IR}}_{\mathrm{A}}$ of $54.86\,\%$ to $22.97\,\%$ $\bar{\mathrm{IR}}_{\mathrm{B}}$. The strongest relative reduction is observed in the set speed IR, which decreases from $39.76\,\%$ $\bar{\mathrm{IR}}_{\mathrm{A}}$ to $12.42\,\%$ $\bar{\mathrm{IR}}_{\mathrm{B}}$. The pedal IR also shows a strong reduction, falling from $22.32\,\%$ $\bar{\mathrm{IR}}_{\mathrm{A}}$ to $12.04\,\%$ $\bar{\mathrm{IR}}_{\mathrm{B}}$.

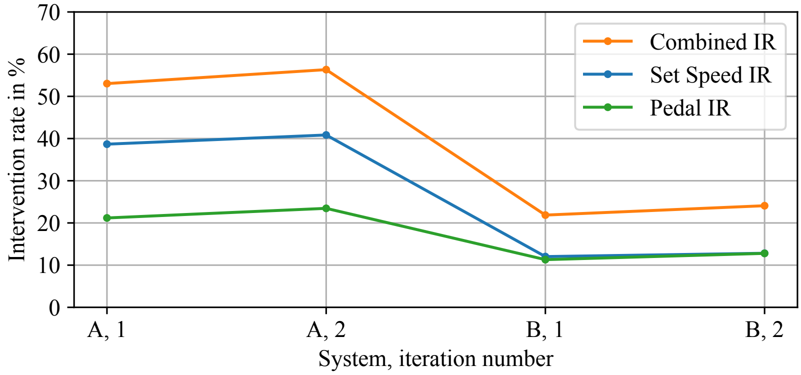


Figure 6.8: Development of the mean IRs for each drive with interventions over the course of the experiment. The x-axis denotes the corresponding system and iteration number. The SPAA is applied after the second iteration of system A and the first iteration of system B.

Notably, the set speed $\bar{\mathrm{IR}}_{\mathrm{A}}$ is considerably higher than the pedal $\bar{\mathrm{IR}}_{\mathrm{A}}$. This can be explained by the fact that set speed changes typically influence the speed profile over large segments of each drive, whereas pedal interventions more often affect only the short-term behavior of the PLDF. Another important observation is that both $\bar{\mathrm{IR}}_{\mathrm{A}}$ and $\bar{\mathrm{IR}}_{\mathrm{B}}$ slightly increase when comparing the first and second iteration of each system. This aligns with the findings in Section 3.3.2, which indicated that drivers typically intervened less frequently during their first drive with the PLDF in the dataset creation test group study compared to subsequent drives. For some drivers, this behavior was noticed by the study operator during the study execution. When asked about their reasoning for intervening less during the first iteration of each system, some drivers reported that they wanted to observe the system's driving behavior more closely before performing additional interventions during the second iteration. A further observation is that, although the IRs decrease substantially over the course of the experiments, the final combined IR does not reach $0\,\%$, and drivers still intervene during the second drive with system B. This indicates that drivers are not fully satisfied with the final speed profile after two iterations, and that future work should investigate the long-term intervention behavior of drivers when using the self-learning PLDF. An analysis of whether the combined IR with system B ever converges to $0\,\%$ over multiple iterations would be particularly relevant.

The statistical significance of the differences in IRs is evaluated using either Student's t-tests for paired samples or Wilcoxon signed-rank tests, depending on the normality of the underlying data. A Shapiro-Wilk test is applied to assess normality, again using a p-value threshold of $5\,\%$. The mean IRs of system A and B, together with the significance test results, are presented in Table 6.5. The results indicate that the reduction in IRs is statistically significant for all three IR types. Thus, Hypothesis 6.2 can be confirmed: The frequency of both pedal and set speed interventions is significantly reduced by the proposed self-learning PLDF compared to its original baseline version.

Table 6.5: Statistical significance test results for the IR reduction between system A and system B. Depending on the normality of the underlying data, either a Student's t-test for paired samples or a Wilcoxon signed-rank test is applied.

IR type	$\bar{\mathrm{IR}}_\mathrm{A}$	$\bar{\mathrm{IR}}_\mathrm{B}$	Test	Statistic	p-value
Pedal	$22.32\,\%$	$12.04\,\%$	Paired t-test	$t(42) = -6.42$	$p = 9.78 \cdot 10^{-8}$
Set speed	$39.76\,\%$	$12.42\,\%$	Wilcoxon signed-rank	$W = 24.0$	$p = 1.38 \cdot 10^{-7}$
Combined	$54.68\,\%$	$22.97\,\%$	Wilcoxon signed-rank	$W = 1.0$	$p = 4.55 \cdot 10^{-13}$

6.3.3 Driver Satisfaction and Intervention Rate Correlation Analysis

In the dataset creation test group study, Hypothesis 3.1 was evaluated, which states that drivers who frequently intervene are less satisfied with the PLDF compared to drivers who intervene less frequently. In Section 3.3.3, a significant negative correlation between the reported driver satisfaction and the intervention frequency, especially within the PLDF's ODD, was found. In this section, this hypothesis is evaluated again based on the higher sample size of 43 drivers from the prototype test group study. The used satisfaction measures are the general satisfaction from the system acceptance scale [VHD97] and the speed profile satisfaction from the custom questionnaire. Both of these satisfaction measures are separately evaluated against the combined IR. Since each participant evaluated both system A and system B, two pairs of satisfaction and IR data are available for each driver. However, these values may not be used in a simple correlation analysis, since they contain repeated measurements within individual drivers. Consequently, observations from the same participant are not independent, as each participant may have an individual baseline satisfaction with both driving functions. To account for this dependency, a linear mixed-effects regression with a random intercept for participants is used. This statistical method allows for the compensation of different baseline satisfaction levels across participants. The results of both conducted linear mixed-effects regressions are shown in Table 6.6.

Table 6.6: Linear mixed-effects regression results for the correlation between driver satisfaction and intervention frequency based on the prototype test group study dataset.

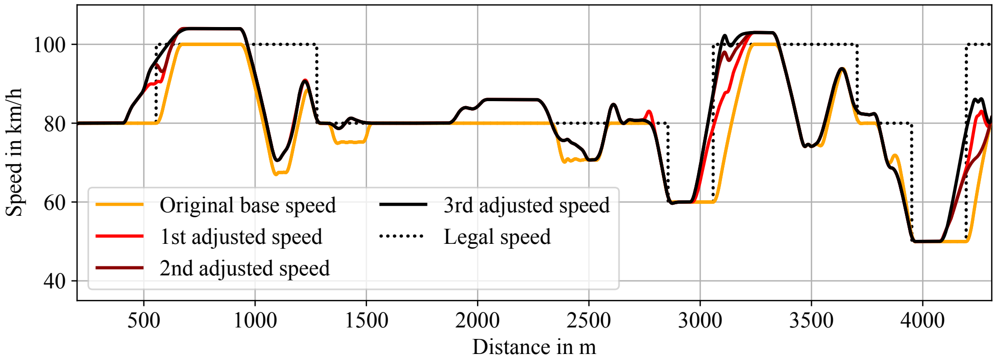
Satisfaction measure	Coefficient	Standard error	p-value	95 % confidence interval
General satisfaction	$\beta = -0.647$	0.388	$p = 0.096$	$[-1.407, 0.114]$
Speed profile satisfaction	$\beta = -0.622$	0.331	$p = 0.060$	$[-1.270, 0.026]$

The estimated regression coefficient β for both satisfaction measures indicates a trend in the dataset suggesting that drivers who intervened less tended to be more satisfied with the driving functions. However, this effect is only marginally significant for both satisfaction measures, as the calculated p-values do not fall below the conventional threshold of 5 %. Therefore, no statistically significant evidence was found for Hypothesis 3.1 based on the underlying dataset. Nonetheless, the direction of the effect is consistent with the hypothesis, even though the statistical evidence is weak. These results suggest that additional factors beyond the intervention frequency influence the driver satisfaction in this study setup, such as the HMI or the general design of the driving functions. The analysis of the participants' qualitative feedback, presented in the Section 6.3.5, aims to investigate these factors more in-depth.

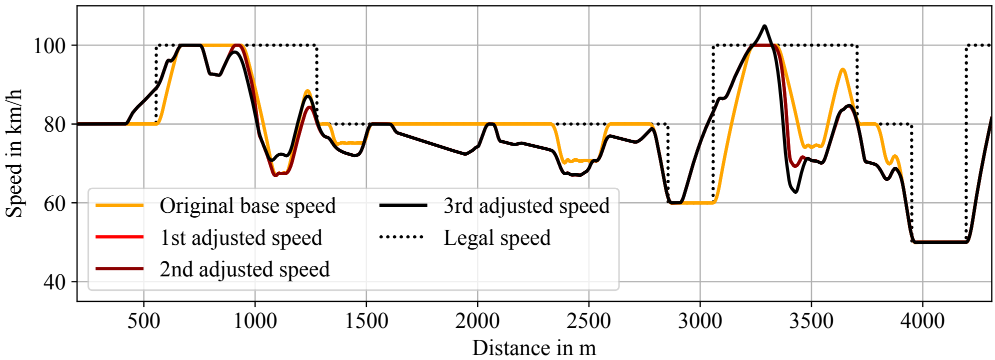
6.3.4 Qualitative Speed Profile Development Analysis

In this section, the development of the self-learning PLDF's baseline speed profile is analyzed for three example participants. The corresponding speed profiles are illustrated in Figure 6.9. In these plots, the original PLDF baseline speed profile is depicted together with the subsequent adjusted speed profiles of the respective participants. The first adjusted speed profile is generated from the driver interventions recorded during the second drive with system A, whereas the following two adjusted baselines are generated from the driver interventions during the first and second system B drives. In Appendix D.5, the three relevant drives of each participant are depicted separately, including the previous baseline, the recorded driver speed, and the resulting adjusted speed profile.

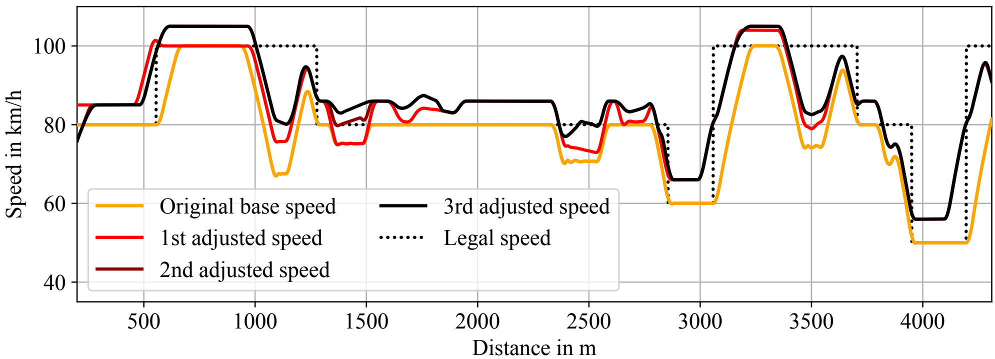
In Figure 6.9a, the speed profile development of participant number 26 is illustrated. The participant primarily intervenes during the first illustrated drive, performing both pedal and set speed interventions. After the SPAA is applied for the first time, the driver only



(a) Baseline speed profile development for participant number 26.



(b) Baseline speed profile development for participant number 13.



(c) Baseline speed profile development for participant number 41.

Figure 6.9: Development of the baseline speed profile over the course of the experiment for three example participants.

performs small adjustments via the gas pedal during the accelerations onto higher speed limits in the subsequent drives. In these situations, the adjustments made by the SPAA appear to be too conservative, and the driver continues to intervene due to still being dissatisfied with the driving behavior in this situation. However, the participant ceases to intervene in all other segments where interventions were previously performed, indicating that they are satisfied with the conducted adjustments by the SPAA. Similar behavior is also observed for the other example participants shown in Figure 6.9b and Figure 6.9c. This is a good example for the SPAA's compensation of overreactions by the drivers. Although only 50 % of the recorded driver intervention speed profile is incorporated into the adjusted speed profile, drivers largely refrain from intervening further in most segments of the track after one SPAA iteration.

The speed profile development during participant number 13's test drives is depicted in Figure 6.9b. This participant frequently intervenes via the brake pedal to reduce the vehicle speed in high-curvature road segments. They also intervene using the gas pedal to accelerate earlier onto upcoming higher speed limits, but never perform any set speed changes. Similarly to participant number 26, they mostly stop intervening after the first SPAA iteration. However, during their second system B drive, they adjust the vehicle speed around two curves again, using both the gas and brake pedal once. This participant's speed profile development serves as an example of an increased intervention frequency during the second system B drive compared to the first, as discussed in Section 6.3.2.

Finally, Figure 6.9c illustrates the speed profile development of participant number 41. This participant frequently intervenes during the second system A drive, using both pedal interventions and set speed adjustments to increase the vehicle speed across most segments of the track. During the second iteration, the driver continues to intervene using both the pedals and set speed changes. However, during the third iteration, the driver only intervenes during one high-curvature segment, indicating that they are largely satisfied with the adjusted speed profile.

The three illustrated examples highlight relatively different individual preferences regarding the PLDF's driving behavior. Despite these differences, the intervention frequency is reduced over the course of the experiment for all three participants, indicating that the applied SPAA effectively adapts the speed profile toward their individual preferences. The significant reduction in intervention frequency after only one application of the SPAA also indicates that it effectively compensates disproportionate and delayed driver intervention behavior, resulting in a more conservative speed profile that still reflects the participant's preferences. However, since most participants continue to intervene, albeit less frequently than at the beginning, the long-term driving behavior should be further investigated to determine whether the intervention frequency converges to 0 % after a sufficient number of iterations.

6.3.5 Qualitative Feedback Analysis

At the end of the study procedure, each participant was asked to provide free-form oral feedback on the experienced driving functions. This feedback was documented by the study operator and subsequently clustered manually according to the addressed subtopics. Finally, the clustered feedback was summarized and analyzed to identify strengths and weaknesses of the proposed self-learning driving function, in order to derive further improvement potential.

Most of the received feedback on system B is generally positive, as reflected in the questionnaire results. Several participants realized the working principle of the self-learning driving function while driving, although many participants also assumed that system A and system B only differ by having distinct static speed profiles. The most frequently stated feedback on the driving experience with system B, reported by 15 participants, emphasizes a notable improvement of the driven speed profile, resulting in a more homogeneous, natural, and human-like driving behavior. In contrast, system A is commonly described as more artificial and machine-like. A frequently mentioned point of criticism, explicitly stated by 11 participants, concerns the static adherence to the legal speed limit in system A, combined with abrupt accelerations and decelerations when approaching new legal speed limits.

The most frequently stated point of criticism of system B, reported by twelve drivers, concerns its lack of communication with the driver. For example, providing the information that system B learns from their driver interventions would have significantly reduced the confusion experienced by some participants. This issue especially includes the communication of the system with the driver through the user interface. For example, when the system correctly detected a legal speed of $70\,\mathrm{km/h}$ and the PLDF's set speed also displays $70\,\mathrm{km/h}$, drivers may become confused if system B then drives above or below this speed without further explanation. Eight participants reported that in such cases they would have preferred, e.g., an explicit indication in the form of an icon or a text display in the instrument cluster, which explains the deviation from the legal speed and the displayed set speed. If this relevant information is not communicated, drivers may feel unsafe and confused when the PLDF behaves in a way they do not understand. It is likely that this lack of communication also contributed to the observation that six participants did not rate their general satisfaction with system B higher than with system A despite reporting a higher satisfaction with system B's speed profile.

Five participants described the behavior of system B as not smooth enough and suggested that additional smoothing mechanisms should be applied. A qualitative analysis of the recorded driving behavior of these drivers revealed a tendency toward abrupt interventions, characterized by high accelerations and decelerations. In addition, these drivers were also found to inconsistently intervene at identical locations across multiple drive-throughs. For example, they would intervene at locations where they had previously not intervened or they would brake at location where they previously accelerated. Accordingly, the self-learning PLDF adopts the abrupt driving behavior from the provided demonstrations. Therefore, mechanisms should be implemented that either limit the maximum changes to the baseline speed profile in one iteration or further arbitration steps should be applied. Such arbitration could, e.g., generate the next speed profile not solely from one but multiple demonstrated interventions, potentially resulting in a smoother and more generalized speed profile.

A final limitation is that the self-learning PLDF used in the study directly learns from each performed driver intervention without applying any classification or arbitration steps. Four participants performed unintended interventions during their drives, e.g., by pressing the ADAS control stalk into the false direction. System B then directly learned from these unintended interventions and adjusted the speed profile for the next iteration accordingly. The affected participants reported difficulties in compensating and removing this unwanted learned behavior again, causing irritation. To avoid such effects, the system should only incorporate consistently performed driver interventions, as outlined in the general self-learning driving function framework introduced in Section 6.1.1.

6.3.6 Limitations and Improvement Potential

In this section, the observed limitations of the self-learning driving function are summarized and potential improvements are derived. In Section 6.3.5, the free-form feedback provided by each participant on the experienced systems is analyzed. The most commonly stated limitation of the evaluated self-learning PLDF is the lack of communication to the driver. To address this limitation, the self-learning PLDF should provide dedicated visual feedback, e.g., in the form of icons or texts in the instrument cluster, to explicitly inform the driver about deviations from the original PLDF baseline speed profile. Additionally, the driver could be informed in the beginning of their commute that the speed profile in some segments was updated, in order to prepare them for a changed function behavior. Moreover, customers should be thoroughly briefed on the system's underlying working principles when activating the feature for the first time in their vehicle.

Another necessary improvement derived from the participant feedback is the application of the complete self-learning PLDF framework as introduced in Section 6.1.1. In particular, this includes the classification of interventions and arbitration across multiple drives, thereby reducing the number of unintended or irrelevant interventions that are learned by the system. Additional arbitration mechanisms could further enhance robustness, e.g., by not generating the first adjusted speed profile directly from a single driver intervention. Instead, multiple driver interventions collected during the primary arbitration step could be aggregated to derive a more generalized and representative adjusted speed profile.

In addition to the limitations of the self-learning PLDF's design, some restrictions also arise from the employed study design. The primary limitation is the restricted study duration, which did not allow for long-term analyses of the driver intervention behavior. As discussed in the preceding sections, the intervention frequency is significantly reduced over the course of the experiment, however, it does not reach values close to 0 %. This indicates that some participants are still somewhat dissatisfied with the resulting speed profile at the end of the experiment. Due to the limited study duration, it cannot be determined whether the intervention frequency would eventually converge to 0 % after a sufficient number of iterations. If so, the ability of the SPAA to generate exactly the ideal desired speed profile for each participant would be confirmed. However, if the intervention frequency converges to levels significantly above 0 %, this could either indicate the need for further refinement of the SPAA or reflect inherently inconsistent long-term driving behavior. In the latter case, no universally satisfactory speed profile can be found, and further analyses would be required to investigate the underlying reasons for the inconsistent driving behavior. Nevertheless, the primary objective of the self-learning PLDF is not to eliminate all interventions, but rather to generally increase driver satisfaction and reduce the intervention frequency. From this perspective, the SPAA's primary objective is already achieved in its current state, as shown in the test group study results.

Another limitation arising from the study design is its fixed experiment order. This fixed order was chosen due to time constraints and because it more realistically reflects the customer experience with the system. However, future studies should employ a non-fixed experiment order to minimize potential learning effects.

Furthermore, the limited yaw rotation of the employed driving simulator did not allow for the evaluation of roundabout and turn scenarios. Thus, future work should evaluate the SPAA's applicability in these scenarios as well.

Although the analyses presented in Chapter 4 identified LI adjustments as the most applicable based on the recorded driver behavior, future work should also evaluate the remaining adjustment strategies in a test group study setting. In particular, the potential benefits of GI adjustments should be examined, as a generalized adjustment of the PLDF's behavior across all similar locations is less complex and more transferable to previously unseen routes than location-specific adjustments. These generalized adjustments to the PLDF's behavior, such as a general speed offset, should be derived based on each driver's individual intervention behavior, and the effect of these GI adjustments on the driver satisfaction should be compared to the LI-focused findings of the present study.

Finally, the present test group study was conducted in a driving simulator rather than under real-world driving conditions. As discussed in Appendix D.2, relative validity can be assumed in this simulator setup, which is sufficient in the context of the present test group study. However, absolute validity can only be partially assumed. To fully validate the study results and to evaluate the optimizations proposed in this section, a test group study under real-world conditions should be conducted during the future development process of the self-learning PLDF.

6.4 Summary and Contributions

In this chapter, an iterative self-learning PLDF framework is introduced based on the results of the preceding chapters. This self-learning PLDF records performed driver interventions and automatically classifies whether the intervention is relevant for a function adjustment. If the intervention is classified as relevant, the long-term driving behavior in the corresponding locations is accumulated and arbitrated. Consistently performed driver interventions are then used as feedback for the personalization of the PLDF.

As part of this framework, the SPAA is introduced, which is used to derive new adjusted speed profiles from the performed driver interventions and the previous baseline speed profile. To evaluate the applicability of the proposed self-learning PLDF and the SPAA in particular, a driving simulator-based test group study is conducted with 43 participants. The study results demonstrate a significant increase in driver satisfaction with the system, as well as a significantly reduced frequency of performed driver interventions compared to the baseline PLDF. Accordingly, the conducted test group study indicates promising potential for a future in-production application of the developed self-learning PLDF. The study results further suggest that utilizing driver-initiated takeovers as feedback for driving function personalization is a valid approach to significantly increase driver satisfaction and reduce the driver intervention frequency.

Based on the study results and the participants' feedback, limitations and potential optimizations of the developed system are identified. The primary limitation of the employed self-learning driving function is the lack of communication with the driver. Dedicated user interface elements should be added to increase the transparency of the self-learning PLDF's behavior and to reduce potential confusion by the drivers. Further restrictions arise from the study design, primarily its limited duration, which did not allow for an investigation of the long-term driving behavior with the self-learning PLDF. Accordingly, future work should focus on analyzing said long-term driving behavior, particularly under real-world driving conditions, using the full proposed self-learning PLDF framework.

To summarize, the research contributions of this chapter are as follows:

- Based on the findings of the preceding chapters, an iterative self-learning PLDF framework is introduced, which continuously learns from driver interventions.
- The SPAA is proposed: an algorithm for the automatic generation of adjusted baseline speed profiles based on performed driver interventions.

- The applicability of the developed self-learning PLDF is demonstrated in a driving simulator-based test group study, which shows a significant increase in driver satisfaction and a significant decrease in the frequency of performed interventions compared to the baseline PLDF.

7 Conclusion

This thesis investigates naturalistic human driving behavior with an in-production PLDF and introduces a framework for the derivation of necessary adjustments to map-based driving functions based on recorded interventions to enhance driver satisfaction. This framework is then applied, and a prototypical personalized PLDF is developed and evaluated, which iteratively learns from driver interventions in a traded control setting.

In related research, driver-initiated takeovers are rarely examined in detail, and no in-depth analysis on the granular reasons behind voluntary driver interventions could be found. Nevertheless, it is generally accepted that driver-initiated takeovers are an integral part of naturalistic driving with Level 1 and Level 2 driving functions [Ger+21; Mor+20; PM08; PF10; Yan+23]. Moreover, a correlation between driver dissatisfaction and intervention frequency is commonly assumed [GJ23; Lee+21; MZ21; Wan+13]. If this correlation is confirmed, driver interventions can serve as valuable feedback for the personalization of driving functions, in a similar way to recent IIL approaches, which utilize human expert interventions to optimize an existing policy. However, related research on personalized longitudinal driving functions rarely leverages driver interventions as feedback and typically focuses on generalized policy adjustments of ACC systems [GJ23; HHW20; Yi+20]. This thesis therefore addresses the resulting research gap by analyzing the granular reasons for driver interventions that occur during naturalistic driving with an assisted longitudinal driving function. Based on these interventions, necessary adjustments to the driving function's behavior are derived to improve driver satisfaction with the system.

First, a test group study is conducted to analyze the intervention behavior of 17 participants using a PLDF during naturalistic real-world driving. Over the course of the study, each performed driver intervention is manually labeled by the drivers via voice recordings. Based on the recorded driver interventions, a hierarchical labeling strategy is introduced, enabling a detailed analysis of the underlying reasons for driver interventions. The results show that more than $50\,\%$ of interventions occur within the PLDF's ODD, indicating a high optimization potential since these interventions solely occur due to deviations in the participants' preferred driving behavior and the PLDF's intended behavior. Moreover, a significant correlation is found between the driver dissatisfaction and the frequency of performed interventions, suggesting that the adjustment of the PLDF's behavior based on driver interventions can potentially increase driver satisfaction.

Based on the study results, a method for the derivation of necessary adjustments to the driving function is introduced. The most common relevant intervention types concern the vehicle speed in specific map locations, such as straight roads, curves, and roundabouts. Accordingly, the proposed method analyzes the consistency of individual driver behavior in these locations to derive applicable adjustment strategies for the driving function. The analysis considers two dimensions of the driving behavior: the individualization component and the location dependency. Applying the method to the test group dataset reveals that no generalized policy adjustment is suitable for all drivers in the dataset. Moreover, personalized general policy adjustments are applicable only for a few drivers in specific location types. This finding contrasts with related research on driving function personalization, which typically assumes that individual driving behavior is sufficiently consistent to be represented by generalized driving policies. However, the method finds that while the global consistency of individual driver behavior is relatively low, the drivers' consistency in specific locations is relatively high. Thus, a personalized and localized adjustment strategy is proposed.

The preceding analyses were only made possible by the manual labeling of driver interventions during and after the study. As a result, they cannot be directly applied to already existing large-scale datasets, such as fleet data, without extensive manual labeling efforts. To address this limitation, an automatic classification pipeline is developed based on the labeled driver intervention dataset using MTSC algorithms. After appropriate preprocessing, three state-of-the-art MTSC models are trained on the dataset, each evaluated with three labeling strategies of differing granularity. The best performance is achieved by the MTSC model ROCKET, which reaches an F1-score of 92.1 % on the binary classification task, distinguishing between relevant and irrelevant intervention types. Subsequently, the trained models are analyzed in detail to identify potential ways to improve the model performance further. The analysis reveals that, in particular, missing information on surrounding traffic participants is a primary cause for inter-binary class confusion. Nevertheless, it is demonstrated that the automatic classification of driver interventions is possible solely using bus signal data of in-production vehicles.

Based on the preceding findings, a prototypical driving function adjustment framework is developed and applied to the PLDF that iteratively learns from driver interventions in a traded control setting. The resulting self-learning driving function incorporates the developed automatic classification pipeline to distinguish relevant and irrelevant driver interventions. For relevant interventions, the proposed method for the derivation of necessary adjustments is applied to locally accumulate the interventions and arbitrate whether the observed behavior is consistent enough to adjust the PLDF's behavior accordingly. For the derivation of new adjusted speed profiles based on recorded driver interventions, the SPAA is proposed. This SPAA is then used iteratively to shift the PLDF's baseline speed profile toward the driver's preferences until they are satisfied with the system's behavior and stop intervening accordingly.

The prototype's applicability is evaluated in a driving simulator-based test group study with 43 participants. The results of the study show that the self-learning driving function significantly increases the driver satisfaction, while significantly reducing the intervention frequency compared to the original PLDF. Thus, the applicability of the self-learning PLDF is confirmed. Finally, the recorded data and the participants' feedback are analyzed to identify further improvement potential for the self-learning PLDF.

To summarize, this thesis demonstrates that driver interventions during assisted driving are a valid source for feedback, and that these interventions can be leveraged to iteratively optimize and personalize the respective driving function. In the context of the examined PLDF, personalized adjustments to specific map locations are proposed to account for each driver's individual preferences. The applicability of the prototypical self-learning PLDF is validated in a test group study. Based on these results, further research and development toward deployment of the self-learning PLDF framework in series vehicles is advocated. The final design and implementation details of the self-learning PLDF as a customer function fall within the scope of series development and must comply with legal regulations and standards.

A Supplements of the Dataset Creation Test Group Study

This appendix contains supplementary materials for Chapter 3. The translated questionnaire used in the study is shown in Section A.1. In Section A.2, all possible labels of the five fields in the annotation tool are listed that were used during the initial dataset annotation. Section A.3 contains multiple tables that illustrate all AAL2 labels in the created dataset. Finally, Section A.4 depicts example plots for selected AAL2 intervention types that are based on deviating personal preferences.

A.1 Dataset Creation Test Group Study Questionnaire

The translated questionnaires used in the dataset creation test group study are depicted in Figure A.1 to Figure A.3. The first page of the questionnaire focuses on demographic information and the driver's experience with ADASs. The second page requests feedback on the driver satisfaction with the PLDF and their frequency of interventions. And the third page contains text fields for more detailed free-form feedback.

A.2 Annotation Labels

During the annotation of the dataset explained in Section 3.2.1, each intervention was described using a combination of five label fields: *system*, *driver input*, *situation*, *reason*, and *desired behavior*. All options used in the annotation tool for ach of these five label fields are depicted in the following tables. Table A.1 shows all possible labels of the *system*, *driver input*, and *desired behavior* field, while Table A.2 depicts the possible labels for the *situation* and *reason* fields.

As can be seen, the situations *give way / stop sign*, *left turn*, *right turn*, and *roundabout* are split into four different subcategories: *approach*, *entering*, *driving through*, and *leaving*. These are the four situations where the main traffic interactions with other vehicles occur. In these situations, the driver is potentially required to intervene and slow down the vehicle in order to yield to potential traffic. The four subcategories describe the phases of driving through these situations during which a driver may potentially intervene. During the *approach* phase, the vehicle starts decelerating and drivers might want to intervene here either to maintain a higher speed for longer or to decelerate

PORSCHE

Before the Study: Demographic Information

The following questionnaire is used to collect some demographic data. We are particularly interested in how much driving experience you have with automated driving functions. This voluntary information is important in order to be able to correctly interpret your test results and assessments in the study.
Thank you!

To be filled out by the study director		
Date	Experiment	Participant nr.

To be filled out by the participant/driver						
Age (in years)	< 20	21 ≤ 30	31 ≤ 40	41 ≤ 50	51 ≤ 60	> 60
	☐	☐	☐	☐	☐	☐
Driving experience (in years)	≤ 5	6 ≤ 10	11 ≤ 20	21 ≤ 40	> 40	
	☐	☐	☐	☐	☐	
How commonly do you use Porsche InnoDrive?	Daily	Weekly	Monthly	Less often	Never	
	☐	☐	☐	☐	☐	
How commonly do you use driver assistance systems in general?	Daily	Weekly	Monthly	Less often	Never	
	☐	☐	☐	☐	☐	
Which driver assistance systems do you use?						

Figure A.1: First page of the dataset creation test group study's questionnaire focusing on demographic information and experience with ADASs.

PORSCHE

After the Study: Assessment of the Driving Function

The following questionnaire serves as a follow-up to the study. We are particularly interested in how satisfied you were with the driving function you experienced. This voluntary information is important in order to be able to correctly interpret your test results and assessments in the study.
Thank you!

To be filled out by the study director		
Date	Experiment	Participant nr.

To be filled out by the participant/driver					
I was satisfied with the driving function in general.	Strongly agree ☐	Agree ☐	Neutral ☐	Disagree ☐	Strongly disagree ☐
I was satisfied with the number of times I had to intervene.	Strongly agree ☐	Agree ☐	Neutral ☐	Disagree ☐	Strongly disagree ☐
A reduction of the number of needed interventions would increase my satisfaction with the system.	Strongly agree ☐	Agree ☐	Neutral ☐	Disagree ☐	Strongly disagree ☐
I was satisfied with the driving function's behavior outside of the scenarios I intervened in.	Strongly agree ☐	Agree ☐	Neutral ☐	Disagree ☐	Strongly disagree ☐
I would use the driving function privately.	Strongly agree ☐	Agree ☐	Neutral ☐	Disagree ☐	Strongly disagree ☐

Figure A.2: Second page of the dataset creation test group study's questionnaire focusing on driver satisfaction with the PLDF and their frequency of interventions.

PORSCHE

To be filled out by the participant/driver	
In which situations did you most commonly intervene?	
In which situations were you dissatisfied with the driving function's behavior without intervening?	
General feedback (free form):	

Figure A.3: Third page of the dataset creation test group study's questionnaire featuring text fields for further written feedback.

Table A.1: Used labels during dataset annotation of the fields *system*, *driver input*, and *desired behavior*.

System	Driver Input	Desired behavior
ACC	Accelerate	ACC higher distance
PLDF	Brake	ACC lower distance
	Cancel	ACC stronger acceleration
	Decrease set speed	Accelerate earlier
	Increase set speed	Accelerate stronger
		Decelerate later
		Higher speed wanted
		Lower speed wanted
		n/a

earlier or stronger. In the *entering* phase, the vehicle is shortly before entering the traffic situation and the driver is required to assess whether they need to yield to other traffic participants. During annotation, the *entering* situation was only used if the reason for the intervention was *traffic*. In the *driving through* phase, the vehicle is inside the traffic location and the driver would usually intervene to either traverse at a higher or lower speed. And finally, in the *leaving* phase, the vehicle has left the traffic situation and accelerates to the set speed again. Here, the only reason why drivers intervened was to accelerate and reach the set speed faster.

Most labels in the *reason* field describe situations that either fall outside of the PLDF's ODD or are caused by incorrect map and perception information. The only exception are *personal preference* and *driver wants to drive manually*. In both cases, the driver intervenes due to a deviating personal preference on how to drive in certain scenarios. The difference is that the *driver wants to drive manually* label was used to mark prolonged segments where the driver decided to drive fully manually, although the study instructions stated not to do so. These manual driving segments are therefore not usable for this thesis, since the goal is to personalize the driving function based on a traded control scenario with driver feedback. These prolonged segments of manual driving would only be usable if the goal was to imitate the manual driving style, which it is not, as explained in the research objectives in Section 2.6.

A.3 Abstraction Level Two Label Taxonomy

Section 3.3.2 explains the four AALs used to summarize and group the original granular annotations in the dataset. However, only the labels and distributions of AAL3 and

Table A.2: Used labels during dataset annotation of the fields *situation* and *reason*.

Situation	**Reason**
Acceleration to current speed limit	ACC target mistakenly lost
Curve	ACC target not detected
Give way / stop sign - Approach	ACC target retained too long
Give way / stop sign - Driving through	ACC wrong target
Give way / stop sign - Entering	Cut-in by other driver
Give way / stop sign - Leaving	Driver wants to drive manually
Left turn - Approach	Lane change
Left turn - Driving through	Make it through green traffic light
Left turn - Entering	Merging
Left turn - Leaving	Unintentional intervention
Narrow street	Other
Other	Overtaking maneuver
Pedestrian crossing	Override navigation system
Railway crossing	Personal preference
Right turn - Approach	Red traffic light
Right turn - Driving through	Start PLDF after long stop
Right turn - Entering	Stop from taking over on right lane
Right turn - Leaving	Stopover
Roundabout - Approach	Traffic
Roundabout - Driving through	U-turn
Roundabout - Entering	Unknown
Roundabout - Leaving	Wrong give way / roundabout in map
Speed limit change to higher speed	Wrong speed limit in map / TSD
Speed limit change to lower speed	
Stop and go	
Traffic light	
n/a	

AAL4 are shown. Therefore, Table A.3 to Table A.6 introduce all 55 abstraction labels of AAL2. Out of these interventions, only the ones which are described as *set speed increase*, *set speed decrease*, or *set speed adjustment* are driver interventions conducted via an adjustment of the set speed. All other interventions are exclusively pedal interventions.

The following paragraphs contain comments on a few labels.

In Table A.5, multiple intervention labels are listed that aim to correct incorrect speed limit information. During annotation, it could not be determined in most cases whether the incorrect speed limit information originated from the TSD or from the map. However, in the case of *acceleration due to incorrect lower speed limit in map*, this could be distinguished. When the PLDF receives the information about an upcoming lower speed limit in the map, it begins to decelerate accordingly. If the driver then accelerates to cancel this deceleration and the TSD does not detect a lower legal speed limit sign afterward, then the incorrect speed limit information definitely originated in the map. Therefore, this label is separate from the *acceleration due to incorrect speed limit* label.

During the annotation process, the labels for many interventions could be clearly assigned. However, in certain cases, the correct label was ambiguous. The highest ambiguity was found during the annotation of accelerations in low velocity situations with surrounding traffic. Here, the labels *stronger acceleration after stop*, *ACC stronger acceleration*, *ACC lower distance*, and *reactivate system after long stop* were especially affected by ambiguities. First, it was often unclear whether the ACC or the PLDF acceleration was currently too slow. Second, both *ACC stronger acceleration* and *ACC lower distance* have the same effect of decreasing the distance to the leading vehicle. And lastly, *reactivate system after long stop* interventions often merged with other interventions, since the drivers commonly activated the PLDF via the gas and then directly performed the acceleration to the set speed. In these ambiguous situations, the voice annotation content was commonly used to decide which annotation to choose. However, a clear distinction of these interventions is sometimes not possible.

A.4 Example Plots of Interventions due to Deviating Personal Preferences

This section provides selected example plots of common and relevant AAL2 interventions based on deviating personal preferences. To limit the complexity of the plots, only relevant signals are depicted in the signal-over-time plots, depending on the intervention type. Figure A.4 contains examples of *speed adjustments on straight roads*. In Figure A.5, examples of *speed adjustments in high-curvature segments* are depicted. Examples for the *adjustment of acceleration timings onto speed limits* are shown in Figure A.6. *Adjustment of acceleration strength* interventions are provided in Figure A.7. Finally, two examples for the *adjustment of ACC distance* are depicted in Figure A.8.

Table A.3: The AAL2 and AAL3 labels in the AAL4 category *personal preference* and their absolute (Abs.) numbers, relative (Rel.), and normalized (Norm.) distributions. AAL3 labels are marked in bold.

Label name	Abs.	Rel.	Norm.
Speed adjustment on straight roads	**968**	**28.92 %**	**25.18 %**
Set speed increase on straight road	703	21.00 %	17.74 %
Straight road acceleration	147	4.39 %	3.79 %
Set speed decrease on straight road	88	2.63 %	2.62 %
Straight road deceleration	30	0.90 %	1.03 %
Speed adjustment in high-curvature segments	**458**	**13.68 %**	**14.38 %**
Roundabout higher speed	190	5.68 %	5.81 %
Curve higher speed	67	2.00 %	1.61 %
Right turn higher speed	53	1.58 %	1.66 %
Right turn lower speed	38	1.14 %	1.35 %
Left turn higher speed	36	1.08 %	0.96 %
Left turn lower speed	34	1.02 %	1.29 %
Curve lower speed	20	0.60 %	0.97 %
Roundabout lower speed	12	0.36 %	0.40 %
Give way sign lower speed	8	0.24 %	0.33 %
Adjustment of acceleration timings onto speed limits	**366**	**10.94 %**	**10.06 %**
Earlier acceleration onto higher speed limit	183	5.47 %	6.29 %
Later deceleration onto lower speed limit	160	4.78 %	3.46 %
Set speed increase before higher speed limit	23	0.69 %	0.32 %
Adjustment of acceleration strength	**79**	**2.36 %**	**2.27 %**
Stronger acceleration after stop	46	1.37 %	1.31 %
Stronger acceleration to current set speed	33	0.99 %	0.96 %
Adjustment of ACC distance	**46**	**1.37 %**	**1.51 %**
ACC stronger acceleration	31	0.93 %	0.94 %
ACC lower distance	14	0.42 %	0.55 %
ACC higher distance	1	0.03 %	0.01 %

Table A.4: The AAL2 and AAL3 labels in the AAL4 category *outside of ODD* and their absolute (Abs.) numbers, relative (Rel.), and normalized (Norm.) distributions. AAL3 labels are marked in bold.

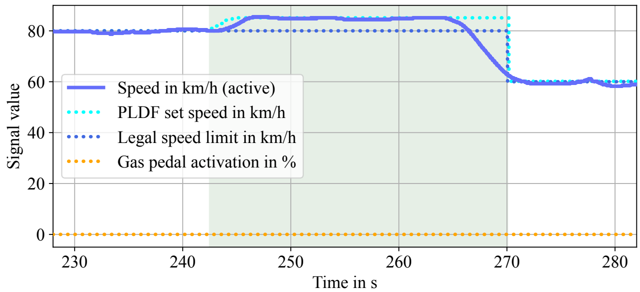
Label name	**Abs.**	**Rel.**	**Norm.**
Right of way interactions	**340**	**10.16 %**	**14.34 %**
Brake due to traffic in left turn	100	2.99 %	4.22 %
Brake due to traffic in narrow street	96	2.87 %	3.72 %
Brake due to traffic in roundabout	67	2.00 %	3.01 %
Brake due to traffic in right turn	50	1.49 %	2.29 %
Brake due to pedestrian crossing	9	0.27 %	0.30 %
Gas due to traffic in narrow street	8	0.24 %	0.27 %
Brake due to traffic at a give way sign	8	0.24 %	0.42 %
Gas due to traffic in roundabout	2	0.06 %	0.11 %
Traffic light interactions	**208**	**6.21 %**	**7.55 %**
Brake at red traffic light	195	5.83 %	7.22 %
Gas to make it through green traffic light	13	0.39 %	0.33 %
Multi lane interactions	**122**	**3.65 %**	**2.49 %**
Overtaking maneuver	48	1.43 %	1.02 %
Lane change	38	1.14 %	0.78 %
Merging	22	0.66 %	0.43 %
Cut-in by other driver	12	0.36 %	0.23 %
Stop from taking over on right lane	2	0.06 %	0.02 %
Reactivate system after long stop	**112**	**3.35 %**	**2.59 %**
Reactivate system after long stop	112	3.35 %	2.59 %
Other traffic interactions	**16**	**0.48 %**	**0.49 %**
Other traffic interactions	16	0.48 %	0.49 %

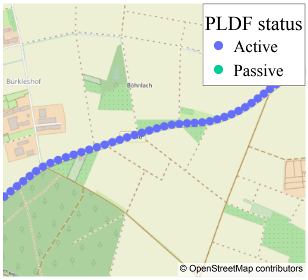
Table A.5: The AAL2 and AAL3 labels in the AAL4 category *incorrect input data* and their absolute (Abs.) numbers, relative (Rel.), and normalized (Norm.) distributions. AAL3 labels are marked in bold.

Label name	**Abs.**	**Rel.**	**Norm.**
Incorrect ACC sensing state	**212**	**6.33 %**	**7.16 %**
ACC target not detected	86	2.57 %	3.33 %
ACC target retained too long	58	1.73 %	1.63 %
ACC wrong target	55	1.64 %	1.67 %
ACC target mistakenly lost	13	0.39 %	0.54 %
Incorrect TSD or map information	**189**	**5.65 %**	**5.53 %**
Acceleration due to incorrect speed limit	71	2.12 %	2.20 %
Set speed increase due to incorrect speed limit	48	1.43 %	1.30 %
Acceleration due to incorrect lower speed limit in map	31	0.93 %	1.03 %
Brake due to incorrect speed limit	20	0.60 %	0.59 %
Acceleration due to incorrect give way sign in map	16	0.48 %	0.32 %
Set speed decrease due to incorrect speed limit	3	0.09 %	0.08 %

Table A.6: The AAL2 and AAL3 labels in the AAL4 category *other* and their absolute (Abs.) numbers, relative (Rel.), and normalized (Norm.) distributions. AAL3 labels are marked in bold.

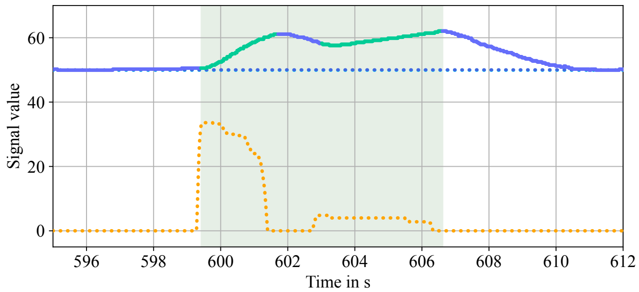
Label name	Abs.	Rel.	Norm.
Missing annotation	**105**	**3.14 %**	**2.54 %**
Missing annotation	105	3.14 %	2.54 %
Unintentional intervention	**78**	**2.33 %**	**2.41 %**
Unintentional set speed adjustment	41	1.23 %	1.19 %
Unintentional pedal intervention	37	1.11 %	1.23 %
Other	**48**	**1.43 %**	**1.50 %**
Other	16	0.48 %	0.54 %
Prolonged manual driving	12	0.36 %	0.37 %
Stopover	11	0.33 %	0.34 %
Acceleration to override navigation system	6	0.18 %	0.10 %
U-turn	3	0.09 %	0.15 %





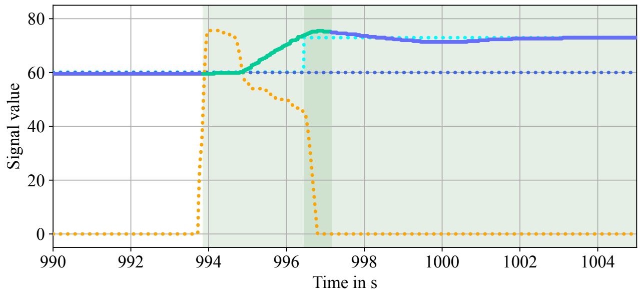
(a) A *set speed increase on straight road* intervention. The driver increases the set speed on a straight road above the legal speed from 80 km/h to 85 km/h.

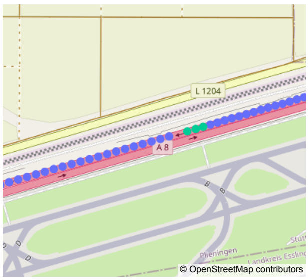
(b) The corresponding GPS traces of the *set speed increase on straight road* intervention.



(c) A *straight road acceleration* intervention. The driver presses the gas pedal twice to drive faster on a straight road segment.

(d) The corresponding GPS traces of the *straight road acceleration* intervention.

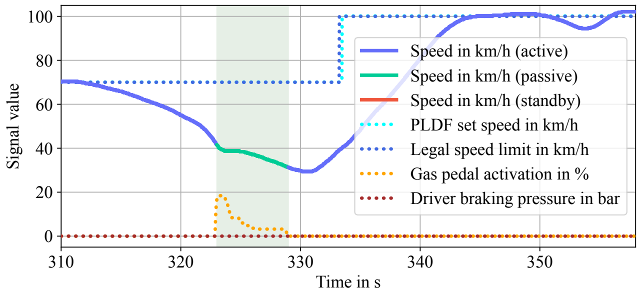


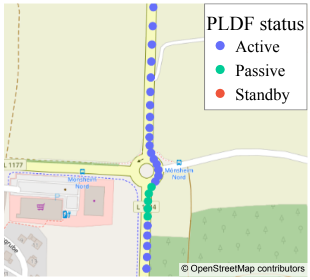


(e) A *straight road acceleration* intervention directly followed by a *set speed increase on straight road* intervention. The driver presses the gas pedal to drive faster on a straight road and increases the set speed shortly after to stay at the higher speed.

(f) The corresponding GPS traces of the *straight road acceleration* and *set speed increase on straight road* interventions.

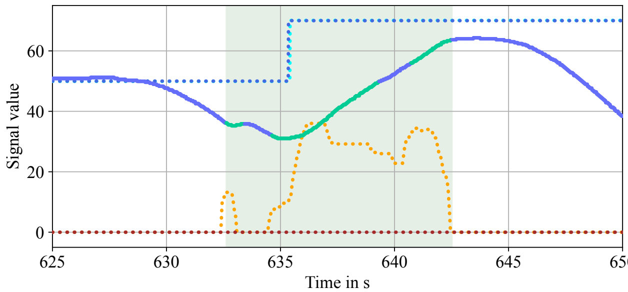
Figure A.4: Signal-over-time plots and the corresponding GPS traces of example *speed adjustment on straight roads* interventions. Interventions are highlighted with a green background. The corresponding AAL2 labels are named in the subfigure captions.

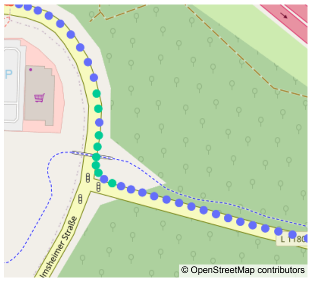




(a) A *roundabout higher speed* intervention. The driver presses the gas while entering the roundabout to traverse it at a higher speed.

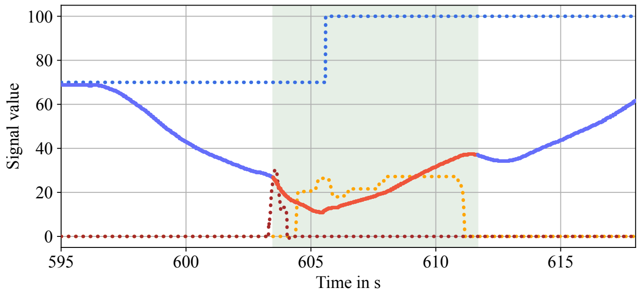
(b) The corresponding GPS traces of the *roundabout higher speed* intervention.





(c) A *right turn higher speed* intervention. The driver increases the vehicle speed before the right turn and while traversing it by pressing the gas pedal.

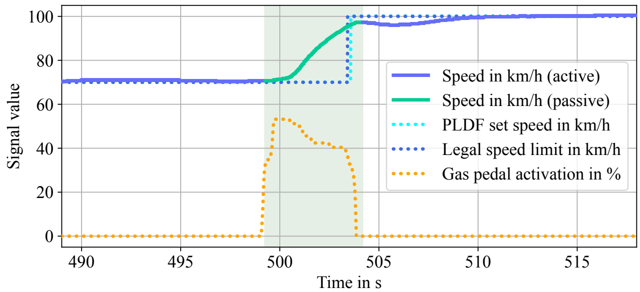
(d) The corresponding GPS traces of the *right turn higher speed* intervention.



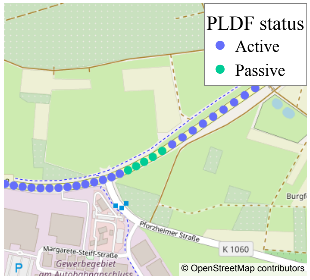
(e) A *right turn lower speed* intervention. The PLDF decreases the vehicle speed before the turn. Shortly before reaching it, the driver presses the brake to further reduce the vehicle speed. After manually traversing the turn, the driver hands back control to the PLDF.

(f) The corresponding GPS traces of the *right turn lower speed* intervention.

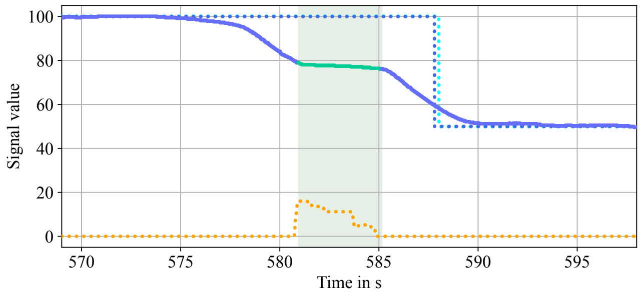
Figure A.5: Signal-over-time plots and the corresponding GPS traces of example *speed adjustment in high-curvature segments* interventions. Interventions are highlighted with a green background. The corresponding AAL2 labels are named in the subfigure captions.



(a) An *earlier acceleration onto higher speed limit* intervention. The driver presses the gas pedal a few seconds before passing a higher speed limit sign in order to accelerate earlier.

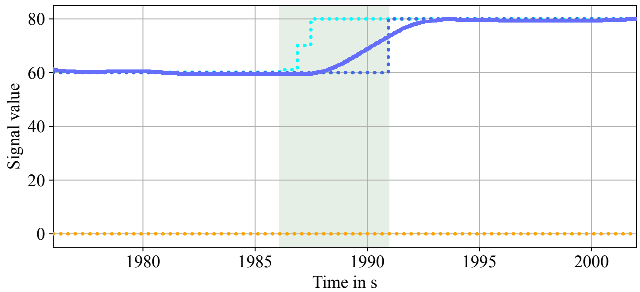


(b) The corresponding GPS traces of the *earlier acceleration onto higher speed limit* intervention.

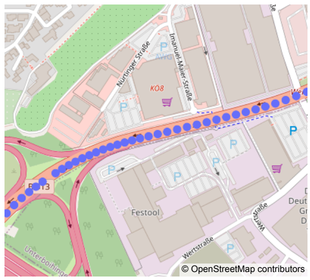


(c) A *later deceleration onto lower speed limit* intervention. The PLDF predictively decelerates onto an upcoming lower speed limit but the driver presses the gas pedal to delay the deceleration by a few seconds.

(d) The corresponding GPS traces of the *later deceleration onto lower speed limit* intervention.

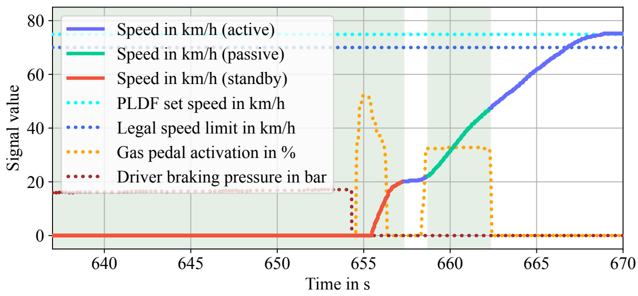


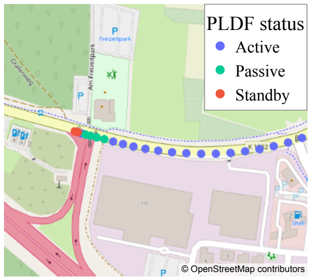
(e) A *set speed increase before higher speed limit* intervention. A few seconds before reaching a higher speed limit sign, the driver increases the set speed in order start the acceleration earlier.



(f) The corresponding GPS traces of the *set speed increase before higher speed limit* intervention.

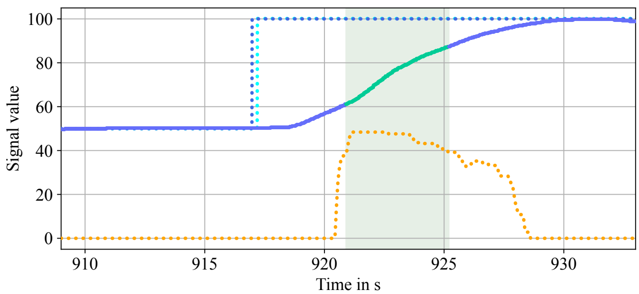
Figure A.6: Signal-over-time plots and the corresponding GPS traces of example *adjustment of acceleration timings onto speed limits* interventions. Interventions are highlighted with a green background. The corresponding AAL2 labels are named in the subfigure captions.





(a) A *stronger acceleration after stop* intervention after a *brake at red traffic light*. The driver deactivated the PLDF via the brake pedal at a red traffic light. After it turns green, the driver accelerates and hands over the vehicle control to the PLDF. Shortly after, the driver intervenes again by pressing the gas to accelerate more strongly to the current set speed.

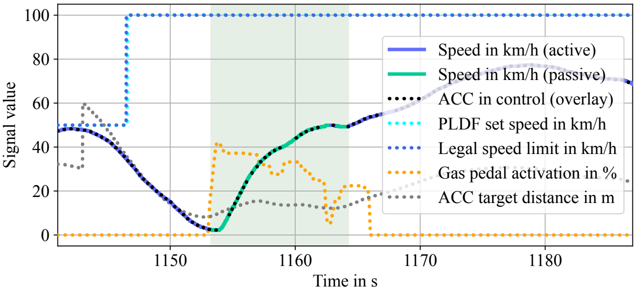
(b) The corresponding GPS traces of the *stronger acceleration after stop* intervention.

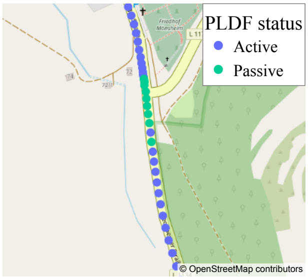


(c) A *stronger acceleration to current set speed* intervention. After passing a higher speed limit sign, the PLDF accelerates accordingly. However, the driver is dissatisfied with the acceleration strength and intervenes by additionally pressing the gas pedal.

(d) The corresponding GPS traces of the *stronger acceleration to current set speed* intervention.

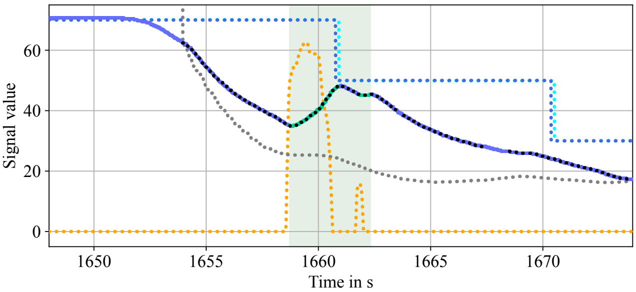
Figure A.7: Signal-over-time plots and the corresponding GPS traces of example *adjustment of acceleration strength* interventions. Interventions are highlighted with a green background. The corresponding AAL2 labels are named in the subfigure captions.

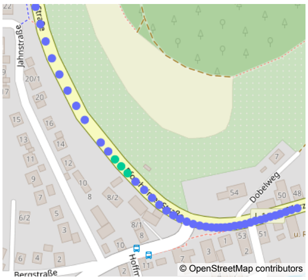




(a) An *ACC stronger acceleration* intervention. The leading vehicle slows down and almost comes to a stop at a traffic light. Shortly after, the leading vehicle accelerates again but the PLDF only accelerates slowly. The driver then intervenes by pressing the gas to accelerate more strongly.

(b) The corresponding GPS traces of the *ACC stronger acceleration* intervention.





(c) An *ACC lower distance* intervention. The leading vehicle slows down and the ACC decelerates the ego vehicle accordingly. However, the driver prefers a lower distance to the leading vehicle and presses the gas pedal twice.

(d) The corresponding GPS traces of the *ACC lower distance* intervention.

Figure A.8: Signal-over-time plots and the corresponding GPS traces of example *adjustment of ACC distance* interventions. Interventions are highlighted with a green background. The black dotted overlay of the vehicle speed signal in the signal-over-time plots signalizes active ACC control. The corresponding AAL2 labels are named in the subfigure captions.

B Supplements of the Derivation of Necessary Adjustments to the Driving Function

This appendix contains supplementary materials for Chapter 4.

B.1 Complete Results of the Global Individual and Local Individual Intervention Rate Analyses

In Section 4.2, the results of the framework for the derivation of necessary adjustments of the PLDF are presented. For the GI and LI analyses, the resulting data was summarized in the corresponding sections to improve readability. Therefore, the tables featuring the complete results of these analyses are presented in this section.

The complete IR results of the GI analysis are shown in Table B.1. There, IR^{+}_{GI} and IR^{-}_{GI} for each driver in each location type are depicted, together with the number of intervention opportunities each driver had for each location type.

In Table 4.7, the distribution of the applicability of unique locations for a LI adjustment are aggregated across all drivers. Therefore, the complete results of the LI applicability analysis for each individual driver in the different location types are depicted in Table B.2 to Table B.5. Due to space limitations, the results were divided into four separate tables. Table B.2 depicts the distribution of locations where a LI adjustment is applicable in each location type for all drivers. Table B.3 depicts the distribution of locations where drivers behaved inconsistently. Table B.4 illustrates the distribution of locations where no adjustment is neccessary. Finally, Table B.5 depicts the total numbers of locations where a LI adjustment was evaluated.

Table B.1: IR^{+}_{GI}, IR^{-}_{GI}, and the total number of intervention opportunities for each driver in each location type.

Driver index	IR^{+}_{GI} / IR^{-}_{GI} / number of intervention opportunities					
	SR	**SLI**	**SLD**	**R**	**C**	**T**
0	10.5 %/6.3 %/143	51.6 %/0.0 %/62	4.8 %/0.0 %/63	83.3 %/0.0 %/54	40.4 %/0.0 %/52	9.4 %/3.1 %/32
1	20.8 %/0.0 %/101	35.1 %/0.0 %/37	12.0 %/0.0 %/25	36.6 %/0.0 %/41	5.3 %/15.8 %/19	25.0 %/6.2 %/16
2	7.5 %/0.0 %/106	9.1 %/0.0 %/55	0.0 %/0.0 %/35	91.4 %/0.0 %/35	25.0 %/0.0 %/52	9.6 %/13.5 %/52
3	74.3 %/0.9 %/113	40.9 %/0.0 %/22	89.6 %/0.0 %/48	93.3 %/0.0 %/15	0.0 %/50.0 %/4	14.3 %/28.6 %/7
4	54.4 %/0.0 %/136	18.5 %/0.0 %/54	11.6 %/0.0 %/43	50.0 %/0.0 %/10	0.0 %/0.0 %/36	18.9 %/13.5 %/37
5	31.4 %/1.4 %/70	25.6 %/0.0 %/43	0.0 %/0.0 %/33	-/-/0	0.0 %/0.0 %/11	66.7 %/0.0 %/9
6	86.7 %/2.7 %/113	86.5 %/0.0 %/52	67.3 %/0.0 %/49	96.3 %/0.0 %/27	14.9 %/0.0 %/47	57.6 %/27.3 %/33
7	1.3 %/0.0 %/76	23.5 %/0.0 %/34	0.0 %/0.0 %/19	9.1 %/0.0 %/11	3.6 %/14.3 %/28	5.9 %/0.0 %/17
8	88.5 %/0.0 %/87	15.0 %/0.0 %/20	50.0 %/0.0 %/16	0.0 %/0.0 %/1	0.0 %/0.0 %/14	20.0 %/20.0 %/10
9	10.7 %/0.6 %/159	3.6 %/0.0 %/84	1.5 %/0.0 %/67	18.8 %/0.0 %/16	0.0 %/15.3 %/72	0.0 %/42.9 %/21
10	47.3 %/0.2 %/442	37.2 %/0.0 %/148	27.3 %/0.0 %/150	86.2 %/3.4 %/29	56.9 %/2.8 %/109	31.6 %/0.0 %/19
11	47.9 %/0.0 %/140	69.5 %/0.0 %/59	57.1 %/0.0 %/56	100.0 %/0.0 %/31	39.0 %/0.0 %/41	80.6 %/12.9 %/31
12	10.9 %/0.3 %/330	3.4 %/0.0 %/147	1.1 %/0.0 %/175	0.0 %/0.0 %/17	0.0 %/0.0 %/123	5.1 %/6.1 %/98
13	47.1 %/0.0 %/104	31.9 %/0.0 %/47	35.9 %/0.0 %/39	86.7 %/0.0 %/30	0.0 %/0.0 %/11	18.9 %/10.8 %/37
14	19.5 %/3.7 %/82	32.4 %/0.0 %/37	0.0 %/0.0 %/28	18.8 %/6.2 %/16	0.0 %/0.0 %/18	0.0 %/13.2 %/38
15	39.3 %/0.0 %/107	16.1 %/0.0 %/62	5.9 %/0.0 %/34	60.9 %/0.0 %/23	7.1 %/0.0 %/99	2.8 %/11.1 %/36
16	4.5 %/39.0 %/154	1.5 %/0.0 %/65	1.3 %/0.0 %/76	0.0 %/50.0 %/6	0.0 %/1.4 %/71	5.7 %/17.1 %/35

Table B.2: Distribution of locations where a LI adjustment is applicable in each location type for all drivers. If no valid location exists for a pairing of driver and location type, a “-” is depicted.

Driver index	LI adjustment applicable						
	SR	**SLI**	**SLD**	**R**	**C**	**T**	**Total**
0	11.11 %	57.14 %	8.33 %	81.82 %	66.67 %	0.00 %	32.89 %
1	13.33 %	33.33 %	0.00 %	22.22 %	0.00 %	50.00 %	17.07 %
2	0.00 %	7.69 %	0.00 %	88.89 %	22.22 %	23.08 %	19.44 %
3	86.36 %	50.00 %	100.00 %	100.00 %	-	100.00 %	87.50 %
4	76.00 %	16.67 %	0.00 %	-	0.00 %	0.00 %	39.62 %
5	13.33 %	11.11 %	0.00 %	-	0.00 %	100.00 %	11.76 %
6	90.91 %	90.00 %	60.00 %	100.00 %	11.11 %	87.50 %	75.38 %
7	0.00 %	14.29 %	0.00 %	-	40.00 %	0.00 %	8.82 %
8	92.86 %	0.00 %	0.00 %	-	0.00 %	-	72.22 %
9	9.52 %	4.55 %	0.00 %	0.00 %	15.38 %	50.00 %	9.18 %
10	39.73 %	37.50 %	19.35 %	100.00 %	63.16 %	0.00 %	40.26 %
11	40.00 %	63.64 %	53.85 %	100.00 %	25.00 %	66.67 %	52.17 %
12	4.92 %	0.00 %	0.00 %	0.00 %	0.00 %	15.79 %	3.66 %
13	37.50 %	25.00 %	44.44 %	100.00 %	0.00 %	25.00 %	40.98 %
14	21.05 %	20.00 %	0.00 %	0.00 %	0.00 %	0.00 %	12.50 %
15	37.50 %	12.50 %	0.00 %	50.00 %	0.00 %	0.00 %	17.81 %
16	45.71 %	0.00 %	0.00 %	100.00 %	0.00 %	20.00 %	21.18 %
Total	33.40 %	23.04 %	18.78 %	70.42 %	18.12 %	26.88 %	29.20 %

Table B.3: Distribution of locations with inconsistent driver behavior on LI level in each location type for all drivers. If no valid location exists for a pairing of driver and location type, a "-" is depicted.

Driver index	Inconsistent driver behavior						
	SR	**SLI**	**SLD**	**R**	**C**	**T**	**Total**
0	7.41 %	7.14 %	0.00 %	18.18 %	0.00 %	33.33 %	9.21 %
1	13.33 %	16.67 %	33.33 %	44.44 %	66.67 %	50.00 %	29.27 %
2	8.33 %	7.69 %	0.00 %	11.11 %	11.11 %	0.00 %	6.94 %
3	4.55 %	25.00 %	0.00 %	0.00 %	-	0.00 %	5.00 %
4	12.00 %	16.67 %	0.00 %	-	0.00 %	60.00 %	15.09 %
5	33.33 %	22.22 %	0.00 %	-	0.00 %	0.00 %	20.59 %
6	0.00 %	0.00 %	30.00 %	0.00 %	11.11 %	0.00 %	6.15 %
7	0.00 %	0.00 %	0.00 %	-	0.00 %	0.00 %	0.00 %
8	7.14 %	50.00 %	100.00 %	-	0.00 %	-	16.67 %
9	7.14 %	0.00 %	0.00 %	33.33 %	7.69 %	0.00 %	5.10 %
10	21.92 %	29.17 %	19.35 %	0.00 %	21.05 %	100.00 %	22.08 %
11	8.00 %	9.09 %	23.08 %	0.00 %	37.50 %	33.33 %	15.94 %
12	6.56 %	10.00 %	0.00 %	0.00 %	0.00 %	5.26 %	4.88 %
13	41.67 %	16.67 %	11.11 %	0.00 %	0.00 %	0.00 %	21.31 %
14	5.26 %	20.00 %	0.00 %	50.00 %	0.00 %	12.50 %	12.50 %
15	8.33 %	12.50 %	0.00 %	50.00 %	15.79 %	0.00 %	12.33 %
16	8.57 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	3.53 %
Total	11.75 %	11.98 %	8.84 %	16.90 %	10.87 %	11.83 %	11.56 %

Table B.4: Distribution of locations where no LI adjustment is neccessary in each location type for all drivers. If no valid location exists for a pairing of driver and location type, a "-" is depicted.

Driver index	No adjustment necessary						
	SR	**SLI**	**SLD**	**R**	**C**	**T**	**Total**
0	81.48 %	35.71 %	91.67 %	0.00 %	33.33 %	66.67 %	57.89 %
1	73.33 %	50.00 %	66.67 %	33.33 %	33.33 %	0.00 %	53.66 %
2	91.67 %	84.62 %	100.00 %	0.00 %	66.67 %	76.92 %	73.61 %
3	9.09 %	25.00 %	0.00 %	0.00 %	-	0.00 %	7.50 %
4	12.00 %	66.67 %	100.00 %	-	100.00 %	40.00 %	45.28 %
5	53.33 %	66.67 %	100.00 %	-	100.00 %	0.00 %	67.65 %
6	9.09 %	10.00 %	10.00 %	0.00 %	77.78 %	12.50 %	18.46 %
7	100.00 %	85.71 %	100.00 %	-	60.00 %	100.00 %	91.18 %
8	0.00 %	50.00 %	0.00 %	-	100.00 %	-	11.11 %
9	83.33 %	95.45 %	100.00 %	66.67 %	76.92 %	50.00 %	85.71 %
10	38.36 %	33.33 %	61.29 %	0.00 %	15.79 %	0.00 %	37.66 %
11	52.00 %	27.27 %	23.08 %	0.00 %	37.50 %	0.00 %	31.88 %
12	88.52 %	90.00 %	100.00 %	100.00 %	100.00 %	78.95 %	91.46 %
13	20.83 %	58.33 %	44.44 %	0.00 %	100.00 %	75.00 %	37.70 %
14	73.68 %	60.00 %	100.00 %	50.00 %	100.00 %	87.50 %	75.00 %
15	54.17 %	75.00 %	100.00 %	0.00 %	84.21 %	100.00 %	69.86 %
16	45.71 %	100.00 %	100.00 %	0.00 %	100.00 %	80.00 %	75.29 %
Total	54.85 %	64.98 %	72.38 %	12.68 %	71.01 %	61.29 %	59.24 %

Table B.5: Total number of locations where a LI adjustment was evaluated in each location type for all drivers.

Driver index	Total number of locations						
	SR	**SLI**	**SLD**	**R**	**C**	**T**	**Total**
0	27	14	12	11	6	6	76
1	15	6	6	9	3	2	41
2	24	13	4	9	9	13	72
3	22	4	10	3	0	1	40
4	25	12	5	0	6	5	53
5	15	9	7	0	2	1	34
6	22	10	10	6	9	8	65
7	18	7	3	0	5	1	34
8	14	2	1	0	1	0	18
9	42	22	14	3	13	4	98
10	73	24	31	6	19	1	154
11	25	11	13	6	8	6	69
12	61	30	33	2	19	19	164
13	24	12	9	7	1	8	61
14	19	10	3	4	4	8	48
15	24	16	5	4	19	5	73
16	35	15	15	1	14	5	85
Total	485	217	181	71	138	93	1185

C Supplements of the Automatic Classification of Driver Interventions

This appendix contains supplementary materials for Chapter 5. In Section C.1, three MTSC candidate models are highlighted, which were not chosen for the classification task. Section C.2 describes the conversion of the original AAL2 labels to the full classification labels. Finally, the complete list of input signals used during classification is highlighted in Section C.3.

C.1 Unused Multivariate Time Series Classification Algorithms

In Section 5.1, the fundamentals of MTSC are explained, and the three algorithms chosen for the evaluation of the automatic classification of driver interventions are highlighted. Additionally, three other potential candidate algorithms for MTSC are briefly mentioned in the section due to their relatively strong performance on MTSC benchmarks [Dha+20; Rui+21]: DTW, HIVE-COTE, and WEASEL+MUSE. However, they were not chosen for the following evaluations in Chapter 5. Therefore, these algorithms are briefly introduced here instead.

C.1.1 Dynamic Time Warping

DTW [SC78] calculates the similarity of two UTS by computing the distances between their observations. In contrast to the Euclidean distance, which maps the observations based on their fixed indices, DTW allows for a one-to-many mapping of observations with different indices [SYWK15]. The shortest distance between the two time series is called the optimal *warping path*. This warping path needs to fulfill three requirements [SC07]:

1. All observations of both time series are included in the warping path.
2. The warping path's indexing is continuous and monotonically increasing.
3. The warping path starts and ends with the first and last observations of both time series.

An example of such a warping path between two time series is depicted in Figure C.1. During the classification process, a 1-nearest-neighbor algorithm is used, where an unknown time series is assigned the same label as the most similar time series from the training dataset. As the base algorithm is only used for UTS data, multiple adjustment strategies for MTS data were proposed [Dha+20; Rui+21; SYWK15].

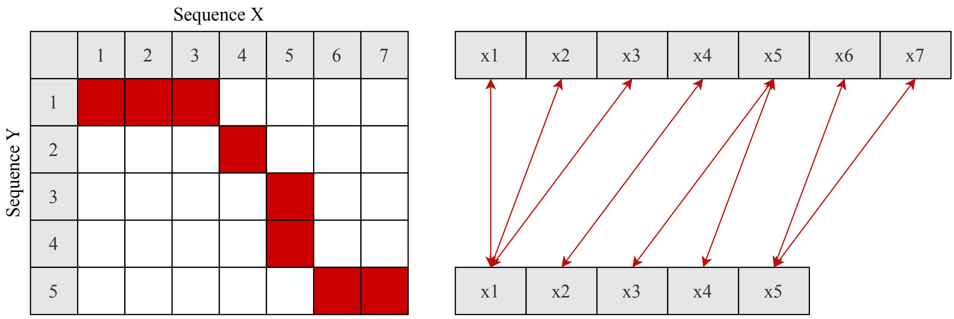


Figure C.1: Example DTW warping path (adapted from Müller [Mul15]).

C.1.2 HIerarchical VotE COllective of Transformation-based Ensemble

HIVE-COTE [LTB16; Bag+20] is an ensemble-based algorithm which combines a multitude of different classifiers from multiple domains. The results from each sub-algorithm in the ensemble are then used in a probabilistic voting scheme to generate the final model predictions. Due to this ensemble-based architecture, HIVE-COTE generally reaches high scores in benchmarks, but it also features a high demand for computational resources and exceptionally long training times. For example, Ruiz *et al.* [Rui+21] report that HIVE-COTE would have required over $10{,}000\,\mathrm{h}$ to sequentially train on all UEA MTSC benchmark tasks. In comparison, ROCKET finished its training after roughly $2\,\mathrm{h}$ while reaching similar accuracies on the benchmarks.

C.1.3 Word ExtrAction for time SEries cLassification plus Multivariate Unsupervised Symbols and dErivatives

WEASEL+MUSE [SL18] is a dictionary-based algorithm which aims to discretize patterns of time series data into words which are then fed into a machine learning classifier. The words are generated by using sliding windows which extract discrete features from the MTS data. The feature vectors are then fed through feature selection, removing non-discriminative features before they are fed into the classifier. However, WEASEL+MUSE also has a high demand for computational resources and features significantly longer training times compared to other algorithms with similar performance, as reported by Dhariyal *et al.* [Dha+20].

C.2 Conversion of Annotation Abstraction Level 2 to Classification Labels

As explained in Section 5.2.2, the used labels for the full classification are based on the AAL2 labels. These labels were modified for the use as classification labels, primarily via the exclusion of specific labels and the merging of similar labels. The following lists depict the excluded AAL2 labels from the classification dataset.

Due to missing annotations:

- *Missing annotation*.

Due to being unintended by the driver:

- *Unintentional set speed adjustment*,
- *Unintentional pedal intervention*.

Due to being a set speed intervention:

- *Set speed increase on straight road*,
- *Set speed decrease on straight road*,
- *Set speed increase before higher speed limit*,
- *Set speed increase due to incorrect speed limit*,
- *Set speed decrease due to incorrect speed limit*.

Additionally, the AAL2 label *Acceleration to override navigation system* is excluded from the dataset, since the reason for the intervention is a conflict between the driver and the navigation system which is unrelated to the PLDF. In this intervention type, the navigation system tells the driver to take a turn at an intersection. Accordingly, the PLDF slows down so that the driver may take the turn at an appropriate speed. But the driver decides to override the navigation system by pressing the gas pedal in order to instead go straight at the intersection. The resulting intervention therefore represents a disagreement of driver and navigation, which is unrelated to the PLDF itself.

The merging and renaming of the AAL2 labels is depicted in Table C.1 to Table C.3. As described in Section 5.2.2, most label names are slightly shortened and some AAL2 intervention types are merged. Additionally, remaining classes that contain fewer than five instances are merged with the AAL2 label *other*, since they are too small for the stratified five-fold cross-validation employed during training.

The resulting final classification labels are depicted in Table C.4 and Table C.5. Table C.4 depicts all classes with a *relevant* binary label, and Table C.5 depicts all classes with *irrelevant* binary labels.

Table C.1: All relevant classification labels and their corresponding AAL2 labels. Merged labels are marked in bold letters.

Classification labels (relevant)	Corresponding AAL2 labels
Curve higher speed	Curve higher speed
Curve lower speed	Curve lower speed
Give way sign lower speed	Give way sign lower speed
Roundabout higher speed	Roundabout higher speed
Roundabout lower speed	Roundabout lower speed
Slower map speed acceleration	Acceleration due to incorrect lower speed limit in map
	Acceleration due to incorrect give way sign in map
Speed limit earlier acceleration	Earlier acceleration onto higher speed limit
Speed limit later deceleration	Later deceleration onto lower speed limit
Speed limit stronger acceleration	Stronger acceleration to current set speed
Straight road higher speed	Straight road acceleration
	Acceleration due to incorrect speed limit
Straight road lower speed	Straight road deceleration
	Brake due to incorrect speed limit
Turn higher speed	Left turn higher speed
	Right turn higher speed
Turn lower speed	Left turn lower speed
	Right turn lower speed

Table C.2: The first half of the irrelevant classification labels and their corresponding AAL2 labels. Merged labels are marked in bold letters.

Classification labels (irrelevant)	Corresponding AAL2 labels
ACC low speed acceleration	ACC lower distance
	ACC stronger acceleration
	Stronger acceleration after stop
ACC target not detected	ACC target mistakenly lost
	ACC target not detected
ACC wrong target	ACC target retained too long
	ACC wrong target
Green traffic light acceleration	Gas to make it through green traffic light
Lane change	Lane change
Manual driving	Prolonged manual driving
Multi-lane traffic	Cut-in by other driver
	Merging
Narrow street traffic acceleration	Gas due to traffic in narrow street

Table C.3: The second half of the irrelevant classification labels and their corresponding AAL2 labels. Merged labels are marked in bold letters.

Classification labels (irrelevant)	Corresponding AAL2 labels
Other	ACChigher distance
	Gas due to traffic in roundabout
	Other
	Stop from taking over on right lane
	U-turn
Other traffic deceleration	Brake due to pedestrian crossing
	Brake due to traffic at a give way sign
	Brake due to traffic in narrow street
	Other traffic interactions
Overtaking	Overtaking maneuver
Reactivate after stop	Reactivate system after long stop
Red traffic light	Brake at red traffic light
Roundabout traffic deceleration	Brake due to traffic in roundabout
Stopover	Stopover
Turn traffic deceleration	Brake due to traffic in left turn
	Brake due to traffic in right turn

Table C.4: Relevant intervention classes for the classification, including their binary classification label, the number of instances in the dataset, and a description.

Class name (relevant)	Instances	Description
Curve higher speed	124	Driver accelerates to drive faster through a curve.
Curve lower speed	20	Driver decelerates to drive slower through a curve.
Give way sign lower speed	8	Driver decelerates to approach a give way sign on a straight road more slowly.
Roundabout higher speed	398	Driver accelerates to drive faster through a roundabout.
Roundabout lower speed	12	Driver decelerates to drive slower through a roundabout.
Slower map speed acceleration	63	Driver accelerates above a wrong predictive map speed.
Speed limit earlier acceleration	266	Driver accelerates earlier onto an upcoming higher speed limit.
Speed limit later deceleration	285	Driver decelerates later onto an upcoming lower speed limit.
Speed limit stronger acceleration	41	Driver accelerates additionally during an acceleration to the current legal speed.
Straight road higher speed	426	Driver accelerates above the current set speed on a straight road.
Straight road lower speed	51	Driver decelerates below the current set speed on a straight road.
Turn higher speed	142	Driver drives faster through a turn at an intersection.
Turn lower speed	70	Driver drives slower through a turn at an intersection.

Table C.5: Irrelevant intervention classes for the classification, including their binary classification label, their number of instances, and a description.

Class name (irrelevant)	Instances	Description
ACC low speed acceleration	142	Acceleration in low speed stop and go scenarios to override too sluggish ACC behavior.
ACC target not detected	98	Deceleration to keep a safe distance to a leading vehicle that was not or too late detected by the ACC.
ACC wrong target	133	Acceleration to override the ACC that detected a wrong target vehicle.
Green traffic light acceleration	14	Acceleration to make it through a green traffic light before it turns red.
Lane change	66	Acceleration or deceleration to change lanes on a multi lane street.
Manual driving	12	Driver turns off the PLDF for a prolonged period of manual driving.
Multi-lane traffic	39	Acceleration or deceleration due to multi-lane traffic, e.g., merging.
Narrow street traffic acceleration	11	Acceleration in a narrow street to clear the lane for oncoming traffic that otherwise may not be able to pass.
Other	24	Combination of irrelevant classes which by themselves do not have enough instances for the conducted stratified cross-fold validation.
Other traffic deceleration	128	Deceleration due to traffic in scenarios not covered by the other classes, e.g., to give way in narrow streets.
Overtaking	78	Acceleration to overtake another vehicle.
Reactivate after stop	124	Acceleration after a long stop in order to reactivate the system.
Red traffic light	195	Deceleration in front of a red traffic light.
Roundabout traffic deceleration	67	Deceleration while entering a roundabout due to traffic.
Stopover	10	Driver stops the vehicle for a prolonged period of time.
Turn traffic deceleration	148	Deceleration while entering a turn due to traffic.

C.3 Signal List

In Section 5.2.3, the preprocessing of the classification dataset is explained, including the preprocessing of the PLDF bus signals. The list of used signals for the classification and a short explanation for each signal is contained in Table C.6 and Table C.7.

In the original bus signals, the general braking pressure is recorded, independent of who presses the brakes, i.e., the driver or the driving function. Accordingly, the *driver brakes flag* is contained in the dataset to distinguish whether the driver or the system pressed the brakes. For the analyses in this thesis, only the portion of the braking pressure that originates from the driver is relevant. Therefore, the original braking pressure signal and the *driver brakes flag* were combined to create the *driver braking pressure* signal, which is also used during classification. The *driver brakes flag* was kept in the classification signals, although its information is mirrored in the *driver braking pressure* signal.

C.4 Hyperparameter Tuning Results

In Section 5.2.4, the model training pipeline is explained, including the employed hyperparameter tuning. Table C.8 depicts the optimal values determined via tha hyperparameter tuning. Each hyperparameter is listed with its respective models, and the range of tested values. The parameters that achieved the highest performance for each model are highlighted in bold with a model-specific marker.

Table C.6: First half of the preprocessed vehicle bus signals used during classification.

Signal name	**Description**	**Unit**
ACC in control flag	Binary indicator whether the ACC currently overrides the PLDF speed.	-
ACC target distance	Distance to the leading vehicle.	m
ACC target relative speed	Relative speed of the leading vehicle compared to the ego vehicle.	km/h
Distance to higher map speed limit	Distance to the closest change to a higher speed limit based on map data, capped at 100m.	m
Distance to higher TSD speed limit	Distance to the closest change to a higher speed limit based on TSD data, capped at 100m.	m
Distance to lower map speed limit	Distance to the closest change to a lower speed limit based on map data, capped at 100m.	m
Distance to lower TSD speed limit	Distance to the closest change to a lower speed limit based on TSD data, capped at 100m.	m
Distance to roundabout	Distance to the closest beginning of a roundabout based on map data, capped at 100m.	m
Distance to traffic light	Distance to the closest traffic light based on map data, capped at 100m.	m
Distance to turn	Distance to the closest turn based on map data, capped at 100m.	m

Table C.7: Second half of the preprocessed vehicle bus signals used during classification.

Signal name	**Description**	**Unit**
Driver braking pressure	Braking pressure applied by the driver via the brake pedal.	bar
Driver brakes flag	Binary indicator whether the brake pedal is pressed or not.	-
Gas pedal activation	Activation of the gas pedal.	%
Legal speed limit (map)	Current legal speed limit retrieved from map data.	km/h
Legal speed limit (TSD)	Current legal speed limit according to the TSD.	km/h
Number of lanes	Number of lanes of the current street.	-
PLDF set speed	Current set speed of the PLDF.	km/h
PLDF status	Categorical, one-hot encoded variable indicating the current PLDF status.	-
Street curvature	Street curvature retrieved from map data.	1/m
Vehicle speed	Current vehicle speed shown in the speedometer.	km/h

Table C.8: Hyperparameter tuning results. The ranges of each tested hyperparameter are shown, and the chosen values are highlighted in bold text with an additional marker depending on the applicable algorithm: $^{\mathrm{I}}$ for InceptionTime, $^{\mathrm{L}}$ for LSTM-FCN, $^{\mathrm{R}}$ for ROCKET.

Hyperparameter	Model	Tested values
Maximum map feature distance	All	$50\,\mathrm{m}$, $\mathbf{100\,m}^{\mathrm{ILR}}$, $150\,\mathrm{m}$, $200\,\mathrm{m}$
Input sequence length $T_{sequence}$	All	$\mathbf{20\,s}^{\mathrm{R}}$, $40\,\mathrm{s}$, $\mathbf{60\,s}^{\mathrm{I}}$, $\mathbf{80\,s}^{\mathrm{L}}$
Cutout window T_{cutout}	All	$0\,\mathrm{s}$, $2.5\,\mathrm{s}$, $5\,\mathrm{s}$, $\mathbf{7.5\,s}^{\mathrm{L}}$, $\mathbf{10\,s}^{\mathrm{IR}}$, $12.5\,\mathrm{s}$, $15\,\mathrm{s}$
Batch size	InceptionTime, LSTM-FCN	**8**$^{\mathrm{L}}$, **16**$^{\mathrm{I}}$, 32, 64, 128
Learning rate	InceptionTime, LSTM-FCN	10^{-5}, $5 \cdot 10^{-5}$, $\mathbf{10^{-4}}^{\mathrm{I}}$, $5 \cdot 10^{-4}$, $\mathbf{10^{-3}}^{\mathrm{L}}$, $5 \cdot 10^{-3}$, 10^{-2}
Kernel size	InceptionTime	20, **40**$^{\mathrm{I}}$, 60, 80
Network depth	InceptionTime	3, **6**$^{\mathrm{I}}$, 9, 12
Kernel sizes	LSTM-FCN	[4, 2, 1], **[8, 5, 3]**$^{\mathrm{L}}$, [16, 10, 6]
Filter sizes	LSTM-FCN	[64, 128, 64], [128, 256, 128], **[256, 512, 256]**$^{\mathrm{L}}$, [512, 1024, 512]
Dropout rate	LSTM-FCN	0.6, **0.7**$^{\mathrm{L}}$, 0.8, 0.9
Number of kernels	ROCKET	5000, 10000, 20000, **40000**$^{\mathrm{R}}$, 60000
Maximum number of dilations per kernel	ROCKET	16, **32**$^{\mathrm{R}}$, 64
Number of features per kernel	ROCKET	2, **4**$^{\mathrm{R}}$, 8, 16

D Supplements of the Prototype Test Group Study

This appendix contains supplementary materials for Chapter 6. In Section D.1, a supplementary explanation of the stretch factor calculation used in the SPAA is provided. A discussion on the validity of the prototype test group study's results when applied to real-world driving scenarios is provided in Section D.2. In Section D.3 and Section D.4, the briefing and questionnaires employed during the prototype study are presented. Finally, Section D.5 contains the complete speed profile development plots of the three example participants analyzed in Section 6.3.4.

D.1 Stretch Factor Calculation

In Section 6.1.2, the implementation of the pedal intervention SPAA is explained. In the first preprocessing step of the driver intervention segments, a *stretch factor* α is introduced to stretch the distance values of the intervention in the negative distance direction. Usually, an initial value of $\alpha_0 = 0.5$ is used in the SPAA. However, this stretch factor is limited by two conditions:

1. A maximum time threshold $T_{\max} = 3\,\mathrm{s}$ is defined that limits the distance between the start of the original and stretched intervention. The corresponding time-limited maximum stretch factor is denoted as $\alpha_{\max}$, with $\alpha_{\max} > 0$.
2. The stretch factor is reduced near high-curvature segments, such as curves and turns, by calculating an adjusted stretch factor α_{curve} based on variations of the vehicle speed around the intervention, with $0 \leq \alpha_{\text{curve}} \leq \alpha_0$.

The final stretch factor is then calculated as the minimum of both stretch factors:

$$\alpha = \min(\alpha_{\max}, \alpha_{\text{curve}}). \tag{D.1}$$

The calculation of both $\alpha_{\max}$ and α_{curve} is explained in the following sections.

D.1.1 Time-Limited Maximum Stretch Factor

Let the original start and end distances of an intervention be denoted by d_0 and d_n, respectively, and let its start time be given by t_0. The spatial length of the intervention is then defined as

$$L_\mathrm{I} = d_n - d_0. \tag{D.2}$$

The earliest admissible start time of the stretched intervention is introduced as

$$t_\mathrm{min} = t_0 - T_\mathrm{max}. \tag{D.3}$$

Let $s_\mathrm{driver}(t)$ denote the cumulative distance traveled as a function of time, derived from the original driver speed profile. The corresponding earliest admissible start distance is then given by

$$d_\mathrm{min} = s_\mathrm{driver}(t_\mathrm{min}). \tag{D.4}$$

The maximum admissible stretch distance is calculated as

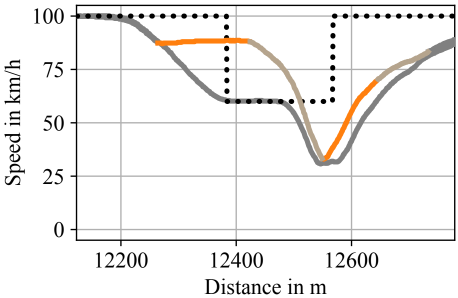
$$L_\mathrm{S}^\mathrm{max} = d_0 - d_\mathrm{min}. \tag{D.5}$$

Finally, the time-limited maximum stretch factor is expressed as

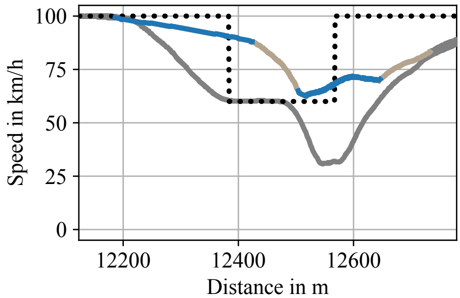
$$\alpha_\mathrm{max} = \frac{L_\mathrm{S}^\mathrm{max}}{L_\mathrm{I}}. \tag{D.6}$$

D.1.2 Curvature-Limited Stretch Factor

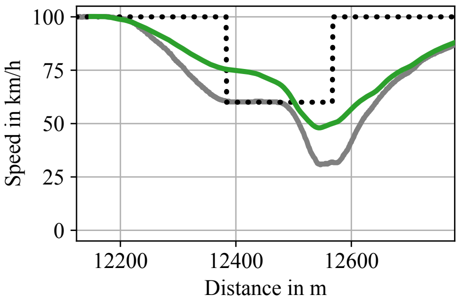
The main drawback of the intervention stretching is that preceding segments of the speed profile that are potentially safety-relevant may be overwritten. This issue is particularly critical for decelerations in high-curvature road sections, such as roundabouts, curves, and turns. An example of the SPAA in a turn scenario without the stretch factor restriction is provided in Figure D.1. The original PLDF baseline speed and the corresponding driver interventions are illustrated in Figure D.1a. Initially, the legal speed limit decreases from $100\,\mathrm{km/h}$ to $60\,\mathrm{km/h}$, and the driver intervenes by pressing the gas pedal in order to delay the deceleration toward the upcoming lower speed limit. The driver then releases the gas pedal, and the PLDF performs a strong deceleration due to an upcoming left turn, which is traversed at $30\,\mathrm{km/h}$ in both the baseline and the driver speed profile. After passing the turn apex, the driver intervenes again by accelerating to exit the turn at a higher speed. Both performed interventions are preprocessed in Figure D.1b by stretching them in the negative distance direction and applying a linear offset to realign them with the surrounding speed profile. As can be seen, this preprocessing works well for the first driver intervention, but it causes significant issues in the second intervention, where critical parts of the deceleration segment preceding the turn are overwritten. As a result, the merged and smoothed new baseline speed in Figure D.1c traverses the turn at approximately $50\,\mathrm{km/h}$ instead of $30\,\mathrm{km/h}$, which poses substantial safety concerns. Thus, the stretch factor applied to the second intervention must be limited to ensure that no safety-critical deceleration segments are overwritten.



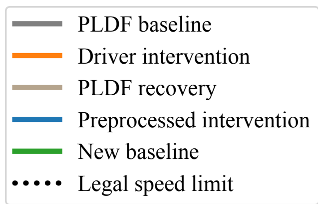
(a) The PLDF's original baseline speed and the performed example driver interventions.



(b) Preprocessing of the driver interventions without a curvature-limited stretch factor: Both intervention segments are stretched in the negative distance direction, and a linearly decreasing offset is applied.



(c) The preprocessed driver interventions and the PLDF baseline speed are merged and smoothed to create the new baseline speed.



(d) Legend of the speed-over-distance plots.

Figure D.1: The SPAA without the curvature-limited stretch factor applied to two example interventions around a turn. The first intervention is labeled as *later deceleration onto lower speed limit*, and the second intervention has the *left turn higher speed* AAL2 label. Each subplot shows the same driver intervention at the same location, but at different stages of the pedal intervention SPAA process. The combined legend for the three speed-over-distance plots is provided in Figure D.1d.

The computation of the curvature-limited stretch factor α_{curve} is detailed in the following paragraphs. The algorithm's implementation is designed to achieve fast computation times and is intended to operate on relatively short intervention segments.

Step 1: Identification of the relevant baseline speed segment. Let an original intervention segment of the driver speed profile be represented by its distances $(d_i)_{i=0}^n$ and corresponding velocities $(v_i)_{i=0}^n$. The spatial length of this intervention segment is defined as

$$L_\mathrm{I} = d_n - d_0. \tag{D.7}$$

By applying the initial stretch factor α_0 to the intervention segment, the spatial length of the segment that is potentially overwritten by the stretched intervention is given by

$$L_\mathrm{S} = (\alpha_0 + 1) \cdot L_\mathrm{I}. \tag{D.8}$$

The original portion of the baseline speed profile that may be overwritten is denoted by the distances $(s_i)_{i=0}^m$ and the corresponding velocities $(u_i)_{i=0}^m$, where

$$s_0 = d_n - L_\mathrm{S}, \qquad s_m = d_n. \tag{D.9}$$

Step 2: Identification of the minimum speed. To assess whether the potentially overwritten segment of the baseline speed profile contains any valley-shaped fluctuations, the minimum speed within the segment is determined and compared to its surrounding values. Let the minimum speed value be denoted by u_min, occurring at index k, which is calculated as

$$u_\mathrm{min} = \min_{0 \le i \le m} u_i, \qquad k = \arg \min_{0 \le i \le m} u_i. \tag{D.10}$$

Step 3: Identification of the segment's surrounding maximum speeds. The maximum speed values within the investigated segment, before and after the previously identified minimum speed, are computed as

$$u_\mathrm{max}^\mathrm{left} = \max_{0 \le i < k} u_i, \qquad u_\mathrm{max}^\mathrm{right} = \max_{k < i \le m} u_i, \tag{D.11}$$

where only the lower of these two maxima is considered for the calculation of the curvature-limited stretch factor:

$$u_\mathrm{max}^- = \min(u_\mathrm{max}^\mathrm{left}, u_\mathrm{max}^\mathrm{right}). \tag{D.12}$$

Step 4: Quantification of valley-shaped speed fluctuations. The so-called *valley factor* F_{valley} is defined as the difference between the minimum speed and the lower of the two surrounding maxima, relative to the spatial length of the potentially overwritten segment:

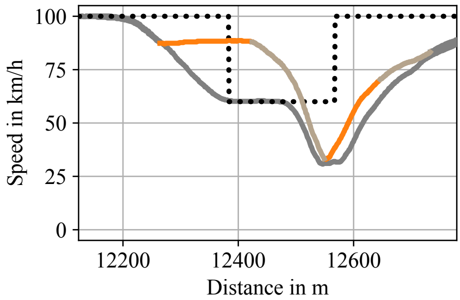
$$F_{\text{valley}} = \frac{u_{\max}^{-} - u_{\min}}{L_{\text{S}}}. \tag{D.13}$$

For speed profiles that do not exhibit valley-shaped fluctuations, such as monotonically increasing or monotonically decreasing profiles, the minimum speed occurs at either the beginning or the end of the segment. In these cases, F_{valley} equals zero. Non-zero values of F_{valley} occur only if a higher speed value is present both before and after the minimum speed, indicating a valley within the segment.

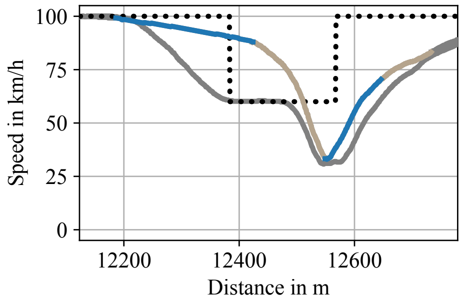
Step 5: Determination of the curvature-limited stretch factor. Using the previously calculated valley factor F_{valley}, the curvature-limited stretch factor α_{curve} is obtained. If F_{valley} equals zero, no limitation of the initial stretch factor is necessary, and α_0 is used. Conversely, if the valley factor reaches the maximum admissible threshold $F_{\text{valley}}^{\max}$, the stretch factor is set to zero. For F_{valley} values between zero and $F_{\text{valley}}^{\max}$, the stretch factor is proportionally reduced. Formally, α_{curve} is defined as

$$\alpha_{\text{curve}} = \begin{cases} \alpha_0, & \text{if } F_{\text{valley}} = 0, \\ \alpha_0 \cdot \left(1 - \dfrac{F_{\text{valley}}}{F_{\text{valley}}^{\max}}\right), & \text{if } 0 < F_{\text{valley}} < F_{\text{valley}}^{\max}, \\ 0, & \text{if } F_{\text{valley}} \geq F_{\text{valley}}^{\max}. \end{cases} \tag{D.14}$$

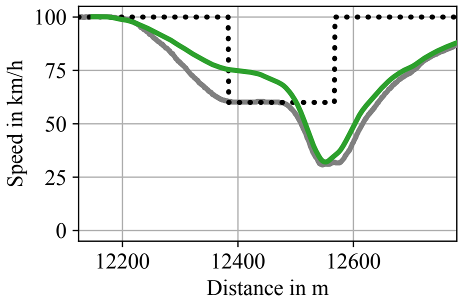
For the analyses conducted in this thesis, a threshold of $F_{\text{valley}}^{\max} = 0.1\,\text{km}/(\text{h}\,\text{m}) \equiv 100\,1/\text{h}$ is used. In Figure D.2, the SPAA with the curvature-limited stretch factor is applied to the same driver interventions as illustrated in Figure D.1. The first intervention is stretched using the base stretch factor $\alpha_0 = 0.5$, since the affected baseline speed profile is monotonically decreasing, and $F_{\text{valley}} = 0\,\text{km}/(\text{h}\,\text{m})$ is calculated. In contrast, the baseline speed profile corresponding to the second intervention reaches a valley factor of $F_{\text{valley}} = 0.47\,\text{km}/(\text{h}\,\text{m})$, which results in a stretch factor of $\alpha = 0$ for the second intervention. Consequently, the safety-critical deceleration segment preceding the turn remains unaffected by the stretched intervention.



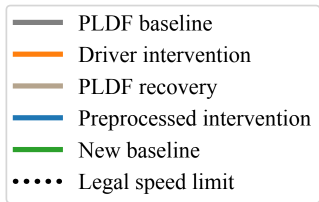
(a) The PLDF's original baseline speed and the performed example driver interventions.



(b) Preprocessing of the driver interventions using the curvature-limited stretch factor: Only the first intervention is preprocessed using the base stretch factor of $\alpha_0 = 0.5$, while no stretching is applied to the second intervention.



(c) The preprocessed driver interventions and the PLDF baseline speed are merged and smoothed to create the new baseline speed.



(d) Legend of the speed-over-distance plots.

Figure D.2: The SPAA using the curvature-limited stretch factor applied to the same example interventions as in Figure D.1. While the first intervention is stretched normally using the base stretch factor of $\alpha_0 = 0.5$, a stretch factor of $\alpha = 0$ is used for the second intervention in order to prevent the overwriting of safety-relevant decelerations preceding the turn. Each subplot shows the same driver intervention at the same location, but at different stages of the pedal intervention SPAA process. The combined legend for the three speed-over-distance plots is provided in Figure D.2d.

D.2 Real-World Validity of Simulation-based Test Group Studies

In Section 6.2.3, the simulation environment used in the prototype test group study is presented. However, as this simulation-based study aims to transfer its results to real-world driving, the applicability of the employed simulation environment and study design to real-world driving conditions must be evaluated.

The correspondence between driver behavior in driving simulators and in real-world driving is referred to as *behavioral validity* [Bla82; Him+23]. This behavioral validity is commonly distinguished into *relative validity* and *absolute validity*. Relative validity indicates that trends observed in the simulator follow the same direction as those in real-world driving, whereas absolute validity refers to the absolute numerical correspondence of measurable behavior between the simulator and the real world [Bla82]. Depending on the specific use case, a test group study may require either relative or absolute validity [Him+25]. In the context of the prototype test group study presented in Chapter 6, relative validity is sufficient, as the objective is to evaluate the differences in driver behavior and satisfaction between the original and the self-learning driving function. Thus, it is sufficient to preserve the trends between these measures, while absolute validity is not strictly necessary, but still desirable.

Wynne, Beanland, and Salmon [WBS19] conducted a systematic literature review of 44 studies examining the validity of driving simulators in representing real-world driving behavior. Their findings indicate that behavioral validity generally depends on the dependent variables considered and the driving simulator used. While numerous studies were found that address the validity of manual driving behavior, only few works were found on the validity of ADAS-related behavior. Across the reviewed works, the measures used to assess manual driving behavior vary considerably. The most commonly analyzed variables include mean speed and speed variation, lateral deviation and lane positioning, driving errors, and overall driving performance. Among these, the validity of the drivers' speed choice is particularly relevant for the prototype test group study. Absolute validity of the speed choice is found in most studies using medium- to high-fidelity driving simulators, whereas relative validity is found in most studies using low-fidelity simulators. According to the scoring system introduced by the authors, the simulator used in this thesis falls into the category of high-fidelity simulators, as it includes a 270 ° field of view, a motion platform with road sensation feedback, and a full vehicular cabin [WBS19].

Himmels *et al.* [Him+25] investigate the behavioral validity of driving simulators specifically in the context of ADASs. Their literature review confirms that only a limited number of studies address this domain, with most focusing on takeover times in response to system-initiated takeover requests [EBS17; Rie10; SB+18; Win+23]. However, the results of these studies are contradictory, as some studies report higher takeover times in driving simulators, whereas other studies report shorter takeover times compared to real-world conditions. Further research has examined collision avoidance scenarios [Zö+15], where no behavioral validity was observed in a static simulator, driving comfort [Bel+17], which demonstrated behavioral validity in some scenarios using two high-fidelity driving simulators, and driving behavior with a traffic light assistant [KYB21], which yielded relative validity for some of the used metrics in a static simulator. Therefore, Himmels *et al.* [Him+25] also come to the conclusion that behavioral validity is highly dependent on the specific use case, the investigated measures, and the driving simulator used.

Himmels *et al.* [Him+25] then evaluate the behavioral validity of the high-fidelity driving simulator owned by the BMW Group through three test group studies. One study was conducted under real-world driving conditions, while the other two were conducted in the evaluated driving simulator using different motion platform configurations. The authors investigate whether human factors related to the use of a Level 2 driving function can be validly assessed in the driving simulator, and how the chosen motion system influences the measured behavioral validity. The human factors regarding the used Level 2 driving function were recorded using questionnaires and include the cognitive load, system usability, acceptance, user experience, trust, and the mental model of the driving function. For the acceptance evaluation, the system acceptance scale by van der Laan, Heino, and De Waard [VHD97] was used, including both the usefulness and satisfaction subscales. As explained in Section 6.2.5, the same questionnaire is also used in the prototype test group study. In addition to these human factors, the usage behavior of the driving function was also evaluated. In particular, this includes the relative proportion of time the system was activated, as well as the locations and situations in which drivers chose to deactivate the driving function. Therefore, this study also addressed driver intervention behavior, similar to the prototype test group study [Him+25].

Each of the three test group studies was conducted with 31 participants, none of whom took part in more than one study. The used driving function in both the real vehicle and the simulator is a Level 2 in-production predictive driving function. Its described functionality and ODD align with the in-production PLDF, except for the lateral steering control, which is absent in the original PLDF. However, in the prototype test group study, the PLDF is coupled with a steering assist, effectively making it a Level 2 driving function as well. The BMW driving simulator features a hexapod similar to the one used in the prototype test group study, but it is additionally mounted on X-Y-rails with a redundant yaw table. In the first simulator configuration, called *simBIG*, the full simulator setup is utilized, whereas the second configuration, *simSMALL*, only uses the hexapod excluding the X-Y-rails and yaw table, resulting in lower acceleration forces and motion fidelity. Therefore, *simSMALL* resembles the simulator setup used in the prototype test group study. Both the simulator-based test group studies and the real-world study were conducted on the same rural route covering a distance of approximately 6 km. For the simulator studies, a precise digital twin of the track was created, including a visualization closely resembling the real-world environment. In addition, simulated traffic participants were introduced to recreate traffic situations encountered by some drivers during the real-world test drives. The route was driven two times in each direction, resulting in a total of four drives. Overall, the study procedures were designed to be nearly identical between the real-world drives and the driving simulator. For the subsequent data analysis, Bayesian ANalysis Of VAriance (BANOVA) was used [Him+25].

The results of the study show absolute validity for cognitive load, usability, acceptance, user experience, and the mental model in the *simBIG* configuration. For *simSMALL*, strong evidence was found only for the absolute validity of usability and the mental model, while weak evidence was found for the absolute validity of cognitive load, satisfaction, and trust. Regarding relative validity, strong evidence was found for usability, acceptance, and the mental model in *simSMALL*, whereas weak evidence was found for the relative validity of cognitive load and user experience. No evidence of validity was observed for driver trust in either simulator configuration. In addition to the human factors, system usage, i.e., driver intervention behavior, was analyzed. The conducted analyses show that drivers deactivated the function at the same locations across all three test environments, and relative validity is found for driving function usage in both simulator setups, while absolute validity was not evaluated [Him+25].

To summarize, Himmels *et al.* [Him+25] conducted a study with a highly similar setup to the prototype test group study in the *simSMALL* configuration. They report strong evidence for the relative validity of both driving function usage patterns and system acceptance, the latter assessed with the same questionnaire by van der Laan, Heino, and De Waard [VHD97]. For the driver satisfaction, a subscale of system acceptance, weak evidence of absolute validity in *simSMALL* is additionally found. Further analyses also confirm that the driver intervention behavior is consistent between the studies conducted in the real world and the simulators. Most related work also found absolute validity for the general speed choice and speed variations when using high-fidelity driving simulators [WBS19]. Based on these findings, validity for the real-world applicability of the results from the prototype test group study is assumed.

D.3 Prototype Study Briefing

The translated briefing used in the prototype test group study is depicted in Figure D.3 and Figure D.4. The briefing is elaborated on in Section 6.2.4.

D.4 Prototype Test Group Study Questionnaire

The translated questionnaires used in the prototype test group study are depicted in Figure D.5 to Figure D.7. The first page of the questionnaire focuses on demographic information and the driver's experience with ADASs. The second page contains the system acceptance scale by van der Laan, Heino, and De Waard [VHD97]. The official German translation of the system acceptance scale [WK] was used in the original German version of the questionnaires. The third page contains a custom questionnaire that inquires about the participant's satisfaction with the PLDF's speed profile specifically in the relevant situation types. The fourth and fifth pages of the questionnaire are exact copies of the second and third pages, as the same questionnaires are used for the separate evaluation of system A and system B.

PORSCHE

Study on Adaptive Driving Functions – Briefing for Participants

Introduction:
You are driving home from work. While driving, you use an assisted driving function that controls the vehicle speed. This driving function resembles Porsche InnoDrive and other similar predictive cruise control driving functions in terms of its operation and functionality.

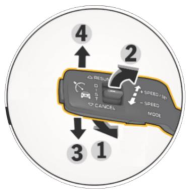
This means that the system includes:

- the automatic maintenance of the current set speed,
- the automatic adoption of speed limits,
- predictive deceleration onto lower speed limits,
- and automatic deceleration in and around curves.

Additionally, a **steering assist** (active lane centering) is active, which actively steers the vehicle to the lane center. This steering assist also stays active when the longitudinal driving function is overruled. Therefore, you can drive both hands-on and hands-off with the system. If you are driving hands-on, you can override the steering manually if you wish. The evaluation of the steering assist is **not** part of the study. Only the velocity control should be evaluated.

Operation of Porsche InnoDrive:
Driver assistance control stalk usage:

1. Pushing the lever away: Deactivate the system
2. Pulling the lever: Activate the system
3. Pushing the lever down: Set speed -1 km/h
   Pushing the lever further down: Set speed -10 km/h
4. Pushing the lever up: Set speed +1 km/h
   Pushing the lever further up: Set speed +10 km/h



Pedal usage:

- Gas: Vehicle accelerates, system stays active afterward
- Brake: Vehicle decelerates, system is deactivated
  (System must be manually reactivated afterward)

Content of the Study:
We prepared two versions of the same driving function for you. These driving functions are adaptive, i.e., they **adjust themselves to your individual driving behavior**. The goal of this study is that you test and evaluate both functions. To do this, you always drive the same route in the simulator. First you drive with system A and then with system B.
To familiarize yourself with the system and the simulator, first take a short test drive (approx. 1-3 minutes), during which you can test the pedals and the driver assistance control stalk.

Figure D.3: First page of the prototype test group study's briefing.

PORSCHE

To familiarize yourself with the route, use system A to drive the full route afterward, **without intervening**. Both the adjustment of the driving function's set speed and the use of the gas or brake pedal are considered to be interventions.
Start the vehicle via the gas pedal and then hand over control to the driving function by pulling the driver assistance control stalk toward yourself (Note: the system only activates after you released the gas pedal). The current set speed of the driving function is now shown in the instrument cluster (in the green number above the speedometer) and the system takes over the control of the vehicle speed. The study director informs you when the drive is finished.
After the introductory drives, you may drive with the active driving function as you would in the real world. **Please adjust the system behavior if the velocity chosen by the system deviates from your desired behavior.** You can adjust either the set speed via the control stalk or adjust the vehicle speed directly via the gas or brake pedals. If you used the brake, we kindly ask you to reactivate the system after your intervention again.

Study procedure:

1. Short test of the buttons and pedals in the driving simulator.
2. Introductory drive with system A without active driver interventions.
3. Test drives with system A.
4. Filling out the first questionnaire.
5. Test drives with system B.
6. Filling out the second questionnaire.

Introduction to the driving simulator:

The driving simulator is placed on a motion platform which is active during the tests. In order to use the motion platform, all doors of the vehicle must remain closed and your seat belt must stay fastened at all times. On startup, the motion platform moves up by approximately 50 cm. You may not notices this, since the visualization also moves upwards accordingly. Please do not leave the vehicle during the tests. Remain seated, keep your seatbelt fastened, and leave the simulator door closed until the instructor tells you that you can leave the simulator. During the experiment, the study director will talk to you on the radio and give you instructions. The study director also hears you speak over the radio. If you want to take a break or feel unwell, please inform us.

Final notes:

- The focus of the study is **only** the evaluation of the driven velocities. The steering assist should **not** be evaluated.
- Please look at the attached **questionnaire** before the experiment to know which aspects to evaluate. (You may already fill in the demographic data.)
- Please override the system during the test drives according to your personal preferences, just as you would in a real vehicle.

Figure D.4: Second page of the prototype test group study's briefing.

PORSCHE

Before the Study: Demographic Information

The following questionnaire is used to collect some demographic data. We are particularly interested in how much driving experience you have with automated driving functions. This voluntary information is important in order to be able to correctly interpret your test results and assessments in the study.
Thank you!

To be filled out by the study director		
Date	Experiment	Participant nr.
__.02.2025	1	__

To be filled out by the participant/driver						
Gender	Male	Female	Diverse	No statement		
	☐	☐	☐	☐		
Age (in years)	< 20	21 ≤ 30	31 ≤ 40	41 ≤ 50	51 ≤ 60	> 60
	☐	☐	☐	☐	☐	☐
Driving experience (in years)	≤ 5	6 ≤ 10	11 ≤ 20	21 ≤ 40	> 40	
	☐	☐	☐	☐	☐	
How commonly do you use Porsche InnoDrive?	Daily	Weekly	Monthly	Less often	Never	
	☐	☐	☐	☐	☐	
How commonly do you use comparable driving functions (e.g., ACC, VW Travel Assist, Mercedes DISTRONIC, etc.)?	Daily	Weekly	Monthly	Less often	Never	
	☐	☐	☐	☐	☐	
Which comparable driving functions do you use?						

Figure D.5: First page of the prototype test group study's questionnaire focusing on demographic information and experience with ADASs.

PORSCHE

After the First Part of the Study: Assessment of System A

The following questionnaire serves as a follow-up to the study. We are particularly interested in how satisfied you were with the driving function you experienced. This voluntary information is important in order to be able to correctly interpret your test results and assessments in the study.
Thank you!

To be filled out by the study director		
Date	Experiment	Participant nr.
__.02.2025	1	__ A

To be filled out by the participant/driver						
Please evaluate the system. Carefully read each pair of words and make a cross in each row.						
Useful	☐	☐	☐	☐	☐	Useless
Pleasant	☐	☐	☐	☐	☐	Unpleasant
Bad	☐	☐	☐	☐	☐	Good
Nice	☐	☐	☐	☐	☐	Annoying
Effective	☐	☐	☐	☐	☐	Superfluous
Irritating	☐	☐	☐	☐	☐	Likeable
Assisting	☐	☐	☐	☐	☐	Worthless
Undesirable	☐	☐	☐	☐	☐	Desirable
Raising Alertness	☐	☐	☐	☐	☐	Sleep-inducing

Figure D.6: Second page of the prototype test group study's questionnaire containing the system acceptance scale by van der Laan, Heino, and De Waard [VHD97]. Two separate copies of this page are used for the evaluation of system A and system B.

PORSCHE

To be filled out by the participant/driver					
I was satisfied with the driving function in general.	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
	☐	☐	☐	☐	☐
I was satisfied with the number of times I had to intervene.	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
	☐	☐	☐	☐	☐
I was satisfied with the system's ...					
Speed on straight road segments.	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
	☐	☐	☐	☐	☐
Acceleration timings onto higher speed limit signs.	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
	☐	☐	☐	☐	☐
Deceleration timings onto lower speed limits signs.	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
	☐	☐	☐	☐	☐
Speed in and around curves.	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
	☐	☐	☐	☐	☐
Acceleration strength.	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
	☐	☐	☐	☐	☐
Deceleration strength.	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
	☐	☐	☐	☐	☐

Figure D.7: Third page of the prototype test group study's questionnaire featuring a custom questionnaire about the driver satisfaction with the PLDF's speed profile in the relevant situations. Two separate copies of this page are used for the evaluation of system A and system B.

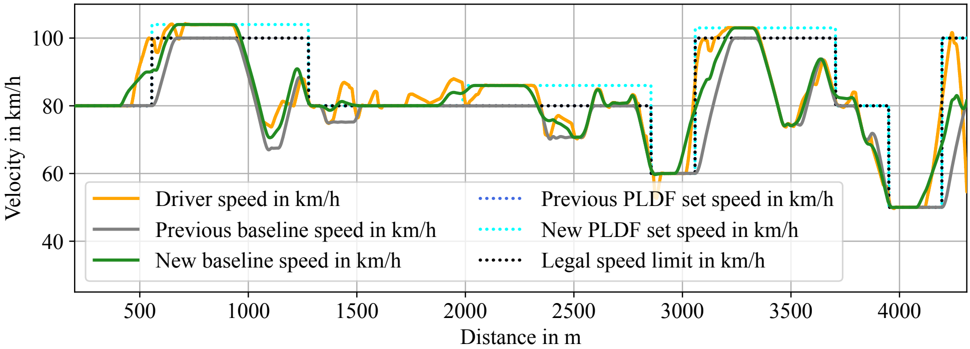
From these questionnaires, the five human factors measures *general acceptance*, *general satisfaction*, *general usefulness*, *intervention frequency satisfaction*, and *speed profile satisfaction* are derived, as described in Section 6.3.1. The first three measures are derived from the system acceptance scale by calculating the mean value of their respective items in the questionnaire, as explained in Section 2.1.7. The intervention frequency satisfaction is directly taken from the custom questionnaire's second item *I was satisfied with the number of times I had to intervene*. The speed profile satisfaction is calculated as the mean of the last six items of the custom questionnaire on the third and fifth pages, which are titled by the phrase *I was satisfied with the system's ...* . The first item of the custom questionnaire, *I was satisfied with the driving function in general*, was not used for the evaluations of the study results.

D.5 Complete Baseline Speed Profile Development Example Plots

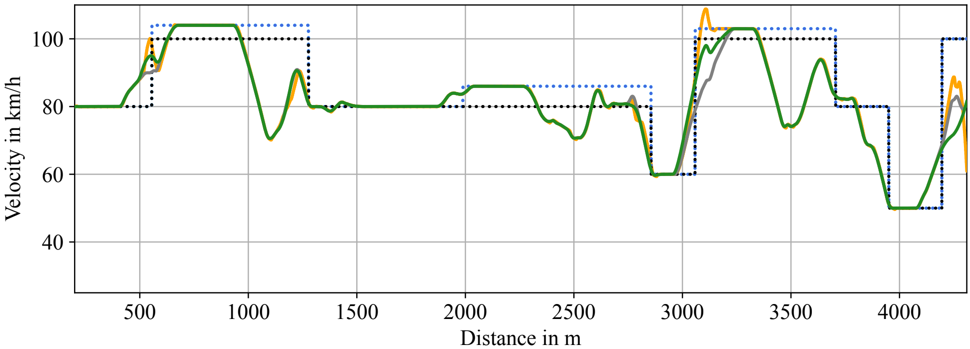
In Section 6.3.4, the development of the baseline speed profiles is analyzed based on three example participants using condensed plots that only feature the baseline speed profiles. In this section, the complete speed profiles of each iteration of the three example participants are illustrated separately, including the driver speed profile used to generate the adjusted speed profiles. The speed profile development plots of participant number 26 are illustrated in Figure D.8. Figure D.9 illustrates the speed profile development of participant number 13. Finally, participant number 41's speed profile development is depicted in Figure D.10.

In each plot, the baseline speed profile generated in the previous iteration is depicted as a gray solid line. In the case of system A drives, the used baseline speed is the original PLDF baseline speed. The recorded driver speed profile is depicted as an orange solid line. In segments where the driver speed deviates from the baseline speed, a driver intervention is conducted. Using the previous baseline speed and the recorded driver speed, the SPAA is used to generate the next adjusted baseline speed. This new speed profile is then used as the baseline speed of the self-learning PLDF in the next iteration. Additionally, the set speed of the previous iteration is depicted by the dark blue dotted line, and the set speed changes conducted in the current iteration are depicted by the light blue dotted line.

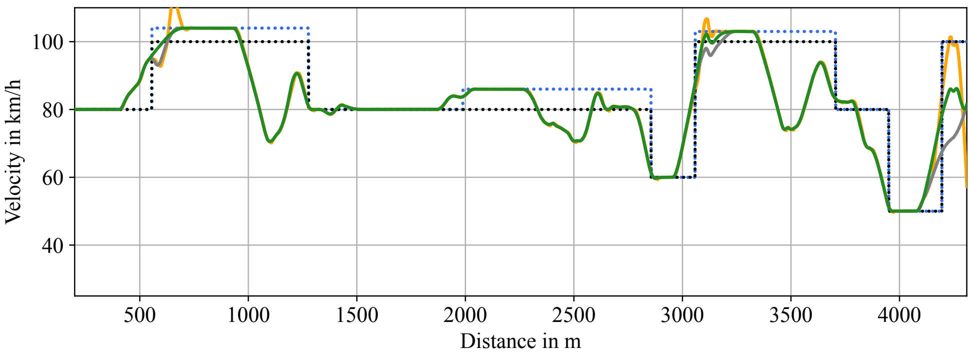
Participant number 13 prematurely stopped the vehicle at the end of the town segment, at a distance of approximately 4200 m, during all test drives. Therefore, the driver speed profile drops to zero at this location.



(a) Second system A drive. The generated new speed profile is used as the baseline in the first system B drive.

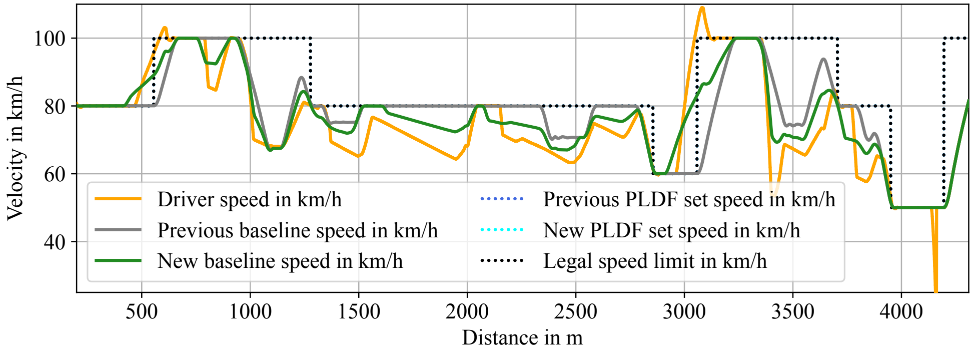


(b) First system B drive. The generated new speed profile is used as the baseline in the second system B drive.

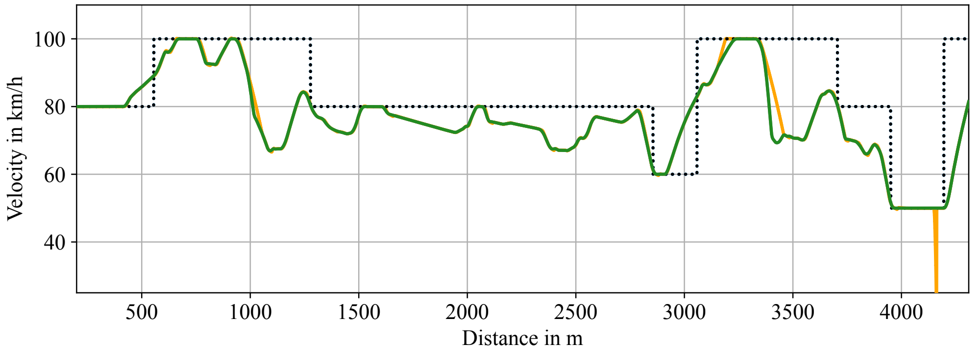


(c) Second system B drive. The generated new speed profile is used as the final baseline in the last system B drive without interventions.

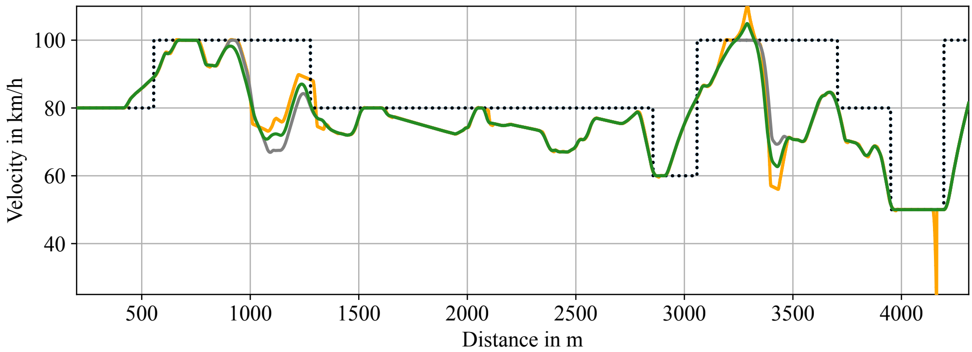
Figure D.8: Complete baseline speed profile development plots for participant number 26.



(a) Second system A drive. The generated new speed profile is used as the baseline in the first system B drive.

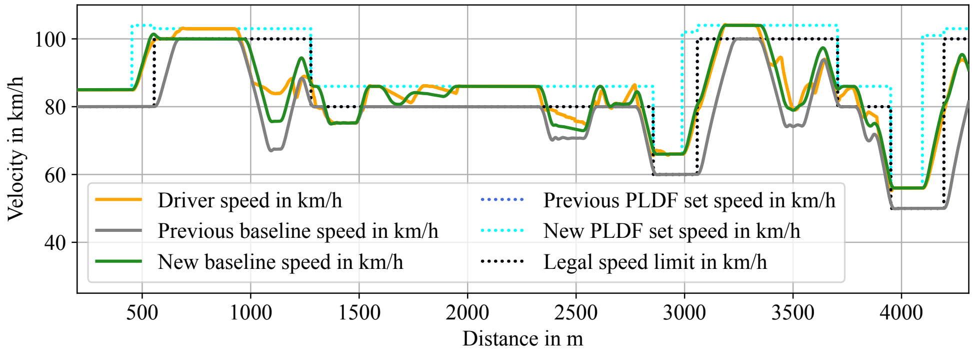


(b) First system B drive. The generated new speed profile is used as the baseline in the second system B drive.

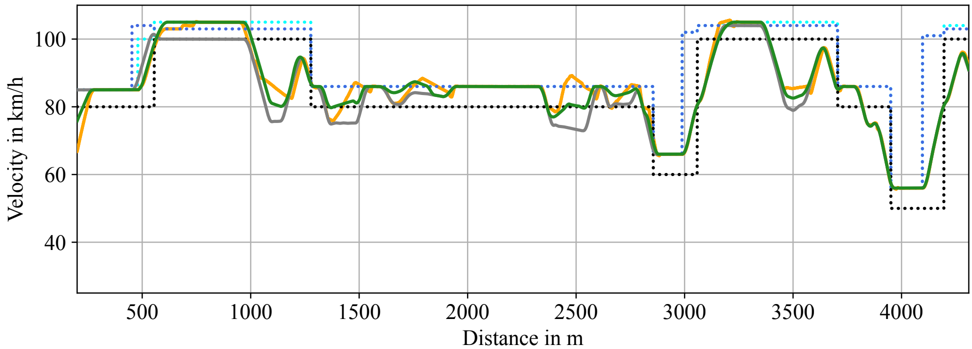


(c) Second system B drive. The generated new speed profile is used as the final baseline in the last system B drive without interventions.

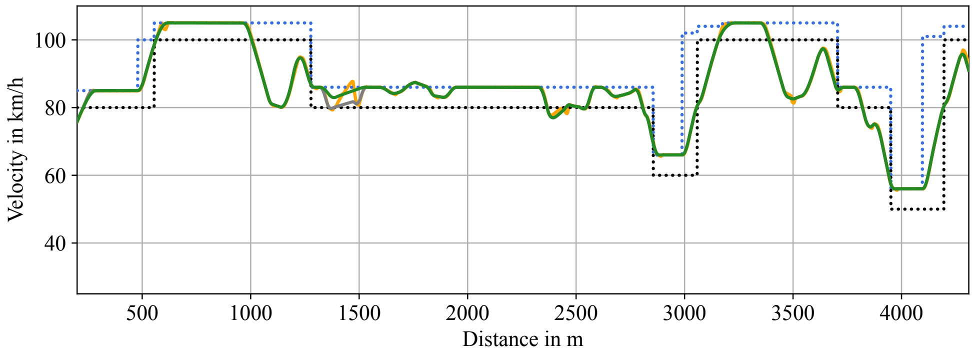
Figure D.9: Complete baseline speed profile development plots for participant number 13.



(a) Second system A drive. The generated new speed profile is used as the baseline in the first system B drive.



(b) The first system B drive. The generated new speed profile is used as the baseline in the second system B drive.



(c) The second system B drive. The generated new speed profile is used as the final baseline in the last system B drive without interventions.

Figure D.10: Complete baseline speed profile development plots for participant number 41.

Abbreviations and Symbols

Abbreviations

AAL	Annotation Abstraction Level
AAL1	Annotation Abstraction Level One
AAL2	Annotation Abstraction Level Two
AAL3	Annotation Abstraction Level Three
AAL4	Annotation Abstraction Level Four
ACC	Adaptive Cruise Control
ADAS	Advanced Driver Assistance System
AP	Auto Pilot
AVT	Advanced Vehicle Technology
BANOVA	Bayesian ANalysis Of VAriance
BC	Behavioral Cloning
BN	Batch Normalization
C	Curve
CNN	Convolutional Neural Network
DAgger	Dataset Aggregation
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DTW	Dynamic Time Warping
ECU	Electronic Control Unit
EIL	Expert Intervention Learning
FCN	Fully Convolutional Network
GC	Global Collective
GI	Global Individual
GPS	Global Positioning System
HIVE-COTE	HIerarchical VotE COllective of Transformation-based Ensemble
HMI	Human-Machine-Interface
IIL	Interactive Imitation Learning
IL	Imitation Learning
IMPC	Inverse Model Predictive Control
IMU	Inertial Measurement Unit
IR	Intervention Rate
IRL	Inverse Reinforcement Learning
LC	Local Collective
LED	Light-Emitting Diode

LI	Local Individual
LSTM	Long Short Term Memory
LSTM-FCN	Long Short Term Memory Fully Convolutional Network
MIT	Massachusetts Institute of Technology
MTS	Multivariate Time Series
MTSC	Multivariate Time Series Classification
ODD	Operational Design Domain
PFI	Permutation Feature Importance
PLDF	Predictive Longitudinal Driving Function
R	Roundabout
ReLU	Rectified Linear Unit
RL	Reinforcement Learning
ROCKET	RandOm Convolutional KErnel Transform
SAE	Society of Automotive Engineers
SC	Super Cruise
SLD	Speed Limit Decrease
SLI	Speed Limit Increase
SPAA	Speed Profile Adjustment Algorithm
SR	Straight Road
T	Turn
THW	Time Headway
ToC	Transition of Control
TSD	Traffic Sign Detection
TTC	Time to Collision
UDP	User Datagram Protocol
UEA	University of East Anglia
UTS	Univariate Time Series
UTSC	Univariate Time Series Classification
WEASEL+MUSE	Word ExtrAction for time SEries cLassification plus Multivariate Unsupervised Symbols and dErivatives

Latin Letters

Symbol	Description
$c^{+}_{j,k}$	Number of drive-throughs with a speed increase by driver j at location k
$c^{-}_{j,k}$	Number of drive-throughs with a speed decrease by driver j at location k
$c^{\text{total}}_{j,k}$	Number of drive-throughs by driver j at location k where an intervention was possible
d_i	Distance value of a driver intervention at index i
d'_i	Distance value of a stretched driver intervention at index i
F_{valley}	Valley factor, used for the calculation of α_{curve}

Symbol	Description
$F_{\text{valley}}^{\max}$	Maximum valley factor above which α_{curve} is set to 0
$\text{IR}_{\square}^{+}$	Higher speed IR for the evaluation of $\square$ adjustments specified by index
$\text{IR}_{\square}^{-}$	Lower speed IR for the evaluation of $\square$ adjustments specified by index
L_{I}	Spatial length of an intervention
L_{S}	Spatial length of a stretched intervention
$L_{\text{S}}^{\max}$	Maximum admissible stretched intervention length
$N_{\mathcal{D}}$	Number of drivers
$N_{\mathcal{L}}$	Number of location clusters
s_i	Distance value of the to-be-overwritten baseline speed profile at index i
T_{cutout}	Cutout window around an intervention during MTSC preprocessing
$T_{\max}$	Maximum time threshold limiting the SPAA stretching
$T_{sequence}$	Total intervention sequence duration during MTSC preprocessing
u_i	Longitudinal speed of the to-be-overwritten baseline speed profile at index i
$v_{\text{baseline}}(d)$	Longitudinal baseline speed of the PLDF over d
$v_{\text{driver}}(d)$	Longitudinal driver speed profile over d
$v_{\text{final}}(d)$	Adjusted longitudinal driver speed profile over d, generated by the SPAA
v_i	Longitudinal speed of an intervention at index i
v_i'	Offset-corrected longitudinal speed of an intervention at index i
$v_{\text{mean}}(d)$	Mean speed profile between $v_{\text{baseline}}(d)$ and $v_{\text{prepro}}(d)$ over d
$v_{\text{prepro}}(d)$	Preprocessed longitudinal driver speed profile over d

Greek Letters

Symbol	Description
α	Stretch factor used in the SPAA
α_0	Initial stretch factor
α_{curve}	Stretch factor limited due to high road curvature
$\alpha_{\max}$	Maximum stretch factor limited by $T_{\max}$
$\tau_{\square}^{\mathrm{high}}$	Upper IR threshold, defining the minimum IR for a $\square$ adjustment
$\tau_{\square}^{\mathrm{low}}$	Lower IR threshold, defining the maximum IR for no $\square$ adjustment

Calligraphic and Blackboard Bold

Symbol	Description
$\mathcal{D}$	Set of driver indices
$\mathcal{L}$	Set of location indices

Indices, Exponents, and Operators

Symbol	Description
$\arg\min$	Argument at minimum
$\max$	Maximum or maximization
$\min$	Minimum or minimization
$\bar{\square}$	Arithmetic mean of $\square$

List of Figures

List of Tables

References

Public References

[ABH07] ALKIM, T. P.; BOOTSMA, G.; HOOGENDOORN, S. P.: Field Operational Test "The Assisted Driver". In: *2007 IEEE Intelligent Vehicles Symposium*. 2007, pp. 1198–1203.

[Ade09] ADELL, E.: Acceptance of driver support systems. In: *Proceedings of the European Conference on Human Centred Design for Intelligent Transport Systems* (Jan. 2009).

[Ajz85] AJZEN, I.: 'From Intentions to Actions: A Theory of Planned Behavior'. In: *Action Control: From Cognition to Behavior*. Ed. by KUHL, J.; BECKMANN, J. Berlin, Heidelberg: Springer Berlin Heidelberg, 1985, pp. 11–39.

[AN04] ABBEEL, P.; NG, A.: Apprenticeship Learning via Inverse Reinforcement Learning. In: *Proceedings, Twenty-First International Conference on Machine Learning, ICML 2004* (Sept. 2004).

[AT05] ADOMAVICIUS, G.; TUZHILIN, A.: Personalization technologies: a process-oriented perspective. In: *Communications of the ACM* 48.10 (2005), pp. 83–90.

[Bag+18] BAGNALL, A. J.; DAU, H. A.; LINES, J.; FLYNN, M.; LARGE, J.; BOSTROM, A.; SOUTHAM, P.; KEOGH, E. J.: The UEA multivariate time series classification archive, 2018. In: *CoRR* abs/1811.00075 (2018). arXiv: 1811.00075.

[Bag+20] BAGNALL, A.; FLYNN, M.; LARGE, J.; LINES, J.; MIDDLEHURST, M.: On the Usage and Performance of the Hierarchical Vote Collective of Transformation-Based Ensembles Version 1.0 (HIVE-COTE v1.0). In: *International Workshop on Advanced Analytics and Learning on Temporal Data* 5 (2020), pp. 3–18.

[Bas+17] BASU, C.; YANG, Q.; HUNGERMAN, D.; SINGHAL, M.; DRAGAN, A. D.: Do You Want Your Autonomous Car To Drive Like You? In: *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. HRI '17. ACM, Mar. 2017.

[BBT24] BECKER, S.; BRANDENBURG, S.; THÜRING, M.: Driver-initiated takeovers during critical evasion maneuvers in automated driving. In: *Accident Analysis & Prevention* 194 (2024), p. 107362.

[Bel+17] BELLEM, H.; KLÜVER, M.; SCHRAUF, M.; SCHÖNER, H.-P.; HECHT, H.; KREMS, J. F.: Can We Study Autonomous Driving Comfort in Moving-Base Driving Simulators? A Validation Study. In: *Human Factors* 59.3 (2017), pp. 442–456.

[Bel+18] BELLEM, H.; THIEL, B.; SCHRAUF, M.; KREMS, J. F.: Comfort in automated driving: An analysis of preferences for different automated driving styles and their dependence on personality traits. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 55 (2018), pp. 90–100.

[BGJ19] BOLDUC, A. P.; GUO, L.; JIA, Y.: Multimodel Approach to Personalized Autonomous Adaptive Cruise Control. In: *IEEE Transactions on Intelligent Vehicles* 4.2 (2019), pp. 321–330.

[Bi+20] BI, J.; DHIMAN, V.; XIAO, T.; XU, C.: Learning from Interventions Using Hierarchical Policies for Safe Learning. In: *Proceedings of the AAAI Conference on Artificial Intelligence* 34 (Apr. 2020), pp. 10352–10360.

[BK13] BEGGIATO, M.; KREMS, J. F.: The evolution of mental model, trust and acceptance of adaptive cruise control in relation to initial information. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 18 (2013), pp. 47–57.

[Bla82] BLAAUW, G. J.: Driving Experience and Task Demands in Simulator and Instrumented Car: A Validation Study. In: *Human Factors* 24.4 (1982), pp. 473–486.

[BMW] BMW AG: *BMW Advanced Driver Assistance Systems Brochure*. URL: https://www.bmwusa.com/content/dam/bmw/marketUS/common/explore/driver-assistance-safety-features/pdf/BMW_ADAS_Reference_Guide.pdf (visited on 31/01/2026).

[BRK17] BRAUNAGEL, C.; ROSENSTIEL, W.; KASNECI, E.: Ready for Take-Over? A New Driver Assistance System for an Automated Classification of Driver Take-Over Readiness. In: *IEEE Intelligent Transportation Systems Magazine* 9.4 (2017), pp. 10–22.

[BS95] BAIN, M.; SAMMUT, C.: A Framework for Behavioural Cloning. In: *Machine Intelligence 15*. 1995, pp. 103–129.

[BSDP08] BIFULCO, G. N.; SIMONELLI, F.; DI PACE, R.: Experiments toward an human-like Adaptive Cruise Control. In: *2008 IEEE Intelligent Vehicles Symposium*. 2008, pp. 919–924.

[Cad] CADILLAC MOTOR CAR DIVISION: *SUPER CRUISE Convenience & Personalization Guide*. URL: https://www.cadillac.com/content/dam/cadillac/na/us/english/index/ownership/technology/supercruise/pdfs/2021-cadillac-super-cruise-reference-guide.pdf (visited on 23/05/2025).

[Cel+22] Celemin, C.; Pérez-Dattari, R.; Chisari, E.; Franzese, G.; Souza Rosa, L. de; Prakash, R.; Ajanović, Z.; Ferraz, M.; Valada, A.; Kober, J.: Interactive Imitation Learning in Robotics: A Survey. In: *Foundations and Trends® in Robotics* 10.1-2 (2022), pp. 1–197.

[Che+17] Chen, X.; Zhai, Y.; Lu, C.; Gong, J.; Wang, G.: A learning model for personalized adaptive cruise control. In: *2017 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, June 2017.

[Chi+22] Chisari, E.; Welschehold, T.; Boedecker, J.; Burgard, W.; Valada, A.: Correct Me If I am Wrong: Interactive Learning for Robotic Manipulation. In: *IEEE Robotics and Automation Letters* 7.2 (2022), pp. 3695–3702.

[Chu+23] Chu, H.; Zhuang, H.; Wang, W.; Na, X.; Guo, L.; Zhang, J.; Gao, B.; Chen, H.: A Review of Driving Style Recognition Methods From Short-Term and Long-Term Perspectives. In: *IEEE Transactions on Intelligent Vehicles* 8.11 (2023), pp. 4599–4612.

[Cle+22] Clement, P.; Veledar, O.; Könczöl, C.; Danzinger, H.; Posch, M.; Eichberger, A.; Macher, G.: Enhancing Acceptance and Trust in Automated Driving trough Virtual Experience on a Driving Simulator. In: *Energies* 15.3 (2022).

[CMM02] Canale, M.; Malan, S.; Murdocco, V.: Personalization of ACC Stop and Go Task based on Human Driver Behaviour Analysis. In: *IFAC Proceedings Volumes* 35.1 (2002), pp. 357–362.

[CO19] Cavalin, P.; Oliveira, L.: Confusion Matrix-Based Building of Hierarchical Classification. In: *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*. Ed. by Vera-Rodriguez, R.; Fierrez, J.; Morales, A. Cham: Springer International Publishing, 2019, pp. 271–278.

[Dav89] Davis, F. D.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. In: *MIS Quarterly* 13.3 (1989), pp. 319–340.

[Dha+20] Dhariyal, B.; Le Nguyen, T.; Gsponer, S.; Ifrim, G.: An Examination of the State-of-the-Art for Multivariate Time Series Classification. In: *2020 International Conference on Data Mining Workshops (ICDMW)*. 2020, pp. 243–250.

[DPW20] Dempster, A.; Petitjean, F.; Webb, G. I.: ROCKET: Exceptionally fast and accurate time series classification using random convolutional kernels. In: *Data Mining and Knowledge Discovery* 34 (2020), 1454–1495.

[Dr.] Dr. Ing. h.c. F. Porsche AG: *Porsche Taycan Owner's Manual - Porsche InnoDrive (PID)*. URL: https://manual.porsche.com/#/wpi/Y1A-2023-20220620-37-RDW-1.0.3/en_GB/topic/id/fec83933bf8c152f0a0114026be3b773 (visited on 12/05/2025).

[EBS17] Eriksson, A.; Banks, V.; Stanton, N.: Transition to manual: Comparing simulator with on-road control transitions. In: *Accident Analysis & Prevention* 102 (2017), pp. 227–234.

[ES17] Eriksson, A.; Stanton, N. A.: Takeover Time in Highly Automated Vehicles: Noncritical Transitions to and From Manual Control. In: *Human Factors* 59.4 (2017), pp. 689–705.

[Faw+20a] Fawaz, H. I.; Lucas, B.; Forestier, G.; Pelletier, C.; Schmidt, D. F.; Weber, J.; Webb, G. I.; Idoumghar, L.; Muller, P.-A.; Petitjean, F.: InceptionTime: Finding AlexNet for time series classification. In: *Data Mining and Knowledge Discovery* 34 (2020), 1936–1962.

[Faw+20b] Fawaz, H. I.; Lucas, B.; Forestier, G.; Pelletier, C.; Schmidt, D. F.; Weber, J.; Webb, G. I.; Idoumghar, L.; Muller, P.-A.; Petitjean, F.: *InceptionTime (GitHub repository)*. 2020. URL: https://github.com/hfawaz/InceptionTime (visited on 05/08/2025).

[FP06] Fan, H.; Poole, M.: What Is Personalization? Perspectives on the Design and Implementation of Personalization in Information Systems. In: *Journal of Organizational Computing and Electronic Commerce - J ORGAN COMPUT ELECTRON COMME* 16 (Jan. 2006), pp. 179–202.

[Fri+19] Fridman, L.; Brown, D. E.; Glazer, M.; Angell, W.; Dodd, S.; Jenik, B.; Terwilliger, J.; Patsekin, A.; Kindelsberger, J.; Ding, L.; Seaman, S.; Mehler, A.; Sipperley, A.; Pettinato, A.; Seppelt, B. D.; Angell, L.; Mehler, B.; Reimer, B.: MIT Advanced Vehicle Technology Study: Large-Scale Naturalistic Driving Study of Driver Behavior and Interaction With Automation. In: *IEEE Access* 7 (2019), pp. 102021–102038.

[Gao+24] Gao, J.; Yu, B.; Chen, Y.; Bao, S.; Gao, K.; Zhang, L.: An ADAS with better driver satisfaction under rear-end near-crash scenarios: A spatio-temporal graph transformer-based prediction framework of evasive behavior and collision risk. In: *Transportation Research Part C: Emerging Technologies* 159 (2024), p. 104491.

[Gas+17] Gasser, T. M.; Frey, A. T.; Seeck, A.; Auerswald, R.: Comprehensive definitions for automated driving and ADAS. In: *25th International Technical Conference on the Enhanced Safety of Vehicles (ESV) National Highway Traffic Safety Administration*. 2017.

[Gel+16] Gelder, E. de; Cara, I.; Uittenbogaard, J.; Kroon, L.; Iersel, S. van; Hogema, J.: toward personalised automated driving: Prediction of preferred ACC behaviour based on manual driving. In: *2016 IEEE Intelligent Vehicles Symposium (IV)*. 2016, pp. 1211–1216.

[Ger+21] Gershon, P.; Seaman, S.; Mehler, B.; Reimer, B.; Coughlin, J.: Driver behavior and the use of automation in real-world driving. In: *Accident Analysis & Prevention* 158 (2021), p. 106217.

[GJ23] GUO, L.; JIA, Y.: Bilateral Adaptation of Longitudinal Control of Automated Vehicles and Human Drivers. In: *IEEE Transactions on Intelligent Transportation Systems* 24.5 (2023), pp. 5663–5671.

[Goe+19] GOECKS, V. G.; GREMILLION, G. M.; LAWHERN, V. J.; VALASEK, J.; WAYTOWICH, N. R.: Efficiently Combining Human Demonstrations and Interventions for Safe Training of Autonomous Systems in Real-Time. In: *Proceedings of the AAAI Conference on Artificial Intelligence* 33.01 (2019), pp. 2462–2470.

[Gri+16] GRIESCHE, S.; NICOLAY, E.; ASSMANN, D.; DOTZAUER, M.; KÄTHNER, D.: Should my car drive as I do? What kind of driving style do drivers prefer for the design of automated driving functions? In: *Proceedings of the 17th Braunschweiger Symposium AAET*. Feb. 2016.

[GSC02] GODBOLE, S.; SARAWAGI, S.; CHAKRABARTI, S.: Scaling multi-class Support Vector Machines using inter-class confusion. In: (Aug. 2002).

[Ham94] HAMILTON, J. D.: *Time Series Analysis*. Princeton University Press, 1994.

[Har+22] HARTMANN, A.; CYBERSKI, C.; SCHÖNFELD, U.; KRZOK, W.; MÜLLER, S.: Effect of Horizontal Acceleration and Seat Orientation on Motion Sickness in Passenger Cars. In: *Proceedings of the 15th International Symposium on Advanced Vehicle Control (AVEC'22)*. Sept. 2022.

[HHW20] HASENJAGER, M.; HECKMANN, M.; WERSING, H.: A Survey of Personalization for Advanced Driver Assistance Systems. In: *IEEE Transactions on Intelligent Vehicles* 5.2 (June 2020), pp. 335–344.

[Him+23] HIMMELS, C.; PARDUZI, A.; FISCHER, M.; RIENER, A.: toward a Common Understanding of Driving Simulator Validity. In: Sept. 2023, pp. 191–196.

[Him+25] HIMMELS, C.; BUCHNER, C.; SCHMITZ, J.; PARDUZI, A.; RIENER, A.: Validity of Driver Assistance Systems in Driving Simulators: A Comparative Study of Real-World Driving and Two Simulator Environments. In: *International Journal of Human-Computer Interaction* (May 2025), pp. 1–18.

[Hoc01] HOC, J.-M.: toward a cognitive approach to human-machine cooperation in dynamic situations. In: *International Journal of Human-Computer Studies* 54.4 (2001), pp. 509–540.

[Hoc+18] HOCK, P.; KRAUS, J.; BABEL, F.; WALCH, M.; RUKZIO, E.; BAUMANN, M.: How to Design Valid Simulator Studies for Investigating User Experience in Automated Driving: Review and Hands-On Considerations. In: *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. AutomotiveUI '18. Toronto, ON, Canada: Association for Computing Machinery, 2018, pp. 105–117.

[IMW20] ITTNER, S.; MÜHLBACHER, D.; WEISSWANGE, T. H.: The Discomfort of Riding Shotgun - Why Many People Don't Like to Be Co-driver. In: *Frontiers in Psychology* Volume 11 - 2020 (2020).

[Itt+23] ITTNER, S.; MÜHLBACHER, D.; VOLLRATH, M.; WEISSWANGE, T. H.: Development and evaluation of passenger assistance system concepts to reduce passenger discomfort. In: *Frontiers in Psychology* Volume 14 - 2023 (2023).

[Kar+17] KARIM, F.; MAJUMDAR, S.; DARABI, H.; CHEN, S.: LSTM Fully Convolutional Networks for Time Series Classification. In: *IEEE Access* 6 (2017), pp. 1662–1669.

[Kar+19] KARIM, F.; MAJUMDAR, S.; DARABI, H.; HARFORD, S.: Multivariate LSTM-FCNs for time series classification. In: *Neural Networks* 116 (2019), pp. 237–245.

[KBB18] KÖRBER, M.; BASELER, E.; BENGLER, K.: Introduction matters: Manipulating trust in automation and reliance in automated driving. In: *Applied Ergonomics* 66 (2018), pp. 18–31.

[Kel+19] KELLY, M.; SIDRANE, C.; DRIGGS-CAMPBELL, K.; KOCHENDERFER, M. J.: HG-DAgger: Interactive Imitation Learning with Human Experts. In: *2019 International Conference on Robotics and Automation (ICRA)*. 2019, pp. 8077–8083.

[KGB15] KUDERER, M.; GULATI, S.; BURGARD, W.: Learning driving styles for autonomous vehicles from demonstration. In: *2015 IEEE International Conference on Robotics and Automation (ICRA)*. 2015, pp. 2641–2646.

[KMS20] KIM, H.; MIN, K.; SUNWOO, M.: Driver Characteristics Oriented Autonomous Longitudinal Driving System in Car-Following Situation. In: *Sensors* 20.21 (2020).

[KYB21] KRAUSE, M.; YILMAZ, L.; BENGLER, K.: Comparison of Real and Simulated Driving for a Static Driving Simulator. In: *International Conference on Applied Human Factors and Ergonomics (AHFE)*. Jan. 2021.

[Lar+15] LARUE, G. S.; RAKOTONIRAINY, A.; HAWORTH, N. L.; DARVELL, M.: Assessing driver acceptance of Intelligent Transport Systems in the context of railway level crossings. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 30 (2015), pp. 1–13.

[Las+17] LASKEY, M.; CHUCK, C.; LEE, J.; MAHLER, J.; KRISHNAN, S.; JAMIESON, K.; DRAGAN, A.; GOLDBERG, K.: Comparing human-centric and robot-centric sampling for robot deep learning from demonstrations. In: *2017 IEEE International Conference on Robotics and Automation (ICRA)*. 2017, pp. 358–365.

[LCB16] LEFEVRE, S.; CARVALHO, A.; BORRELLI, F.: A Learning-Based Framework for Velocity Control in Autonomous Driving. In: *IEEE Transactions on Automation Science and Engineering* 13.1 (Jan. 2016), pp. 32–42.

[Lee+20] Lee, C. W.; Nayeer, N.; Garcia, D. E.; Agrawal, A.; Liu, B.: Identifying the Operational Design Domain for an Automated Driving System through Assessed Risk. In: *2020 IEEE Intelligent Vehicles Symposium (IV)*. 2020, pp. 1317–1322.

[Lee+21] Lee, J. D.; Liu, S.-Y.; Domeyer, J.; DinparastDjadid, A.: Assessing Drivers' Trust of Automated Vehicle Driving Styles With a Two-Part Mixed Model of Intervention Tendency and Magnitude. In: *Human Factors* 63.2 (2021), pp. 197–209.

[Lik32] Likert, R.: A technique for the measurement of attitudes. In: *Archives of Psychology* 22.140 (1932), p. 55.

[LS04] Lee, J. D.; See, K. A.: Trust in Automation: Designing for Appropriate Reliance. In: *Human Factors* 46.1 (2004), pp. 50–80.

[LTB16] Lines, J.; Taylor, S.; Bagnall, A.: HIVE-COTE: The Hierarchical Vote Collective of Transformation-Based Ensembles for Time Series Classification. In: *2016 IEEE 16th International Conference on Data Mining (ICDM)*. 2016, pp. 1041–1046.

[Lu+16] Lu, Z.; Happee, R.; Cabrall, C. D.; Kyriakidis, M.; de Winter, J. C.: Human factors of transitions in automated driving: A general framework and literature survey. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 43 (2016), pp. 183–198.

[MA+20] Morales-Alvarez, W.; Sipele, O.; Léberon, R.; Tadjine, H. H.; Olaverri-Monreal, C.: Automated Driving: A Literature Review of the Take over Request in Conditional Automation. In: *Electronics* 9.12 (2020).

[Mag+22] Maggi, D.; Romano, R.; Carsten, O.; Winter, J. C. D.: When terminology hinders research: the colloquialisms of transitions of control in automated driving. In: *Cognition, Technology & Work* 24.3 (2022), pp. 509–520.

[Man+20] Mandlekar, A.; Xu, D.; Martin-Martin, R.; Zhu, Y.; Fei-Fei, L.; Savarese, S.: *Human-in-the-Loop Imitation Learning using Remote Teleoperation*. Dec. 2020.

[Mar] Market Growth Reports: *Advanced Driver Assistance Systems (ADAS) Market Size, Share, Growth, Trends, and Industry Analysis, By Type (Adaptive Cruise Control (ACC), Lane Departure Warning (LDW) System, Park Assist, Blind Spot Detection, Others), By Application (Passenger Car, Light Commercial Vehicle (LCV), Heavy Commercial Vehicle (HCV)), Regional Insights and Forecast From 2025 To 2033*. url: https://www.marketgrowthreports.com/market-reports/advanced-driver-assistance-systems-adas-market-100281 (visited on 08/09/2025).

[Mar+08] MARTENS, M.; PAUWELUSSEN, J.; SCHIEBEN, A.; FLEMISCH, F.; MERAT, N.; JAMSON, S.; CACI, R.: Human Factors' aspects in automated and semi-automatic transport systems: State of the art. In: (2008).

[Mar+20] MARCANO, M.; DÍAZ, S.; PÉREZ, J.; IRIGOYEN, E.: A Review of Shared Control for Automated Vehicles: Theory and Applications. In: *IEEE Transactions on Human-Machine Systems* 50.6 (2020), pp. 475–491.

[Mer] MERCEDES-BENZ AG: *Mercedes-Benz EQE Owner's Manual*. URL: https://www.mercedes-benz.co.uk/passengercars/owners-manuals/en-gb/pdf/mercedes-eqe-saloon-2025-september-v295-mbux-owners-manual-1.pdf (visited on 31/01/2026).

[Mic85] MICHON, J. A.: 'A Critical View of Driver Behavior Models: What Do We Know, What Should We Do?' In: *Human Behavior and Traffic Safety*. Springer US, 1985, pp. 485–524.

[Mor+20] MORANDO, A.; GERSHON, P.; MEHLER, B.; REIMER, B.: Driver-initiated Tesla Autopilot Disengagements in Naturalistic Driving. In: *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. AutomotiveUI '20. Virtual Event, DC, USA: Association for Computing Machinery, 2020, pp. 570–65.

[MS20] MECACCI, G.; SIO, F. SANTONI DE: Meaningful human control as reason-responsiveness: the case of dual-mode vehicles. In: *Ethics and Information Technology* (June 2020).

[Mul15] MULLER, M.: *Fundamentals of music processing*. 1st ed. Cham, Switzerland: Springer International Publishing, 2015.

[MZ21] MA, Z.; ZHANG, Y.: Drivers trust, acceptance, and takeover behaviors in fully automated vehicles: Effects of automated driving styles and driver's driving styles. In: *Accident Analysis & Prevention* 159 (Sept. 2021), p. 106238.

[NW23] NORDHOFF, S.; WINTER, J. DE: *Why do drivers and automation disengage the automation? Results from a study among Tesla users*. Mar. 2023.

[OF+21] OLIVEIRA FARIA, N. DE; MERENDA, C.; GREATBATCH, R.; TANOUS, K.; SUGA, C.; AKASH, K.; MISU, T.; GABBARD, J.: The Effect of Augmented Reality Cues on Glance Behavior and Driver-Initiated Takeover on SAE Level 2 Automated-Driving. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 65.1 (2021), pp. 1342–1346.

[Osa+18] OSA, T.; PAJARINEN, J.; NEUMANN, G.; BAGNELL, J.; ABBEEL, P.; PETERS, J.: An Algorithmic Perspective on Imitation Learning. In: *Foundations and Trends in Robotics* 7 (Nov. 2018), pp. 1–179.

[Oth21] OTHMAN, K.: Public acceptance and perception of autonomous vehicles: a comprehensive review. In: *AI and Ethics* 1 (Feb. 2021), pp. 355–387.

[PC19] PONOMAREV, A.; CHERNYSHEVA, A.: Adaptation and Personalization in Driver Assistance Systems. In: *2019 24th Conference of Open Innovations Association (FRUCT)*. 2019, pp. 335–344.

[Pen+24] PENG, C.; MERAT, N.; ROMANO, R.; HAJISEYEDJAVADI, F.; PASCHALIDIS, E.; WEI, C.; RADHAKRISHNAN, V.; SOLERNOU, A.; FORSTER, D.; BOER, E.: Drivers' evaluation of different automated driving styles: Is it both comfortable and natural? In: *Human Factors* 66.3 (2024), pp. 787–806.

[PF10] PAUWELUSSEN, J.; FEENSTRA, P. J.: Driver Behavior Analysis During ACC Activation and Deactivation in a Real Traffic Environment. In: *IEEE Transactions on Intelligent Transportation Systems* 11.2 (2010), pp. 329–338.

[PGR23] PNINA GERSHON, B. M.; REIMER, B.: Driver response and recovery following automation initiated disengagement in real-world hands-free driving. In: *Traffic Injury Prevention* 24.4 (2023), pp. 356–361.

[PM08] PAUWELUSSEN, J.; MINDERHOUD, M.: The effects of deactivation and (re)activation of ACC on driver behaviour analyzed in real traffic. In: *2008 IEEE Intelligent Vehicles Symposium*. 2008, pp. 257–262.

[Rah+17] RAHMAN, M. M.; LESCH, M. F.; HORREY, W. J.; STRAWDERMAN, L.: Assessing the utility of TAM, TPB, and UTAUT for advanced driver assistance systems. In: *Accident Analysis & Prevention* 108 (2017), pp. 361–373.

[Rah+18] RAHMAN, M. M.; STRAWDERMAN, L.; LESCH, M. F.; HORREY, W. J.; BABSKI-REEVES, K.; GARRISON, T.: Modelling driver acceptance of driver support systems. In: *Accident Analysis & Prevention* 121 (2018), pp. 134–147.

[Ram+17] RAMYAR, S.; HOMAIFAR, A.; SALAKEN, S. M.; NAHAVANDI, S.; KURT, A.: A personalized highway driving assistance system. In: *2017 IEEE Intelligent Vehicles Symposium (IV)*. 2017, pp. 1596–1601.

[Rei93] REINSEL, G. C.: *Elements of Multivariate Time Series Analysis*. Springer New York, 1993.

[RGB11] ROSS, S.; GORDON, G.; BAGNELL, D.: A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning. In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. Ed. by GORDON, G.; DUNSON, D.; DUDÍK, M. Vol. 15. Proceedings of Machine Learning Research. Fort Lauderdale, FL, USA: PMLR, 2011, pp. 627–635.

[Rie10] RIENER, A.: Simulating On-the-Road Behavior Using a Driving Simulator. In: *2010 Third International Conference on Advances in Computer-Human Interactions*. 2010, pp. 25–31.

[Rui+21] RUIZ, A. P.; FLYNN, M.; LARGE, J.; MIDDLEHURST, M.; BAGNALL, A.: The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances. In: *Data Mining and Knowledge Discovery* 35.2 (2021), pp. 401–449.

[Ryd+23] RYDSTRÖM, A.; MULLAART, M. S.; NOVAKAZI, F.; JOHANSSON, M.; ERIKSSON, A.: Drivers' Performance in Non-critical Take-Overs From an Automated Driving System—An On-Road Study. In: *Human Factors* 65.8 (2023), pp. 1841–1857.

[Sae] *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*. Society of Automotive Engineers, 2021.

[Sag+15] SAGBERG, F.; SELPI; BIANCHI PICCININI, G. F.; ENGSTRÖM, J.: A Review of Research on Driving Styles and Road Safety. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 57.7 (June 2015), pp. 1248–1275.

[Sal+22] SALEHINEJAD, H.; WANG, Y.; YU, Y.; JIN, T.; VALAEE, S.: S-Rocket: Selective random convolution kernels for time series classification. In: (2022).

[SB+18] SADEGHIAN BOROJENI, S.; BOLL, S.; HEUTEN, W.; BÜLTHOFF, H.; CHUANG, L.: Feel the Movement: Real Motion Influences Responses to Take-over Requests in Highly Automated Vehicles. In: Apr. 2018, pp. 1–13.

[SC07] SALVADOR, S.; CHAN, P.: Toward accurate dynamic time warping in linear time and space. In: *Intelligent Data Analysis* 11.5 (2007), pp. 561–580.

[SC78] SAKOE, H.; CHIBA, S.: Dynamic programming algorithm optimization for spoken word recognition. In: *IEEE Transactions on Acoustics, Speech, and Signal Processings* 26 (1978), pp. 43–49.

[Sch12] SCHAAP, N.: *Driving behaviour in unexpected situations. A study into drivers' compensation behaviour to safety-ciritical situations and the effects of mental workload, event urgency and task prioritization*, phdthesis. Feb. 2012.

[SL18] SCHÄFER, P.; LESER, U.: Multivariate Time Series Classification with WEASEL+MUSE. In: *ECML/PKDD Workshop on Advanced Analytics and Learning on Temporal Data (AALTD18)* (2018), pp. 1–11.

[Spe+22] SPENCER, J.; CHOUDHURY, S.; BARNES, M.; SCHMITTLE, M.; CHIANG, M.; RAMADGE, P.; SRINIVASA, S.: Expert Intervention Learning: An online framework for robot learning from explicit and implicit human feedback. In: *Autonomous Robots* 46 (Jan. 2022).

[SVB78] SHERIDAN, T. B.; VERPLANK, W. L.; BROOKS, T. L.: *Human and Computer Control of Undersea Teleoperators*. Technical Report MIT-SG-79-20 (Sea Grant Report). Cambridge, MA: Massachusetts Institute of Technology, Man-Machine Systems Laboratory, 1978.

[SYWK15] Shokoohi-Yekta, M.; Wang, J.; Keogh, E.: 'On the Non-Trivial Generalization of Dynamic Time Warping to the Multi-Dimensional Case'. In: *Proceedings of the 2015 SIAM International Conference on Data Mining (SDM)*. 2015, pp. 289–297.

[Sze+17] Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.: Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. In: *Proceedings of the AAAI Conference on Artificial Intelligence* 31.1 (2017).

[TBAMG04] Taubman-Ben-Ari, O.; Mikulincer, M.; Gillath, O.: The multidimensional driving style inventory—scale construct and validation. In: *Accident Analysis & Prevention* 36.3 (May 2004), pp. 323–332.

[Tof+09] Toffetti, A.; Wilschut, E. S.; Martens, M. H.; Schieben, A.; Rambaldini, A.; Merat, N.; Flemisch, F.: CityMobil: Human Factor Issues Regarding Highly Automated Vehicles on eLane. In: *Transportation Research Record* 2110.1 (2009), pp. 1–8.

[Tuz09] Tuzhilin, A.: Personalization: The state of the art and future directions. In: *Business computing* 3.3 (2009), pp. 3–43.

[Var+15] Varotto, S. F.; Hoogendoorn, R.; Arem, B.; Hoogendoorn, S.: Empirical Longitudinal Driving Behavior in Authority Transitions Between Adaptive Cruise Control and Manual Driving. In: *Transportation Research Record Journal of the Transportation Research Board* 2489 (Jan. 2015), pp. 105–114.

[Vas+23] Vasile, L.; Seitz, B.; Staab, V.; Liebherr, M.; Däsch, C.; Schramm, D.: Influences of Personal Driving Styles and Experienced System Characteristics on Driving Style Preferences in Automated Driving. In: *Applied Sciences* 13.15 (July 2023), p. 8855.

[Ven+03] Venkatesh, V.; Morris, M. G.; Davis, G. B.; Davis, F. D.: User Acceptance of Information Technology: Toward a Unified View. In: *MIS Quarterly* 27.3 (2003), pp. 425–478.

[VHD97] Van Der Laan, J. D.; Heino, A.; De Waard, D.: A simple procedure for the assessment of acceptance of advanced transport telematics. In: *Transportation Research Part C: Emerging Technologies* 5.1 (1997), pp. 1–10.

[VLZ14] Vaitkus, V.; Lengvenis, P.; Zylius, G.: Driving style classification using long-term accelerometer information. In: *2014 19th International Conference on Methods and Models in Automation and Robotics (MMAR)*. 2014, pp. 641–644.

[Wan+13] Wang, J.; Zhang, L.; Zhang, D.; Li, K.: An Adaptive Longitudinal Driving Assistance System Based on Driver Characteristics. In: *IEEE Transactions on Intelligent Transportation Systems* 14.1 (Mar. 2013), pp. 1–12.

[WBS19] Wynne, R. A.; Beanland, V.; Salmon, P. M.: Systematic review of driving simulator validation studies. In: *Safety Science* 117 (2019), pp. 138–151.

[Win+15] WINNER, H.; HAKULI, S.; LOTZ, F.; SINGER, C.: *Handbuch Fahrerassistenzsysteme [Handbook Driver Assistance Systems]*. Springer, 2015.

[Win+23] WINTERSBERGER, P.; SCHARTMÜLLER, C.; SADEGHIAN, S.; FRISON, A.-K.; RIENER, A.: Evaluation of Imminent Take-Over Requests With Real Automation on a Test Track. In: *Human Factors* 65.8 (2023), pp. 1776–1792.

[WK] WAARD, D. DE; KONDZIOR, M.: *Akzeptanzskala: Methode zur Erfassung der Akzeptanz eines Systems*. URL: https://www.hfes-europe.org/accept/accept_de.htm (visited on 29/08/2025).

[WPA23] WINTER, J. C. F. DE; PETERMEIJER, S. M.; ABBINK, D. A.: Shared control versus traded control in driving: a debate around automation pitfalls. In: *Ergonomics* 66.10 (2023), pp. 1494–1520.

[WXC14] WANG, W.; XI, J.; CHEN, H.: Modeling and Recognizing Driver Behavior Based on Driving Data: A Survey. In: *Mathematical Problems in Engineering* 2014 (2014), pp. 1–20.

[Yan] YANO RESEARCH INSTITUTE LTD.: *Global Autonomous Driving Systems Market: Key Research Findings 2024*. URL: https://www.yanoresearch.com/en/press-release/show/press_id/3693 (visited on 08/09/2025).

[Yan+23] YANG, S.; MCKERRAL, A.; MULHALL, M. D.; LENNÉ, M. G.; REIMER, B.; GERSHON, P.: Takeover Context Matters: Characterising Context of Takeovers in Naturalistic Driving using Super Cruise and Autopilot. In: *Proceedings of the 15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. AutomotiveUI '23. Ingolstadt, Germany: Association for Computing Machinery, 2023, pp. 112–122.

[Yi+20] YI, D.; SU, J.; HU, L.; LIU, C.; QUDDUS, M.; DIANATI, M.; CHEN, W.-H.: Implicit Personalization in Driving Assistance: State-of-the-Art and Open Issues. In: *IEEE Transactions on Intelligent Vehicles* 5.3 (2020), pp. 397–413.

[Yus+16] YUSOF, N. M.; KARJANTO, J.; TERKEN, J.; DELBRESSINE, F.; HASSAN, M. Z.; RAUTERBERG, M.: The Exploration of Autonomous Vehicle Driving Styles: Preferred Longitudinal, Lateral, and Vertical Accelerations. In: *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. AutomotiveUI'16. ACM, Oct. 2016.

[Zö+15] ZÖLLER, I.; BETZ, A.; MAUTES, N.; SCHOLZ, L.; ABENDROTH, B.; BRUDER, R.; WINNER, H.: Valid representation of a highly dynamic collision avoidance scenario in a driving simulator. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 31 (2015), pp. 54–66.

Own Publications and Conference Proceedings

[Sch+24a] **SCHWAGER, R.**; GRIMM, M.; LIU, X.; EWECKER, L.; BRÜHL, T.; SOHN, T. S.; HOHMANN, S.: An Analysis of Driver-Initiated Takeovers during

Assisted Driving and their Effect on Driver Satisfaction. In: *2024 IEEE Intelligent Vehicles Symposium (IV)*. 2024, pp. 1907–1914.

[Sch+24b] **SCHWAGER, R.**; SVARAL, M.; GRIMM, M.; HARTMANN, S.; EWECKER, L.; BRÜHL, T.; SOHN, T. S.; HOHMANN, S.: Adjustment of an Assisted Driving Function Based on Driver-Initiated Takeovers. In: *2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC)*. 2024, pp. 3844–3851.

[Sch+26a] **SCHWAGER, R.**; ANASTASIO, A.; HARTMANN, S.; RONELLENFITSCH, A.; GRIMM, M.; BRÜHL, T.; SOHN, T. S.; EBERHARDT, T. D.; HOHMANN, S.: Improving Driver Satisfaction with a Driving Function Learning from Implicit Human Feedback - a Test Group Study. *arXiv preprint: https://doi.org/10.48550/arXiv.2602.13733*. 2026.

[Sch+26b] **SCHWAGER, R.**; SCHICK, L.; SVARAL, M.; GRIMM, M.; BRÜHL, T.; SOHN, T. S.; EBERHARDT, T. D.; HOHMANN, S.: Automatic Classification of Longitudinal Driver-Initiated Takeovers during Assisted Driving. In: *2026 IEEE Intelligent Vehicles Symposium (IV)*. 2026. In press.

Supervised Student Theses

[Dal23] DALKE, L. M.: *Entwicklung eines Frameworks zur Visualisierung und Analyse mehrfach befahrener Strecken aus cloudbasierten Erprobungsdaten*, Bachelor's thesis. Faculty of Economics, Baden-Wuerttemberg Cooperative State University (DHBW), 2023.

[Liu24] LIU, X.: *Analysis of Driver Interventions on Assisted Driving Functions*, Diploma thesis. Faculty of Transport and Traffic Sciences, Dresden University of Technology (TUD), 2024.

[Que24] QUERFURTH, L.: *Conceptual Design of a Method for Adaptation and Individualization of Target Trajectories Based on Driver Feedback using Machine Learning Methods*, Master's thesis. Institute of Vehicle System Technology, Karlsruhe Institute of Technology (KIT), 2024.

[Sch24] SCHICK, L.: *Automatic Classification of Short-term Driver-initiated Takeovers during Assisted Driving*, Master's thesis. Faculty of Economics and Social Sciences, University of Tuebingen, 2024.