Deliverable **D7.3** /

Pilot Evaluation Results

Version: 1.0 Final

Dissemination level: PU

Lead contractor: ika

Due date: 31.10.2021

Version date: 29.09.2021

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 723051.

Document information

Authors

Hendrik Weber – ika
Johannes Hiller – ika
Lutz Eckstein – ika
Barbara Metz – WIVW GmbH
Andreas Landau – WIVW GmbH
Yee Mun Lee – University of Leeds
Tyron Louw – University of Leeds
Ruth Madigan – University of Leeds
Natasha Merat – University Leeds
Esko Lehtonen – VTT
Henri Sintonen – VTT
Satu Innamaa – VTT
Thomas Streubel – Chalmers
Linda Pipkorn – Chalmers
Erik Svanberg – SAFER
Marijke van Weperen – TNO
Jeroen Hogema – TNO
Anastasia Bolovinou – ICCS
Anastasios Rigos – ICCS
Marek Junghans – DLR
Meng Zhang – DLR
Juan Pastor Trullos – DLR
Alexander Zerbe – BAST
Roland Schindhelm – BAST
Yves Page – Renault
Walter Hagleitner – ADAS
Adrian Zlocki – fka

Coordinator

Aria Etemad
Volkswagen Group Innovation
Hermann-Münch-Str. 1
38440 Wolfsburg
Germany

Phone: +49-5361-9-13654
Email: aria.etemad@volkswagen.de

Project funding

Horizon 2020
ART-02-2016 – Automation pilots for passenger cars
Contract number 723051
www.L3Pilot.eu

Legal Disclaimer

Table of contents

List of Figures

List of Tables

Executive Summary

This deliverable presents the results of the Technical and Traffic Evaluation, and the User and Acceptance Evaluation of the L3Pilot project. For these evaluations, vehicle and questionnaire data were collected during the piloting operations of automated driving functions (ADFs) in L3Pilot. Results were generated separately for different types of ADF distinguished by three driving domains: *motorway*, *urban,* and *parking*.

The goal of the evaluation was to derive the impact of SAE Level 3 ADF on various aspects related to the vehicle's behaviour (i.e., own driving behaviour and interaction behaviour with other road users) in traffic, as well as aspects related to the user's attitude towards the systems. Thus, it was decided that data should not be evaluated based on individual implementations, but on a merged dataset consisting of the data from all Pilot sites which perform harmonised piloting operations. To answer the research questions, data was collected from trips where the ADF could be activated and from baseline trips in which the ADF were switched off. However, the same setup was used for the trips active systems as well as for the baseline data. Given that the piloted vehicles were still prototypical systems, in most cases they had to be driven by professional drivers who could pre-empt safety-critical situations.

For the Technical and Traffic Evaluation, vehicle data was collected at 14 pilot sites and then provided to the research partners within the project. At each test site, raw data was transformed into the Common Data Format developed by L3Pilot. Research partners analysed the driving data to derive data-based answers to the research questions. For this, driving data was segmented into instances of driving scenarios such as *Uninfluenced Driving*, *Following*, *Lane Change* or *Cut-In*. For each instance of a driving scenario detected in the data, specific performance indicators were derived by an automated toolchain. Handling the data in a common format made it possible to utilise the same data processing toolchain for all pilot sites, thus ensuring that the calculation of performance indicators was identical for each Pilot site. For each individual pilot site, the derived performance indicators were uploaded to the consolidated database towards answering the research questions. To avoid any benchmarking between the systems by the different manufacturers, the database was set up in a way that it did not allow to identify which pilot site contributed which entries in the database. Only the partners providing the data could identify and, if necessary, modify the data they supplied. Parking data were evaluated individually per study and the results were combined.

For the User and Acceptance Evaluation, harmonised questionnaires have been used which were translated into the required languages for the different pilot sites and managed using an online tool. The answers of participants were coded in a dedicated format and uploaded to a shared database, which was anonymised similarly to the vehicle data. Evaluation of user data was carried out by type of function and type of driver operating it:

- Urban (professional drivers operating the vehicle and ordinary drivers as passengers)
- Motorway with professional drivers

- Motorway with non-professional drivers (including Wizard-of-Oz operations)

Besides the questionnaire, videos of the vehicle and the driver inside were analysed to assess their interaction with the ADF. For the analysis of take-over requests a video-based coding was applied.

For **motorway ADF,** 2175 h of driving data within the operational design domain of the ADF and 294 questionnaires (58 professional driver, 236 non-professional) were analysed, from which several main findings regarding technical and user-related research questions can be reported:

- Automated vehicles drove at slower speeds compared to baseline across all evaluated scenarios.
- While in automated mode, the vehicles kept significantly larger headway distances compared to baseline.
- The lane keeping behaviour of the automated vehicles was found to be more stable than in baseline.
- Automated vehicles spent more time in stable driving scenarios such as uninfluenced driving or following. However, this may partially be linked to the prototypical state of the evaluated ADF.
- Drivers were generally positive about the ADF.
- No motion sickness was reported, and drivers found the systems comfortable.
- Professional drivers tended to be less positive about the system. However, system familiarity and driver type did not affect willingness to use the system.
- In more than 60% of take-over situations it took less than 4 seconds before drivers reacted to the take-over requests and deactivated the function. The reaction time in 99% of situations was under 10 seconds. All the take-over requests happened in everyday driving situations.

For **urban ADF,** the quantity of piloting data (1120 h) was small compared to the amount of data needed to cover the large variability of interactions at intersections, which made it difficult to evaluate the impact of automated driving at intersections. The main findings for urban ADF include:

- For intersection scenarios, automated vehicles spent more time travelling through the intersection, suggesting a more careful behaviour while passing through or turning at an intersection.
- Urban vehicles did not drive above the speed limit, whereas drivers in baseline did slightly.
- Behaviour while following a lead vehicle did not differ significantly from the behaviour of human drivers in baseline, suggesting that automated vehicles would not interfere with the flow of traffic in urban areas.
- In general, driving dynamics—both longitudinal and lateral accelerations—could be reduced while driving with an active ADF.

- Users were generally positive about the urban system, but slightly less so than users of the motorway system.
- No motion sickness was reported by users experiencing the system.
- The more users were familiar with urban automated driving, the more they were willing to use the system.

For **parking ADF**, 3823 parking manoeuvres and 109 questionnaires were analysed. It was found that the systems took longer to complete the parking manoeuvres and they drove at a slower speed compared to manual parking. Still, drivers were in favour of the systems and considered them to be safe and useful.

Besides the findings reported in this deliverable, the analysed data served as input for the impact assessments and socio-economic impact assessment reported in L3Pilot Deliverable D7.4 – Impact Evaluation Results.

The applied methodology made it possible to analyse the behaviour of automated vehicles in traffic and how users accept the systems and interact with them. This analysis was carried out already at a stage before their market introduction, while avoiding benchmarking between the individual piloted systems. Such anonymised data merging and evaluation required substantial effort to ensure data quality already at the beginning of the data evaluation toolchain.

Future work involving on-road tests with automated vehicles is recommended to focus more on long-term effects related to users of SAE L3 ADF by executing multiple drives per driver over a longer time span. Furthermore, an in-depth analysis should be carried out for situations at the edge of the operational design domain of such systems or in unexpected situations during their operation. For this, processes need to be established which allow sharing disaggregated and time-series data within the project. However, at the same time, the processes need to be established such that they still prevent reengineering of systems or benchmarking between systems to ensure competitiveness and transparency within the automotive industry.

1 Introduction

This chapter presents is an introduction to the L3Pilot project, starting with its motivation and objectives. Based on this, the scope and structure of the evaluation subproject is outlined and the relation between this and other L3Pilot deliverables is explained. Section 1.4 introduces the automated driving functions analysed. Since the COVID 19 pandemic occurred during the latter third of the projects, its impact on the piloting and the evaluation work is outlined in section 1.5.

1.1 Motivation and Objectives for the L3Pilot Project

Over the years, numerous projects have paved the way for automated driving (AD), allowing the driver to take his or her eyes off the road for at least parts of the journey. Significant progress has been made, but AD is not yet ready for market introduction. However, the technology is rapidly advancing and today is at a stage that justifies automated driving tests in large-scale Pilots. L3Pilot is taking the final steps before the introduction of automated cars in daily traffic. Drivers are used to Advanced Driver Assistance Systems (ADAS), and numerous vehicles are equipped with ADAS and active safety systems.

Automation is not solved simply by integrating more advanced technology. This topic needs above all a focus on user behaviour as well as experience with automated driving systems. User acceptance is one of the keys to the success of AD on the market, as is an understanding of the legal restrictions which need to be discussed and solved first on a broader level. Furthermore, the operation of automated vehicles in mixed traffic needs to be studied to identify the impacts on safety or traffic flow, as deviating behaviour compared to the surrounding vehicles – even if it is still within the law – has an impact on the interactions with other traffic participants.

The idea of vehicles controlling themselves by computer raises fears among the global population comparable to those in the late 1800s when motor vehicles were introduced. The lack of acceptance may hinder the introduction of automated driving despite its expectable potential for improving safety and efficiency. To overcome public concerns, automated vehicles (AV) need to be designed according to user needs; otherwise, they will not be accepted. AD systems will influence societies and peoples' lives more than previous automotive innovations after the introduction of series production cars over a hundred years ago. L3Pilot plans to contribute to this change.

The overall objective of the L3Pilot project was to test and study the viability of automated driving as a safe and efficient means of transportation, as well as explore and promote new service concepts to provide inclusive mobility. AD technology has matured to a sufficient level to motivate a final phase of road tests to answer the key questions before market introduction. These recently attained levels of maturity ensured an appropriate assessment of the impact of AD, what is happening both inside and outside the vehicles, how vehicle security can be ensured, and an evaluation of societal impacts and future business models.

Recent work indicates how driver assistance systems and AD functions can best be validated by means of extensive road tests, with a sufficiently long operation time, to allow extensive interaction between the driver and testable functions. The project carried out large-scale testing and piloting of AD with developed SAE Level 3 (L3) functions (Figure 1.1) exposed to different users and mixed traffic environments, including conventional vehicles and vulnerable road users (VRU), along different road networks.

The data collected in these extensive pilots supported the main aims of the project to:

- Lay the foundation for the design of future, user accepted, L3 and L4 systems, to ensure their commercial success. This will be achieved by assessing user reactions, experiences, and preferences for the AD systems' functionalities.
- Enable non-automotive stakeholders, such as authorities and certification bodies, to prepare measures that will support the uptake of AD, including updated regulations for the certification of vehicle functions with a higher degree of automation, as well as incentives for the user.
- Create unified de-facto standardised methods to ensure further development of AD applications (Code of Practice).

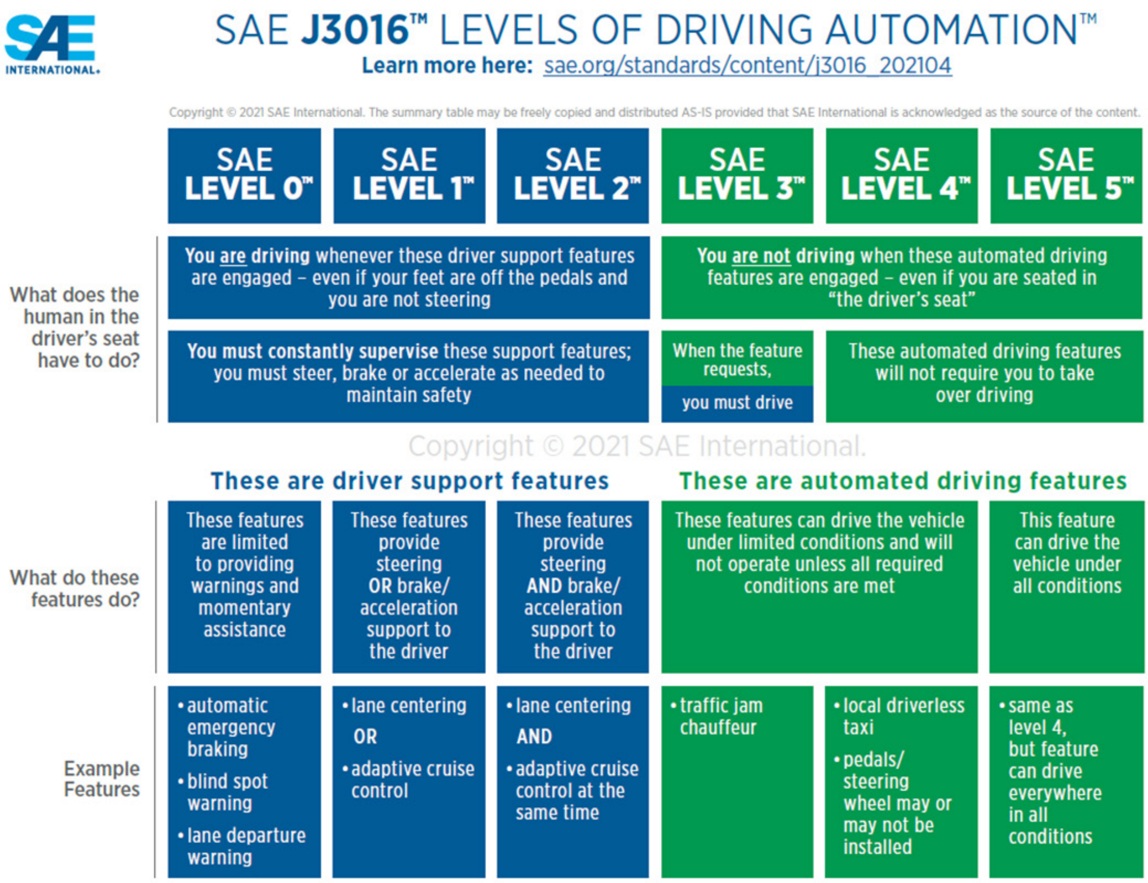


Figure 1.1: SAE Levels of Driving Automation J3016 (Copyright 2021 SAE International).

The consortium addressed four major technical and scientific objectives listed below:

- Create a standardised Europe-wide piloting environment for automated driving.
- Coordinate activities across the piloting community to acquire the required data.
- Pilot, test and evaluate automated driving functions.
- Innovate and promote AD for wider awareness and market introduction.

1.2 Scope of the Subproject - Evaluation

The L3Pilot project focused on large-scale piloting of ADFs, primarily L3 functions. The key in testing is to ensure that the functionality of the systems used is exposed to variable conditions, and performance is consistent, reliable, and predictable. This increases the probability to create a positive user experience. A good experience of using AD will accelerate acceptance and adoption of the technology and improve the business cases for different stakeholders.

The L3Pilot consortium brings together stakeholders from the whole value chain, including OEMs, suppliers, academic institutes, research institutes, infrastructure operators, governmental agencies, the insurance sector, and user groups. More than 750 users tested 70 vehicles across Europe with bases in seven European countries: Belgium, France, Germany, Italy, Luxembourg, Sweden, and the United Kingdom. The project lasted for 50 months, road tests started in April 2019, and Piloting on variable road conditions took two years.

The evaluation subproject had the responsibility of combining the data and insights gathered in the other subprojects and generating the project evaluation results from those inputs. Within the subproject, four different areas of evaluation were considered. These are:

- **Technical and Traffic Evaluation,** focusing on aspects concerning technical aspects which can be measured from the vehicle, like the quality of staying in lane, as well as parameters which characterize how the vehicles behaves in traffic and interacts with other road users.
- **User and Acceptance Evaluation**, dealing with the user's interaction and experience with the tested ADF as well as user aspects which are beyond the piloting operations; this was investigated by executing dedicated studies on aspects like long-term behaviour or the public's attitudes towards AD.
- **Impact Assessment,** which analyses effects the introduction of automated driving will have on the areas of safety, efficiency and mobility, considering that automated driving needs to be introduced in mixed traffic.
- **Socio-Economic Impact Evaluation**, which translates the assessed impacts of AD into monetary costs or benefits for society, deriving the overall balance between costs and benefits linked to the introduction of automated driving.

The aspects of AD for the different areas of evaluation and the related levels of data aggregations are shown in Figure 1.2. Moving from the technical and user-related assessment to societal impacts initially requires analysis of data collected from single users and vehicles. To derive the effects of automated driving in general, the findings need to be merged to fleet level allowing an aggregated data analysis. These aggregated findings can then be evaluated on a societal level, within Europe to derive impacts in the areas of safety, environment and mobility, which can then be merged to the overall cost benefit analysis for the introduction of AD in Europe.

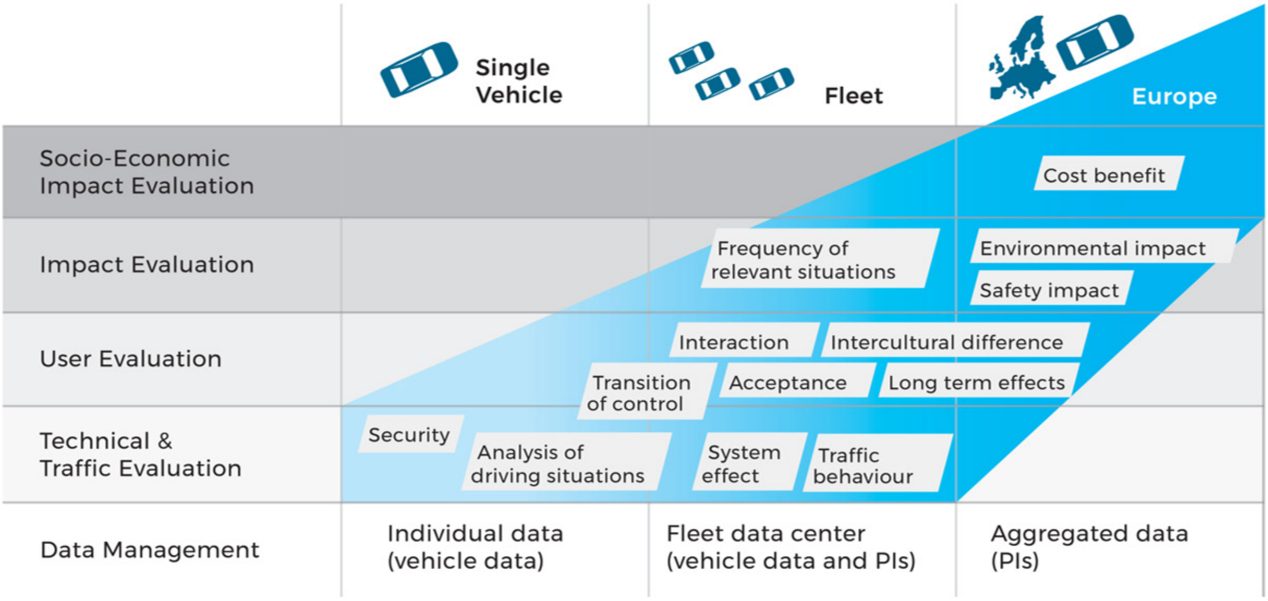


Figure 1.2: Fields of evaluation and aggregation level of data.

1.3 Content of Deliverable and Relation to Other Deliverables

This deliverable focuses on the evaluation of data acquired during the piloting activities within L3Pilot. To achieve this, several steps needed to be taken which have been described in several deliverables:

- The aim is to answer the research questions (RQs) defined in **D3.1 – From research questions to logging needs** (Hibbert et al., 2018). The list of RQs compiled was used as a basis for defining which signals need to be logged from the piloted vehicles to answer the RQs. At this stage, it was not yet certain whether all the RQs defined could actually be answered with the data collected. Therefore, after the handover of the first batches of piloting data, the feasibility of the different RQs was reconsidered and is reported in this deliverable.
- As the vehicles operated in the Pilot were not yet market ready, the study design needed to be adapted to the maturity and the experimental nature of the vehicles. To harmonise the study design across the different Pilot sites, **D3.2 – Experimental Procedure** (Penttinen et al., 2019) was generated, which provided guidelines and recommendations for each Pilot, enabling a consistent merging of the data collected from different sites.

- A summary of the piloting activities executed in L3Pilot is given in **D6.5 – Reporting Outcomes** (Andreone et al., 2021), providing insight into the piloted systems, the implementation of the study design, and the data collected.
- The methods for the evaluation to be carried out in the Evaluation subproject were defined in **D3.3 – Evaluation methods** (Metz et al., 2019).
- Building up on the available methods, a specific plan was defined in **D3.4 – Evaluation Plan** (Innamaa et al., 2020), which also provided further details and updates on the methodology. Together with D3.3 – Evaluation methods (Metz et al., 2019), it forms the direct basis for the assessment work performed in L3Pilot.
- For the implementation of the methodology, we used a selection of tools developed during L3Pilot as described in **D5.1 – Pilot Tools for L3Pilot** (Nagy et al., 2018). Signals were transformed into a common data format (Nagy et al., 2018 & Hiller et al., 2019) which allowed the creation of a common toolchain for all partners working with data from the piloted vehicles (Nagy et al., 2018 & Hiller et al., 2020).
- The experience gathered from the technical teams working on piloting data, quality checks, manipulation, and processing was also reflected in Deliverable **D5.2 – Guidelines and Lessons Learned** (Christen et al. 2021) - Guidelines and Lessons Learned on Pilot Tools and Data, where lessons learnt were gathered and clustered in three main chapters: Data Logging, Data Management and Data Analysis.
- Data for answering the project RQs is stored and shared via the consolidated database (CDB), which enables pseudonymised sharing of data not allowing the identification of individual Pilot side. This process is described in **D6.2 – Database for data Collection: Evaluation Format & common data set for future Research** (Bellotti et al., 2019).

This deliverable focuses on answering RQs concerned with technical aspects and the vehicle's behaviour in traffic as well as user-related topics employing data collected during the piloting operations. Besides the piloting, further studies were executed focusing on relevant topics to create a comprehensive picture of the impacts of automated driving on traffic, user behaviour and society.

- **D7.1 – Annual quantitative Survey about user acceptance towards ADAS and Vehicle Automation** (Nordhoff et al., 2021) describes the findings from a survey on conditionally automated driving (SAE L3) involving participants from countries all over the world. This study allows comparing potential users' attitudes and acceptance towards automated driving in different countries.
- **D7.2 – L3/L4 Long-term study about user experiences** (Metz et al., 2021) describes studies on the behavioural adaptation of users of automated driving functions utilising driving simulator and Wizard-of-Oz studies, which allow focussing on user-related aspects that could not have been studied in regular on-road piloting activities.

- **D7.4 – Impact Evaluation results** (Bjorvatn et al, 2021) reports the results of the impact assessment within the areas safety, mobility, and efficiency and environment, as well as the results from the socio-economic impact assessment by means of a cost-benefit analysis. The results from D7.4 utilise findings from the other mentioned deliverables, as well as this one wherever possible, to justify assumptions made for the impact assessment.

1.4 Evaluated ADF

SAE Level 3 automation does not require the driver to supervise the driving task, but the driver needs to be available as a fall-back layer for the system within a limited time span when the system issues a take-over request (TOR). If the driver does not respond properly to the TOR, the vehicle will perform a minimal risk manoeuvre.

In the following, the different ADF evaluated in L3Pilot are described. These high-level descriptions of the evaluated systems depict the common bases of the systems. While all systems follow the description, the individual layout implementation and the resulting behaviour of the system may differ slightly. The detailed functionalities of the systems are undisclosed.

1.4.1 Motorway Chauffeur & Traffic Jam Chauffeur

L3Pilot considers two different ADF operated on motorways. One of them is an SAE Level 3 Traffic Jam Chauffeur, which allows the driver to hand over the driving task to the ADF without the need to supervise. The Traffic Jam Chauffeur operates on motorways (controlled access) and similar roads up to a speed of 60 km/h. Operation of the traffic jam ADF requires a leading vehicle to be present. In case a slow vehicle is in front of the ego-vehicle, the ADF can execute a lane change to a lane with faster flowing traffic.

In contrast, the SAE Level 3 motorway chauffeur covers a speed range of up to 130 km/h on motorways and similar roads. The motorway chauffeur may either follow a leading vehicle or keep a speed below the speed limit. Depending on the system design, the motorway ADF may execute lane changes in order to drive at its desired speed.

The evaluation in L3Pilot does not make a distinction between the Traffic Jam and Motorway ADF on a system level. Instead, if an evaluated system is currently in a driving situation which may be considered a traffic jam suitable for a Traffic Jam Chauffeur, the situation will be considered as a situation relevant for a Traffic Jam Chauffeur, even if the system would also allow for a full speed range operation on motorways. Hence, in the following no distinction will be made between these systems. Both are considered to be motorway ADFs, while a distinction between traffic jam situations and normal motorway driving is made on the driving scenario level (see Section 2.5.1).

1.4.2 Urban Chauffeur

The Urban Chauffeur targets stress-free driving in urban areas. With the Urban Chauffeur, the vehicle automatically follows the lane, starts and stops and handles lane changes – either for overtaking or to fulfil the navigation task – within cities. When coming to a crossing, the car handles right and left turns, recognises oncoming traffic and VRUs, and selects the correct crossing path, even if no lane marking is present.

1.4.3 Parking Chauffeur

The Parking Chauffeur is a vehicle function that allows the user to request their vehicle to complete manoeuvring into and out of garages and driveways. The car either learns a fixed trajectory from the entrance of the house to the home garage and vice versa or determines a suitable manoeuvre to enter or pull out of a nearby parking position. This automated driving feature relieves the driver from recurring parking manoeuvres. Depending on the Operational Design Domain (ODD), the Parking Chauffeur has also been tested at SAE L3 or L4, i.e., without the need to hand over control of the vehicle to the human driver in case a fall-back manoeuvre is required.

1.5 The Effect of COVID

Most of the tests within L3Pilot were delayed due to the COVID-19 pandemic. These delays influenced the progress of the evaluation work. In particular, the performance of several tools for data evaluation had been planned to be validated during the development stage using preliminary Pilot data. Because of delays in admitting tested vehicles onto public roads, data was not available in this development phase for most Pilot sites. By the time Pilot operations were scheduled, the COVID-19 pandemic had severely impacted the work procedures within Europe and all operations were put on hold. After some time, depending on the local situation, Pilot operations could be continued under precautionary measures[1].

For several sites, data delivery to the evaluation subproject was delayed of several months, posing significant challenges.

The data analysis process applied a common toolchain and a harmonised data format (see Hiller et al., 2019). Because this procedure had to work with data from all Pilot sites, extensive testing of the evaluation toolchain was required, especially as different partners had to agree on the methods for data anonymisation. The final round of testing could only be completed once data was available for all the Pilot sites, which could only be achieved by the end of 2020. This timing was still sufficient for ensuring the effectiveness of algorithms and tools and for validating the overall data quality. However, it was not possible to revise decisions on the data formats or to adapt the setup of the data sharing process, which were defined at a stage when only very limited amounts of preliminary data were available.

[1] For pilots conducted during the period of COVID-19 pandemic, safety precautions and measures were ensured. These included wearing face masks, allowing only two persons per vehicle, disinfection of the vehicle between drives, air conditioning switched on and windows lowered to encourage ventilation, driving breaks outside of the vehicles, and ensuring the participants were comfortable with continuing.

2 Methodology

The following chapter describes the methodology behind the results presented in this deliverable. Therefore, it briefly repeats the methodology applied within L3Pilot, before going into detail about the technical & traffic assessment and necessary data processing, and detailing the applied methods for user and acceptance analysis. The applied methodology builds upon the results of the dedicated subprojects, summarised in the following deliverables:

- D3.1 – From Research Questions to Logging Needs (Hibberd et al. 2018)
- D3.2 – Experimental Procedure (Penttinen et al. 2019)
- D3.3 – Evaluation Methods (Metz et al. 2019)
- D3.4 – Evaluation plan (Innamaa et al. 2020)

The following describes the most relevant aspects of the L3Pilot methodology and any changes to the approaches described in the deliverables.

2.1 Overall Methodology in L3Pilot

L3Pilot evaluation is based on the FESTA methodology (see the latest version of the FESTA Handbook, FOT-Net & CARTRE (2018)). However, as this methodology was designed to be applied to field-operational tests (FOTs) with market-ready products, it did not fully apply to studies with prototypical ADFs. Thus, some adjustment of the original “V” structure was needed to accommodate testing of prototype functionalities, such as ADFs, in real traffic. The Pilot nature of the tests in L3Pilot brings some practical and ethical limitations regarding the use of the automated vehicles and limits any firm conclusions drawn about their implementation in the real world, or their expected impacts. To generate valid conclusions regarding the impacts of the ADFs, it is important to consider all the principles used to collect the evaluation data and to carefully consider any ensuing conclusions. Consequently, L3Pilot adapted the original *FESTA V* to better describe the key steps of the project (Figure 2.1). Details of the changes made and reasons them, as well as a detailed description of each, step can be found in Deliverable D3.4 – Evaluation Plan (Innamaa et al., 2020).

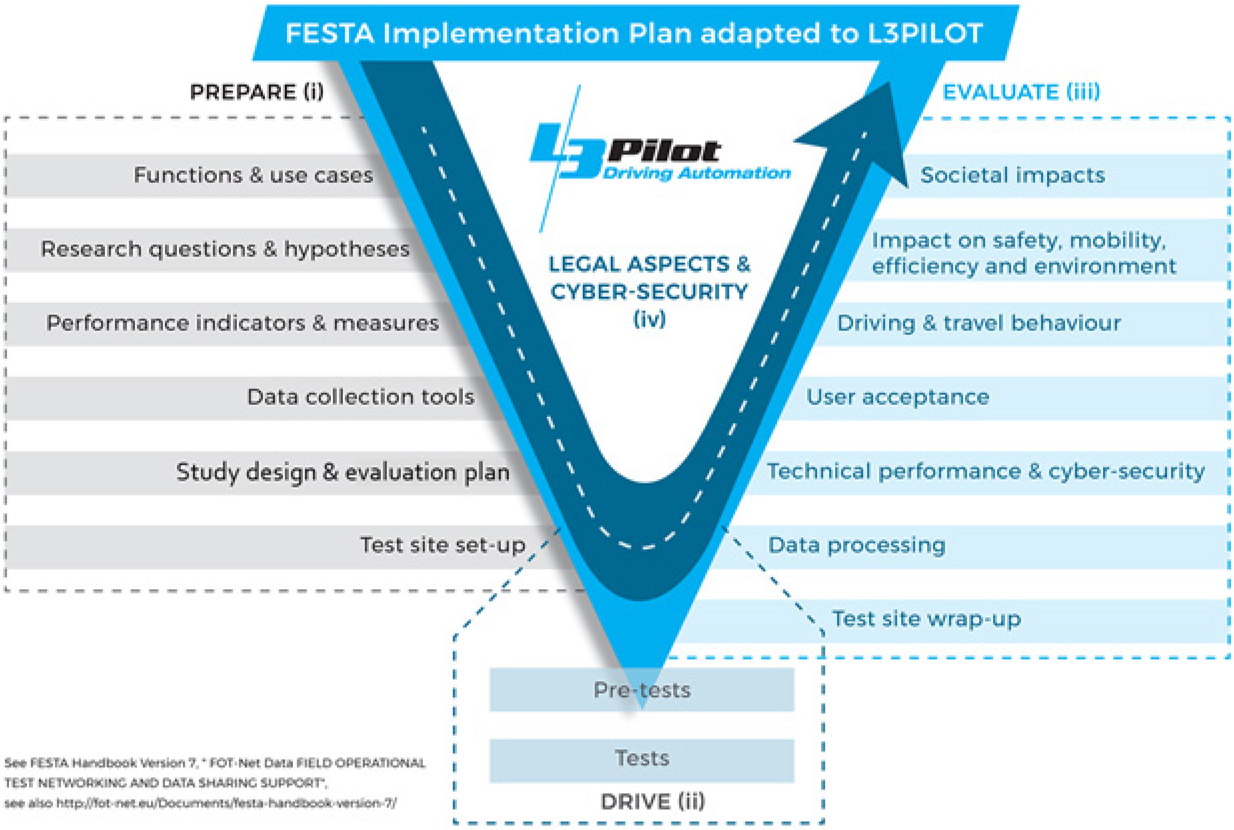


Figure 2.1: FESTA V adapted to the L3Pilot project (Innamaa et al., 2020).

The presented deliverable sums up the first four steps taken in the EVALUATE phase of the FESTA V, which starts with the *test site wrap-up* delivering the collected data and metadata for evaluation. Before the delivery, the Pilot sites handled the *data processing* converting their raw data into a common data format. Dedicated evaluation partners processed this data according to commonly agreed principles and tools. They also uploaded data to a CDB to be used later by the other evaluation partners for different evaluation areas.

The evaluation of *technical performance & cybersecurity* aimed to understand the system as experienced by users in the field tests. For this, measures were evaluated which give an insight on technical aspects such as lane keeping performance. The results related to the technical performance are presented in this deliverable (Section 3). Due to the prototype nature of these products, cybersecurity was not evaluated.

The evaluation of *user acceptance* aimed to understand users' experience and acceptance of the tested ADF. The results related to user and acceptance which were derived from the piloting operations are presented in this deliverable (Section 4).

Driving & travel behaviour evaluation aimed to understand the changes that the introduction and use of ADFs will lead to. These changes were reflected in the following phases of evaluation. Driving behaviour is addressed in this deliverable by means of an analysis of measures characterising the vehicle's behaviour in traffic (Section 4). At a general level, travel behaviour evaluation is part of Deliverable D7.4 – Impact Evaluation Results Mobility impact assessment (Bjorvatn et al. 2021).

The next phase of the evaluation assessed the *impacts on safety, mobility, efficiency and environment* and scaled them up to EU level utilising data and results from the *driving & travel behaviour* evaluation and the *user acceptance* evaluation. Finally, *societal impacts* were assessed using the results of the previous step in cost-benefit analysis. L3Pilot Deliverable D7.4 – Impact Evaluation Results (Bjorvatn et al. 2021) also reports these results.

2.2 Data sharing

The sharing of data between partners piloting the ADF and researchers answering the selected RQs is the central element for a pilot study involving several vehicle manufacturers. In the following, the given constraints for setting up such data sharing and the implemented solution are discussed.

2.2.1 Constraints

L3Pilot deals with systems enabling automated driving that have not yet been introduced to the market. At this stage, the systems are still confidential. Nevertheless, L3Pilot wants to ensure a data evaluation that is as thorough as possible to enable understanding of the effects automated driving will have on users, traffic and society already at this development stage. Consequently, certain measures need to be adopted to guarantee that no confidential information about the system is shared among the vehicle manufacturers and suppliers participating in the piloting or with third parties outside the project. Requirements for data sharing can be summarised as follows:

- The data shared should not facilitate benchmarking between the piloted systems.
- The data shared should not allow reverse engineering of ADF parameters.
- Data available for answering the RQs should not be linked to individual Pilot sites.
- For some PIs, special requirements of confidentiality might apply that restrict sharing information on a disaggregated level within the project.

For setting up the data evaluation process two options were considered, which are also described in Deliverable D3.3 – Evaluation Methods (Metz et al 2019): the first one involves “merging of results” and consists of applying statistical tests on data sets regarding the individual Pilot sites for which a meta-analysis is then carried out. In contrast, the second approach “merging of PIs” would combine the available PIs per RQ in a common database and apply the statistical test on this collective dataset.

As the harmonised study design applied at the individual Pilot sites allows for a merging of their data, Metz et al. (2019) recommended merging of PIs wherever possible. Merging of results was considered as a fall-back option in case certain constraints do not allow for the previous approach. For the User and Acceptance analysis, merging of PIs could be implemented by directly sharing the collected data, i.e., the individual participant’s answers to the questionnaire items. For the

Technical & Traffic analysis, the requirements for data confidentiality led to the decision that no time-series data from the individual sites could be shared with the entire consortium. PIs thus needed to be derived from the time series data which aggregate information per defined segments of trips (driving scenarios) or entire trips. All the PIs considered in the project, as well as their relation to the RQs, are described in D3.4 – Evaluation plan (Innamaa et al., 2020).

2.2.2 Implemented data sharing process

As recommended in Metz et al. (2019), merging of PIs was implemented for motorway and urban ADF, both for vehicle data and for questionnaire data.

The chosen approach for merging data across the different Pilot sites required establishing a data handling and sharing process that met the requirements for not making Pilot sites identifiable and not facilitating benchmarking and reverse engineering. In the data evaluation process three different roles were defined for the partners involved in the data acquisition and data evaluation process:

Pilot leaders are operators of Pilot sites who implement the study design, execute the pilots, and implement the recording of data from the piloted vehicles.

Pilot data processing partners are partners involved in the evaluation of data who have the dedicated role of working in close collaboration with one or multiple Pilot leaders, which allows the sharing of required disaggregated or time-series data. Between Pilot leaders and data processing partners, individual non-disclosure agreements may have been set up to meet requirements for personal data protection and confidentiality. In general, the task of the data processing partners is to process and aggregate the piloting data to a stage at which it could be shared with other partners involved in the evaluation.

Evaluation partners work with the aggregated data that has been merged across the different Pilot sites. They do not have access to any information that could reveal the identity of Pilot sites that contributed to individual entries in the general dataset.

Between these roles, a data sharing and a merging process was established. A central tool for this process was the consolidated database, which allowed for controlled merging of data while at the same time hiding the identity of the individual Pilot sites to the evaluation partners. The structure and interfaces for the consolidated database are described in D6.2 – Database for data collection: evaluation format & common data set for future research (Bellotti et al., 2019) and (Hiller et al., 2019). Nearly all partners involved in the evaluation took the role of a Pilot data processing partner for a limited number of Pilot leaders and of evaluation partner working with the data available in the consolidated database.

The data processing and evaluation process can be summed up as shown in Figure 2.2. The start of the process is the acquisition of the raw vehicle data at the different Pilot sites. For the data acquisition, the individual Pilot sites implement the study design considerations elaborated by Pentinnen et al. (2019), which allows the merging of data across Pilot sites. The data gathered can be split into two items: vehicle data logged from the CAN-bus and other data communication between subsystems of the piloted vehicles, as well as video data, and questionnaire data consisting of the participants' answers to the questionnaire items. All Pilot site questionnaires are reported in Deliverable D3.3 – Evaluation Methods (Metz et al., 2019).

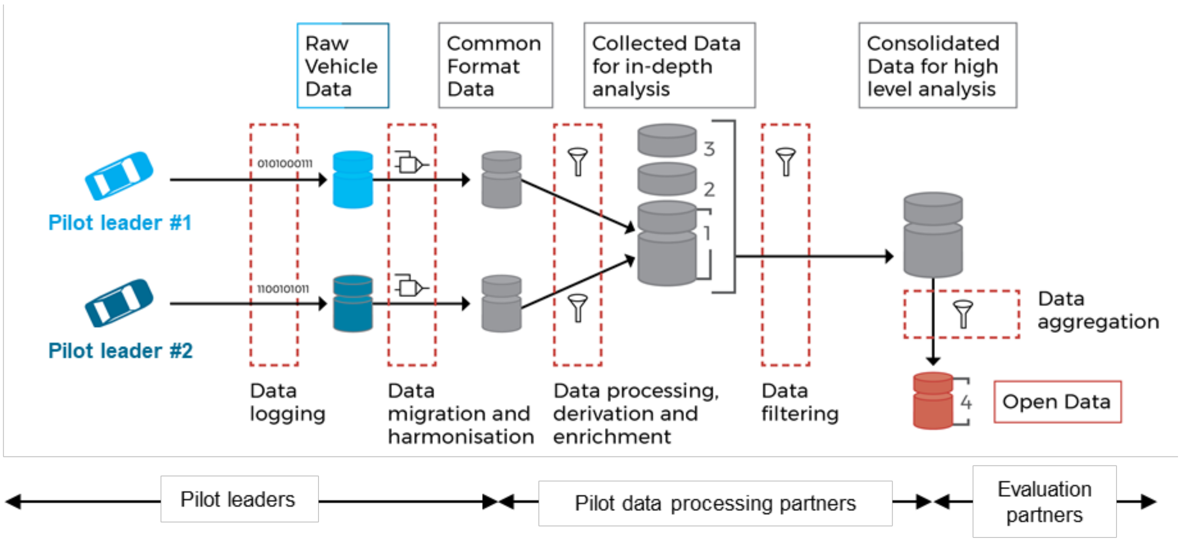


Figure 2.2: Stages for data processing in L3Pilot.

At this stage, vehicle data is recorded in formats that are proprietary to the individual Pilot sites, as the data is directly logged from the automated driving function. Data management is handled based on the internal processes at the Pilot leader. Before handing over the collected data to the data processing partner, it is converted to a common data format (CDF). Using the same data format for all Pilot sites allows for more harmonised processes between the data processing partners, as well as an easier interpretation of the results by the evaluation partners. The CDF for data logged from the piloted vehicles is described in D5.1 – Pilot tools for L3Pilot (Nagy et al., 2018) and by Hiller et al. (2019). Questionnaire data is handled in tabular format. Conversion from the proprietary format to CDF is done by Pilot leaders, who also apply initial quality checks to the data delivered to the data processing partner. A second check of the data quality is then applied by the data processing partner, who examines whether the data can be successfully processed by the evaluation toolchain.

The evaluation toolchain consists of a collection of MATLAB scripts hosted in a shared code repository. The initial version of the scripts was prepared by a dedicated team in L3Pilot. Since by the end of the initial tool development phase the entire toolchain had not been extensively tested with data from the different Pilot site, development was continued within the Evaluation subproject.

The data uploaded to the CDB could then be queried by evaluation partners either via a graphical user interface or an application programming interface (API). The queried datasets contained the relevant PIs for vehicle data and questionnaire answers for user data. Based on these, the defined RQs could be answered.

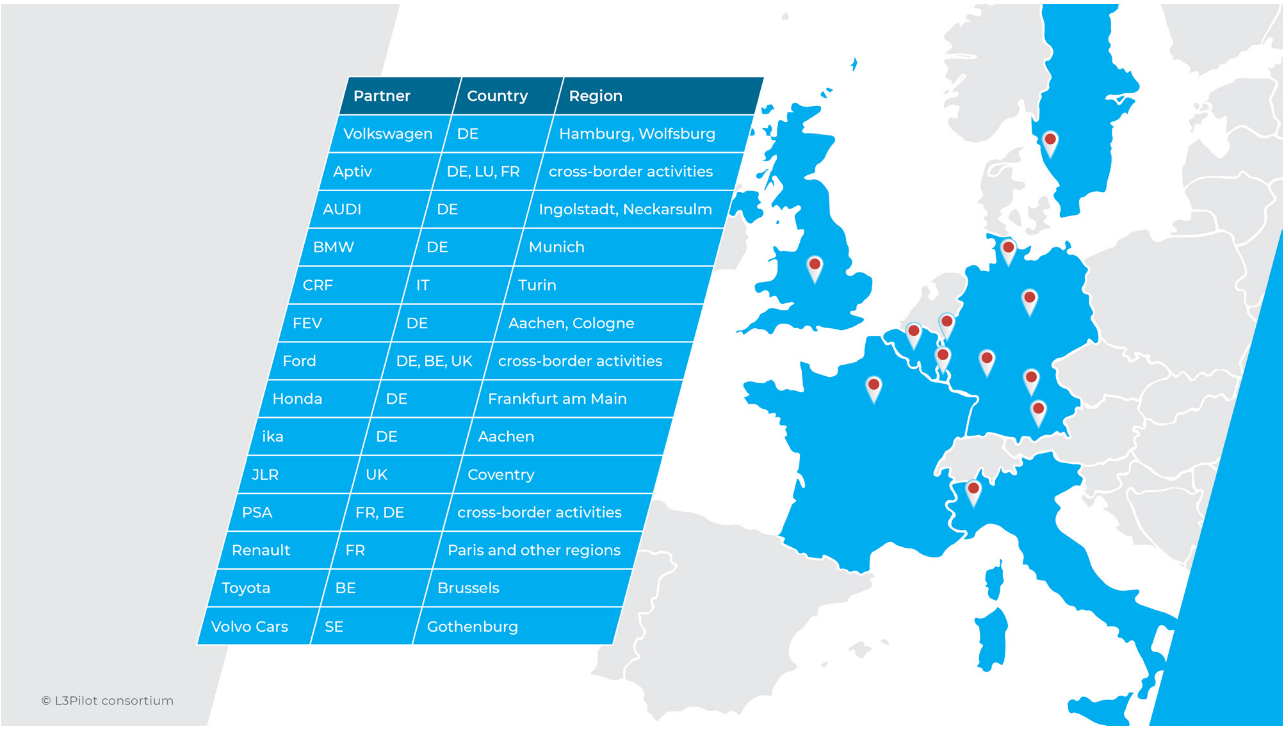
As urban ADF have only been piloted by three Pilot sites and the chosen study designs resulted in considerable differences in the amount of data to be merged, which created a risk of an imbalanced dataset for evaluation as well as a greater risk of exposing the individual systems, a further step for obfuscating data ownership by means of bootstrapping was introduced. This process is further described in Section 2.5.1 and Annex 4.

For parking systems, merging of results was considered, as each parking study would test different trajectories for the different parking systems, such that the disaggregated PIs would differ substantially between studies. Given that some of the parking studies had to be delayed until April 2021, merging of results was also chosen for questionnaire data, simplifying the evaluation process.

Lessons learnt from the data sharing processes established are reported in D5.2 – Guidelines and lessons learned (Christen et al., 2021).

2.2.3 Available piloting data

Piloting efforts in L3Pilot resulted in a unique and extensive basis for the evaluation of L3 ADFs. Piloting operations were executed at 14 Pilot sites operating in seven different European countries and recording data from more than 750 test subjects testing 70 vehicles. Some Pilot sites also tested cross-border operation of the evaluated ADF. An overview of the Pilot sites is shown in Figure 2.3.



Partner	Country	Region
Volkswagen	DE	Hamburg, Wolfsburg
Aptiv	DE, LU, FR	cross-border activities
AUDI	DE	Ingolstadt, Neckarsulm
BMW	DE	Munich
CRF	IT	Turin
FEV	DE	Aachen, Cologne
Ford	DE, BE, UK	cross-border activities
Honda	DE	Frankfurt am Main
ika	DE	Aachen
JLR	UK	Coventry
PSA	FR, DE	cross-border activities
Renault	FR	Paris and other regions
Toyota	BE	Brussels
Volvo Cars	SE	Gothenburg

Figure 2.3: Pilot sites within L3Pilot.

The resulting data set can be characterised by the following statistics for the different piloted systems:

Table 2.1: Vehicles and users per type of ADF.

	Motorway ADF	Urban ADF	Parking ADF
Piloted vehicles	**70**	**6**	**8**
Test subjects *of which:*	**481**	**177**	**92**
Professional & safety drivers	143	17	4
Ordinary drivers	338	0	85
Users in a passenger seat	0	160	3

- The piloting efforts for Motorway ADF comprise 400,000 km driven on motorway, roughly half in as baseline.
  - This data includes kilometres driven on motorways for data acquisition for dedicated traffic jam systems. (Especially, given that most of the piloting was done during the COVID-19 pandemic, which resulted in overall lower traffic levels, traffic jam events were difficult to find for data acquisition.)
  - Data collection executed with traffic jam-only systems could not contribute to the baseline dataset, as the sensors and system setups were made specifically for low speed scenarios.
  - This exclusion of data resulted in 2267 h of motorway data that could be delivered from the Pilot leaders to the data processing partners.
  - From this data, 1808 h could be processed for upload to the CDB. Data was either not uploaded because it was out of the ODD of the piloted ADF, or issues with data quality did not make it possible to derive the required PIs.
- For **Urban ADF**, 1120 h were driven within urban environments, including 130 h of baseline data collection
  - The delivered dataset comprised 638 h of data which were used for data evaluation.
  - An additional step for data processing was implemented, which ensured that all Pilot sites with urban ADF where weighted equally in the evaluated dataset, even though one of the Pilot sites contributed a significantly larger part of the overall dataset (see Section 2.5.1, Annex 4).
- For **Parking ADF**, five experiments were executed at three different Pilot sites, adding up to 3823 analysable parking manoeuvres.

2.3 Answering the Research Questions

The first step according to the FESTA methodology was the definition of RQs, complemented as a key element by the specification of methods for answering. The process of defining these methods also required re-evaluating the feasibility of several of these RQs.

2.3.1 Motorway & Urban ADF

For urban and motorway ADF, the RQs were analysed based on the data provided by the CDB. Most RQs were answered by comparing performance indicators for situations labelled as baseline driving with situations where the ADF was active. Since many indicators did not fulfil the requirements of parametrical tests like ANOVAs or t-tests (e.g., no Gaussian distribution), non-parametrical tests were used. For significant effects, additional information on the effect size is provided. Annex 1 lists all RQs for motorway and urban ADF and provides information on whether and how the RQs were analysed.

For the urban use case, take-over requests (TORs) were not part of the Pilot site setup, so RQs regarding these were not analysed. As for incidents and energy consumption, this was not shared by all Pilot sites, so an analysis in this context was not possible.

2.3.2 Parking

For parking ADF, an aggregation of data that would allow for the use of the common database was applied, since the tested systems and studies differed somewhat. Instead, the common answer to the RQs is based on a combination of results coming from the different Pilot sites i.e. experiments. As described in 2.6.2.2, the experiments were fully analysed by the Pilot data processing partners, including statistical testing. Parking with ADF active and manual parking (i.e., baseline driving) are compared in the analysis.

To combine results across Pilot sites, information on statistical significance and on effect size including a relative change was collected from the different experiments. Based on that information, the RQs were answered on the project level. Annex 1 lists the RQs relevant for parking ADF.

2.3.3 User and Acceptance Evaluation Motorway and Urban ADF

The User and Acceptance RQs were organised into several key themes, including user acceptance and trust of the systems, willingness to use and pay for the functionalities, measures of driver state (stress, distraction, fatigue, workload), user risk perception, driver engagement in non-driving related tasks, user behaviour during and after take-over situations, and user motion sickness. The sub-set of RQs for the User and Acceptance Evaluation are shown in Annex 1, with more detailed derivatives of each presented in Deliverable D3.4 – Evaluation Plan (Innamaa et al., 2020).

Based on the specific RQs, and the fact that Pilot studies cannot provide the data to answer all the RQs, the project developed a multifaceted assessment approach to form a holistic view of users' behaviours with, and acceptance, of the ADFs. These include data from a combination of quantitative and qualitative data collection methodologies centred primarily around the Pilot studies, including user questionnaires, videos of the driving scene, recordings of the drivers' head, hands, and posture during the pilot, and vehicle-based data. Data was also collected from supplementary studies, including driving simulator and Wizard-of-Oz studies, and a large-scale international survey. Each of these methods of data collection is introduced and discussed in the sections below.

This multifaceted data collection and analysis approach is used regularly in FOTs or Naturalistic Driving Studies (NDS) investigating user behaviour. For example, UDRIVE (Lai et al., 2013; van Nes, Bärgman, Christoph, & van Schagen, 2019), SHRP2 (Dingus et al., 2015), ecoDRIVER (Jamson, Kappe, & Louw, 2014) and DRIVEC2X (Brizzolara et al., 2014) all relied on both subjective (attitude and behaviour questionnaires) and objective (vehicle and video) data in their evaluation, which were supplemented by interviews, focus groups, or self-confrontation sessions.

One of the primary sources of data for the User and Acceptance Evaluation within L3Pilot was a Pilot site questionnaire, which gathered subjective data from participants at the 13 Pilot sites (for the full questionnaire see Deliverable D3.3 – Evaluation Methods (Metz et al., 2020)). This is a unique contribution of the L3Pilot project, as participants have had real-world experience with these ADFs, whereas previously, subjective data has been collected from participants either with experience only in simulated environments (cf. Madigan, Louw, & Merat, 2018), or with no hands-on experience at all (cf. Kyriakidis, Happee, & de Winter, 2015).

2.4 Experimental Procedure for the Pilots

When designing the experimental procedures for a Pilot study, the difference between FOTs of close-to, or on-the-market products and pilots of prototype systems was acknowledged. These differences and how to take them into account were detailed in L3Pilot Deliverable D3.3 – Evaluation Methods (Metz et al., 2020). Experimental procedures, which are presented in detail in L3Pilot Deliverable D3.2 – Experimental Procedure (Penttinen et al., 2019), were developed to provide a solid base for the evaluation methodology and to ensure that the results from tests across all Pilot sites can lead to an L3Pilot-wide evaluation, considering the practical limitations of their implementation. Furthermore, the aim was to harmonise the evaluation criteria by providing detailed recommendations for the pilots, creating holistic evaluation results for the L3Pilot project.

Finally, a set of practical recommendations was listed for the Pilot sites to finalise their preparations for the Pilot tests. These were reported in L3Pilot Deliverable D3.2 – Experimental procedure (Penttinen et al., 2019), and an update was made to the recommendations based on the feedback and remarks from the Pilot sites when the pilots were already ongoing (February–March 2020) in L3Pilot Deliverable D3.4 – Evaluation Plan (Innamaa et al. 2020).

Given the constraints to the operation of the piloting of motorway systems, the experimental procedures applied in L3Pilot can be summarised in three categories for systems operating on the motorway or in urban areas:

- Studies on motorway ADF involving ordinary drivers in the driver's seat

- Studies on motorway ADF with safety drivers in the driver's seat
- Studies on urban ADF with ordinary drivers as passengers

All these studies contributed to the vehicle data for Technical and Traffic evaluation and to the questionnaire data for User and Acceptance Evaluation.

Apart from the Piloting, Wizard-of-Oz (WoZ) studies have been executed primarily contributing to the studies on D7.2 – L3/L4 Long-term study About user experiences (Metz et al. 2021). As drivers filled in the same questionnaires as for the piloted motorway ADF, it was possible to merge WoZ questionnaires with motorway ADF questionnaires for ordinary drivers. Simulator studies reported by Metz et al. (2021) involved filling in the same questionnaire as well, but due to the nature of these studies, answers relating simulator studies were treated separately.

2.4.1 Experimental procedure for Motorway ADF with ordinary drivers

Most of the studies involving ordinary drivers in the driver's seat included a driving time of 1 to 1.5 hours per drive (ranging from 60 to 133 km). However, some drives were as short as 30 minutes, and some were as long as 6 hours. All studies were conducted in daylight with clear, cloudy, or light rain weather; there was no testing in extreme weather, heavy rain, or snow conditions, with limited trials conducted at night.

In some studies, participants were allowed to take their eyes, hands, and mind off the driving task during the automated drive and to engage in a non-driving related task (NDRT), but in other studies this was not possible. Drivers were required to take-over control when prompted when they reached the end of the ODD of the piloted system. To request drivers to take-over, the vehicle provided both auditory and visual HMIs in all cases, although the designs of HMIs varied across systems, and the time at which the drivers were informed of an upcoming take-over request also varied. In addition, drivers were also informed when AD was available, again using both visual and auditory cues.

Pre-experimental briefings were conducted in all studies, during which drivers were informed about the organisation of the experiment, system functions and limitations, how to activate and deactivate the ADF, and the route. They were instructed to respect the rules of the highway code during manual driving, and to keep safe and regulatory distances to the surrounding traffic. Where applicable, drivers were informed about the cameras installed in the vehicles. They were also informed about the role of safety drivers and whether they were allowed to perform an NDRT. In most studies, practice drives were conducted before the experimental drives. The objectives of the practice drives were to familiarise drivers with the dynamics of the vehicle, the activation and deactivation of the system, the manual drive and ADF, and to understand the capabilities and limitation of the vehicle. Practice drives lasted typically from 15 to 30 minutes, some on the motorway and others on a test track.

The number of experimental drives per participant varied from one to three depending on different studies, where the time between studies ranged from days to months. Drivers were told that they had full responsibility for the vehicle during manual driving and should follow traffic regulations.

When the ADF was available, they should activate the system. In some studies, drivers were allowed to engage in a non-driving related secondary task but were asked to take-over control when prompted. However, in some other studies, due to this not being permitted by the regulations because the tested vehicles were still in the prototype phase, drivers were asked not to engage in the secondary tasks but to monitor the road. All drivers were compensated for taking part in the study (unless the participant was a safety driver, see Section 2.4.2).

Safety drivers were present in the vehicle unless it was Wizard-of-Oz study. When safety drivers were present, they usually sat in the front passenger seat, and in some cases a technician who monitored the system with screens was sitting in the back. The role of the safety drivers included monitoring for hazards, prompting the driver to take-over during critical situations, and in some cases, taking over control themselves (i.e., when the technician informed the safety driver that the system was no longer working), monitoring the system, and supervising the safety and appropriate conduct of the study. Otherwise, the safety drivers were asked not to converse with the drivers to minimise interruptions and distractions. Safety drivers only intervened in dangerous situations and technical failures, in which participants were instructed not to touch the driving controls and to let the safety driver drive.

2.4.2 Experimental procedure for motorway studies with professional drivers

For Pilot sites where it was not possible to operate the vehicle with an ordinary driver, the vehicle was driven and supervised by a professional driver. When the automated mode of the vehicle was activated, the safety driver continued supervising the vehicle to override the system if critical situations become imminent. For some systems, when executing lane changes during operation it was necessary for the safety driver to confirm the lane change decision of the vehicle before the lane change was executed. NDRT could not be executed by safety drivers during automated operation of the vehicle.

For most of the Pilot sites it was also necessary to collect baseline data with professional drivers. In such cases safety drivers were not instructed to follow any particular driving style. Still, it should be noted that the safety driver collected baseline data with a prototype automated vehicle, so it was top priority not to be involved in critical situations.

2.4.3 Experimental procedure for urban ADF

For the urban ADFs, studies took place between September and November 2020. Studies were conducted on busy and non-busy multiple lane urban roads, including signalised and non-signalised intersections, pedestrian crossings, traffic lights, and bicycle lanes. The speed limit of urban roads was 50 km/h. The urban ADF was able to detect VRUs such as pedestrians and cyclists. Studies were also conducted in daylight, cloudy, sunny, and light rain conditions, but not in extreme weather.

The testing locations included Brussels, Aachen and Hamburg. The length of the test routes was 2.4 km to 2.8 km per route, and the duration of the drives was 10 – 40 minutes (one or two laps). As most of the participants were passengers, visual and auditory signals and messages were presented to the safety drivers for taking over controls and to inform the safety drivers when the AD system was available.

Participants were either sitting in the front passenger seat or in one of the rear seat and were asked to focus, observe, and experience the urban ADF, but in some studies they were allowed to engage in a secondary task because they were passengers and not the driver. The pre-experimental briefing informed the participants about their roles, test routes, and the duration of the study. They received a brief description of the urban ADF including capabilities (i.e. to detect VRUs), limitations (i.e., still a prototype and not at production level, not working in extreme weather). In some studies, participants were could ask questions during the experiment but were allowed to do so at the end. In some studies, participants were also asked to imagine that they were sitting in the driver's seat and must be aware of the take-over requests from the vehicles. Practice drives were not applicable.

Safety drivers in these studies were seated in the driver's seat with a similar role to that of safety drivers while testing the Motorway ADF. However, as the participants were passengers, the safety drivers did not have to warn the participant or take-over control from the participant. The driver simply took over control of the car when requested to do so by the system, or when they felt that it was appropriate.

2.5 Method for Technical & Traffic Assessment

2.5.1 Motorway & Urban ADF

For technical and traffic analysis, driving scenarios are the basic unit of analysis concerning driving behaviour. A driving scenario is a short period of driving defined by its main driving task (e.g., car following, lane change) or triggered by an event (e.g., an obstacle in the lane). A *driving situation* represents a single segment in time that is assigned to a certain driving scenario (Innamaa et al., 2020). A Driving situation can be considered an instance of a driving scenario. Driving situations within different driving scenarios differ fundamentally, whereas situations of the same driving scenarios are similar.

This means that for motorway and urban ADF, all time-series data logged during the on-road tests are divided into driving situations, which all belong to one of the defined driving scenarios. Multiple driving situations of one driving scenario can occur within a single log of driving data (Figure 2.4). The trips are separated into driving scenarios based on the logged signals. The main sources of information which are used for scenario detection are lane position, speed, surrounding vehicles and the lead vehicle. Figure 2.5 shows an example of the relation between measured lane position and the time headway to the lead vehicle and the detected driving scenarios. There is a repeated switch between the scenario Lane change, Following and Uninfluenced driving.

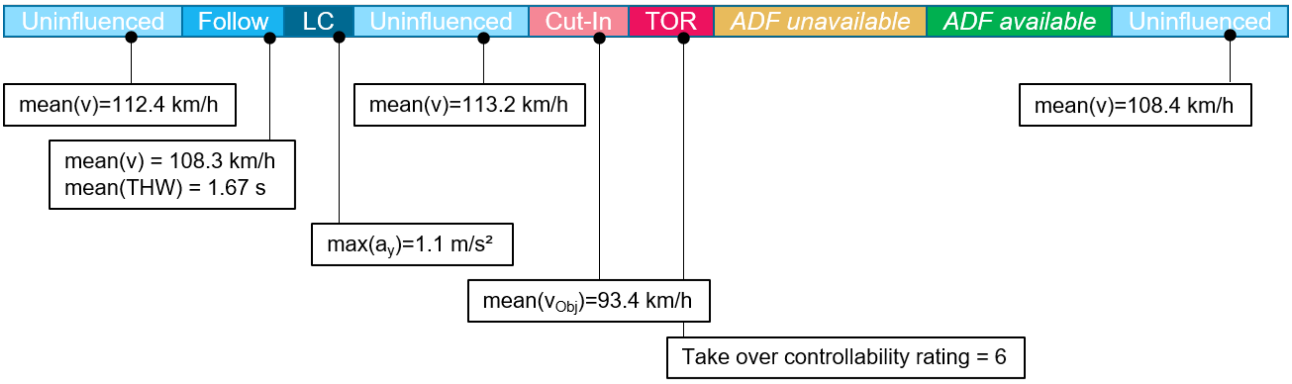


Figure 2.4: Exemplary sequence of scenarios within a trip.

Performance Indicators (PIs) are defined for each driving scenario, and they describe driving behaviour in the scenario in a meaningful way. The defined PIs are calculated for every driving situation identified in the data. Figure 2.4 shows an exemplary sequence of driving scenarios within one trip with an example of one related PI for each scenario.

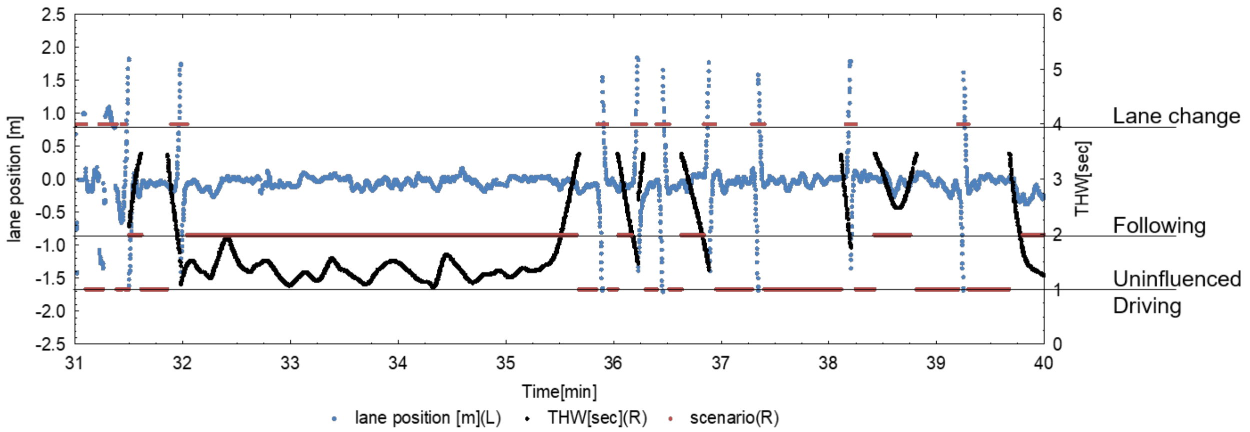


Figure 2.5: Example of the relation between the measured lane position and time headway to the lead vehicle and the detected driving scenarios Lane change, Following and Uninfluenced driving. The example is derived from a trip on the motorway.

The driving scenarios listed in Table.2.2 are used to cover driving on motorways and in urban areas. A detailed description of this approach and of the scenarios is available in Annex 2. As defined within D3.3 – Evaluation Methods (Metz et al., 2019), driving scenarios are mutually exclusive. In order to assure this for the motorway scenarios, a graph-based approach taking into account possible transitions is implemented. For this, scenarios are given a general priority, such that for instance a traffic jam situation will always be treated as a traffic jam, although it may also fit the criteria for following. The priority of scenarios is incorporated with the switches and minimum durations for scenarios to find the optimal sequence of scenarios in a trip. The algorithm is described in detail in Annex A2.1.6.

The scenarios *Uninfluenced driving* and *Following a lead vehicle* have the potential to vary substantially in their duration. Theoretically, the duration of uninfluenced driving can range from a few seconds up to more than an hour on an empty highway. There are several reasons why unwanted side effects of this wide range of scenario duration on the results should be avoided in the analysis:

- As shown by Dozza et al. (2013), there is a direct impact of the duration of an analysed sequence on PIs measuring variation of a measure.
- Without controlling for the duration, a scenario of a few seconds would have the same impact on the overall results as a scenario of a few hours.
- It might be that the scenario duration varies systematically between conditions; in that case impacts due to changes in scenario duration and direct impacts on the indicators could not be differentiated.

To ensure that the impact of scenario duration on the results is minimised, a process called chunking is applied: uninfluenced driving and car following scenarios are divided into sections of 10 seconds' duration and indicators are calculated per section. Figure 2.6 shows how instances of uninfluenced driving are cut-into several chunks with the same duration. In the end one piece remains, the duration of which differs in size from the other chunks. However, this difference is small and is therefore not expected to have an impact on the calculated PIs.

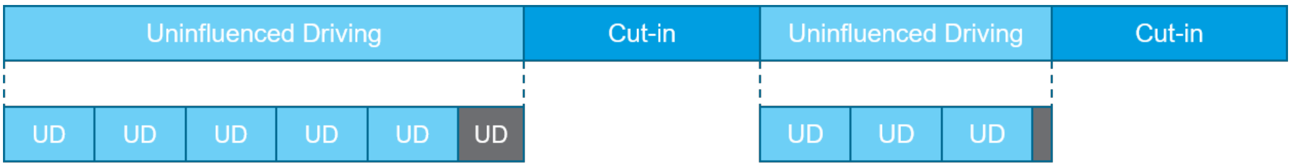


Figure 2.6: Example of an applied chunking procedure.

Table.2.2: Driving scenarios implemented to analyse driving on motorways and in urban areas.

Scenario	Definition	Motorway	Urban
Uninfluenced driving	The ego-vehicle is following its path without being influenced by objects located in or moving into its path. *Uninfluenced driving* is classified when no lead object is detected, or if the time THW between the ego-vehicle and lead object is more than 3.5 s. Also, *uninfluenced driving* is classified if the THW between the two vehicles is between 2 and 3.5 s and the lead object is driving faster than the ego-vehicle. The ego-vehicle's speed must be higher than 5.56 m/s and all of these conditions must be met for more than 2 consecutive seconds. Uninfluenced driving scenarios are divided into sections of 10 s duration and PIs are calculated per section.	X	X
Approaching a lead object	The ego-vehicle is approaching an object located in its path, travelling at a lower speed. *Approaching a lead object* is classified when the THW between the ego-vehicle and a lead	X	X

Scenario	Definition	Motorway	Urban
	object is less than 3.5 s and the ego-vehicle is travelling at a higher speed (>1.4 m/s) than the lead object. If a lead object is moving very slowly or standing still (less than 1 m/s) then the scenario is classified as *approaching a static object*. The THW must then be less than 2 s to ensure that the quality of the lead object measurement is valid.		
Following a lead object	The ego-vehicle is following a lead object. *Following a lead object* is classified when THW between the two vehicles is less than 2 s. Additionally, following a lead object is classified if the THW between the two vehicles is between 2 and 3.5 s and the speed difference between the two vehicles is within ±1.4 m/s. The ego-vehicle speed must be higher than 5.56 m/s and all of these conditions must be met for more than 2 consecutive seconds. Following scenarios are divided into sections of 10 s duration and PIs are calculated per section.	X	X
Approaching a traffic jam	The ego-vehicle is approaching a queue of vehicles in its lane travelling at a low speed. The 20 s before a traffic jam are used to classify this scenario.	X	
Driving in a traffic jam	The ego-vehicle is following a queue of vehicles travelling at a low speed (<60 km/h) for at least 180 s.	X	
Lane change	The ego-vehicle changes lane to the left or right. Lane changes of the ego-vehicle are derived from the lateral position of the ego-vehicle with respect to the position of the lane markings. When the left or right marking is crossed, a lane change is detected and its start and end points are determined. The starting point of the lane change is the point at which the car starts moving in the direction of the lane marking before crossing the marking. The end point of the lane change is the point at which the car stops moving away from the lane marking after crossing the marking. A maximum window size of 10 s before and after crossing the marking is set to limit start and end point respectively. Left and right lane changes are coded separately.	X	X
Cut-in	An object changes lane (or initiates a lane change) into the lane of the ego-vehicle such that the resulting scenario is following or approaching a lead object (cut-ins from the left and right lane are considered).	X	X
Crossing (without conflict)	The ego-vehicle is travelling across an intersection without being influenced by another object.		X
Crossing with static object	The ego-vehicle is travelling across an intersection with a static object located in its desired path.		X
Crossing with lead object	The ego-vehicle is travelling across an intersection while being influenced by a lead object.		X
Crossing with laterally moving object	The ego-vehicle is travelling across an intersection approaching a conflict zone, which it has in common with another object travelling laterally towards the path of the ego-vehicle.		X
Turning (without conflict)	The ego-vehicle is turning at an intersection without being influenced by another object.		X

Scenario	Definition	Motorway	Urban
Turning with static object	The ego-vehicle is turning at an intersection with a static object located in its desired path.		X
Turning with lead object	The ego-vehicle is turning at an intersection while being influenced by a lead object.		X
Turning with laterally moving object	The ego-vehicle is turning at an intersection approaching a conflict zone, which it has in common with another object travelling laterally towards the path of the ego-vehicle.		X
Overtaking of oncoming traffic (passive)	The ego-vehicle is following its lane while a vehicle from the oncoming lane changes into the lane of the ego-vehicle with the intention to change back to its initial lane.		X
Overtaking on oncoming lane (active)	The ego-vehicle changes into the lane of the oncoming traffic, overtakes some obstacle and changes back to its own lane.		X

Urban scenarios can depend significantly more on external factors (such as infrastructure) compared to those on motorways. This leads to the necessity of strictly sorting and prioritising scenarios in the urban context, which is directly coupled to the infrastructural circumstances. Keeping the principle of mutually exclusive scenarios, this means that the scenarios need to be prioritised in the detection as well as in the evaluation. Whereas the easiest approach would be to prioritise using the first detected scenario occurring at a specific intersection, within L3Pilot the focus is on those scenarios at intersections where an interaction in general, or more specifically with vulnerable road users (VRUs) such as pedestrians and cyclists, happens.

For the algorithms within the urban evaluation, this leads to the highest priority of scenarios of crossing or laterally moving objects, as these create the highest amount of interaction and often include VRUs. Scenarios with static objects follow in the priority ranking. Static objects are often not VRUs; however, interaction of the ego-vehicle with these objects often requires strong reactions or moving out of one's own lane, which makes them interesting from an evaluation point of view. The next scenarios in the priority ranking are crossing or turning with lead objects. They are not that highly ranked, because compared to the others, interaction with lead objects is often rather straightforward. If none of these scenarios occurs, the lowest prioritised scenarios of crossing or tuning without conflict are checked. Details on this can be found in A2.2.1.5.

Independent of the urban or motorway use case for every scenario, PIs like average or maximum speed are extracted from the data. Besides those rather simple indicators that describe lag or variation of signals logged during the drive, there are a few indicators that require a more detailed explanation.

Due to confidentiality and the fact that engines and types of fuel differed among pilot sites, the RQ on efficiency cannot be answered based on measured fuel consumption. Instead, a theoretical measure for the energy demand [kWh] based on driving resistance forces is considered. In the applied formula in Equation 2.1, measured speed v_{ego} and accelerations a_{ego} determine the estimated energy demand E. This estimate is independent of the characteristics of the vehicles used during the pilots. The focus is on estimating the effect of a changed driving style on the energy demand of the vehicle. The energy consumption for the trip or a section is then calculated by computing the cumulative energy demand for the given distance.

Equation 2.1: Energy demand E calculated from the respective forces F.

$$E = \int (F * v_{ego})dt = \int \left((F_{air} + F_{roll} + F_{acc}) * v_{ego}\right) dt$$

Equation 2.2: Air resistance of the vehicle.

$$F_{air} = \frac{1}{2} * c_w * A * \rho_{air} * v_{ego}^2$$

Equation 2.3: Rolling resistance of the vehicle.

$$F_{roll} = f_r * m * g$$

Equation 2.4: Force needed to accelerate the vehicle.

$$F_{acc} = e_i * m * a_{ego}$$

The parameters used are listed in Table 2.3.

Table 2.3: Parameters used for the energy demand calculation.

Parameter	Definition	Value
$c_w * A$	Drag coefficient * reference area	$0.83\ m^2$
ρ_{air}	Air density	$1.2 \frac{kg}{m^3}$
f_r	Rolling resistance factor	0.015
m	Vehicle mass	$1400\ kg$
e_i	Factor for rotary masses	1.05

There are several RQs that address the impact of ADF use on the frequency of critical driving situations. Such types of events are also called incidents or near crashes. In this analysis, potentially critical situations are detected in the continuous driving data by applying certain objective thresholds (cf. Benmimoun et al., 2011). The frequency of such situations is analysed for selected driving scenarios. The situations identified based on objective driving data are not verified via video. Table 2.4 shows the applied thresholds. Situations detected via the threshold for lateral dynamic incidents are not included in the analysis, because there were too many detected events in the database to be a realistic estimate for critical situations.

Table 2.4: Incidents and their definitions (Benmimoun et al., 2011).

Incident type		**Criteria**
Distance incident	Front	Forward THW < 0.35 s and v < 20 km/h Forward THW < 0.5 s and v > 20 km/h Forward TTC < 1.75 s
	Side	Distance to side vehicle < 0.5 m
Dynamic incident	Longitudinal	Speed dependent ax between –6 m/s2 and –4m/s2
	Lateral	Speed dependent abs(ay) between 2.5m/s2 and 9m/s2

For motorway ADF, critical situations that occurred due to distances to rear traffic are not analysed based on the overall database but rather on an in-depth analysis for one selected Pilot site. This is done because of data quality issues in signals measuring the speed and position of rear traffic at some test sites. Due to the anonymised origin of the data, no reliable information can be derived for the rear traffic from the CDB. For the same reason, the RQ on driving behaviour of the rear traffic is also addressed by an in-depth analysis with data from one selected Pilot site only.

In the analysis, the PIs calculated per trip or per driving scenario are used to compare the behaviour between manual driving (baseline) and driving with ADF active (treatment). This is done separately per scenario type. Other potentially influencing factors like driver type or speed limit are not considered in the analysis.

As many of the analysed indicators do not fulfil the requirements of parametrical statistical tests, non-parametrical tests are used to compare baseline driving and driving with the ADF. Since such tests do not allow to include potential confounding factors like situational variables into the test, separate tests are calculated per indicator and driving scenario. Other situational factors are not included. Due to the overall large number of scenario instances, the statistical power is high and even minor effects can reach the 5% significance level. To provide information on the size of the reported effects, effect size calculated as Cohen's D and the percentual change between ADF and baseline are reported. For Cohen's D the following formula is used:

Equation 2.5: Cohen's D and the pooled standard deviation.

$$D = \frac{\mu_1 - \mu_2}{\sigma} \; with \; estiamted \; \sigma \; as \; \sigma = \sqrt{\frac{(n_1 - 1) * s_1^2 + (n_2 - 1) * s_2^2}{n_1 + n_2 - 2}}$$

And the change is calculated as:

Equation 2.6: Change in percent between baseline and ADF.

$$Change = \frac{m(ADF) - m(Baseline)}{|m(Baseline)|}$$

For the urban analysis a bootstrapping was done to ensure a more balanced representation of all the Pilot sites while still ensuring confidentiality. The need for this step originated from the fact that there were only three Pilot sites, of which one collected substantially more driving hours than the other two combined. Simply pooling the data, as was done with the motorway, would have led to a situation where the results from a single Pilot site would have dominated the results. In the bootstrapping process, samples were drawn from each of the Pilot sites multiple times by the respective Pilot data processing partners, creating multiple datasets. The sample sizes were set so that after pooling the data samples the data represented all the three Pilot sites in a more balanced way. Because the sampled data might have revealed the source of the data, as the records of the smaller Pilot sites would have been repeated at much higher probability in the final data, a small amount of noise was added to the measured variables.

The bootstrapping step ensured more balanced and confidential processing of the data but also complicated the interpretation of the results. Statistical testing of ADF vs. baseline differences had to be conducted based on the bootstrapped samples. Because of the balanced sampling, the variance in the bootstrapped estimates became larger and were no longer unbiased. In effect, this means that the tests are more conservative at detecting differences between ADF and baseline. On the other hand, if statistically significant differences were found, they should also be present in the original data. A detailed discussion on the process is given in Annex 4.

Due to this bootstrapping process, additional statistical testing was necessary. Simply pooling samples for testing the difference between baseline and ADF would violate the assumption of independent observations. Consequently, the degrees of freedom in the non-parametric test would be large and the p-values possibly too small, exaggerating the type I error. Therefore, the difference between baseline and ADF with bootstrapped data is done using the mean values of each sampling round. These are than compared using a Mann-Whitney-U-Test. The change is also given using the change between the mean values over the means of all sampling rounds, thereby preserving the character of the bootstrapping approach.

For the effect sizes, Cohen's D is used because the overall means and standard deviations are not affected by pooling. For visualisation of the results, histograms are used (similar to the motorway analysis) which contain the data from all the sampling rounds.

For quantifying the interaction between traffic participants, several measures are available. However, these measures are described on an absolute grid of an intersection as shown at the top of Figure 2.7. This data is not available from the recorded data at the urban Pilot sites as it is recorded from the perspective of the ego-vehicle. However, using the positions of the objects and the movement of the ego-vehicle (position, speed, heading angle), an absolute representation of the interaction between objects at intersections can be calculated.

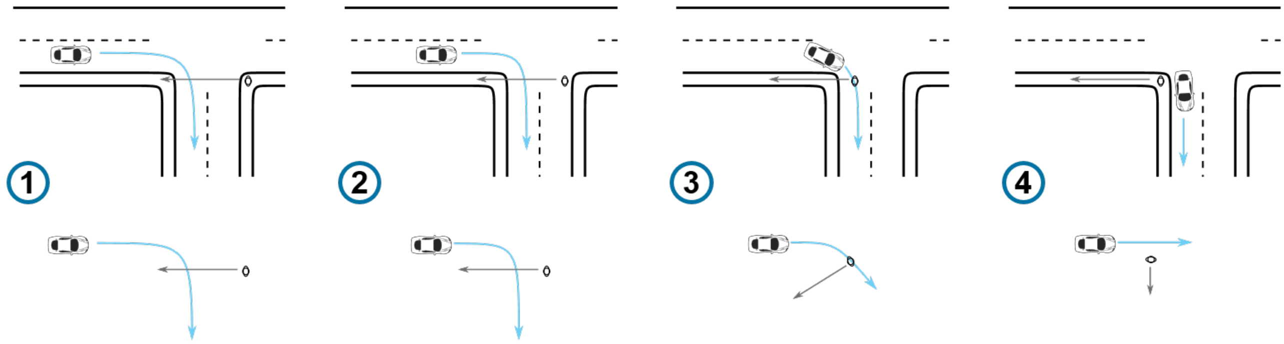


Figure 2.7: Absolute (top) and relative (bottom) trajectories of four distinct interaction steps between ego-vehicle and VRU at an intersection.

Since this process is computation intensive, a pre-filtering is done involving a coarse check for trajectories intersecting with the driven trajectory of the ego-vehicle. Only if an intersecting trajectory (or close pass-by) is found is the complete and accurate calculation carried out.

Using the absolute representation of the intersection interaction achieved through this calculation, the measures representing this interaction can then be calculated.

2.5.2 Parking

For parking ADF, the analysis differs from what has been described for motorway and urban driving. The main reason for this is that there was quite some variation in the parking systems tested in the pilots. There were ADFs restricted to parking on private property (home zone parking), whereas other ADFs were capable of selecting suitable parking spots and parking alongside public roads. This resulted in different experimental approaches and different addressed parking manoeuvres. The evaluation of parking ADF was therefore not based on a large database collected at various test sites. Instead, the different test sites were treated as separate studies in which acceptance and functioning of parking ADFs were assessed with experimental approaches tailored to the specificities of the tested parking system. As a consequence, no common definitions or scripts were used to calculate PIs that describe driving behaviour while parking. Instead, the Pilot data processing partners implemented the scripts needed to compare the measured driving behaviour during parking for parking with the ADF active and manual driving. This was done individually per Pilot site, because only three Pilot sites contributed to the analysis of parking ADF and the experimental setups (e.g., tested parking manoeuvres, experimental environment) differ fundamentally between Pilot sites. Furthermore, the Pilot data processing partner also chose the appropriate statistical test procedure, as studies on parking also differed with respect to the chosen experimental approach.

To merge the results across studies, for all analysed PIs information on the p-value, effect size and change in percent compared to baseline driving was shared based on a commonly used template. The overall results for parking ADF are based on this information.

2.5.3 Baseline for Technical and Traffic Evaluation

To evaluate the impacts of AD, it is necessary to compare the relevant PIs with an appropriate baseline. The baseline should allow deriving a clear picture on the impacts if automated driving were introduced on public roads. The baseline for comparison should make up a representative driver population, collecting naturalistic driver data under the same operating conditions as for the collection of data with the piloted ADF (e.g., in terms of locations and driving behaviour).

Automated driving is realised through state-of-the-art technology. The most relevant factor for this is greatly improved capabilities for environment perception using sophisticated sensors such as LIDAR or computer vision. This capability for environment perception allows for more in-depth analysis of the data recorded, such as the segmentation of data recorded in the driving scenario and the computation of more sophisticated PIs for evaluation. The consequence is that baseline data also needs to be recorded with vehicles with the same environment perception. As pre-series automated driving technology is costly and its operation in public traffic is bound to legal constraints, some restrictions affected the collection of baseline data.

The requirements for the collection of baseline data have been defined by Penttinen et al., (2019):

- A sufficient amount of baseline data should be collected covering all relevant driving scenarios for analysis, as well as environment conditions experienced in treatment.
- If possible, baseline data should be collected with not too small a group of ordinary drivers
- Vehicles for baseline collection should be the same or comparable to those for treatment.
- The format of the logged dataset should be comparable.
- Continuously active ADAS like ACC should be inactive for baseline collection.

Two practical options for baseline collection were considered:

1. Collection of baseline data as part of a single study with participants contributing to both baseline and treatment data collection.
2. Separate collection treatment and baseline data; baseline data collection preferably with non-professional drivers.

Depending on the Pilot site, both options were considered. Given the legal constraints for the operation of vehicles with AD capabilities and environment perception, several Pilot sites were only able to use professional drivers for the acquisition of baseline data. As a result, the driving may not be completely representative of how average human drivers would drive. The safety driver’s task is to operate the vehicle safely to avoid any damage to the vehicle or others. Thus, it is likely that the safety driver will respond very quickly to safety-critical situations, possibly even pre-empting such a situation before it becomes imminent. A further relevant question is whether normal driving behaviour of safety drivers differs considerably. Safety drivers did not receive any special instructions, apart from those relevant to safety on any particular behaviour. The implemented pseudonymised approach for vehicle data, which was implemented to avoid benchmarking, does not make it possible to compare professional drivers’ baseline behaviour to that of ordinary drivers.

Questionnaire answers by professional drivers are separated from those provided by ordinary drivers.

Of the 400,000 kilometres worth of data collected for motorway systems, half was recorded as baseline. This includes driving efforts which were undertaken to collect data with traffic jam-only systems. Data outside the ODD of traffic-jam specific systems could not be used as general baseline data, given the test setup and constraints to the systems' capabilities outside the ODD. For urban systems, of the 1120 hours of driving data, 130 hours of baseline data was collected. For parking systems, the number of baseline experiments is reported with the results.

2.5.4 Processing of Vehicle Data

In the following, the approaches to processing the data collected in the pilots are given separately for motorway as well as urban ADF and parking ADF.

2.5.4.1 Motorway & Urban ADF

For the Technical & Traffic Evaluation, the data logged in a single vehicle (CAN-data, videos) was analysed stepwise. The evaluation workflow is illustrated in Figure 2.8.

The starting point for this process was time-series data logged during the test drives. At the Pilot sites, this data was converted into a data format commonly defined and used in L3Pilot (see D5.1 – Pilot Tools for L3Pilot, Nagy et al., 2018). Then, the data was transferred to the Pilot data processing partners. Here, after data quality checks, scripts developed within L3Pilot were run on the data. These scripts calculated derived measures and identified driving scenarios and further relevant situations in a harmonised way (see D5.1, Nagy et al., 2018). By using the same scripts on data logged at various Pilot sites, it was ensured that the identified driving scenarios and the analysed PIs would be comparable across Pilot sites.

For motorway and urban ADFs, after processing the data by the Pilot data processing partners, relevant performance indicators were extracted from the data per trip and uploaded to a cloud-based database. The database contained tables with indicators calculated per trip (like average energy consumption) and indicators calculated per driving scenarios (like average speed). The database was defined in a way that it was not possible to trace from which Pilot site the data originated. By this, it was ensured that confidentiality was kept and that it was not possible to compare results among Pilot sites.

Urban data processing implemented an additional bootstrapping step before the data upload (cf. Section 2.5.1.). This meant that a few further steps were performed before the upload of the urban data. Firstly, the sample sizes had to be derived for each Pilot site. Secondly, using these aggregated sample sizes, the bootstrapping process could be performed locally at each Pilot data processing partner and finally, the bootstrapped data could be uploaded.

After upload of all data by the Pilot data processing partners, the complete data set is downloaded and used for Technical & Traffic analyses and as input to impact assessment.

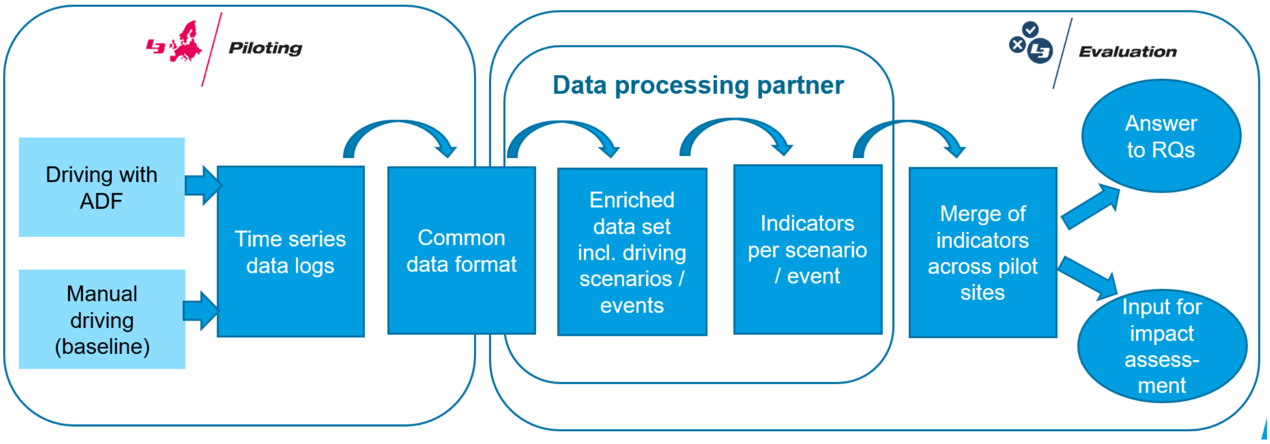


Figure 2.8: The overall workflow for Technical and Traffic evaluation in L3Pilot.

During the described process there were three data sets that differed with regards to level of detail, respectively the amount of information that they contained and regarding access rights. The process started with the dataset logged in the vehicles. These datasets differed between Pilot sites with regard to e.g. format and structure and containing multiple signals but also video and relevant meta data. The raw data sets were stored on the premises of the vehicle owners. The raw time series data logs were converted into the common data format by the vehicle owners and transferred to the Pilot data processing partners. This yielded the second data set. Finally, as the third set, processed data were stored in a CDB, accessible to the evaluation partners. The complete dataset for the combined analysis for the motorway ADF contained data from 12 Pilot sites across Europe, resulting in over 2175 hours of motorway data within the ODD for the analysis of the ADFs.

2.5.4.2 Parking

For parking ADF, the process and the analysis differ compared to Motorway and Urban. There are several reasons for this:

- The data is logged per parking manoeuvre and not per trip with multiple scenarios.
- The functions are tested in dedicated experiments and not during experimental trips with or without the function active.
- The signals available for analysis (e.g., to evaluate the precision of the manoeuvres) vary between Pilot sites and depends on the function.
- The setups of the experiments (e.g., the types of parking manoeuvres included) vary between Pilot sites.

The data logged in the experiments on parking were provided to the evaluation partners responsible for the analysis. This was not done in a common data format; instead the format was agreed on individually between vehicle owner and evaluation partner. The evaluation partner analysed the parking data with respect to the RQs. This includes the full process from data pre-processing to statistical testing. The details of how this was done (e.g., how the start and end point of a manoeuvre is defined) was decided by the evaluation partners based, for instance, on the functionality, investigated manoeuvre types and on the experimental setup.

2.5.5 Database and Data Filtering

All the results are based on data collected at various Pilot sites, pre-processed by the Pilot data processing partners and uploaded to the CDB. Table 2.5 gives an overview of the dataset used for the analyses.

As described earlier, the process for scenario detection and PI calculation for motorway and urban ADF is completely automatised. On the highest level and for all analyses, the complete dataset is filtered in such a way that only the relevant road category (motorway or urban area) and the conditions (baseline and ADF active) are included.

Furthermore, filtering criteria are implemented at the level of the driving scenario to avoid including obviously unreliable scenario instances. The main inclusion and exclusion criteria are set, first based on the duration of the detected scenarios, allowing unrealistically short instances to be excluded. Table 2.5 lists the different scenarios, used inclusion / exclusion criteria, and the number of instances remaining after filtering. The number of scenario instances listed in the result sections can be lower than the number shown in Table 2.5 because not all PIs are available for every scenario instance (e.g., TTC is not always defined).

Throughout the analysis, unrealistically short scenario instances are excluded from the data. For this, the following filter criteria are used:

- Trip-based indicators: minimum overall duration of included trip sections is 5 minutes.
- Cut-in scenarios: minimum duration of included scenario instances is 0.9 seconds.
- Approaching a traffic jam and driving in a traffic jam: minimum duration of included scenario instances is 20 seconds.
- All other scenario types: included scenario instances have a duration of at least 2.0 seconds.

Table 2.5: Impact of applied filter criteria on the database for motorway ADF.

	N instances Baseline			N instances ADF			N instances Total		
	Before filtering	After filtering	% kept	Before filtering	After filtering	% kept	Before filtering	After filtering	% kept
Uninfluenced driving	122,234	99,757	**81.6**	232,517	190,979	**82.1**	354,751	290,736	**82.0**
Following	78,995	64,740	**82.0**	178,247	150,066	**84.2**	25,7242	214,806	**83.5**
Approaching a lead vehicle	17,927	17,859	**99.6**	16,257	16,184	**99.6**	34,184	34,043	**99.6**
Cut-In	2,394	2,391	**99.9**	4229	4,227	**100.0**	6,623	6,618	**99.9**
Lane change	30,288	26,732	**88.3**	27,151	25,141	**92.6**	57,439	51,873	**90.3**
Approaching a traffic jam	126	125	**99.2**	266	266	**100.0**	392	391	**99.7**
Driving in traffic jam	487	390	**80.1**	2,968	1,976	**66.6**	3,455	2,366	**68.5**
Trip section	1,460	1,191	**81.6**	3,816	3,007	**78.8**	5,276	4,198	**79.6**

Similar filtering criteria were applied for the urban use case and analysis. The following criteria were used:

- Trip-based indicators: minimum overall duration of included trip sections is 10 seconds
- Cut-in scenarios: minimum duration of included scenario instances is 0.9 seconds.
- Intersection scenarios: minimum duration of included scenario instances is 0.5 seconds.
- All other scenario types: minimum duration of included scenario instances is 2 seconds.

Table 2.6: Impact of applied filter criteria on the database for urban ADF.

	N instances Baseline			N instances ADF			N instances Total		
	before filter	after filter	% kept	before filter	after filter	% kept	before filter	after filter	% kept
Approaching a lead vehicle	7,051	6,761	**95.9**	5,775	5,634	**97.6**	12,826	12,395	**96.6**
Approaching static object	93,702	14,077	**15.0**	5,526	1,009	**18.3**	99,228	15,086	**15.2**
Crossing without Conflict	14,913	14,913	**100**	1,237	1,237	**100**	16,150	16,150	**100**
Crossing with lead object	60,058	59,991	**99.9**	1,1606	11,579	**99.8**	71,664	71,570	**99.9**
Crossing with laterally moving object	96,412	96,176	**99.8**	111,807	111,557	**99.8**	208,219	207,733	**99.8**

	N instances Baseline			N instances ADF			N instances Total		
Following a lead object	5,517	55,17	**100**	1,157	1,155	**99.8**	6,674	6,672	**99.9**
Lane change	21,301	17,217	**80.8**	16,085	14,125	**87.8**	37,386	31,342	**83.8**
Overtaking with oncoming traffic active	30,759	28,375	**92.3**	4,990	4,169	**83.6**	35,749	32,544	**91.0**
Overtaking with oncoming traffic passive	23,741	685	**2.9**	5,720	93	**1.6**	29,461	778	**2.6**
Turning without conflict	351	351	**100**	423	399	**94.3**	774	750	**96.9**
Turning with lead object	11,378	11,378	**100**	1,347	1,347	**100**	12,725	12,725	**100**
Turning with laterally moving object	15,477	15,472	**99.9**	2,451	2,450	**99.9**	17,928	17,922	**99.9**
Uninfluenced driving	19,219	19,219	**100**	15,136	15,136	**100**	34,355	34,355	**100**

For the analysis of parking ADF, data from five studies conducted at three different Pilot sites can be used. The experimental setup at each Pilot site was specific to the parking scenario being tested and in principle differed a lot among studies (ego-vehicle's initial approaching speed, parking slot sizes, presence/absence of other parked vehicles adjacent to the parking slot, parallel/perpendicular parking manoeuvre relative to the parking slot longitudinal axis, etc.). Table 2.7 gives an overview of the studies included in the analysis.

Table 2.7: Overview of parking studies.

	Study1	Study2	Study3	Study4	Study5
N Driver	65	3	21	20	
Driver type	Non-professional	Professional	19% professional	5% professional	
Age	41 (sd = 10.8)	29 (sd = 3.6)	47 (sd = 15.9)	39 (sd = 11)	
% Female	23%	0%	0%	30%	
N Manoeuvre Total	692	51	21	1309	1750

2.5.6 Additional Data Evaluation using AIM Mobile Traffic Acquisition

The project L3Pilot aims to further develop, mature and assess current ADF regarding traffic safety and efficiency. This includes ADF behaving more like human drivers but without their erroneous behaviours that cause so many deaths and severe injuries worldwide. One aspect is to measure how SAE L3 vehicles behave and how they interact with other manually driven or vulnerable road users. Another is to measure and quantify the normal behaviour of human drivers, which can be seen as a baseline for how ADF should work. Here, the AIM (Application Platform for Intelligent

Mobility) methodology is briefly described, as applied to a roundabout and a signalised intersection.

The AIM Mobile Traffic Acquisition system is part of the AIM test field in Braunschweig, Germany. It consists mainly of one or two combinable mobile stereo camera poles (see Figure 2.9, left), which can be used for the detection, classification, and tracking of road users and generates road user trajectories and augmented video data that form the basis of road user behaviour analysis. The trajectories and video data provide road user ID, time stamp, position, speed, acceleration, heading, and object size for each road user at 25 frames per second.

The trajectories and converted video images provided by the cameras are stored in a local database, in line with GDPR regulations, for offline behaviour analysis and validation. The data from the stereo cameras are anonymised online by converting the high-resolution images to low resolution and augmenting the scene with 3D-bounding boxes. Three camera poles of the AIM mobile units were installed at a roundabout as shown in Figure 2.9 (right).

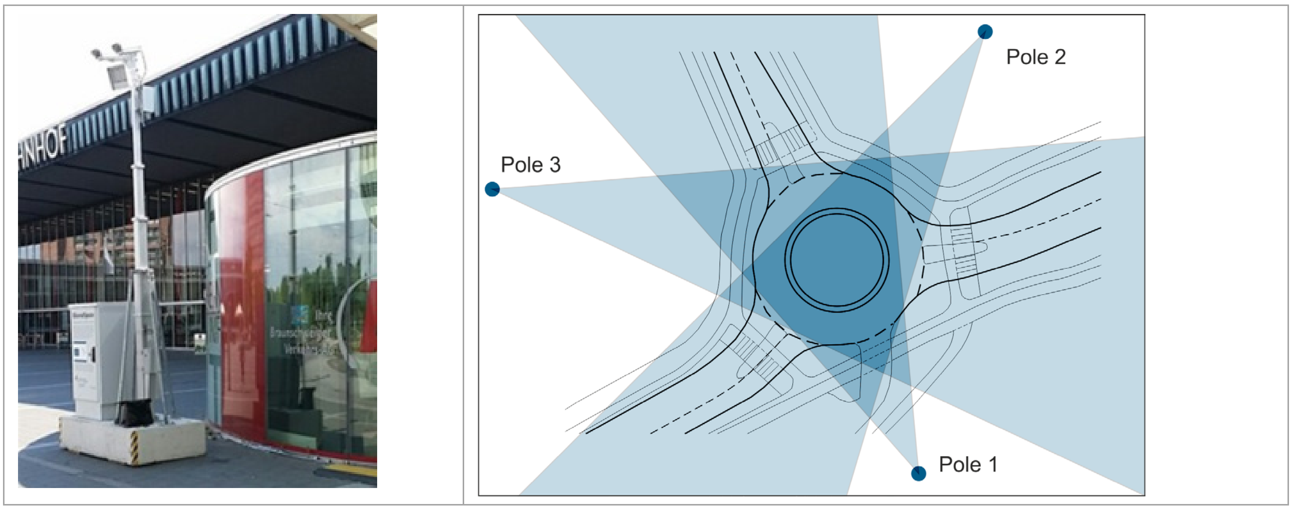


Figure 2.9: AIM Mobile Traffic Acquisition mobile measurement unit (left); selected roundabout and schematic diagram of three installed AIM mobile units with their fields of view (right).

2.5.6.1 Obtaining road user behaviour

Understanding the behaviour of road users with regard to their own driving profile in terms of traffic efficiency, as well as in any safety-related interactions with other road users, is of key interest when speaking about maturing automated driving functions (ADF). On the one hand, ADF should “behave” as manual drivers, but on the other they should be more efficient and much safer than manual drivers. Therefore, measuring current road user behaviour considers two goals: (i) measure and understand how ADF behave regarding their own driving behaviour, and how ADF behave in case of interactions with manual road users, so as to yield quantified statements on the matureness of ADF; and (ii) measure and understand what the nominal driving behaviour of “manual” road users is and how they interact with other “manual” road users, so as to yield quantified statements on how ADF should behave.

Based on the numerical road user trajectories provided by the AIM Mobile Traffic Acquisition units, different variables of road user behaviour can be obtained. The variables considered were those that describe driving behaviour, road user kinematics (speed, acceleration), manoeuvre precision, and journey times. Other variables considered were those that describe road users interacting with other road users in car-following, crossing or merging scenarios in terms of post encroachment time (PET), predicted post encroachment time (pPET), alternatively called time advantage (TAdv), time to collision (TTC), time headway (THW), minimum distance between two interacting road users, and number of encounters and number of critical encounters (near-crashes).

Altogether, four weeks' worth of augmented video data (for verification purposes) and trajectory data (for analysis) of all road users were recorded at the roundabout. Since the roundabout is a special traffic infrastructure with properties that differ somewhat from intersections, it made sense not only to consider driving scenarios such as free/undisturbed driving, merging or crossing, and car-following situations, but also to look at the different phases of the roundabout, which are entering, circling and exiting. Relevant and feasible RQs could thus be asked on this basis. The whole dataset was reduced to the relevant situations, i.e., the trajectories between 7 am and 7 pm. The ADF vehicle was driven only between 3 pm and 7 pm. The trajectories were separated into several subsets to distinguish between the relevant phases defined above, then filtered to detect outliers (e.g., due to acceleration and position noise), ID changes, and trajectory losses. For instance, the Unscented Kalman Filter (UKF) was applied to reduce the lateral position noise of the ADF and baseline trajectories.

The remaining trajectory data was analysed by applying the metrics to answer the relevant RQs. Additionally, thresholds were set to further reduce the amount of data. The baseline of the manually driven vehicles was selected randomly, while maintaining similar and comparable conditions to the drives with ADF; i.e., the baseline situations should have the same paths and overlap the ADF drives by ±30. All ADF situations and baseline situations were verified manually by considering the video scenes. Where statements were based on the whole remaining dataset, some baseline situations were verified randomly. For comparison of ADF and baseline driving and interaction behaviour, the relevant data was analysed by applying inference statistical methods, including e.g., Bonferroni correction of the level of significance.

2.5.6.2 Specific Methods for Data Analysis Relevant for Answering RQs

Manoeuvre Precision

From an external, subjective point of view, the intention of vehicles with ADF besides entering, circling, and exiting the roundabout is not clear. Therefore, in this RQ, manoeuvre precision was measured by variably intersecting certain points on the roundabout. Four virtual loops were placed where they would be crossed by all baseline and ADF vehicles at certain spots on the roundabout. The intersection points were mapped on histograms showing the distance d_i of each vehicle i to the mean distance $m(d)$ of each loop.

Lane Keeping Performance

Lane keeping is considered as variability of each trajectory of interest from a prototypal trajectory *R*., which can be defined as the middle of the lane (georeferenced by the given lane-mark information) or the most representative path adopted by the average of all road users, etc. Here, the most representative path *R* was computed from the whole dataset of manually driven vehicles. The directional variability, which is the standard deviation of the lateral distance *sd*(*ld*), between *R* and any other trajectory T_i, *i = 1...N*, at each single trajectory point, thus reflects lane keeping. The directed Hausdorff distance $d_H(R,T_i)$ is a well-suited measure for this purpose when given two sets of points of *R* and T_i with the norm ||.||:

Equation 2.7: Directed Hausdorff distance.

$$d_H(R,T_i) = \underset{x \backslash in\, R}{max}\{ \underset{y \backslash in\, T_i}{min}\{|| x,y ||\}\}$$

The resulting undirected Hausdorff distance $D_H(R,T_i)$ is given by a resulting *N-x-N* matrix:

Equation 2.8: Undirected Hausdorff distance.

$$D_H(R,T_i) \ = \ max\{d_H(R,T_i), d_H(T_i,R)\},$$

for which its smallest sum of distances to any other trajectory reflects *R* for comparing ADF vehicles with the baseline:

Equation 2.9: Most representative path R.

$$R \ = \ min\,\{sum_{i=1..N}\ D_{Hi}(R_1,T_i),...,sum_{i=1..N}\ D_H(R_N,T_i)\}.$$

The lateral distance *ld* between the corresponding data points of any T_i and *R* is the relevant parameter with which to analyse the lane keeping performance of both the ADF vehicle and the baseline, which can be derived by computing the scalar projection of the distance vector Δ*d* to the orthonormal vector *n* to the heading of *R*, where *θ* is the angle between *n* and *Δd*:

Equation 2.10: Lateral distance between corresponding data points.

$$ld \ = \ ||\Delta d|| \ \cdot \ cos\,\theta.$$

Traffic Flow / Journey times

Due to the fact that exactly one ADF vehicle was driving in the roundabout at the same time, its impact on traffic flow is close to zero in a view of macroscopic traffic parameters. Instead, the ADF vehicle is–due to safety requirements–expected to have a higher journey time *JT* (time needed to get from A to B given the road infrastructure) driving through the roundabout than manually driven vehicles. Therefore, the journey times of the manually driven vehicles and the vehicles with ADF were measured from a defined entering position in the North of the roundabout to a defined exiting position in the East of the roundabout. Additionally, the expected difference between the journey times of the ADF vehicles and the baseline vehicles can be considered as loss time, which quantifies the “loss” of the current maturing level of the ADF in comparison to the baseline. The journey time *JT* between entering the roundabout in the North and exiting it in the East was calculated on the basis of sum of the journey times of three sections “entering”, “circling” and “exiting”:

Equation 2.11 Journey time.

$$JT = JT(\text{entering}) + JT(\text{circling}) + JT(\text{exiting}).$$

Interaction Scenarios in the Roundabout

Due to the lack of interaction between the ADF vehicle with other road users on the roundabout, the researchers focused on a detailed analysis of human drivers interacting with other road users. For the scenarios car-following, VRU crossing and merging, car-following scenarios were divided into the phases “entering”, “circling” and “exiting”. The thresholds of THW were set to six seconds each and all situations above this value were ignored. VRU crossing scenarios were considered by the sub-scenarios “yielding” and “non-yielding”. The PET threshold was set to five seconds and all interaction situations above this limit were ignored.

Merging scenarios on the roundabout did not take place at an ideal 90° degree angle but at much lower angles, evolving towards car-following. According to the yielding behaviour of the ego-vehicle (i.e., the entering vehicle), the merging scenarios were divided into “yielding” and “non-yielding”. The PET threshold was set to six seconds to consider all interacting road users entering and circling below this threshold.

2.6 Method for User and Acceptance Evaluation

In order to carry out the User and Acceptance Evaluation for the three different ADF considered for the evaluation in L3Pilot (see section 1.4), three Pilot site questionnaires were designed, one for each environment, with function-specific questions for ADF operating in each environment. This method allowed us to collect responses that are context and ADF specific.

The questionnaire was in two parts (included as Annex in D3.4 – Evaluation Plan (Innamaa et al. 2020), the first of which was administered before the Pilot drives commenced. The first part included questions related to socio-demographic factors (age, gender, country of residence, education level, employment status, income, and family size), vehicle use and purchasing decisions, driving history, in-vehicle system usage, activities while driving, trip choices, and mobility patterns. The data was then used to create different user groups for the evaluation, and to understand the impact of various socio-demographic factors on participants’ acceptance and perception of the ADFs.

The second part of the questionnaire was administered immediately after the Pilot drive concluded, or the final Pilot drive if a participant participated in more than one drive. It examined participants’ initial reactions to a given ADF, including acceptance, safety and comfort. To examine whether participants felt they would change any of their behaviours should they have access to that ADF in their daily life, they were re-asked questions about vehicle use and purchasing decisions, driving history, in-vehicle system usage, engagement with non-driving tasks, trip choices, and mobility patterns. The questions in this section were phrased to address the specific ADF under investigation, the only exception being motorway and traffic jam ADF, which utilise the same questions, because they have similar ODD.

As an optional additional section, where feasible, users' controllability and performance during and after a take-over was evaluated mid-drive, following any need to resume manual control from the ADF. For this analysis, drivers were asked immediately after a take-over scenario to rate the criticality of the preceding situation as a whole on a ten-point scale, ranging from harmless (1) to uncontrollable (10). The scale is based on that by Neukum et al. (2008) and allows a direct comparison of drivers' own evaluation of the take-over and the post-drive evaluation by expert raters. This data was collected for ordinary drivers and at Pilot sites where the safety protocol permitted mid-drive evaluations.

2.6.1 Questionnaire Methodology

In total, data from 354 unique drivers was collected for the Motorway Pilot Site Questionnaire from the CDB. The data was further tabulated into three groups. Table 2.8 lists the demographic details of the participants in the three driver and test type groups: professional drivers from the Pilot sites, ordinary drivers from the Pilot sites (some including Wizard-of-Oz studies conducted on test tracks), and ordinary drivers from simulator studies. In total, data from 175 participants was collected for the Urban Pilot Site Questionnaire from the CDB. The data consisted of 15 professional drivers and 160 passengers; the data was analysed without separating it into different groups. Table 2.8 shows the demographic information of the participants of the Urban Pilot Site Questionnaire.

Table 2.8: Demographic details of participants in the motorway and urban Pilot site questionnaires.

	Professional Drivers from Motorway Real Pilot Site (N = 58)	Non-Professional Drivers from Motorway Real Pilot Site (N = 236)	Non-Professional Drivers from Motorway Simulator Studies (N = 60)	Urban Pilot Site Questionnaire (N = 175)
Gender	• 47 Male (81%) • 9 Female • 1 Other • 1 Prefer not to say	• 171 Male (72%) • 48 Female • 1 Other • 16 missing data	• 31 Male (52%) • 29 Female	• 115 Male (78%) • 60 Female (22%)
Age	• Range: 23-57 years • M = 40.11 • SD = 11.26	• 22-70 years • M = 40.72 • SD = 11.28 • 16 missing data	• 22-62 years • M = 39.25 • SD = 11.88	• 20-68 years • M = 39.47 • SD = 11.29

	Professional Drivers from Motorway Real Pilot Site (N = 58)	Non-Professional Drivers from Motorway Real Pilot Site (N = 236)	Non-Professional Drivers from Motorway Simulator Studies (N = 60)	Urban Pilot Site Questionnaire (N = 175)
Driving Experience (years)	• < 1 year (0%) • 1-2 years (0%) • 2-10 years (28%) • > 10 years (72%)	• < 1 year (0%) • 1-2 years (0%) • 2-10 years (18%) • >10 years (82%)	• < 1 year (0%) • 1-2 years (2%) • 2-10 years (23%) • >10 years (75%)	• < 1 year (1%) • 1-2 years (1%) • 2-10 years (25%) • > 10 years (74%)
Driving Experience (distance in km)	• < 2000 (3%) • 2000-5000 (3%) • 5000-10000 (22%) • 10000-15000 (12%) • 15000-20000 (14%) • 20000-50000 (34%) • > 50000 (10%)	• < 2000 (4%) • 2000-5000 (8%) • 5000-10000 (14%) • 10000-15000 (20%) • 15000-20000 (30%) • 20000-50000 (21%) • > 50000 (2%)	• < 2000 (8%) • 2000-5000 (18%) • 5000-10000 (18%) • 10000-15000 (17%) • 15000-20000 (17%) • 20000-50000 (18%) • > 50000 (3%)	• < 2000 (6%) • 2000-5000 (16%) • 5000-10000 (18%) • 10000-15000 (17%) • 15000-20000 (21%) • 20000-50000 (21%) • > 50000 (2%)

For parking, questionnaire data was collected in three studies conducted at three different test sites. Since the studies on parking ADF differed substantially between test sites, e.g., with regard to tested manoeuvres, test environment and experimental approach, the subjective data for the evaluation of parking ADFs is not merged on the level of single questionnaires but on the level of studies. All studies are analysed separately, then the results from the different studies are merged in such a way that each study contributed one data point to the overall results. Table 2.7 is a summary of the database for the parking ADF.

As shown in the previous section, there were five RQs related to User's Acceptance and Awareness. To answer each of these RQs, questions were administered using a six-point scale, unless otherwise stated, whereby 1 = Strongly disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly agree and 6 = Don't know.

To answer the L3Pilot User and Acceptance RQs, each relevant questionnaire item is explored according to the driver/test type. To provide an overview of the spread of participants' responses, the percentages of each response are presented in figures. In addition, findings are described to answer each RQ under each figure. Finally, Hierarchical Regression was conducted to investigate how user experience, driver type, and system familiarity affect willingness to use the respective systems. All the results are included in Section 4.3.

2.6.2 Video Analysis

While the main methodology for the User and Acceptance research area was the developed questionnaire, some behavioural aspects can be studied in-depth by analysing videos of the drivers during exposure. For this purpose, several cameras were installed in the cabin to observe the driver interacting with the automated driving function. Especially the transition phases are of particular interest, since the behavioural process of handing over control to the vehicle and taking it back has implications not only for safety but also HMI design. For several RQs, e.g., “How do drivers respond when they are required to retake control?”, video data can provide an insight into the actual response process. For this purpose, observable measures that are objectively identifiable need to be determined and collected in a code book which is used for annotations. The behavioural features are then labelled frame by frame by annotators.

To evaluate the controllability and safety of take-over situations within L3Pilot, a video-based procedure was used on data from two Pilot sites. In this method, expert raters judged the criticality of the preceding take-over situation as a whole on a ten-point scale ranging from harmless (1) to uncontrollable (10). This method, called the take-over-controllability-rating (TOC-rating, Naujoks et al., 2018, www.toc-rating.de/en), provides a uniform and easy to understand approach to evaluating take-over situations. The TOC-rating was developed to provide a more holistic assessment of take-over situations that goes beyond vehicle parameters, such as the deviation of speed or lateral control, but also considers traffic violations (such as missing safety-related glances or absent indicator use) as well as the observed emotions of the driver. One advantage of the TOC-rating that makes it especially suitable for the needs of L3Pilot is that an objective (i.e., standardised and common) rating can be applied across situations and drivers.

For the in-depth analysis of video data, a code book was prepared that contains all relevant features that relate to the RQs of interest. Due to various technical and privacy limitations, video data could not be collected at the main Pilot sites. Therefore, the video data was collected in a Wizard-of-Oz study on public roads with 30 ordinary drivers. Thus, the automation is simulated by a wizard driver while the participant is under the impression that the vehicle is driving automated. The main focus was on the transition phases between human driver and system (hand-over) and the other way around (take-over), 30 seconds before and after the transitions were annotated. For comparison purposes, data from a manual baseline drive was also collected on the same route as the Wizard-of-Oz study.

3 Results of the Technical and Traffic Evaluation

Following the description of applied methods and the data processing, this chapter presents the results of the Technical and Traffic evaluations performed within L3Pilot. The results from the evaluation for the motorway, urban and parking ADF are presented. The most common way to present the results is visualisation by means of a histogram per PI and driving scenario. To answer the RQs, the histograms give the results of the statistical tests, the relative change, and the effect size.

In general, when interpreting the presented results, it should be noted, that the evaluated data does not represent a single system. While all the systems were set up to facilitate the use cases of motorway driving, urban driving, and parking and therefore share broad similarities, individual parameterisation per system - and thus the resulting behaviour - may differ. The strict requirement for avoidance of benchmarking between the systems does not allow dedicated analysis of an individual system. Effects found in the merged data for one ADF category may not be present in some of the systems evaluated. Furthermore, it needs to be emphasised that the piloted systems are at a pre-series stage. Operating them at this stage may require a safer parameterisation than for the operation of a series production AV. Additionally, operation of the systems may be constrained by special legal requirements which may influence the parameterisation of the systems.

3.1 Motorway

In the following, the RQs for the motorway use-case presented within Section 2.3.1 are answered using the data from the Pilot sites. For all RQs presented in this chapter, the focus is on the behaviour of the vehicle fitted with an ADF. It is not analysed how overall traffic changes through introducing an ADF (analysed in Bjorvatn et al., Deliverable D7.4 – Impact Evaluation Results 2021). Due to the prototype nature of the tested ADFs and impact of the required testing conditions (e.g., presence of a safety driver), RQs in the field of technical and traffic analysis sometimes cannot be fully answered. This is especially the case when it comes to the evaluation of ADF in rare events or boundary conditions.

The RQs are in order of their respective number. Unless otherwise stated, the graphs show the distribution of the analysed indicators, always comparing baseline driving with driving with the ADF active.

3.1.1 RQ-T1 - RQ-T3: Behaviour of the ADF

The technical research questions RQ-T1 to RQ-T3 aim at evaluating the functionality of the ADF, i.e., whether the tested ADFs work as intended. Since the tested ADF sometimes relied especially in boundary conditions on the safety driver, aspects like the expected frequency of take-over requests or the duration of sections with ADF being available might differ from what can be expected from a market ready ADF. However, a descriptive analysis is provided that gives information on the availability of the tested ADFs, on the frequency of take-over requests, and on the time the ADF could be used in real traffic.

In total, there are 2805 trips for which the proportion of time the ADF was available and the average duration of section with ADF active can be calculated. In Figure 3.1, in parts of a trip marked as being on the motorway and not baseline, the ADF was on average available 87% of the time (standard deviation sd=19.4%, median = 96%). The average duration of trip sections with ADF active was about 3.9 minutes (sd=3.2 minutes). The maximum duration of a section where the ADF was continuously active was 77 minutes. In the interpretation of these figures, it should be kept in mind that some of the tested ADFs did not support automated lane changes. This means that every lane change initiated by the driver of such an ADF ended the continuously active section.

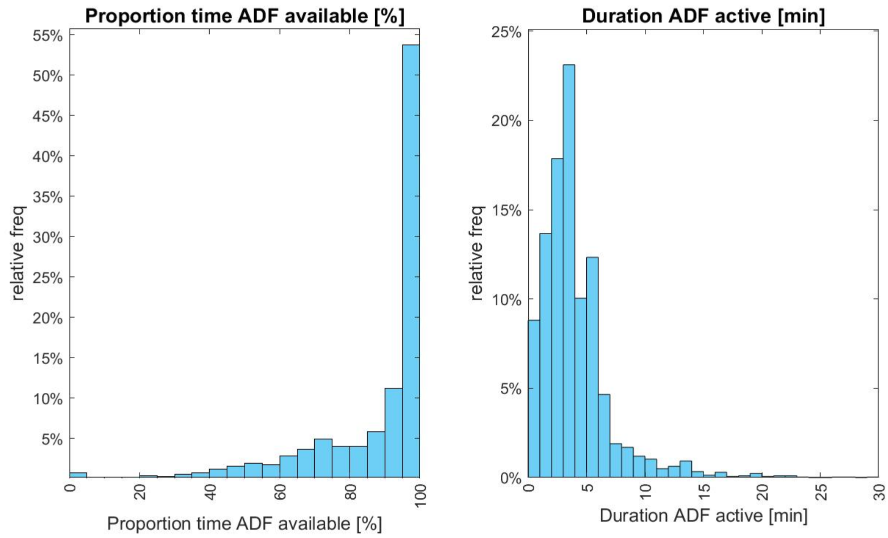


Figure 3.1: Distribution of the duration of a section where the ADF was available (left) and active (right).

The frequency of take-over requests is another indicator that needs to be interpreted carefully. In Figure 3.2, there is a large proportion of trips (54%) during which no take-over requests were reported. One reason for this might be the safety drivers who were instructed to take control back before anything unusual could happen. During 65% of trips there was less than one take-over per hour. For the cumulative distribution shown in Figure 3.2, the trips have been weighed based on their duration. The weighted average frequency of take-over requests per hour is 5.1 (sd = 19.0), which corresponds to one take-over request every 12 minutes.

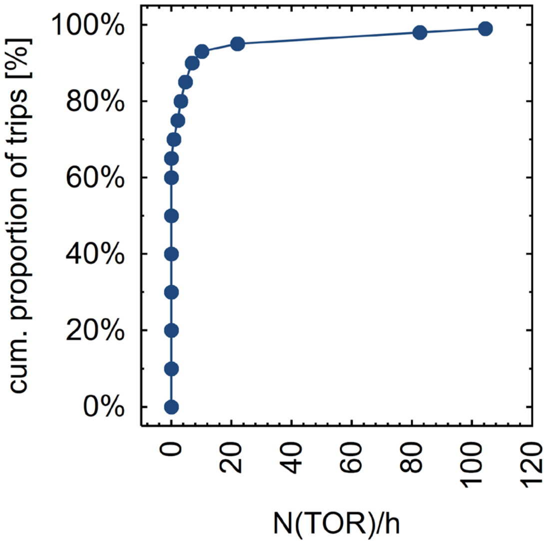


Figure 3.2: Cumulative distribution of frequency of take-over requests (TOR) per hour calculated for trip sections where ADF was available.

3.1.2 RQ-T6: What Is the Impact of ADF on Vehicle Dynamics in Defined Driving Situations?

To evaluate the impact on vehicle dynamics, longitudinal (ax) (Figure 3.3) and lateral accelerations (ay) (Figure 3.4) were analysed. The following parameters were derived: minimum longitudinal acceleration (that is maximum deceleration), min(ax), maximum longitudinal acceleration (max(ax)), standard deviation of longitudinal acceleration (sd(ax)), maximum absolute lateral acceleration (max(abs(ay)), and standard deviation of lateral acceleration (sd(ay)). For indicators describing the effects of longitudinal dynamics, the results show reduced maximum accelerations when driving with the ADF for most driving scenarios except car following.

For the maximum deceleration (min(ax)) the results are more varied: during car following, approaching a lead vehicle, cut-ins, and lane changes, the ADF decelerates more strongly than a driver does in manual driving. However, during uninfluenced driving, approaching a traffic jam and driving in a traffic jam, the deceleration is reduced. Consequently, also the results in the variation of longitudinal acceleration are mixed across driving scenarios. The results of statistical tests are presented in Table 3.1 in terms of Z-score, p-Value, relative change, and effect size.

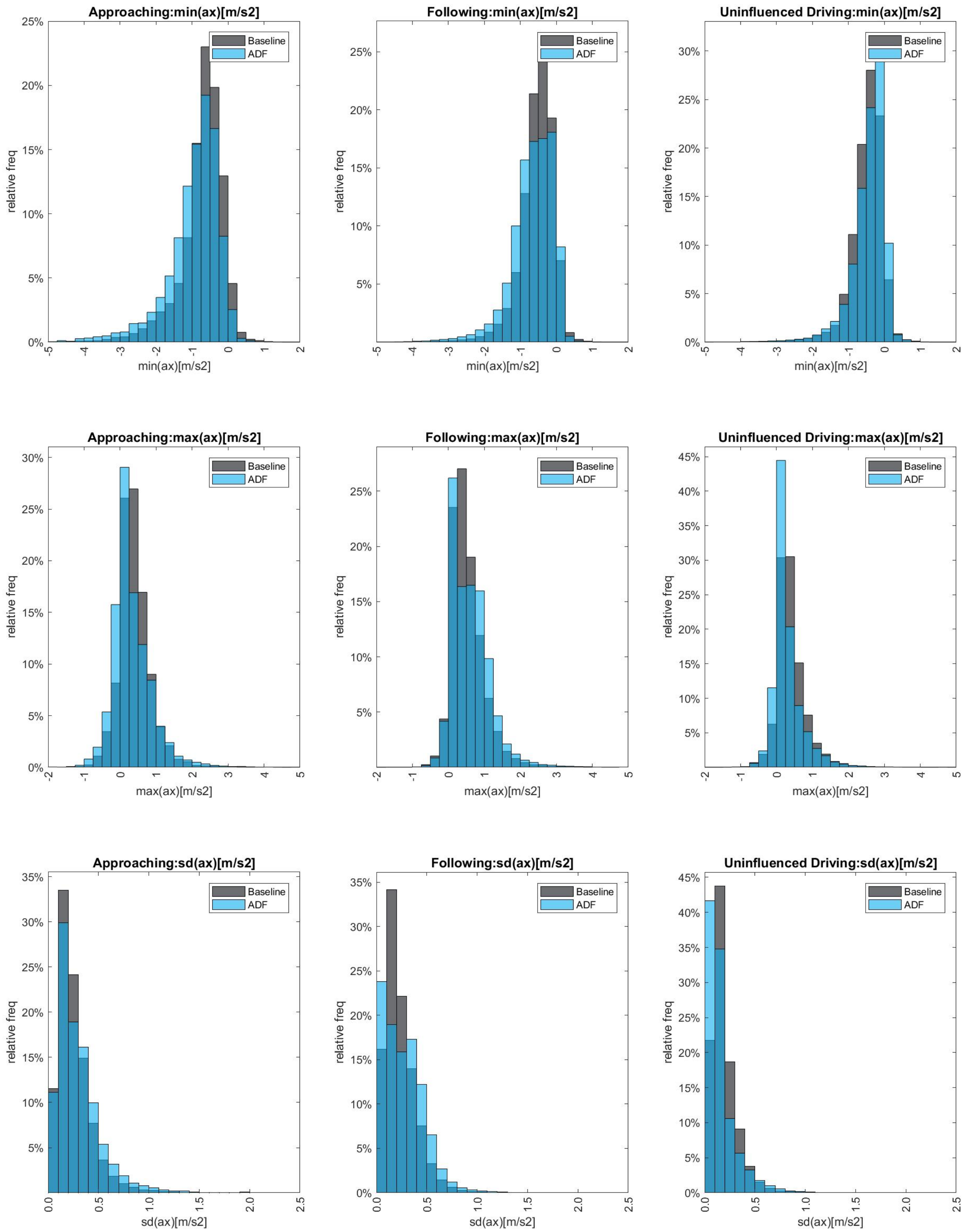


Figure 3.3: Distribution of indicators of longitudinal acceleration for the scenarios approaching a lead vehicle, following a lead vehicle, and uninfluenced driving. Min(ax) in top row, max(ax) in middle, and sd(ax) at bottom.

Maximum absolute lateral acceleration is reduced with ADF during most scenarios except lane changes.

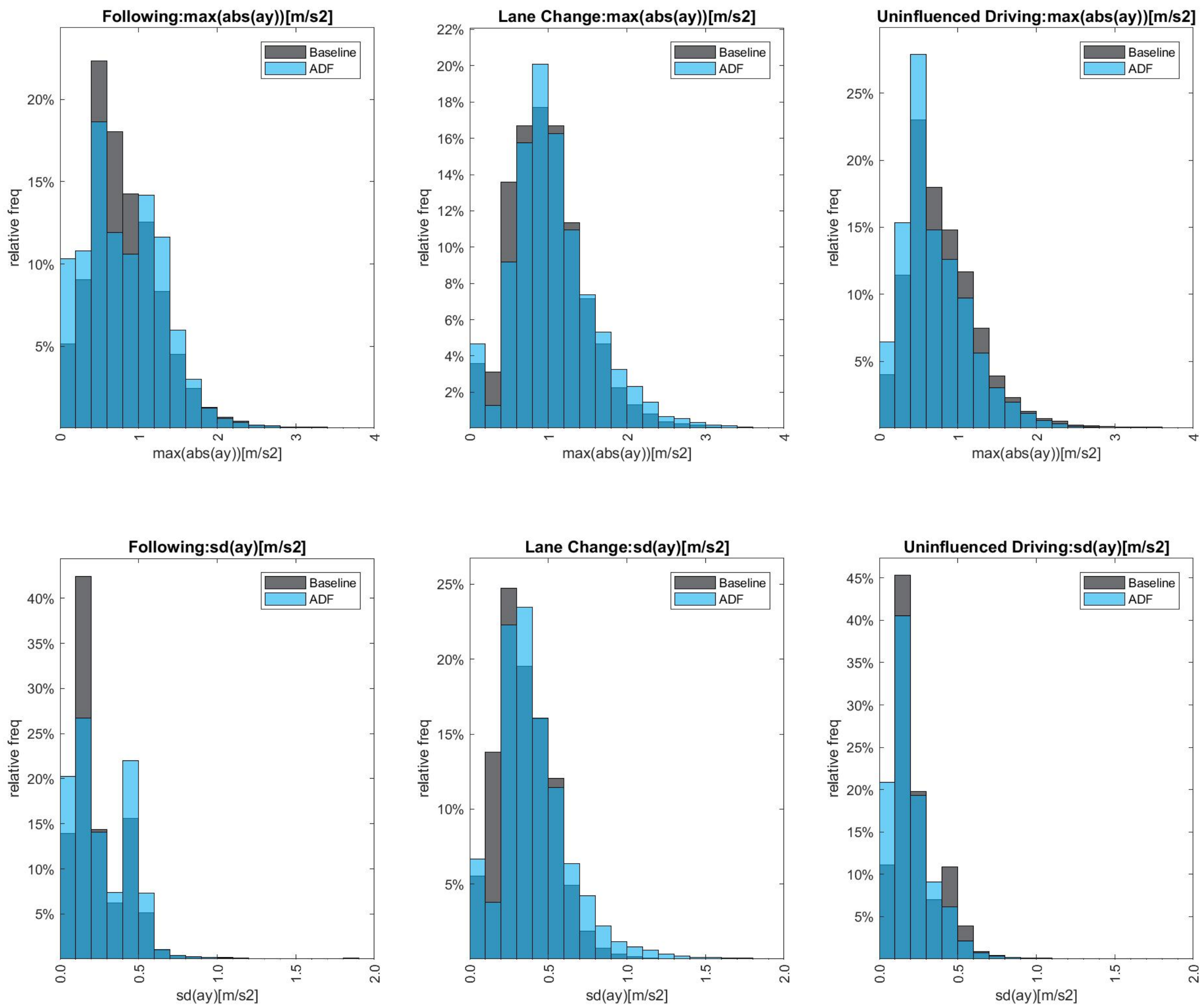


Figure 3.4: Distribution of indicators of lateral acceleration for the scenarios following a lead vehicle, lane change, and uninfluenced driving. Max(abs(ay)) in top row, and sd(ay) at bottom.

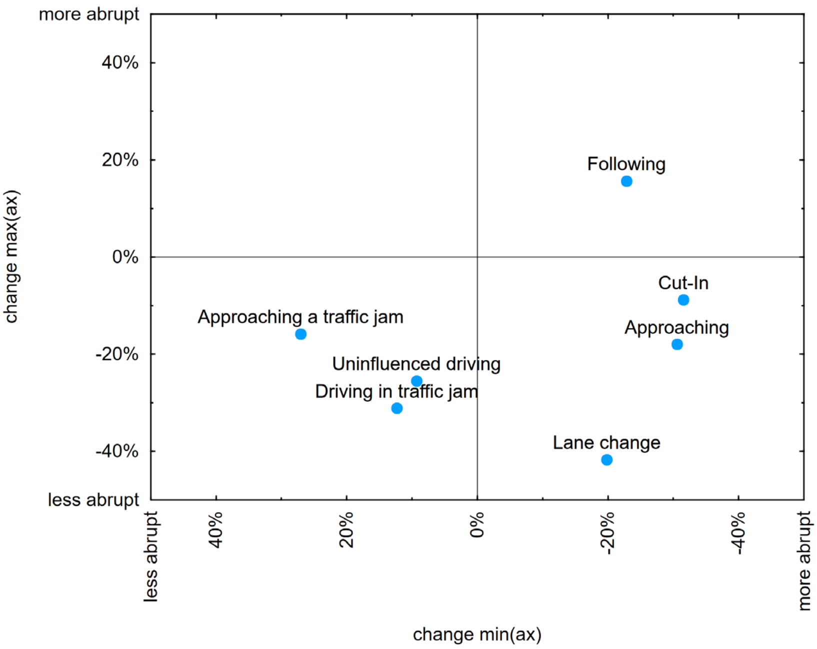


Figure 3.5: Relation between change of minimum and maximum longitudinal acceleration for the different driving scenarios.

Table 3.1: Detailed results for indicators of vehicle dynamics.

Indicator	Scenario	Z	p	Change	Effect size
min(ax)	Uninfluenced driving	-52.4	0.000	9%	-0.09
	Following	41.8	0.000	-23%	0.22
	Approaching a lead vehicle	30.9	0.000	-31%	0.30
	Cut-In	11.2	0.000	-32%	0.29
	Lane change	15.3	0.000	-20%	0.16
	Approaching a traffic jam	-3.6	0.000	27%	-0.35
	Driving in traffic jam	-3.9	0.000	12%	-0.21
max(ax)	Uninfluenced driving	97.8	0.000	-25%	0.21
	Following	-25.9	0.000	16%	-0.15
	Approaching a lead vehicle	22.7	0.000	-18%	0.14
	Cut-In	4.4	0.000	-9%	0.08
	Lane change	52.5	0.000	-42%	0.37
	Approaching a traffic jam	1.7	0.085	-16%	0.16
	Driving in traffic jam	12.4	0.000	-31%	0.73
sd(ax)	Uninfluenced driving	112.3	0.000	-17%	0.21
	Following	-26.8	0.000	14%	-0.18

Indicator	Scenario	Z	p	Change	Effect size
	Approaching a lead vehicle	-13.3	0.000	15%	-0.18
	Cut-In	-9.2	0.000	23%	-0.25
	Lane change	25.6	0.000	-4%	0.04
	Approaching a traffic jam	3.8	0.000	-26%	0.34
	Driving in traffic jam	0.7	0.482	-6%	0.12
max(abs(ay))	Uninfluenced driving	57.0	0.000	-10%	0.14
	Following	1.1	0.278	0%	0.00
	Approaching a lead vehicle	24.3	0.000	-12%	0.21
	Cut-In	17.1	0.000	-23%	0.39
	Lane change	-17.9	0.000	11%	-0.18
	Approaching a traffic jam	1.8	0.068	-16%	0.17
	Driving in traffic jam	8.0	0.000	-40%	0.39
sd(ay)	Uninfluenced driving	56.0	0.000	-11%	0.15
	Following	-21.4	0.000	9%	-0.12
	Approaching a lead vehicle	18.0	0.000	-11%	0.17
	Cut-In	16.8	0.000	-28%	0.44
	Lane change	-31.3	0.000	20%	-0.29
	Approaching a traffic jam	0.0	0.980	-9%	0.07
	Driving in traffic jam	9.2	0.000	-53%	0.34

3.1.3 RQ-T7: What Is the Impact of ADF on the Accuracy of Driving?

To analyse the accuracy of lane keeping, the mean deviation from the lane centre (m(lat pos)) and the variation of lane position (sd(lat pos) are investigated. Across scenarios, vehicles with ADF active drive closer to the centre of the lane and vary less in lateral position. In other words, with ADF lane keeping is more stable.

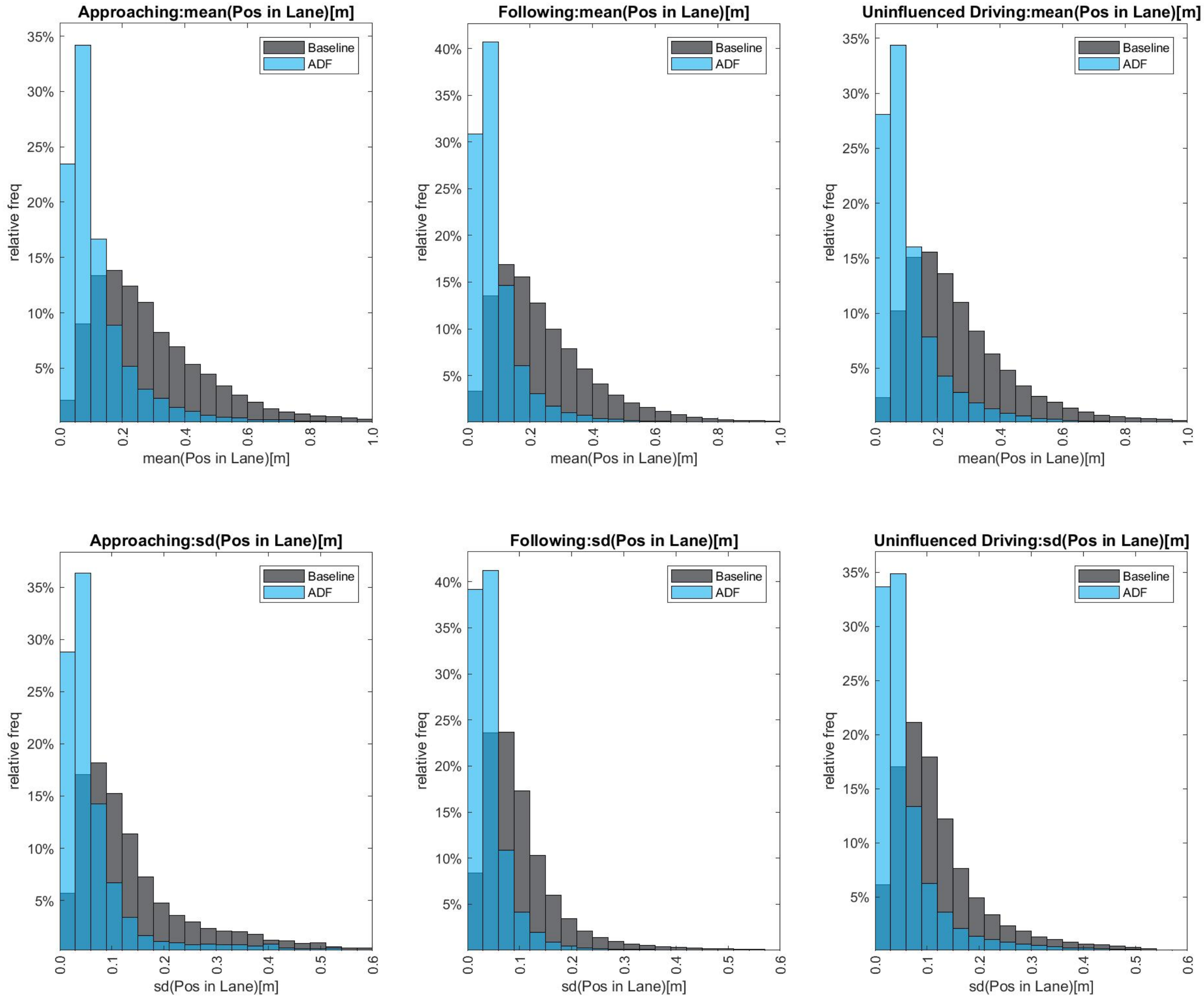


Figure 3.6: Distribution of indicators of lane keeping performance for the scenarios approaching a lead vehicle, following a lead vehicle and uninfluenced driving. Mean(position in lane) in top row, sd(position in lane) at bottom.

As indicators for longitudinal regulation, the variation of speed (sd(v)) is analysed. During stable lane-bound scenarios, during low-speed scenarios and during lane changes the variation of speed is reduced with ADF. During the more dynamic scenarios Cut-in and Approaching a lead vehicle, speed varies more while driving with ADF active.

To get a better overview on how longitudinal regulation changes with ADF, effects for variation of speed and variation of longitudinal acceleration are brought together. For the more dynamic scenarios Cut-in and Approaching a lead vehicle, regulation is less smooth with ADF because speed, as well as longitudinal acceleration, vary more than in manual driving. During low-speed scenarios and uninfluenced driving, longitudinal regulation is smoother with ADF as both variations of speed and of acceleration are reduced. During Following, the ADF manages to reduce the variation of speed. However, this is related to an increase in variation of longitudinal acceleration.

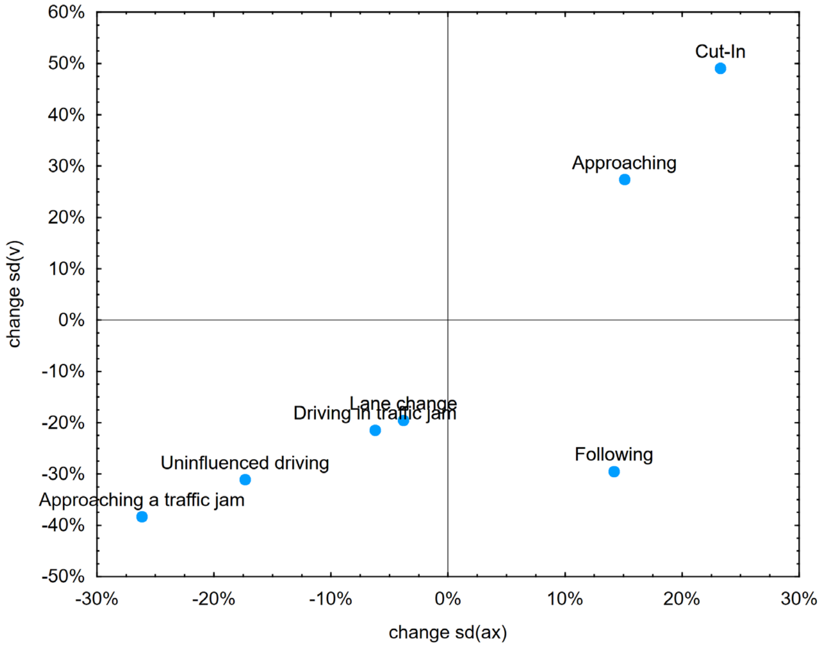


Figure 3.7: Relation between change of variation of speed and change of variation of longitudinal acceleration for the different driving scenarios.

Table 3.2: Detailed results for indicators for precision of driving.

Indicator	Scenario	Z	p	Change	Effect size
sd(lat pos)	Uninfluenced driving	239.9	0.000	-47%	0.71
	Following	222.1	0.000	-54%	0.94
	Approaching a lead vehicle	81.9	0.000	-46%	0.60
	Cut-In	29.4	0.000	-43%	0.68
	Approaching a traffic jam	8.6	0.000	-34%	0.46
	Driving in traffic jam	10.8	0.000	-29%	0.32
mean(lat pos)	Uninfluenced driving	285.7	0.000	-57%	0.94
	Following	255.8	0.000	-62%	1.28
	Approaching a lead vehicle	98.1	0.000	-55%	0.76
	Cut-In	42.9	0.000	-59%	1.16
	Approaching a traffic jam	10.1	0.000	-54%	1.19
	Driving in traffic jam	4.6	0.000	-9%	0.10
sd(v)	Uninfluenced driving	179.0	0.000	-31%	0.25
	Following	103.3	0.000	-29%	0.33
	Approaching a lead vehicle	-11.9	0.000	27%	-0.21
	Cut-In	-14.6	0.000	49%	-0.36

Indicator	Scenario	Z	p	Change	Effect size
	Lane change	43.7	0.000	-20%	0.17
	Approaching a traffic jam	6.4	0.000	-38%	0.70
	Driving in traffic jam	9.6	0.000	-21%	0.50

3.1.4 RQ-T8: What Is the Impact of ADF on the Driven Speed?

Initial explorative analyses showed that, as expected, there was a strong influence of the speed limit on the average speed as well as on the maximum speed. Ignoring this would incorporate a lot of noise in the data. On the other hand, explicitly including speed limit as a factor in the analysis was not a straightforward way to go either. First of all, in more than 40% of the observations, the speed limit was hidden (because it would have made it possible to link entries in the database with a single Pilot site.) or it was unlimited (as can happen on German motorways). Second, separating the analysis by speed limit would introduce the risk of revealing the origin of certain parts of the data, for instance because some limits only occur on certain test sites.

To use as much of the data as possible, and at the same time stay close to the analysis method used for the other PIs, the following approach was used. First, the data was separated by speed limit. For each speed limit, the distribution of the PI was derived, for Baseline (BL) as well as for ADF. Also, the median of the combined distribution (pooled over BL and ADF) was determined. This is illustrated in Figure 3.8.

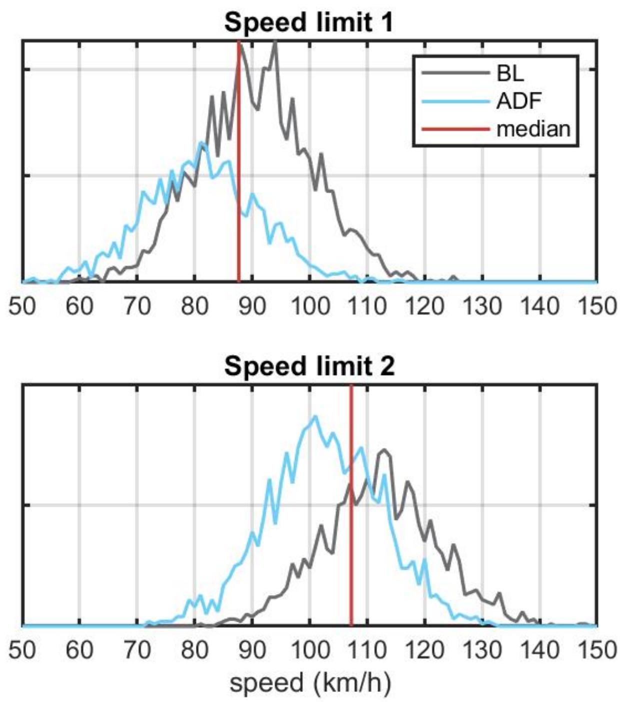


Figure 3.8: Distribution of average speeds as a function of speed limit and experimental condition (artificial data).

Next, all distributions were shifted, such that for each speed limit the median was 0. This is illustrated in Figure 3.9: the shape of the entire distribution remains unchanged, as well as any difference that might exist between BL and ADF.

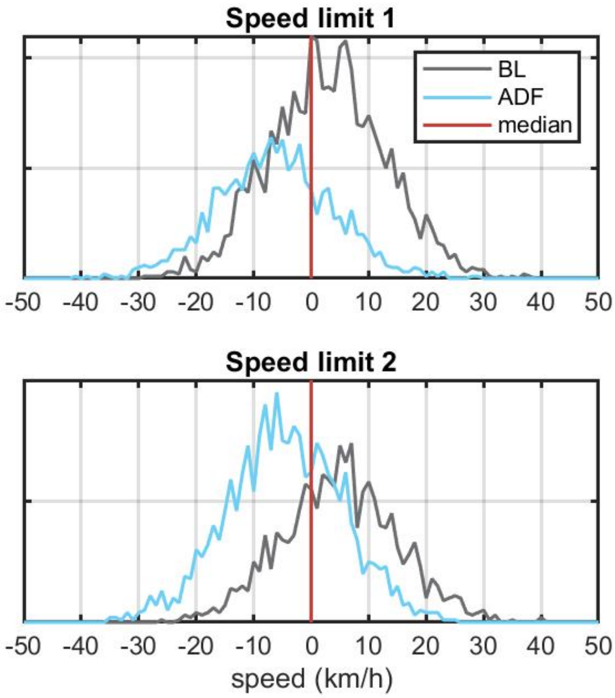


Figure 3.9: Distribution of average speeds as a function of speed limit and experimental condition (artificial data).

As the final pre-processing step, the data were pooled over all speed limits, ending up with two distributions: one for BL and one for ADF (see Figure 3.10). After this, the same non-parametric test could be applied as for the other PIs. In that process, the Cohen effect size could be calculated as well. Only the effect size in % could not be calculated, because the original absolute values of the speeds were lost in the alignments of the medians.

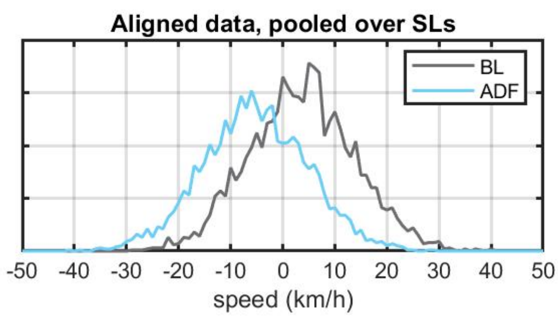


Figure 3.10: Distribution of average speeds as a function experimental condition, after aligning the medians and then pooling over all speed limits (artificial data).

For all analysed scenarios, there was a significant reduction of maximum speed while driving with the ADF, and for all except driving in traffic jam also a reduction of average speed.

Table 3.3: Detailed results for indicators of vehicle speed.

Indicator	Scenario	Z	p	Change	Effect size
mean(v)	Uninfluenced driving	55.4	0.000	N.A.	-0.12
	Following	51.9	0.000	N.A.	-0.13
	Approaching a lead vehicle	29.0	0.000	N.A.	-0.26
	Cut-In	12.0	0.000	N.A.	-0.24
	Approaching a traffic jam	3.1	0.002	N.A.	-0.25
	Driving in traffic jam	1.9	0.058	N.A.	-0.06
max(v)	Uninfluenced driving	62.2	0.000	N.A.	-0.14
	Following	55.9	0.000	N.A.	-0.15
	Approaching a lead vehicle	27.8	0.000	N.A.	-0.25
	Cut-In	10.1	0.000	N.A.	-0.20
	Approaching a traffic jam	4.2	0.000	N.A.	-0.33
	Driving in traffic jam	8.6	0.000	N.A.	-0.41

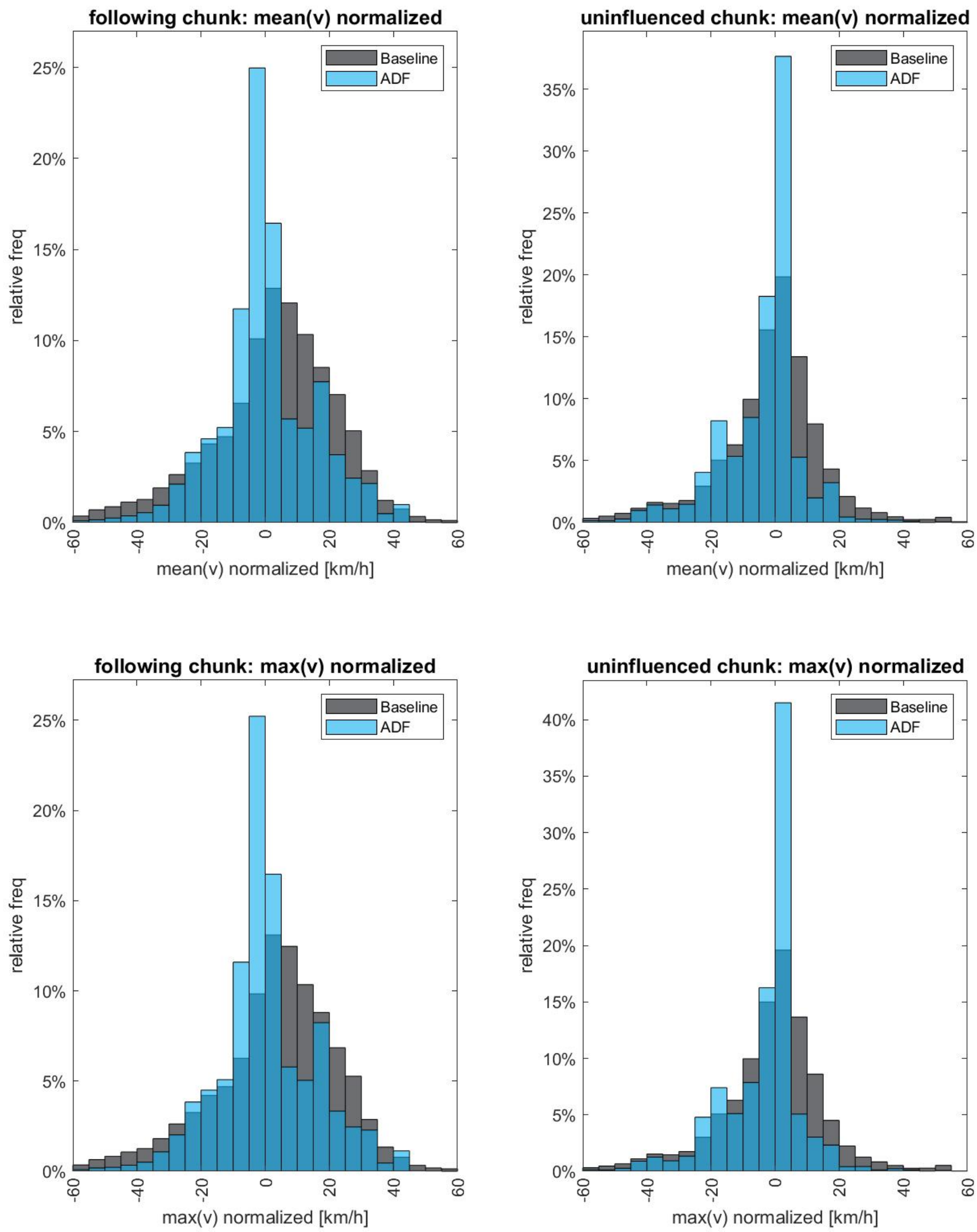


Figure 3.11: Distribution of average speed and maximum speed (normalised) in the scenarios Following and Uninfluenced driving. Mean(speed) in top row, and max(speed) at bottom.

3.1.5 RQ-T9: What Are the Impacts of ADF on Energy Efficiency?

The impact on energy efficiency is not based on measured fuel consumption but on calculated energy demand (cf. Section 2.5.1). As can be seen in the graphs (Figure 3.12), there is a bimodal distribution if this indicator is analysed for all trips and for trips without traffic jams. For trips consisting of traffic jams more than 50% of the driving time, only the lower peak remains. This indicates that the peaks mainly relate to vehicle speed, one showing the distribution for higher speeds and one for lower speeds, e.g., traffic jams and sections on urban motorways with lower speed limits (e.g. 70 kph). Independent of the included trips, there is a significant decrease of calculated energy demand when driving with the ADF active due to changes in driving style.

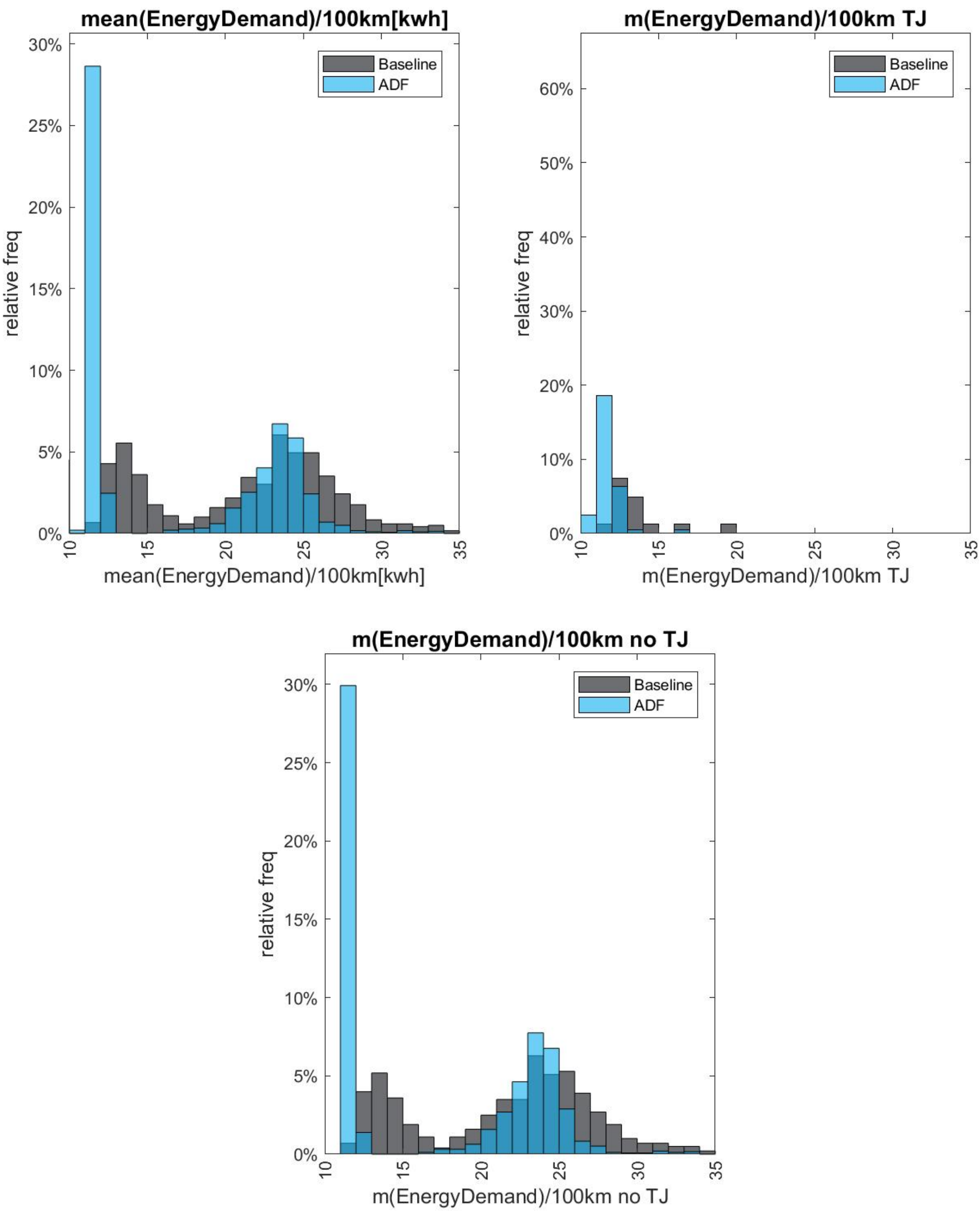


Figure 3.12: Distribution of calculated energy demand in kWh during all trips, trips without traffic jam and trips with mainly traffic jams (>50% of driving time on motorways).

Table 3.4: Detailed results for indicators of energy demand.

Indicator	Scenario	Z	p	Change in %	Effect size
mean(Energy demand)	All trips	17.4	0.000	-20%	0.68
	Trips without traffic jams	-15.6	0.000	-19%	0.66
	Trips with mainly traffic jams	7.6	0.000	-12%	0.89

3.1.6 RQ-T10 / RQ-T14: What Is the Impact of ADF on the Frequency of Near-Crashes / Incidents?

To assess the impact of driving with ADF on the frequency of potentially critical driving situations, the proportion of driving scenarios with defined incident types is evaluated. As can be seen in Figure 3.13, there is a reduction of very short distances to the lead vehicle with the ADF in all analysed scenario types. Lateral distance incidents to the side occur less frequently and their frequency does not change while driving with the ADF. Dynamic incidents, i.e., harsh braking, happened only rarely. For dynamic incidents, a decrease of incidents during following scenarios and an increase during lane changes with ADF can be observed.

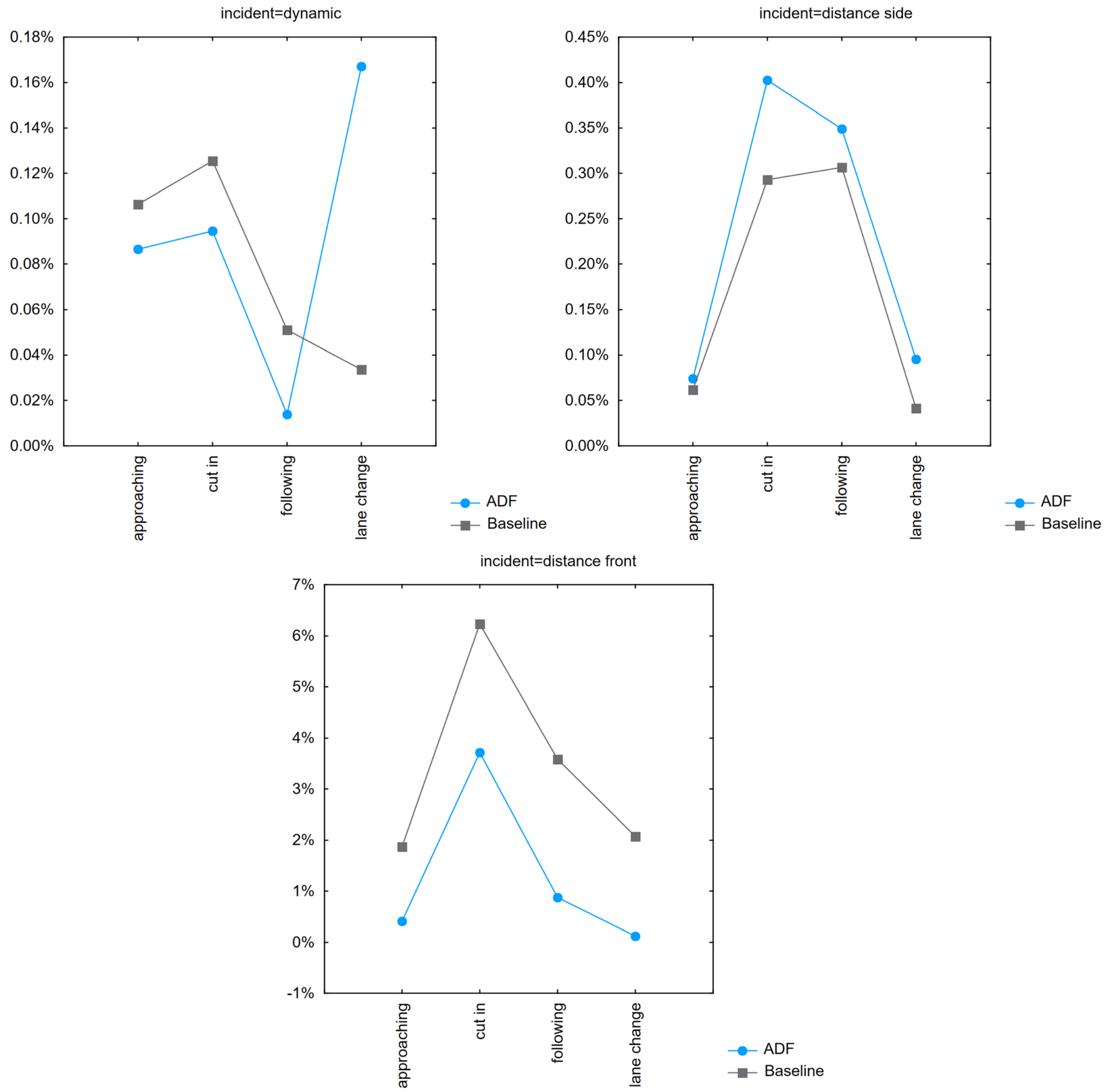


Figure 3.13: Frequency of analysed incident types in different driving scenarios in driving with ADF active and in baseline driving.

Table 3.5: Detailed results on incident frequency.

Scenario	Condition	Distance incident	Dynamic incident		Sum		
		No	**Front**	**Side**	**No**	**Yes**	
Cut-in	ADF active	4053	157	17	4223	4	4227
	What is here? %	95.88%	3.71%	0.40%	99.91%	0.09%	
	Baseline	2235	149	7	2388	3	2391

Scenario	Condition	Distance incident	Dynamic incident		Sum		
		No	Front	Side	No	Yes	
	What is here? %	93.48%	6.23%	0.29%	99.87%	0.13%	
	Sum	6288	306	24	6611	7	6618
Approaching	ADF active	16106	66	12	16170	14	16184
	What is here? %	99.52%	0.41%	0.07%	99.91%	0.09%	
	Baseline	17514	334	11	17840	19	17859
	What is here? %	98.07%	1.87%	0.06%	99.89%	0.11%	
	Sum	33620	400	23	34010	33	34043
Lane change	ADF active	25087	30	24	25099	42	25141
	What is here? %	99.79%	0.12%	0.10%	99.83%	0.17%	
	Baseline	26166	555	11	26723	9	26732
	What is here? %	97.88%	2.08%	0.04%	99.97%	0.03%	
	Sum	51253	585	35	51822	51	51873
Following	ADF active	57492	509	203	58196	8	58204
	What is here? %	98.78%	0.87%	0.35%	99.99%	0.01%	
	Baseline	28218	1053	90	29346	15	29361
	What is here? %	96.11%	3.59%	0.31%	99.95%	0.05%	
	Sum	85710	1562	293	87542	23	87565

Table 3.6: Results of Chi-square tests on incident frequency.

	Pearson's Chi-Square	
	Distance incident	Dynamic incident
Cut-in	22.3688, FG=2, p=.000014	137475, FG=1, n.s.
Approaching	156.535, FG=2, p=0.00000	.346614, FG=1, n.s.
Lane-change	450.324, FG=2, p=0.00000	23.4703, FG=1, p=0.000001
Following	819.857, FG=2, p=0.00000	10.3643, FG=1, p=0.001285

3.1.7 RQ-T11: What Is the Impact of ADF on the Frequency of Certain Events?

To analyse the frequency with which certain driving scenarios occur, the frequency of scenarios per hour and the overall proportion of driving time spent in the driving scenarios are analysed. With both indicators, there is a significant reduction of the frequency of approaching scenarios and of lane changes with ADF. At the same time, the frequency of car following increases. For uninfluenced driving scenarios, there is a decrease of the scenario frequency while the proportion of time spent in that scenario increases. Other than expected, there is no increase in the frequency of cut-in scenarios. Overall, driving seems to be more lane-bound and less dynamic with ADF active.

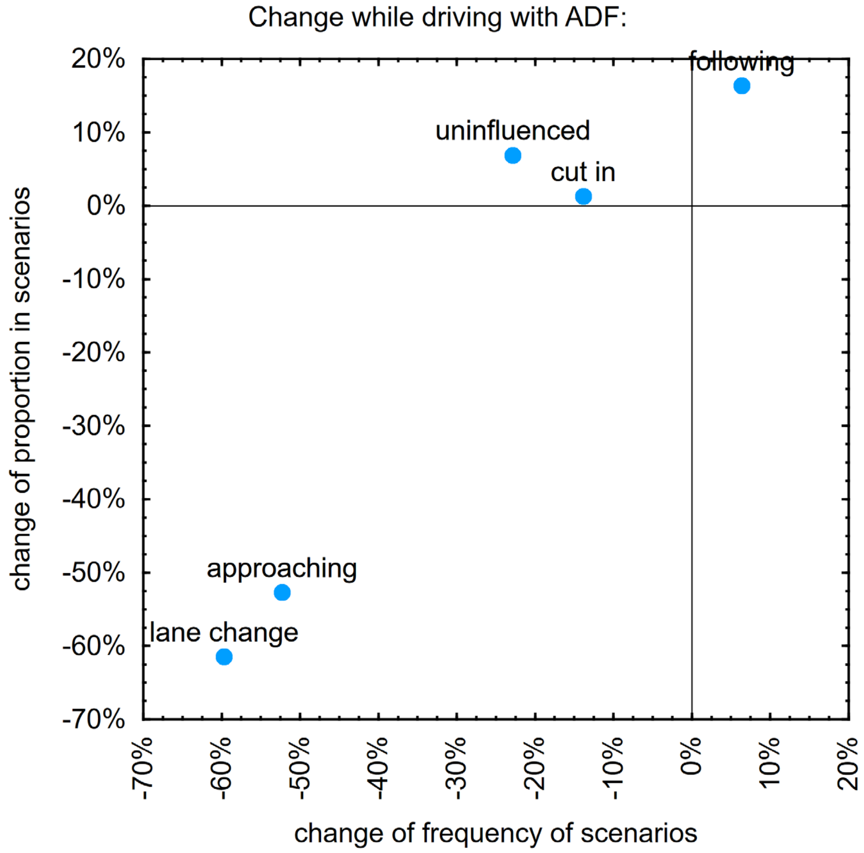


Figure 3.14: Change of scenario frequency while driving with ADF active. Values in Table 3.7.

Table 3.7: Detailed results for indicators of frequency of driving scenarios.

Indicator	Scenario	Z	p	Change in %	Effect size
Frequency	Lane change	28.0	0.000	-60%	1.08
	Uninfluenced	13.8	0.000	-23%	0.48
	Following	5.0	0.000	6%	-0.08
	Approaching	23.9	0.000	-52%	1.01

Indicator	Scenario	Z	p	Change in %	Effect size
	Cut-in	3.7	0.000	-14%	0.13
Proportion	Lane change	29.0	0.000	-61%	1.19
	Uninfluenced	-4.0	0.000	7%	-0.11
	Following	-2.8	0.004	16%	-0.20
	Approaching	22.2	0.000	-53%	0.94
	Cut-in	2.1	0.039	1%	-0.01

The occurrence of cut-ins is looked at in detail for a subset of the data using video review to validate the correct detection and identify reasons for other vehicles to cut in. The motivation for cut-ins is categorised in discretionary (cut-in after overtaking) and mandatory (cut-in for entry ramp, exit ramp, avoiding a work zone or obstacle, end of lane). Overall, 1560 cut-ins (378 baseline, 1182 ADF) were annotated and analysed. There is a significant difference in the cut-in categories. While most cut-ins in baseline are discretionary (70.6%), cut-ins during ADF are leaning towards the mandatory category (53.0%). Especially cut-ins from entry ramps occur much more in ADF.

3.1.8 RQ-T12: What Is the Impact of ADF on Car Following Behaviour?

Parameters like average chosen time headway and variation of time headway are a meaningful indicator of driving behaviour only for scenarios with stable following behaviour like *Car following* and *Driving in a traffic jam*. In those scenarios, the average time headway when following a lead vehicle (mean(THW)) is significantly larger with ADF. Also, the variation of time headway (sd(THW)) is reduced while driving with the ADF.

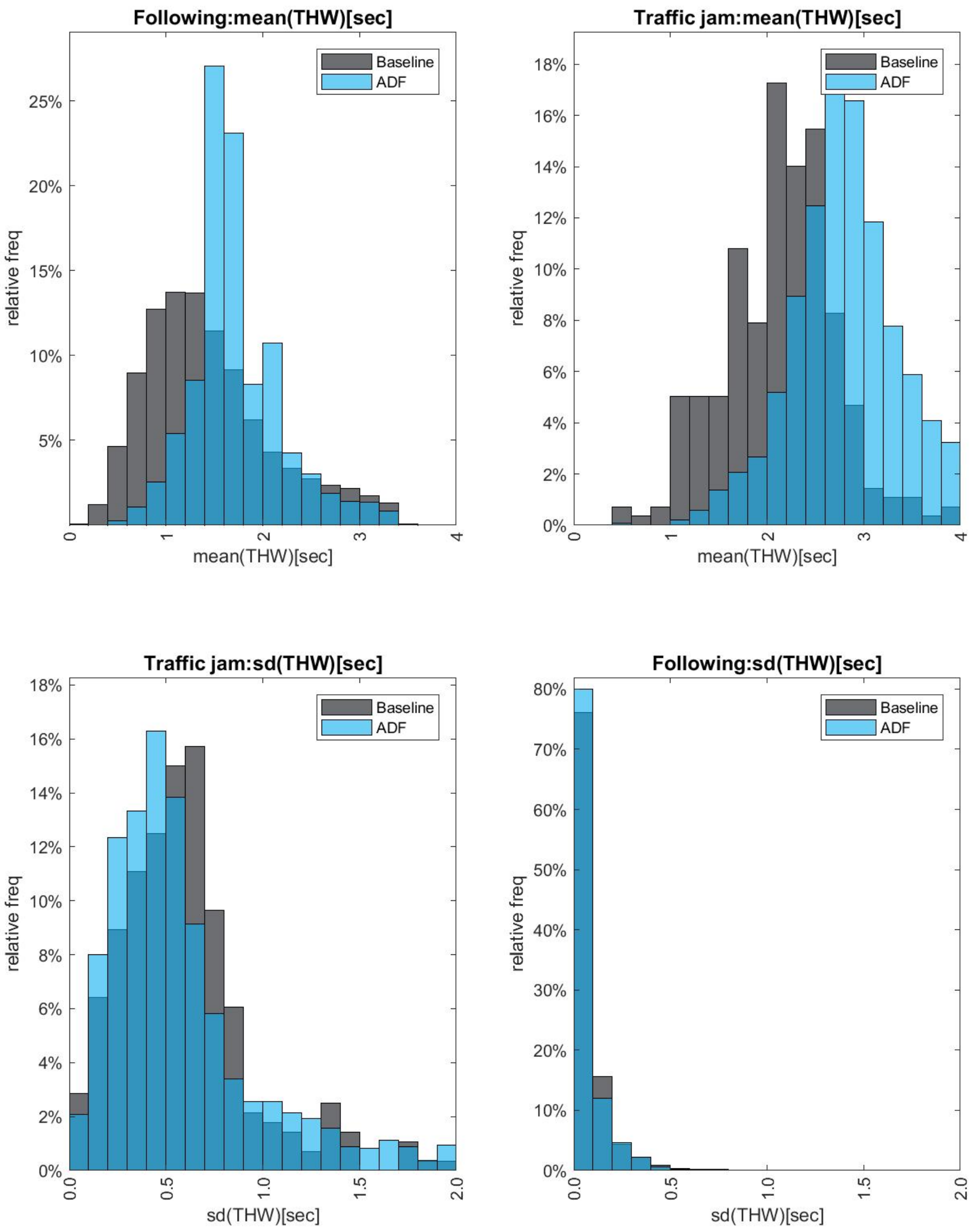


Figure 3.15: Distribution of average time headway (upper graphs) and variation of time headway (lower graphs) in the scenarios following and driving in a traffic jam.

For all scenarios during which a lead vehicle was present, there is a significant increase of minimum time headway (min(THW)) and minimum time-to-collision (min(TTC)) when driving with ADF active.

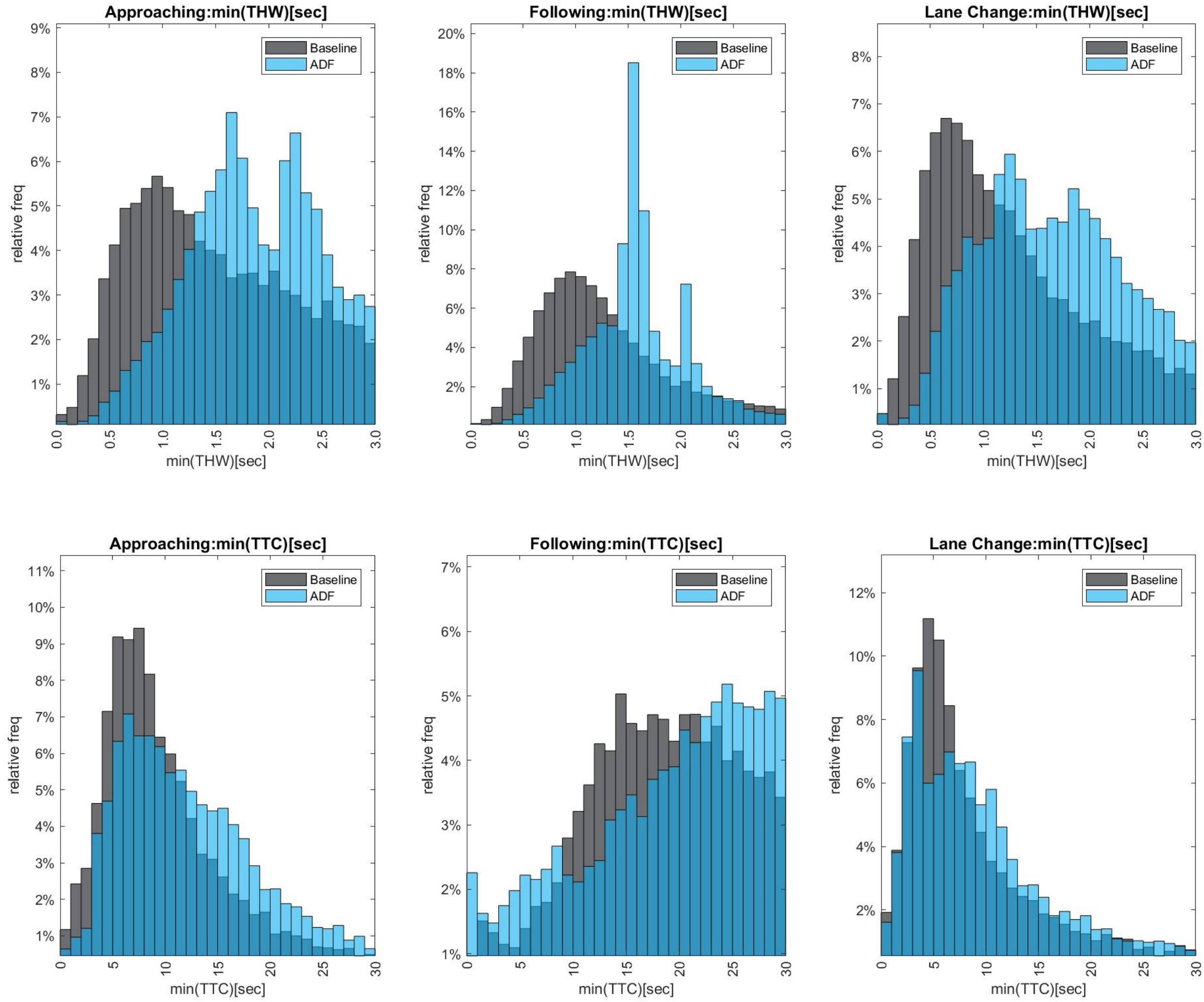


Figure 3.16: Distribution of minimum time headway and minimum time to collision in the scenarios Approaching a lead vehicle, Following and Lane change. Min(THW) in top row, and min(TTC) at bottom.

Table 3.8: Detailed results for indicators for car following behaviour.

Indicator	Scenario	Z	p	Change in %	Effect size
mean(THW)	Following	-127.8	0.000	20%	-0.53
	Driving in traffic jam	-15.3	0.000	31%	-0.66
sd(THW)	Following	68.3	0.000	-16%	0.14
	Driving in traffic jam	0.4	0.725	27%	-0.15
min(THW)	Following	-126.2	0.000	22%	-0.51
	Approaching a lead vehicle	-49.0	0.000	25%	-0.54
	Cut-In	-6.5	0.000	8%	-0.17
	Lane change	-50.9	0.000	39%	-0.33

Indicator	Scenario	Z	p	Change in %	Effect size
	Approaching a traffic jam	-3.8	0.000	19%	-0.40
	Driving in traffic jam	-15.7	0.000	69%	-0.34
min(TTC)	Following	-67.6	0.000	25%	-0.49
	Approaching a lead vehicle	-29.3	0.000	26%	-0.30
	Cut-In	-7.0	0.000	25%	-0.26
	Lane change	-10.3	0.000	9%	-0.08
	Approaching a traffic jam	-4.0	0.000	45%	-0.41
	Driving in traffic jam	-13.5	0.000	80%	-0.41

3.1.9 RQ-T15: How Does the ADF Influence the Behaviour of Subsequent Vehicles?

A data subset was analysed in depth for interactions with rear vehicles, since there was no defined driving scenario capturing that. The following results are based on 52 042 events with a subsequent vehicle (of which 12 116 are in baseline) on a city motorway. Some conditions were excluded, as sensor errors can lead to false data, e.g., driving through tunnels, speeds below 50 km/h and lateral speed of subsequent vehicle above 5.5 m/s to reduce the amounts of ghost objects. The area of interest behind the ego-vehicle is reduced to 40 metres and the rear vehicle's time headway should be higher than 0.3 s and lower than 10 s.

The relevant indicators that were investigated are the minimum distance kept by the subsequent vehicle, the minimum acceleration below zero (max brake values) of the subsequent vehicle, the average relative velocity above zero (representing an approach of the subsequent vehicle), minimum THW of the subsequent vehicle. Table 3.9 shows the results. While there is a small effect on the distance kept (slightly more with ADF compared to baseline) and on the min acceleration of subsequent vehicles (slightly higher for ADF, which means less severe braking on average), there is no significant difference in the relative velocity or in the time headway kept by the subsequent vehicle.

Table 3.9: Analysed indicators and results for RQ-T15.

Indicator	Z	p	Change in %	Effect size
min(distance)	-14.8	0.000	4%	-0.15
min(acc)	-14.7	0.000	-12%	-0.18
m(rel. velocity)	-0.7	0.49	-	-
min(THW)	0.9	0.35	-	-

3.1.10 RQ-T16: How Does the ADF Influence the Behaviour of Preceding Vehicles?

To evaluate the behaviour of the lead vehicle, the indicators average speed of the lead vehicle (mean(v lead vehicle) and standard deviation of speed of the lead vehicle (sd(v lead vehicle)) were derived from the data. For all analysed scenarios, there is a significant reduction of speed of the lead vehicle while driving with ADF active. This is probably because the ADF drives significantly slower than in manual driving and thus, slower vehicles are selected as lead vehicles. To control that effect, the difference between the average speed of the ego-vehicle and the average speed of the lead vehicle was calculated (diff(v lead vehicle). There are significant but small changes of this difference for the scenarios Following and Approaching a lead vehicle. For cut-ins, the speed difference to the lead vehicle is significantly larger compared to baseline. As can be seen in Figure 3.17, in baseline driving there is a larger proportion of vehicles cutting in with a lower speed than the ego-vehicle. While driving with ADF, the speeds between the two vehicles are more similar and in a larger proportion of scenarios the vehicle cutting in is considerably faster. Furthermore, there is a significant reduction of the variation of speed of the lead vehicle for all scenarios while driving with ADF. With ADF, the frequency of lane changes made by the lead vehicle drops by 33%.

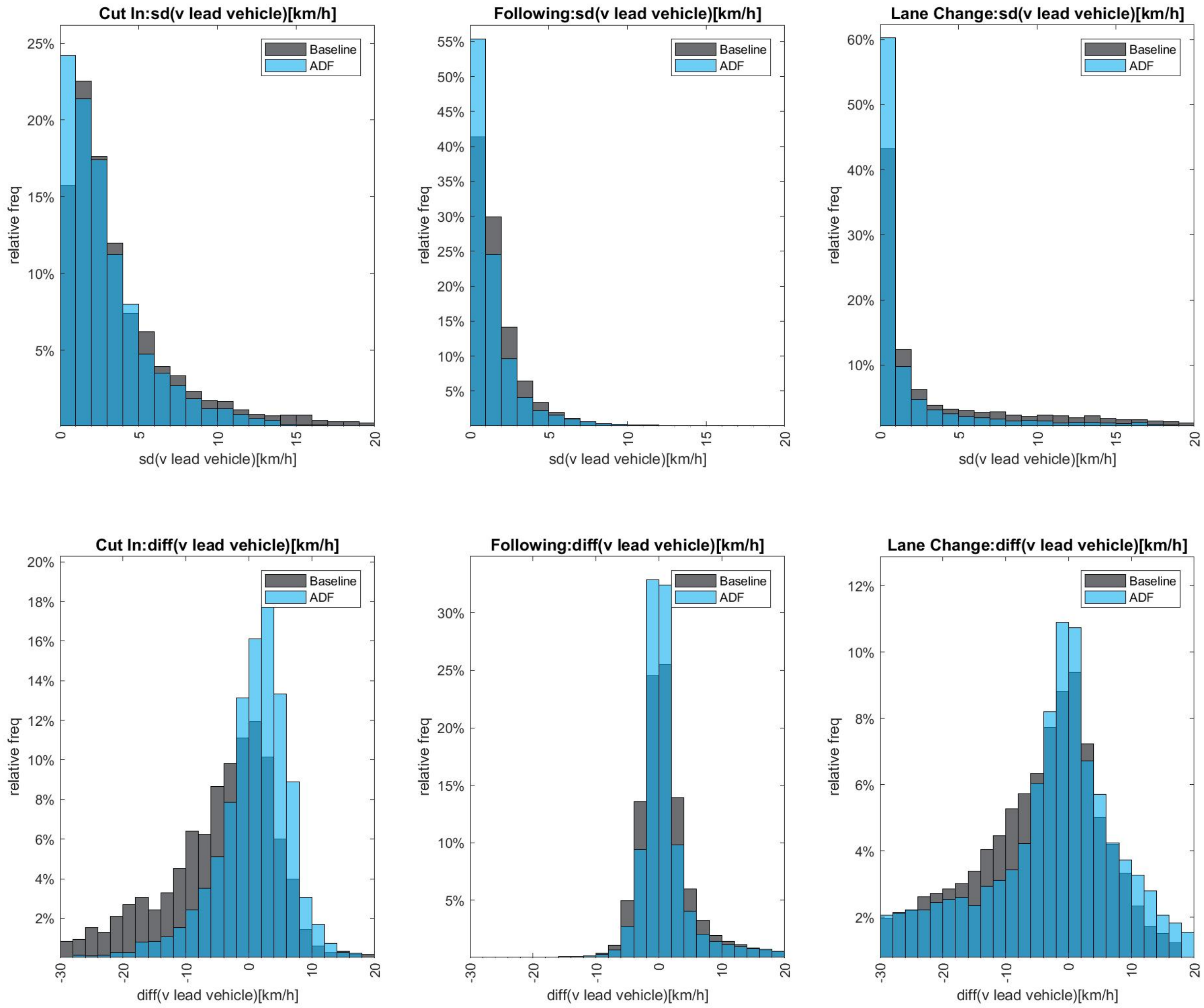


Figure 3.17: Distribution of variation of speed of the lead vehicle and speed difference to the lead vehicle in the scenarios Cut-in, Following and Lane change. Sd(v lead vehicle) in top row, diff(v lead vehicle) at bottom.

Table 3.10: Detailed results for indicators for behaviour of preceding vehicle.

Indicator	Scenario	Z	p	Change in %	Effect size
mean(v lead vehicle)	Following	92.3	0.000	-9%	0.45
	Approaching a lead vehicle	26.4	0.000	-5%	0.28
	Cut-In	8.0	0.000	-5%	0.21
	Lane change	26.1	0.000	-6%	0.26
	Approaching a traffic jam	2.0	0.046	-3%	0.16
	Driving in traffic jam	5.2	0.000	-13%	0.25
Diff(v lead vehicle)	Following	6.7	0.000	-11%	0.03
	Approaching a lead vehicle	-2.3	0.021	1%	-0.01
	Cut-In	-26.1	0.000	103%	-0.61
	Lane change	0.5	0.586	-15%	0.06
	Approaching a traffic jam	-0.5	0.596	29%	-0.14
	Driving in traffic jam	0.8	0.446	-33%	0.12
sd(v lead vehicle)	Following	73.3	0.000	-21%	0.21
	Approaching a lead vehicle	31.2	0.000	-23%	0.21
	Cut-In	8.9	0.000	-20%	0.23
	Lane change	35.7	0.000	-33%	0.27
	Approaching a traffic jam	2.4	0.018	-19%	0.31
	Driving in traffic jam	9.8	0.000	-23%	0.52
N(LaneChange_LeadVeh)	Trip	3.1	0.002	-33%	0.43

3.1.11 RQ-T17: What Is the Impact of ADF on the Number of Near-Crashes / Incidents of Other Traffic Participants?

This RQ was approached using a data subset to look in detail into potential incidents. The data used came from one motorway Pilot site. All recorded incidents were video reviewed. There have been no recorded near-crashes in either the baseline drives or the ADF drives. As the results for THW indicate, the ADF is not likely to cause a conflict with the preceding vehicle. However, a safety driver was present to prevent serious conflicts with other vehicles. There were 10 front incidents recorded, of which seven were dismissed. Two of the remaining were in ADF mode, one was an approach to a sudden traffic jam and required a safety driver intervention, and the other was an extremly close cut-in by another vehicle (no safety driver intervention but alert). In baseline, a driver merged very close to another vehicle an incident type that the ADF would not create.

Table 3.11: Number of analysed incidents.

	Dismiss	Baseline	Treatment	Total
Front distance incident	7	1	2	10
Rear distance incidents	29	12	13	54

The rear incidents caused by subsequent vehicles were 54 overall but 29 were dismissed after review. The remaining are equally distributed between baseline and treatment. While there was about three times as much data collected in treatment, it could be expected that the incidents are reduced in the interaction with an ADF. However, the amount of data does not allow any quantitative statistical analysis and therefore no clear conclusion can be drawn. A more detailed view is provided in the safety impact assessment (Deliverable D7.4 – Impact Evaluation Results Bjorvatn et al., 2021).

3.1.12 Summary, Motorway ADF

Interpretation of the results on system availability and stability should be approached with care. It must be kept in mind that the tested ADFs were still on a prototype level. Furthermore, it is likely that the circumstances of data collection (e.g., prototype ADFs, interventions of safety drivers) have an impact on the results as well, so that a direct conclusion on driving with a market ready mature ADF is challenging. For instance, the frequency of take-over requests might be underestimated on trips where safety drivers prevented unusual situations by taking back control before the end of the ODD actually resulted in a take-over request. On some trips that were part of experimental test drives during which non-professional drivers had the possibility to get to know the ADF, the number of take-over requests might be overestimated because those trips were designed in such a way that take-over requests took place frequently during every drive.

Regarding the measured impact of ADF on driving behaviour, some stable effects can be observed. Overall, in the analysed driving scenarios, driving with the ADF leads to

- a significant decrease of speed,
- a significant increase of distances to the lead vehicle,
- a significantly more stable lane keeping behaviour.

Compared to that, results on vehicle dynamics and longitudinal regulation differ across scenarios. During approaching, cut-ins and lane changes, the ADF decelerates more strongly than a manual driver does, while in car following it decelerates and accelerates more strongly. If we look at the variation of speed and acceleration, longitudinal regulation is more stable with the ADF during uninfluenced driving and low speed scenarios, while in more dynamic scenarios like approaching and cut-ins, it is more abrupt. In summary, it seems that in scenarios that require continuous reaction to other vehicles and that might also benefit from anticipation of situational development, driving with an ADF is related to more pronounced longitudinal regulation than it would be in manual driving.

Lateral vehicle dynamics are reduced with the ADF for all scenarios except lane changes. Overall, driving with the ADF becomes more stable and more lane-bound due to a reduction of lane changes and approaching situations. This results in a higher proportion of driving time spent, especially in following scenarios.

Interpretation of results on the frequency of potentially critical driving situations needs, again, to be treated cautiously. Overall, results indicate that especially the frequency of close distances to the lead vehicle is reduced. One explanation is that the ADF on average keeps larger distances to the vehicle and follows less closely. Furthermore, these results might be directly influenced by the nature of testing prototype functions: safety drivers were present during all drives, and it was their task to intervene before critical situations occurred. This might directly reduce the frequency of critical situations. Furthermore, the overall analysis is based entirely on objective thresholds applied to vehicle data, with no verification based on video recording. Especially indicators assessing rare events like critical situations are prone to be influenced by unusual events, outliers, but also sensor errors like ghost objects. For instance, the in-depth analysis done for a subset of data on critical situations with relation to surrounding traffic revealed that most of the objectively detected events were false positive events. In the end, for that subset of data the absolute number of verified critical situations was too small to draw any conclusions on the impact of ADF on event frequency.

The analysis of calculated energy demand reveals a surprisingly high reduction of energy demand with the ADF. The reported effect is based on changes in driving behaviour (lower speed, more stable driving scenarios, ...). Effects of other influencing factors like vehicle type and energy demand of additional equipment are not considered. The overall reduction is probably partly influenced by trips taking place at a lower speed range on an urban highway. Here, it needs to be known that some drives repeatedly took place on urban motorways with a speed limit of 70km/h. A not completely balanced proportion of those trips taking place in baseline and ADF condition might impact the overall results on energy demand. Nevertheless, also for trips that mainly consisted of driving in traffic jams, a reduction of energy demand of 12% on average is found.

The following pages show two summaries of all the results, one of changes in relation to baseline values and the other of reported effect sizes. PIs for which no significant effect could be found are set to zero in both summaries. Negative values in Table 3.12 show a decrease of the PI (e.g., lower speed), positive values an increase (e.g. larger distances) compared to baseline driving. With the colour coding it becomes easily visible whether the direction and size of effects are similar across scenarios or whether they vary between scenarios. (The colours do no judge whether the change has any positive or negative implications.)

Table 3.12: Percentual changes in PIs when comparing ADF-driven vehicle with manual vehicle in the baseline condition for different RQs. Blue indicates a reduction of the indicator compared to baseline and red an increase. The depth of the colouring codes the size of the effect.

Research question		Performance indicator	Uninfluenced driving	Following	Approaching lead vehicle	Cut-In	Lane change	Approaching traffic jam	Driving in traffic jam	trip	trip - no TJ	trip - TJ
RQ-T6	vehicle dynamics	min(ax)	9%	-23%	-31%	-32%	-20%	27%	12%			
		max(ax)	-25%	16%	-18%	-9%	-42%	0%	-31%			
		sd(ax)	-17%	14%	15%	23%	-4%	-26%	0%			
		max(abs(ay))	-10%	0%	-12%	-23%	11%	0%	-40%			
		sd(ay)	-11%	9%	-11%	-28%	20%	0%	-53%			
RQ-T7	accuracy of driving	m(lat Pos)	-57%	-62%	-55%	-59%	0%	-54%	-9%			
		sd(lat Pos)	-47%	-54%	-46%	-43%	0%	-34%	-29%			
		sd(v)	-31%	-29%	27%	49%	-20%	-38%	-21%			
RQ-T9	energ. Consump.	m(Energy consum.)								-20%	-19%	-12%
RQ-T11	frequency of events	N(scenario/h)	-23%	6%	-52%	-14%	-60%					
		%scenario	7%	16%	-53%	1%	-61%					
RQ-T12	distance to lead vehicle	m(THW)		20%					31%			
		sd(THW)		-16%					0%			
		min(THW)		22%	25%	8%	39%	19%	69%			
		min(TTC)		25%	26%	25%	9%	45%	80%			
RQ-T16	behaviour of preceding traffic	m(v_LeadVeh)		-9%	-5%	-5%	-6%	-3%	-13%			
		diff(v_LeadVeh)		-11%	1%	103%	0%	0%	0%			
		sd(v_LeadVeh)		-21%	-23%	-20%	-33%	-19%	-23%			
		N(LaneChange_LeadVeh)								-33%		

Table 3.13: Effect sizes for PIs when comparing ADF driven vehicle with manual vehicle in the baseline condition for different RQs. Blue indicates a reduction of the indicator compared to baseline and red an increase. The depth of the colouring codes the size of the effect.

Research question		Performance indicator	Uninfluenced driving	Following	Approaching lead vehicle	Cut-In	Lane change	Approaching traffic jam	Driving in traffic jam	trip	trip - no TJ	trip - TJ
RQ-T6	**vehicle dynamics**	**min(ax)**	0.09	-0.22	-0.30	-0.29	-0.16	0.35	0.21			
		max(ax)	-0.21	0.15	-0.14	-0.08	-0.37	0	-0.73			
		sd(ax)	-0.21	0.18	0.18	0.25	-0.04	-0.34	0			
		max(abs(ay))	-0.14	0.00	-0.21	-0.39	0.18	0	-0.39			
		sd(ay)	-0.15	0.12	-0.17	-0.44	0.29	0	-0.34			
RQ-T7	**accuracy of driving**	**m(lat Pos)**	-0.94	-1.28	-0.76	-1.16		-1.19	-0.10			
		sd(lat Pos)	-0.71	-0.94	-0.60	-0.68		-0.46	-0.32			
		sd(v)	-0.25	-0.33	0.21	0.36	-0.17	-0.70	-0.50			
RQ-T8	**speed**	**m(v)**	-0.12	-0.13	-0.26	-0.24		-0.25	0			
		max(v)	-0.14	-0.15	-0.25	-0.20		-0.33	-0.41			
RQ-T9	**energ. Consump.**	**m(Energy consum.)**								-0.68	-0.66	-0.89
RQ-T11	**frequency of events**	**N(scenario/h)**	-0.48	0.08	-1.01	-0.13	-1.08					
		%scenario	0.11	0.20	-0.94	0.01	-1.19					
RQ-T12	**distance to lead vehicle**	**m(THW)**		0.53					0.66			
		sd(THW)		-0.14					0.00			
		min(THW)		0.51	0.54	0.17	0.33	0.40	0.34			
		min(TTC)		0.49	0.30	0.26	0.08	0.41	0.41			
RQ-T16	**behaviour of preceding traffic**	**m(v_LeadVeh)**		-0.45	-0.28	-0.21	-0.26	-0.16	-0.25			
		diff(v_LeadVeh)		-0.03	0.01	0.61	0	0	0			
		sd(v_LeadVeh)		-0.21	-0.21	-0.23	-0.27	-0.31	-0.52			
		N(LaneChange_LeadVeh)								-0.43		

3.2 Urban

The following section describes the results for the urban ADF and answers the RQs specified in Section 2.3 and Annex 1. As mentioned in Section 2.5.1, it must be kept in mind that these results are all based on bootstrapped data and that this bootstrapping process will also have an influence on the results. Most importantly, it must be kept in mind that the bootstrapping process added a small amount of noise to the observations to make it harder to trace from which Pilot site an observation was originating. Therefore, extreme values in the histograms can be exaggerated due to the added noise.

As mentioned in Section 2.5, the urban setting offered many possibilities for the variation of the Pilot site setting. Differences occurred in the setting of the Pilot site within the respective cities, the routing used for the pilots, as well as with the analysed ADFs, to name but three. Therefore, the results presented in this section should always be viewed with these limitations in mind.

The results presented in the following sections, are discussed separately for **intersections** and for **lane-bound scenarios** (traffic following the lanes outside of intersections). Additionally, **scenarios involving lane changes** were analysed for the relevant RQs. These include lane change, cut-in and overtaking with both passive and active oncoming traffic.

Finally, a series of results regarding roundabouts – generated using the Application Platform for Intelligent Mobility (AIM) (see section 2.5.6) – are presented and discussed at the end of this chapter.

3.2.1 RQ-T1: How reliable is system performance in a given driving and traffic scenario?

Due to the differences among the analysed functions within the urban context, a quantitative analysis of the performance of the functions within scenarios was not done for all scenarios, as the ODD of the functions differed. For some functions, ODD limits were reached multiple times per trip. Consequently, the function had to be deactivated and therefore the trips were split into multiple sections with ADF active for ADF per trip and one section per trip for baseline driving. This limitation of the analysed functions can also be seen in Figure 3.18, which shows the duration of sections where the ADF is active and the length of the complete trip for baseline. This can also be seen when looking at the median duration of trip sections (Table 3.14).

Table 3.14: Median duration of trip sections:

Condition	Median Duration [min]
Baseline	19.14
ADF	11.07

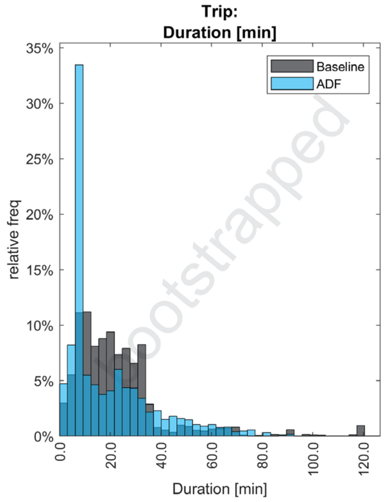


Figure 3.18: Trip sections. For baseline, this is mostly the complete trip . For ADF, this is corresponding to sections with ADF active.

3.2.2 RQ-T6: What Is the Impact of ADF on Vehicle Dynamics in Defined Driving Situations?

For evaluation of the vehicle dynamics, the longitudinal and lateral accelerations were analysed for all three categories of scenarios.

Lane-Bound Scenarios

For lane-bound scenarios, the changes between baseline and ADF were small (Figure 3.19). The ADF tended to drive more smoothly in general, which was reflected in the smaller maximum respective larger minimum accelerations. For the case of minimum accelerations, one explanation could be the more cautious approach of the ADF reacting to vehicles and obstacles earlier compared to the human driver.

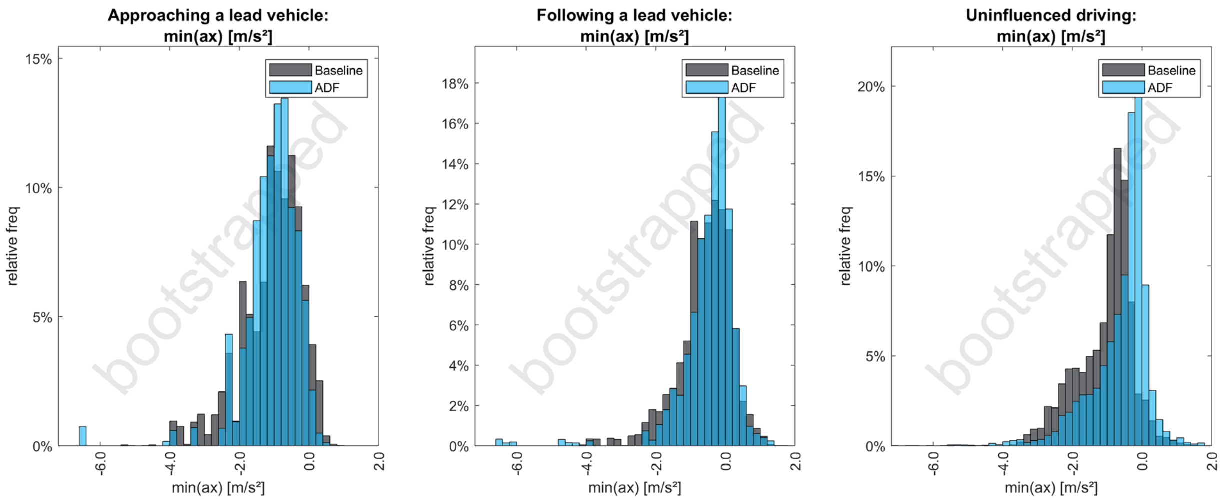


Figure 3.19: Minimum accelerations within lane-bound scenarios.

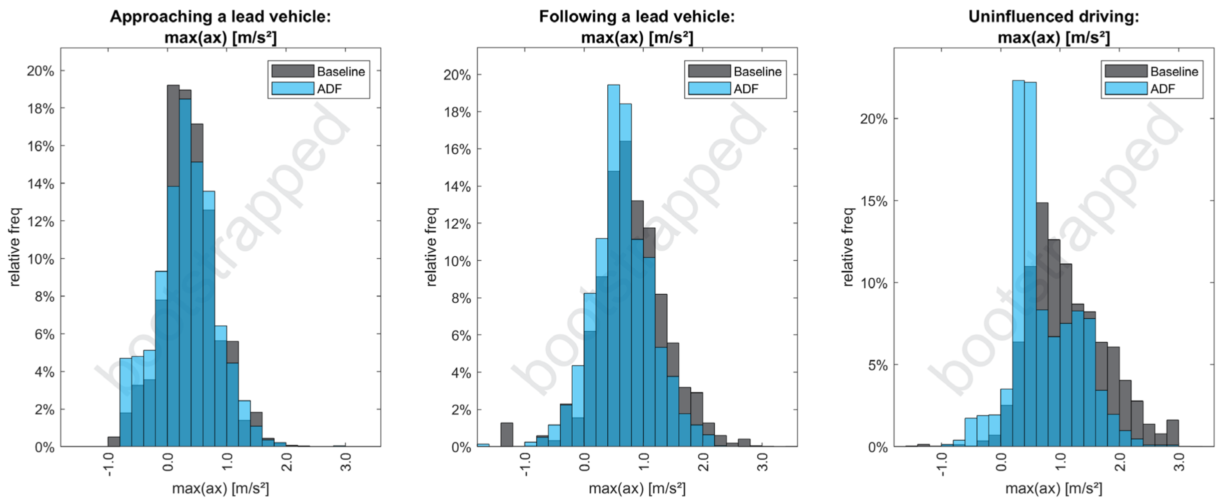


Figure 3.20: Maximum accelerations within lane-bound scenarios.

The standard deviations of the longitudinal accelerations within these scenarios were smaller for the ADF for all lane-bound scenarios (Figure 3.20). This can be interpreted as smoother acceleration behaviour, as would be expected by an ADF. A medium effect can, however, only be seen for uninfluenced driving, as this scenario allows adjusting the accelerations without influence where the smoother regulation of the ADF comes into play. A summary of driving dynamics in lane-bound scenarios is given in Table 3.15.

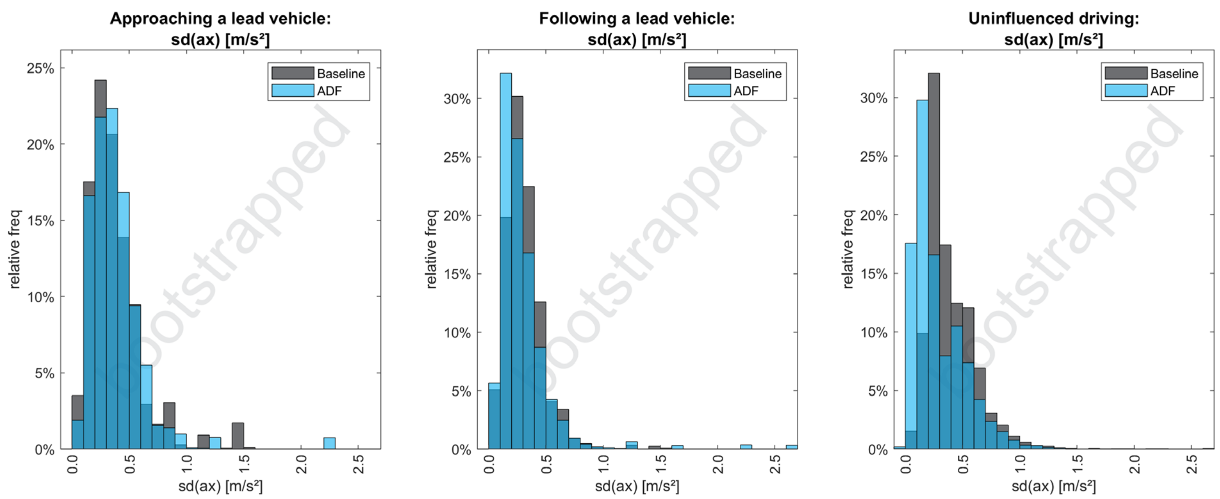


Figure 3.21: Standard deviations of longitudinal accelerations within lane-bound scenarios.

Table 3.15: Detailed results for the indicators of vehicle dynamics for lane-bound scenarios.

Indicator	Scenario	Z	p	Change in %	Effect Size
max(|ay|)	Approaching a lead vehicle	-1.8	0.080	-1%	-0.02
	Following a lead vehicle	-6.6	0.000	-5%	-0.09
	Uninfluenced driving	-8.6	0.000	-25%	-0.50
max(ax)	Approaching a lead vehicle	-5.3	0.000	-13%	-0.10
	Following a lead vehicle	-8.6	0.000	-19%	-0.28
	Uninfluenced driving	-8.6	0.000	-34%	-0.65
mean(ax)	Approaching a lead vehicle	-4.9	0.000	-11%	-0.08
	Following a lead vehicle	1.3	0.190	8%	0.01
	Uninfluenced driving	8.6	0.000	467%	0.23
mean(ay)	Approaching a lead vehicle	1.9	0.060	4%	0.03
	Following a lead vehicle	7.4	0.000	31%	0.11
	Uninfluenced driving	8.6	0.000	295%	0.68
min(ax)	Approaching a lead vehicle	-1.0	0.324	-1%	-0.02
	Following a lead vehicle	8.0	0.000	19%	0.14
	Uninfluenced driving	8.6	0.000	43%	0.53
sd(ax)	Approaching a lead vehicle	1.7	0.089	2%	0.03
	Following a lead vehicle	-4.9	0.000	-5%	-0.06
	Uninfluenced driving	-8.6	0.000	-23%	-0.4
sd(ay)	Approaching a lead vehicle	4.4	0.000	4%	0.06
	Following a lead vehicle	5.3	0.000	4%	0.06
	Uninfluenced driving	-8.6	0.000	-20%	-0.40

Scenarios Involving Lane Changes

For both lane-change and cut-in scenarios, the ADF decreased the standard deviation of longitudinal and lateral accelerations. The ADF also decreased the maximum absolute lateral accelerations in both scenarios (Figure 3.23) as well as the maximum longitudinal accelerations (Figure 3.24). However, for the lane-change scenario the decrease of maximum longitudinal acceleration was not statistically significant, possibly due to the larger variance with baseline values: With ADF the values were mostly between 0 to 1 m/s2, but baseline had a longer negative tail and more values around zero. The decrease in maximum acceleration with ADF was in line with the lane-bound scenarios. The mean and minimum longitudinal and lateral accelerations were not straightforward to interpret for the current scenarios – particularly since the sign of lateral accelerations changed during lane changes.

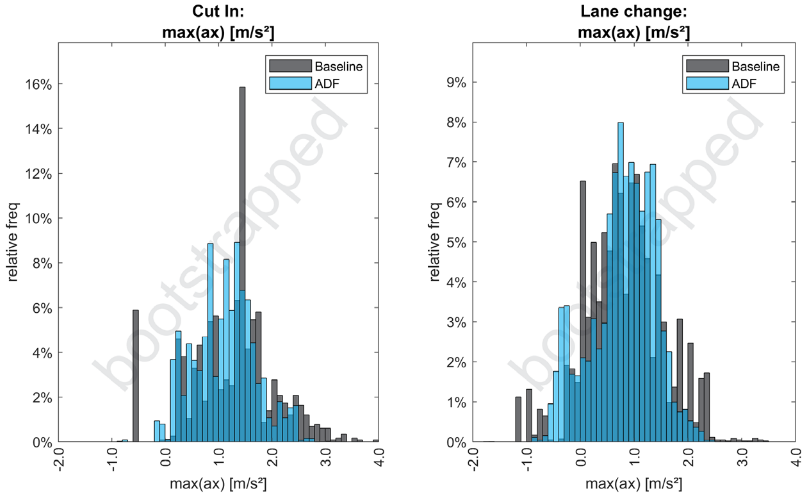


Figure 3.22: Maximum longitudinal accelerations within scenarios involving lane changes.

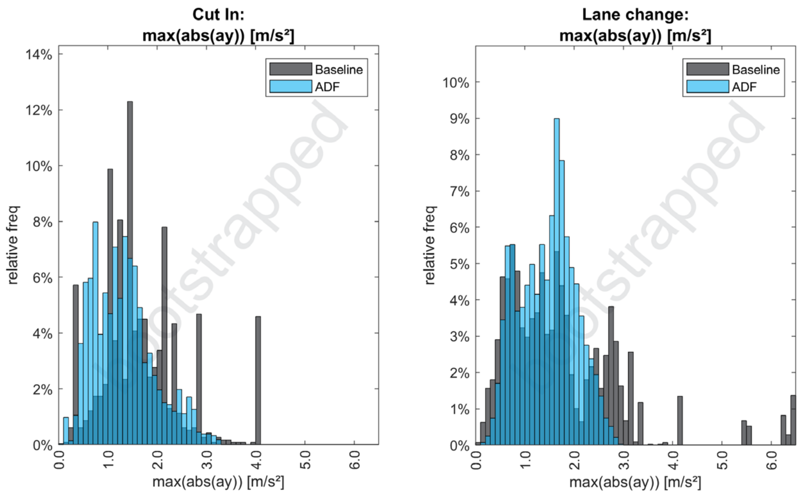


Figure 3.23: Maximum absolute lateral accelerations within scenarios involving lane changes.

Table 3.16: Detailed results for the indicators of vehicle dynamics for scenarios involving lane changes.

Indicator	Scenario	Z	p	Change in %	Effect Size
max(|ay|)	Cut-in	-8.2	0.000	-22%	-0.55
	Lane change	-8.6	0.000	-25%	-0.64
max(ax)	Cut-in	-5.5	0.000	-11%	-0.22
	Lane change	-1.1	0.261	-2%	-0.03
mean(ax)	Cut-in	4.9	0.000	37%	0.19
	Lane change	-3.4	0.001	-67%	-0.07
mean(ay)	Cut-in	8.6	0.000	396%	0.65

Indicator	Scenario	Z	p	Change in %	Effect Size
	Lane change	8.6	0.000	172%	0.37
min(ax)	Cut-in	1.5	0.128	3%	0.03
	Lane change	-3.0	0.003	-5%	-0.05
sd(ax)	Cut-in	-5.1	0.000	-9%	-0.19
	Lane change	-4.3	0.000	-5%	-0.09
sd(ay)	Cut-in	-7.6	0.000	-18%	-0.40
	Lane change	-4.7	0.000	-4%	-0.09

Intersections

For intersections, effects were especially present for the lateral accelerations (cf. Table 3.17), which were mostly positive except for crossing with a laterally moving object, where the effect was negative. As would be expected, these effects were larger in general for turning at intersections compared to simply crossing them. This was in line with the more cautious approach of the ADF to challenging scenarios.

For longitudinal accelerations, the effects were mixed (positive and negative) and small to medium at most, as is shown in the distributions in Figure 3.24, Figure 3.25, and Figure 3.26. This was to be expected, as the main influence on driving dynamics within intersections was probably the road network, and traffic regulations left less space for differences between ADF and human driving.

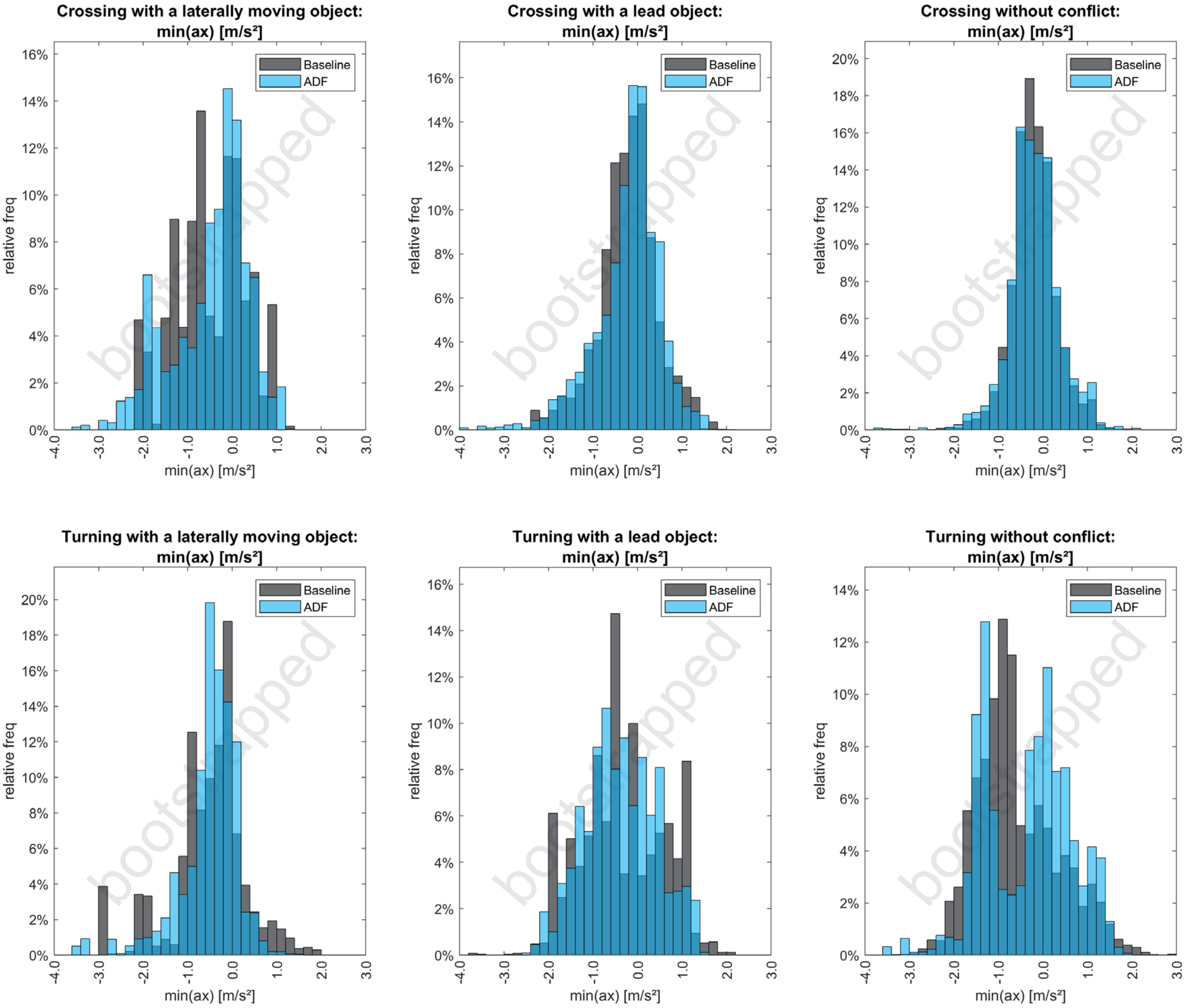


Figure 3.24: Minimum longitudinal accelerations within intersection scenarios.

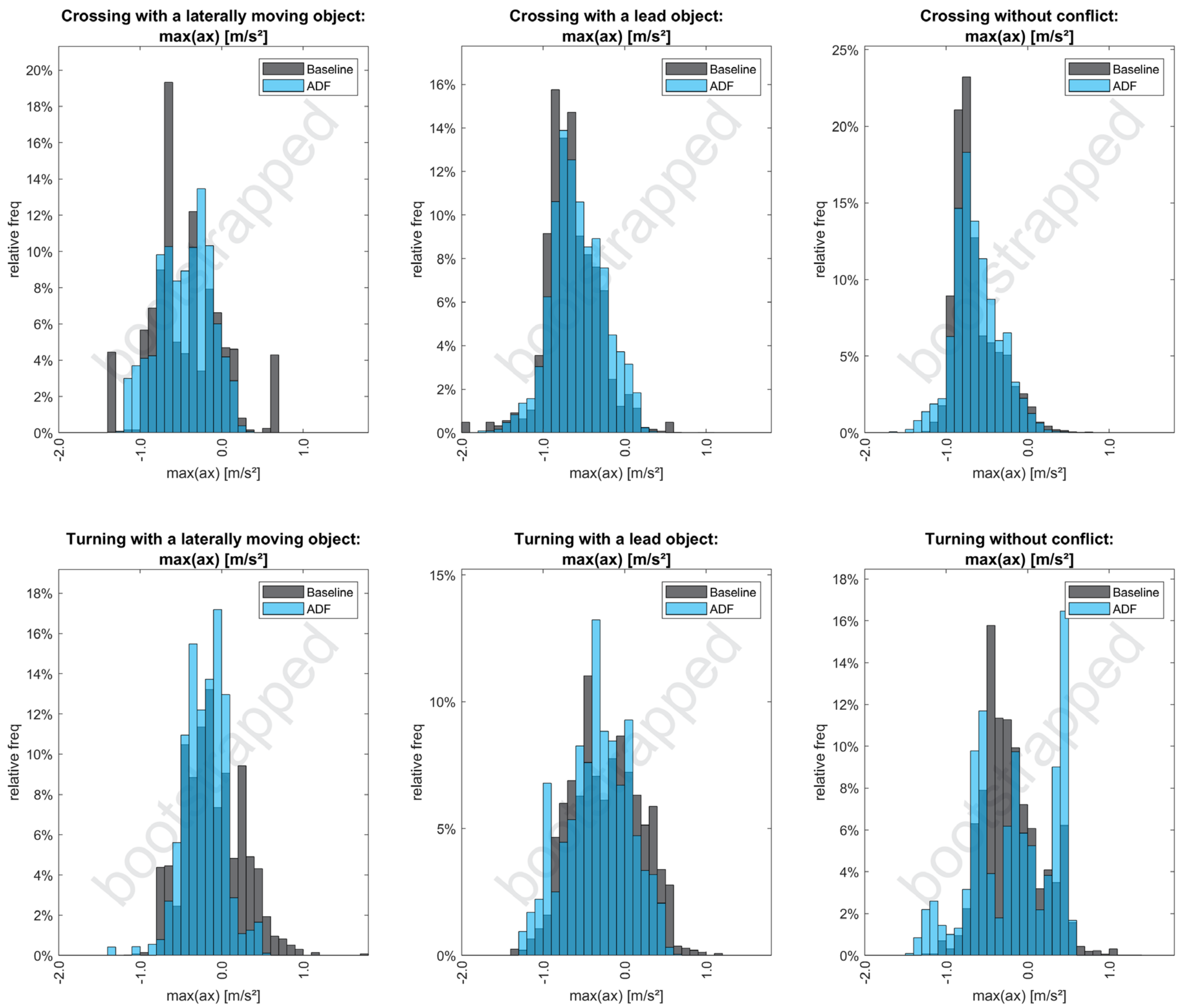


Figure 3.25: Maximum longitudinal accelerations within intersection scenarios.

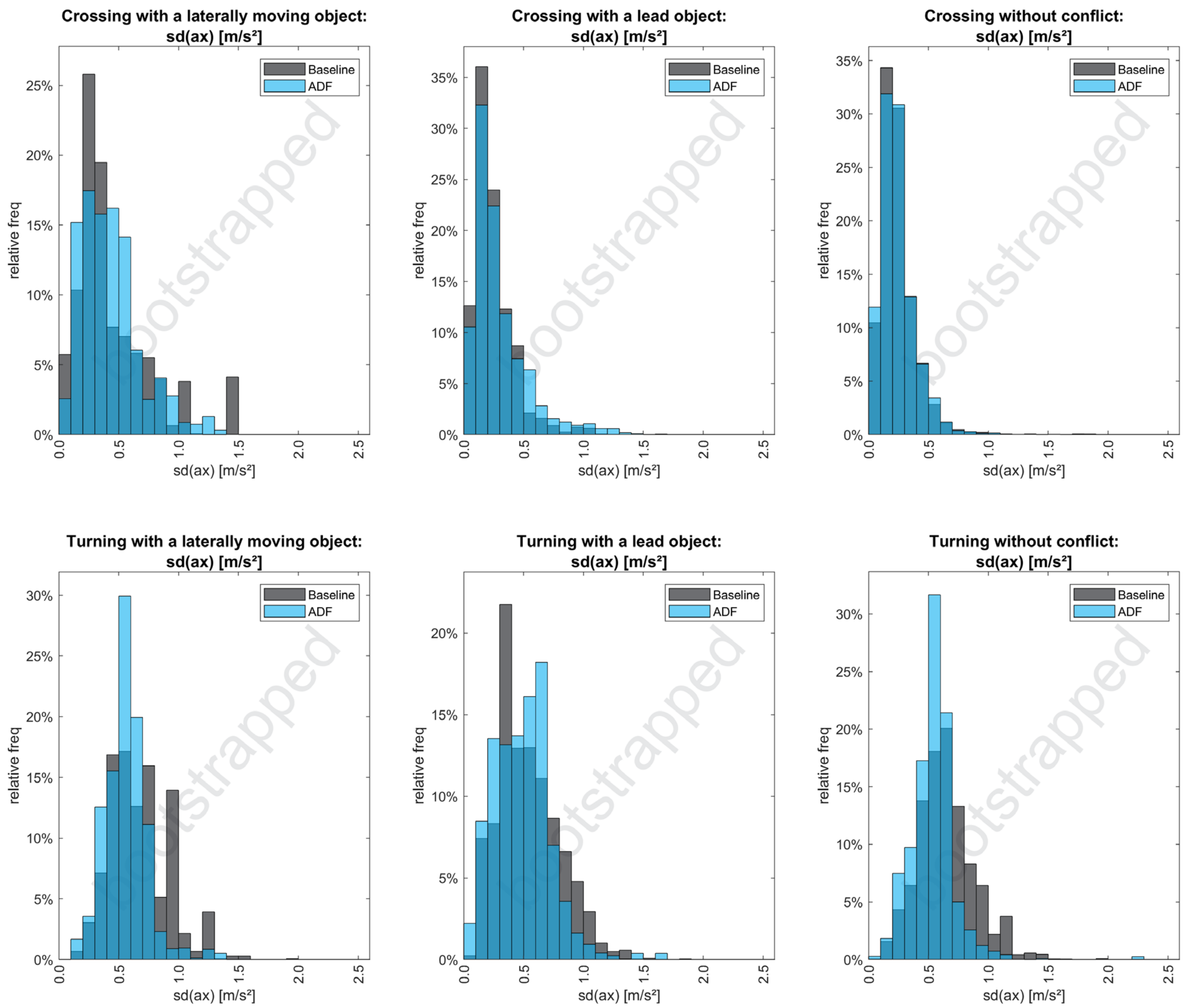


Figure 3.26: Standard deviations of the longitudinal accelerations within intersection scenarios.

Table 3.17: Detailed results for the indicators of vehicle dynamics for intersection scenarios.

Indicator	Scenario	Z	p	Change in %	Effect Size
max(|ay|)	Crossing with laterally moving object	-3.8	0.000	-11%	-0.21
	Crossing with lead object	8.6	0.000	14%	0.19
	Crossing without conflict	8.6	0.000	13%	0.20
	Turning with laterally moving object	-8.4	0.000	-14%	-0.56
	Turning with lead object	-8.5	0.000	-16%	-0.34
	Turning without conflict	-8.6	0.000	-29%	-0.93
max(ax)	Crossing with laterally moving object	-1.9	0.057	-3%	-0.06
	Crossing with lead object	8.6	0.000	19%	0.21
	Crossing without conflict	7.2	0.000	3%	0.04

Indicator	Scenario	Z	p	Change in %	Effect Size
	Turning with laterally moving object	-7.9	0.000	-12%	-0.39
	Turning with lead object	-7.9	0.000	-13%	-0.25
	Turning without conflict	5.3	0.000	2%	0.04
mean(ax)	Crossing with laterally moving object	2.1	0.037	9%	0.05
	Crossing with lead object	8.5	0.000	38%	0.16
	Crossing without conflict	7.6	0.000	9%	0.05
	Turning with laterally moving object	-4.9	0.000	-13%	-0.21
	Turning with lead object	-4.8	0.000	-16%	-0.14
	Turning without conflict	8.6	0.000	28%	0.17
mean(ay)	Crossing with laterally moving object	-7.5	0.000	-87%	-0.45
	Crossing with lead object	6.2	0.000	43%	0.09
	Crossing without conflict	8.6	0.000	144%	0.26
	Turning with laterally moving object	7.4	0.000	1699%	0.28
	Turning with lead object	8.6	0.000	208%	0.32
	Turning without conflict	8.6	0.000	110%	0.89
min(ax)	Crossing with laterally moving object	0.3	0.756	1%	0.00
	Crossing with lead object	-1.9	0.052	-5%	-0.01
	Crossing without conflict	3.6	0.000	4%	0.01
	Turning with laterally moving object	0.1	0.896	1%	0.01
	Turning with lead object	-4.1	0.000	-26%	-0.09
	Turning without conflict	8.6	0.000	39%	0.24
sd(ax)	Crossing with laterally moving object	-0.5	0.596	-2%	-0.03
	Crossing with lead object	8.6	0.000	20%	0.22
	Crossing without conflict	3.5	0.001	1%	0.02
	Turning with laterally moving object	-8.6	0.000	-16%	-0.58
	Turning with lead object	-6.7	0.000	-8%	-0.18
	Turning without conflict	-8.6	0.000	-17%	-0.53
sd(ay)	Crossing with laterally moving object	1.6	0.110	3%	0.05
	Crossing with lead object	8.6	0.000	15%	0.18
	Crossing without conflict	8.6	0.000	12%	0.17
	Turning with laterally moving object	-7.6	0.000	-13%	-0.45
	Turning with lead object	-2.7	0.008	-5%	-0.09
	Turning without conflict	-1.7	0.090	-1%	-0.02

3.2.3 RQ-T7: What Is the Impact of ADF on the Accuracy of Driving?

For intersections it was not possible to analyse the accuracy of driving in the same way as for lane-bound scenarios because lanes were not that clearly defined within intersections. Therefore, positioning in the lane could not be analysed.

The same was true of scenarios involving lane changes, where an analysis of deviation from the lane centre or movement within the lane could not be done. The movement conducted when changing lanes is already assessed when looking at the vehicle dynamics (cf. RQ-T6, Section 3.2.2).

Lane-Bound Scenarios

To analyse the accuracy of driving, the mean position within the lane was analysed (Figure 3.27 and Figure 3.28). This indicator specified how far the vehicles deviates from the middle of the current lane. Overall, it seemed that the ADF is better at driving in the lane centre, which was both expected and consistent with the results from the vehicle dynamics. The effect, however, was small.

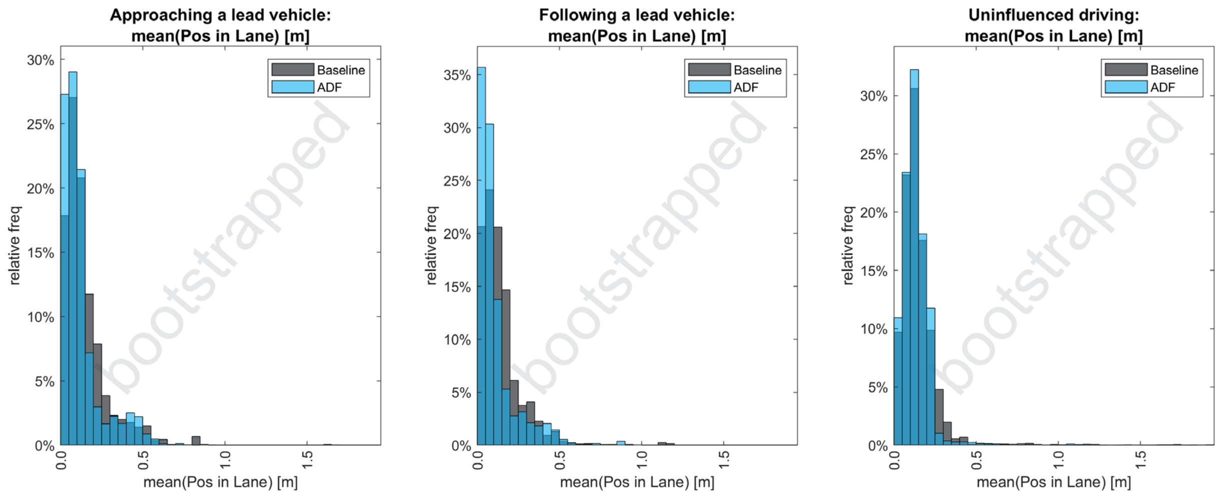


Figure 3.27: Mean position in lane in lane-bound scenarios.

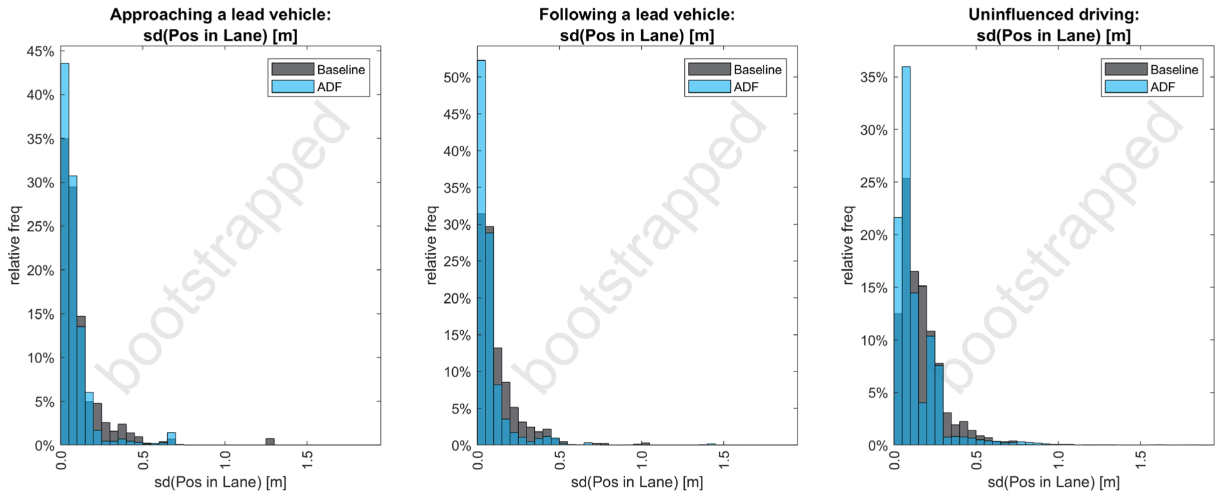


Figure 3.28: Standard deviation of the position in lane in lane-bound scenarios.

Table 3.18: Detailed results for the indicators for accuracy of driving.

Indicator	Scenario	Z	p	Change in %	Effect Size
mean(Position in lane)	Approaching a lead vehicle	-7.6	0.000	-18%	-0.20
	Following a lead vehicle	-8.6	0.000	-20%	-0.23
	Uninfluenced driving	-4.9	0.000	-2%	-0.02
sd(Position in lane)	Approaching a lead vehicle	4.2	0.000	41%	0.09
	Following a lead vehicle	-4.6	0.000	-20%	-0.05
	Uninfluenced driving	8.6	0.000	18%	0.06

3.2.4 RQ-T8: What Is the Impact of ADF on the Driven Speed?

For the impact on the driven speed, the mean and maximum driven speeds within scenario instances were analysed.

Lane-Bound Scenarios

The ADF had an impact on the driven speed (Figure 3.29). The effect was largest in uninfluenced driving because the ADF adhered to the speed limit of 50 km/h within cities, while human drivers did not necessarily do so. In scenarios where the ADF is influenced by other vehicles, the differences were smaller, as the ADF was then driving along with the other vehicles. The strict adherence to the speed limit, however, had implications for the scenarios encountered on these trips (cf. RQ-T11 (Section 3.2.5) & RQ-T12 (Section 3.2.6)).

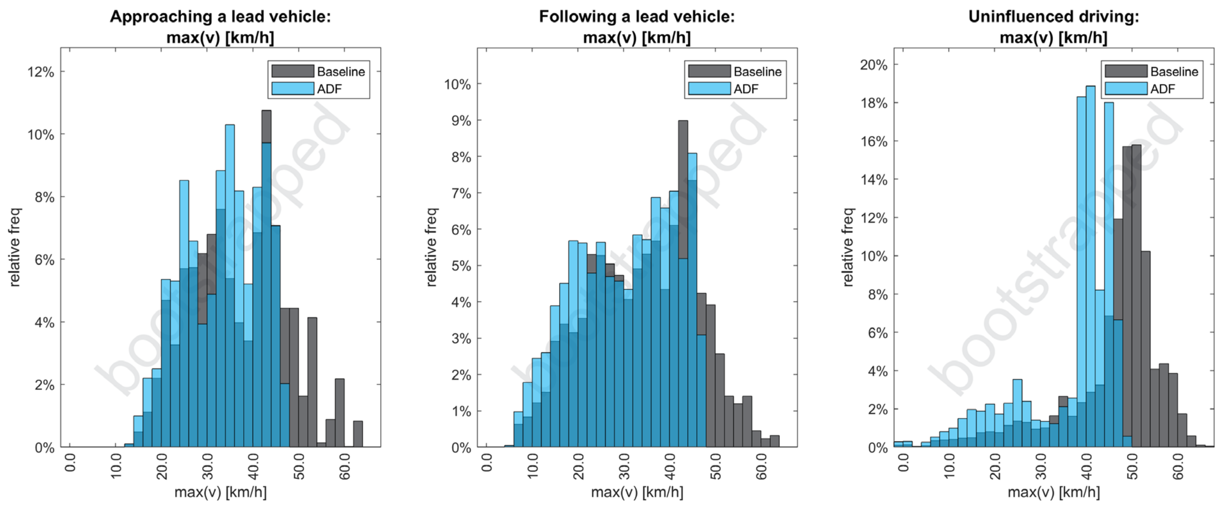


Figure 3.29: Maximum driven speeds within lane-bound scenarios.

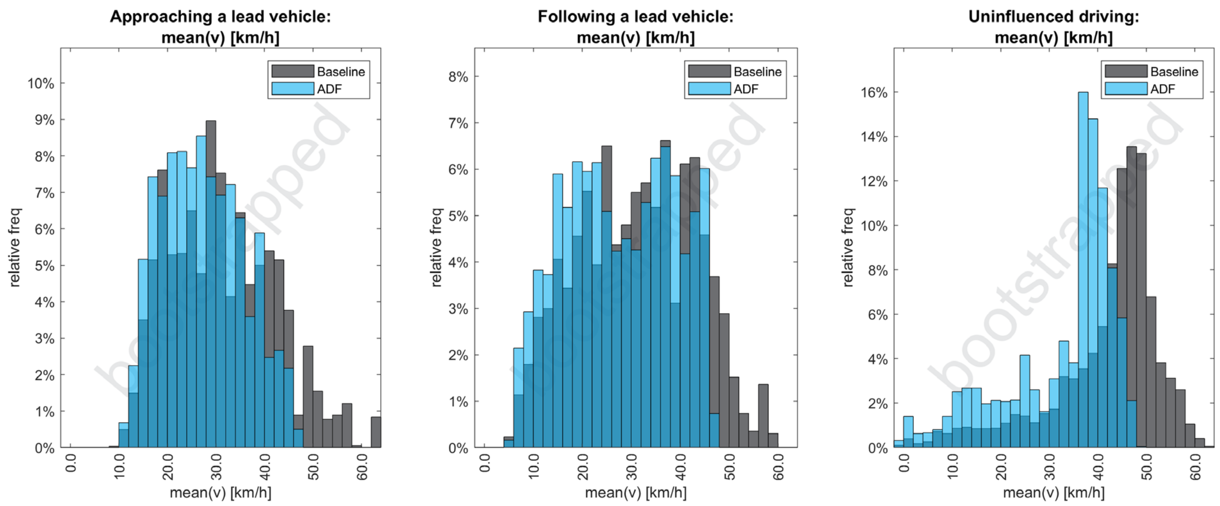


Figure 3.30: Mean driven speeds within lane-bound scenarios.

The standard deviation of the longitudinal acceleration (cf. RQ-T6, Section 3.2.2) in combination with the standard deviation of the driven speed (Figure 3.31) shows that changes in speed diminished for following and uninfluenced scenarios compared to baseline. For approaching scenarios, the standard deviations were slightly higher.

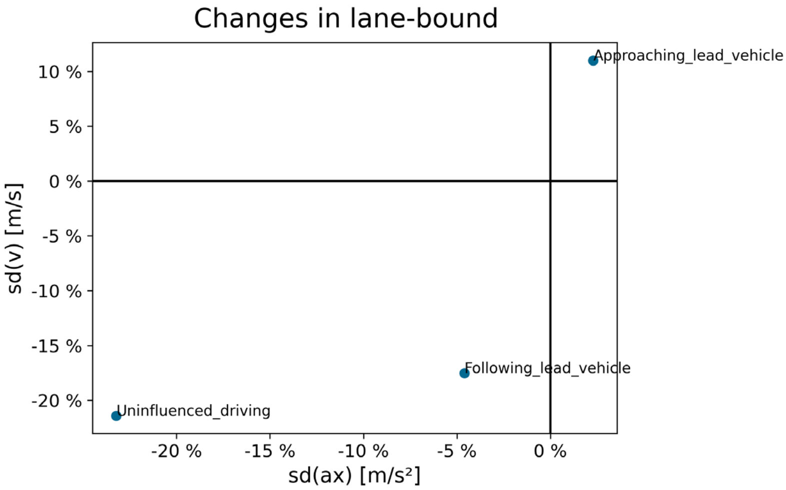


Figure 3.31: Changes in the standard deviations of lane-bound scenarios.

Table 3.19: Detailed results for the indicators for driven speed in lane-bound scenarios.

Indicator	Scenario	Z	p	Change in %	Effect Size
max(v)	Approaching a lead vehicle	-8.6	0.000	-10%	-0.38
	Following a lead vehicle	-8.6	0.000	-12%	-0.36
	Uninfluenced driving	-8.6	0.000	-20%	-0.91
mean(v)	Approaching a lead vehicle	-8.6	0.000	-12%	-0.4
	Following a lead vehicle	-8.6	0.000	-11%	-0.31
	Uninfluenced driving	-8.6	0.000	-22%	-0.82
sd(v)	Approaching a lead vehicle	7.1	0.000	11%	0.16
	Following a lead vehicle	-8.5	0.000	-18%	-0.22
	Uninfluenced driving	-8.6	0.000	-21%	-0.24

Scenarios Involving Lane Changes

For scenarios involving lane changes, the maximum (Figure 3.32) and mean speeds (Figure 3.33) both showed differences when comparing ADF to baseline driving (cf. Table 3.20). For lane changes, the speeds were lower with ADF active, which was consistent with the overall driven speed of the ADF. Considering that the ADFs mostly only changed lane for navigational purposes, this result was conclusive. Cut-ins happened at higher speeds of the ADF compared to baseline, which is in line with the overall higher driven speed by the other vehicles. Therefore, cut-ins that influence the ADF mostly occurred when it was travelling at higher speeds. Cut-ins occurring when the ADF was travelling at a lower speed did not have (or more seldom had) an influence, as the other vehicle is mostly faster.

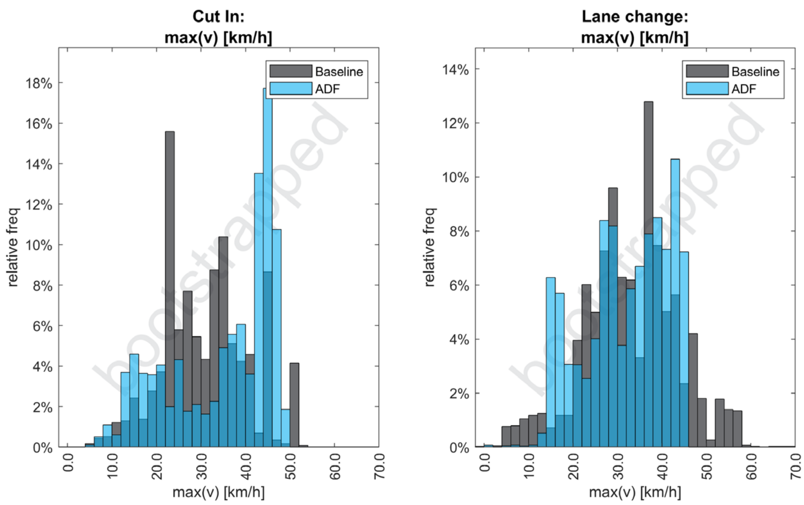


Figure 3.32: Maximum driven speeds within scenarios involving lane changes.

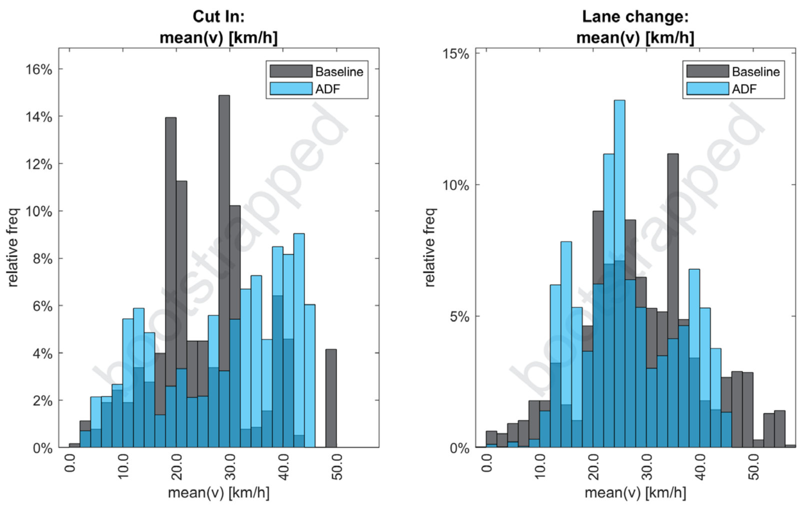


Figure 3.33: Mean driven speeds within scenarios involving lane changes.

The standard deviation of the driven speed against the standard deviation of the longitudinal acceleration (cf. RQ-T6, Section 3.2.2) shows that the regulation of the driven speed was less stable compared to baseline, whilst the regulation of the acceleration was slightly more stable.

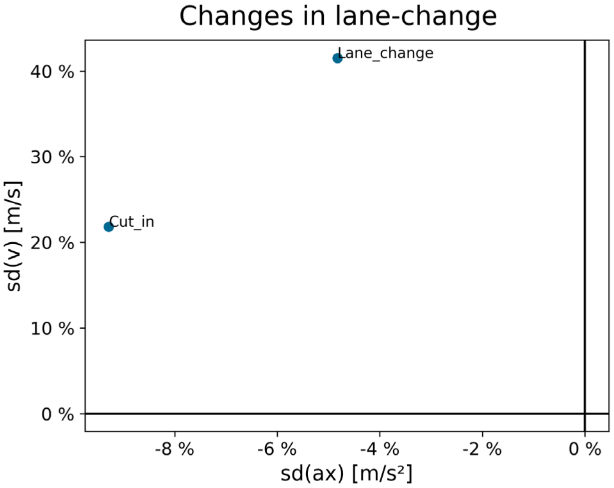


Figure 3.34: Changes in the standard deviations of scenarios involving lane changes.

Table 3.20: Detailed results for the indicators for driven speed in scenarios involving lane changes.

Indicator	Scenario	Z	p	Change in %	Effect Size
max(v)	Cut-in	8.2	0.000	13%	0.35
	Lane change	-6.0	0.000	-4%	-0.12
mean(v)	Cut-in	8.0	0.000	13%	0.29
	Lane change	-8.3	0.000	-8%	-0.26
sd(v)	Cut-in	7.1	0.000	22%	0.26
	Lane change	8.6	0.000	41%	0.40

Intersections

For intersection scenarios, the same holds true as for lane-bound scenarios: The speed limit had more influence on ADFs than it has on human drivers. Especially in the scenarios of crossing and turning without conflict, it can clearly be seen that the speed limit made a difference in the travelled speeds across intersections. The effect and the percentual change were almost the same for crossing and turning, although the travelled speeds were higher for crossing scenarios. For the other scenarios, the influence of the other road users again had an influence on the driven speed for both ADF and baseline.

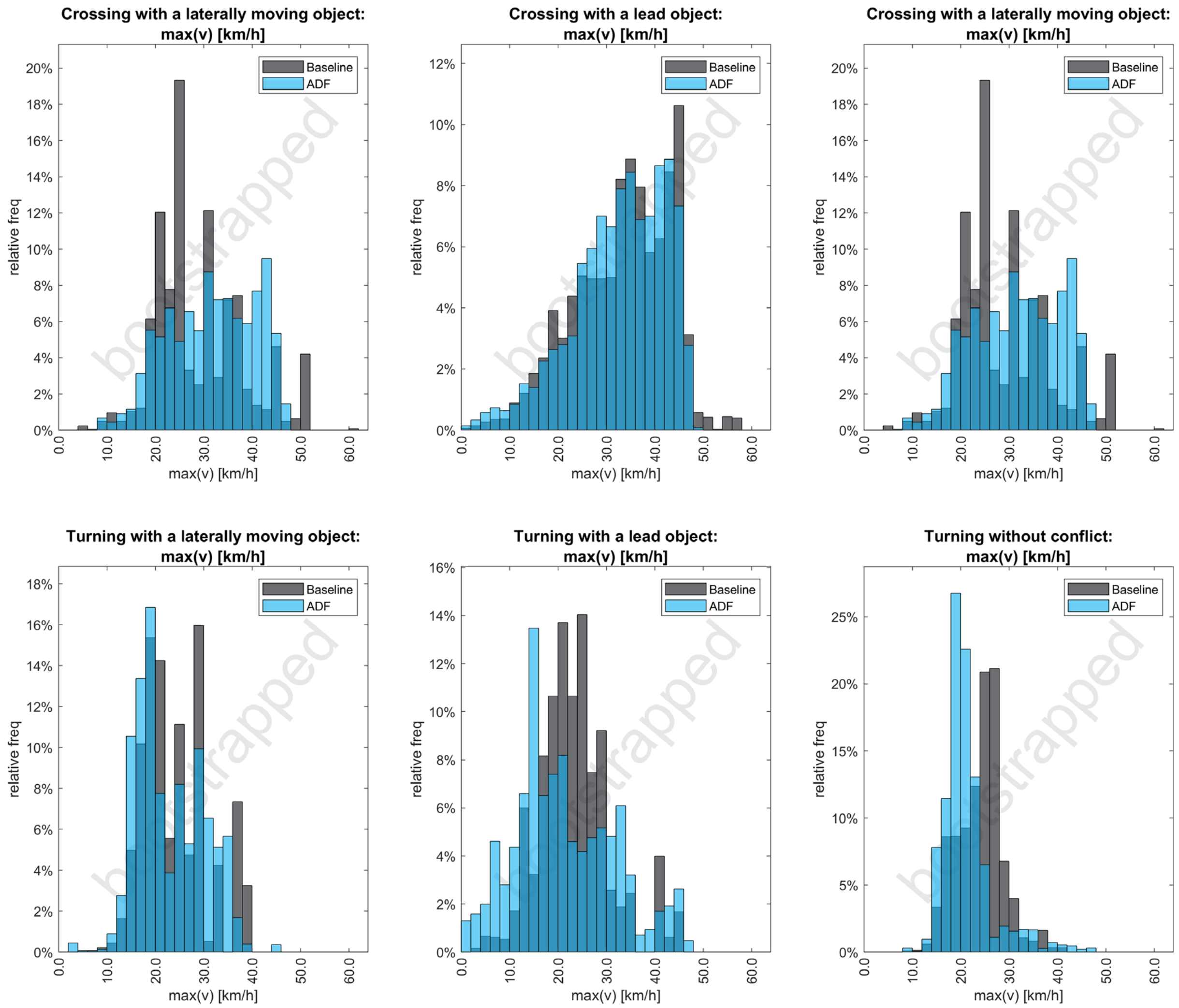


Figure 3.35: Maximum driven speeds within intersection scenarios.

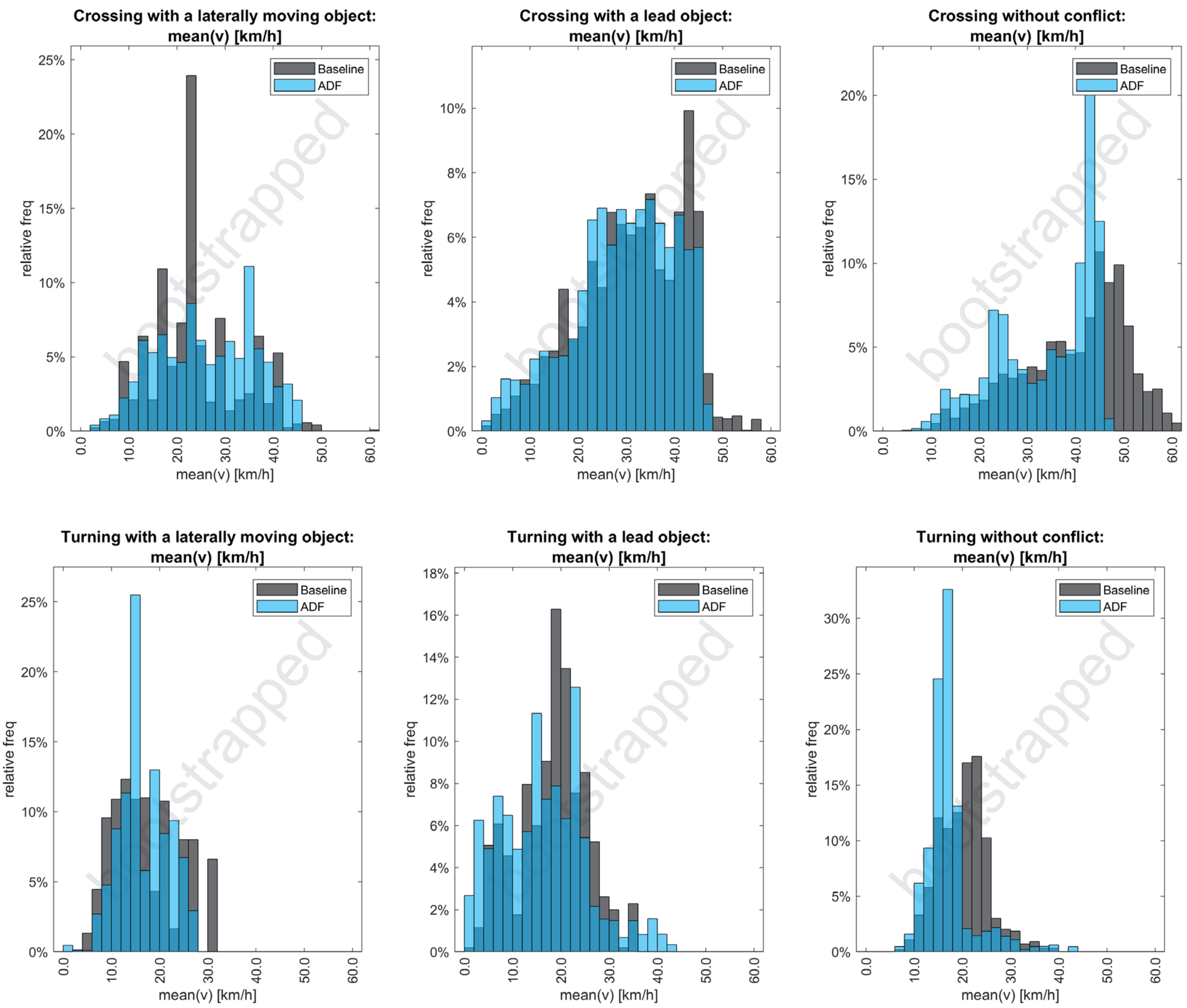


Figure 3.36: Mean driven speeds within intersection scenarios.

As for the other two scenario categories (lane-bound, lane-change), the standard deviations of the longitudinal acceleration (cf. RQ-T6, Section 3.2.2) were compared with those of the driven speed (Figure 3.37). The results for intersections were more mixed, showing better regulation of the accelerations for all turning scenarios, while the speed regulations for the turning case were mixed. For crossing, the regulation of speed was less stable with ADF, whereas the regulation of accelerations showed mixed results.

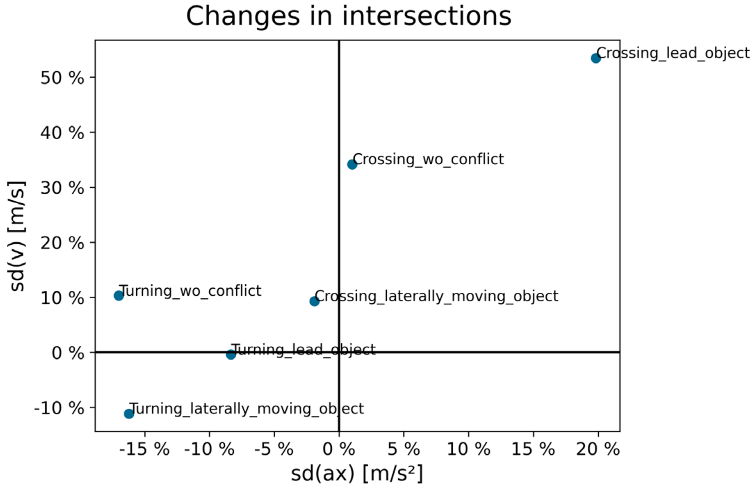


Figure 3.37: Changes in the standard deviation for intersection scenarios.

Table 3.21: Detailed results for the indicators for driven speed in intersection scenarios.

Indicator	Scenario	Z	p	Change in %	Effect Size
max(v)	Crossing with a laterally moving object	7.7	0.000	9%	0.29
	Crossing with a lead object	-6.3	0.000	-2%	-0.07
	Crossing without conflict	-8.6	0.000	-13%	-0.58
	Turning with a laterally moving object	-7.7	0.000	-5%	-0.17
	Turning with a lead object	-8.4	0.000	-9%	-0.20
	Turning without conflict	-8.6	0.000	-12%	-0.56
mean(v)	Crossing with a laterally moving object	7.8	0.000	11%	0.26
	Crossing with a lead object	-8.5	0.000	-6%	-0.16
	Crossing without conflict	-8.6	0.000	-15%	-0.56
	Turning with a laterally moving object	-3.3	0.001	-3%	-0.10
	Turning with a lead object	-7.7	0.000	-10%	-0.21
	Turning without conflict	-8.6	0.000	-16%	-0.63
sd(v)	Crossing with a laterally moving object	4.0	0.000	9%	0.12
	Crossing with a lead object	8.6	0.000	53%	0.37
	Crossing without conflict	8.6	0.000	34%	0.21
	Turning with a laterally moving object	-5.0	0.000	-11%	-0.20
	Turning with a lead object	0.0	0.995	-1%	-0.01
	Turning without conflict	8.3	0.000	10%	0.13

3.2.5 RQ-T11: What Is the Impact of ADF on the Frequency of Certain Events?

The analysis of this RQ is split into two parts. First, the analysis of some events is described on trip level, independent of scenario. Afterwards, the analysis is described on scenario-level.

On trip level, the frequency and duration of stillstands was analysed. As can be seen (Figure 3.38, Table 3.22), ADF had a small effect on these two indicators. The duration of stillstands slightly decreased, while their frequency slightly increased. As for many indicators in the urban environment, it can again be concluded that the overall influences imposed on ADF and baseline by the urban environment equally had more effect on the driving and traffic behaviour compared to the behaviour implemented within the ADFs.

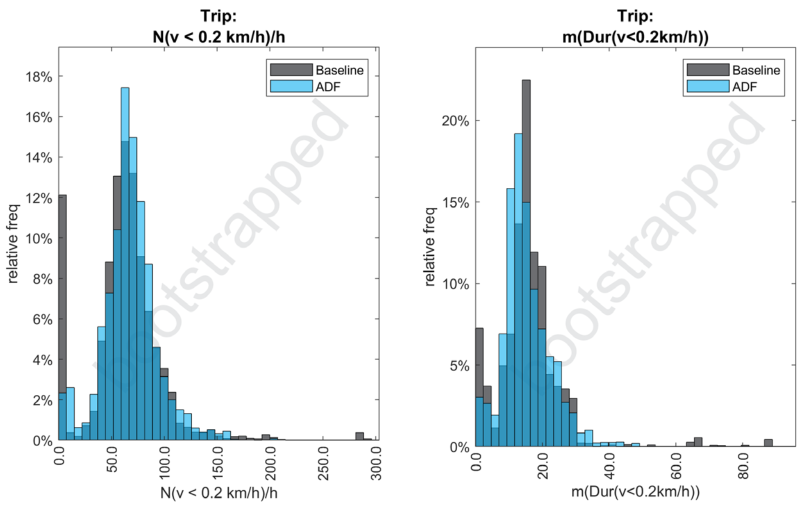


Figure 3.38: Frequency and duration of stillstands on trip level.

Table 3.22: Detailed results for frequency and duration of stillstands on trip level.

Indicator	Z	p	Change in %	Effect Size
m(Dur(v<0,2km/h))	-4.3	0.000	-3%	-0.07
N(v<0,2km/h)/h	7.7	0.000	8%	0.16

Lane-Bound Scenarios

For the lane-bound scenarios, it can be observed that almost all scenarios were overall shorter with ADF active. Approaching a lead object was the only scenario that decreased in frequency. This can be explained by the lower travelled speed of the ADF leading to more uninfluenced driving, which can also be seen when looking at the frequency of uninfluenced driving. On the other hand, uninfluenced driving scenarios were shorter, which means that the uninfluenced state was often interrupted by other traffic participants. This can also be seen for the following scenario, which almost doubled in frequency but was shorter compared to baseline. Again, the explanation lies within the lower speed driven by the ADF, resulting in leading vehicles leaving the influencing zone in front of the ADF faster compared to baseline.

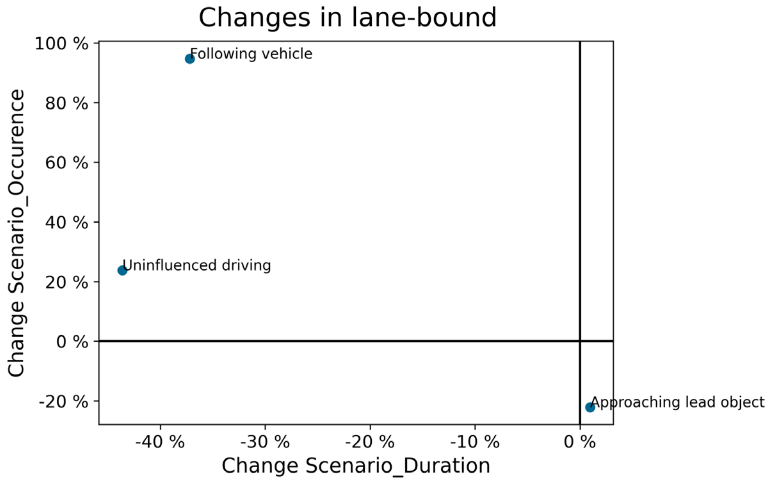


Figure 3.39: Change in the frequency and duration of lane-bound scenarios.

Table 3.23: Detailed results for the indicators for the frequency of scenarios.

Indicator	Scenario	Z	p	Change in %	Effect Size
Frequency	Approaching a lead object	-8.3	0.000	-22%	-0.20
	Following a lead vehicle	8.6	0.000	95%	0.37
	Following a lead VRU	7.3	0.000	29%	0.17
	Uninfluenced driving	8.6	0.000	24%	0.59
Duration	Approaching a lead object	1.7	0.082	1%	0.02
	Following a lead vehicle	-8.6	0.000	-37%	-0.57
	Uninfluenced driving	-8.6	0.000	-44%	-0.60

Scenarios Involving Lane Changes

For scenarios involving lane changes, Figure 3.40 shows that both scenarios had a longer duration with ADF active compared to baseline trips. Lane changes were longer and less frequent. A possible explanation could be the tendency of the ADF to choose its desired lane quite early (based on the configured route) and stick to it for the lane change. Distances kept to other vehicles and the overall lower travelled speed are again the explanation for fewer cut-ins occurring when travelling with ADF active. As shown in RQ-T8 (Section 3.2.4), the difference in travelled speed was rather small for cut-ins and resulted in smaller changes of cut-in duration.

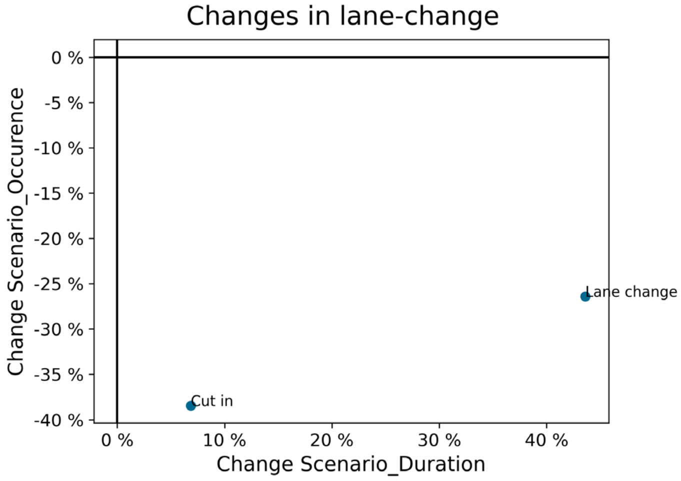


Figure 3.40: Changes in the duration and frequency of scenarios involving lane changes.

Table 3.24: Detailed results for the indicators for the frequency of scenarios involving lane changes.

Indicator	Scenario	Z	p	Change in %	Effect Size
Frequency	Cut-in	-8.6	0.000	-38%	-0.45
	Lane change	-8.6	0.000	-26%	-0.43
Duration	Cut-in	5.0	0.000	7%	0.15
	Lane change	8.6	0.000	44%	0.80

Intersections

At intersections, most scenarios had a longer duration with ADF active (cf. Figure 3.41 and Table 3.25). This was consistent with the lower driven speed within intersection (cf. RQ-T8, Section 3.2.4). The only outlier was the scenario Crossing with laterally moving objects, which was shorter and occurred more often with ADF active.

The increase in scenarios involving laterally moving objects was probably a result of the lower speeds of the ADF. Additionally, gaps kept to the lead vehicle were mostly larger. This probably encouraged other road users to utilise that gap and to drive into the ADF's lane or to cross the intersection in that larger gap.

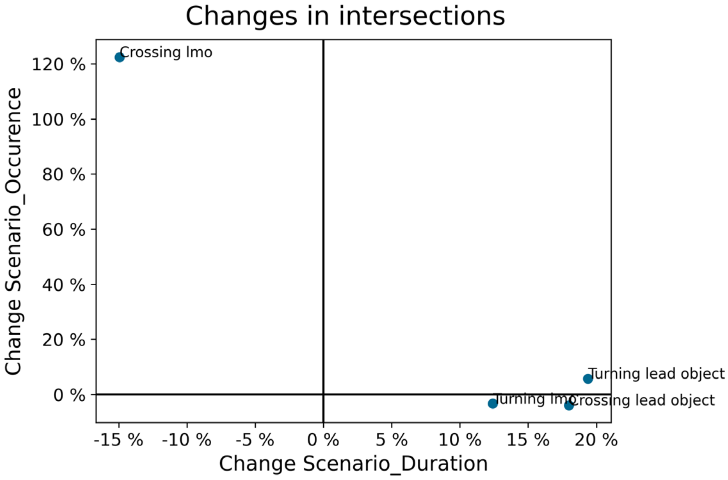


Figure 3.41: Change in the duration and frequency of intersection scenarios.

Table 3.25: Detailed results for the indicators for the frequency of scenarios.

Indicator	Scenarios	Z	p	Change in %	Effect Size
N(Crossing with laterally moving non VRU object)/N(Crossing)	Crossing with laterally moving object	8.6	0.000	123%	0.43
N(Crossing with laterally moving pedestrian)/N(Crossing)	Crossing with laterally moving object	7.6	0.000	62%	0.13
N(Crossing with lead object)/N(Crossing)	Crossing with lead object	-3.4	0.001	-4%	-0.06
N(Turning with laterally moving object)/N(Turning)	Turning with laterally moving object	-2.2	0.028	-3%	-0.04
N(Turning with laterally moving VRU)/N(Turning)	Turning with laterally moving object	7.3	0.000	29%	0.10
N(Turning with lead object)/N(Turning)	Turning with lead object	4.4	0.000	6%	0.05
Duration	Crossing with laterally moving object	-7.6	0.000	-15%	-0.36
	Crossing with lead object	8.6	0.000	18%	0.25
	Crossing without conflict	8.6	0.000	12%	0.21
	Turning with laterally moving object	8.3	0.000	12%	0.38
	Turning with lead object	8.6	0.000	19%	0.39
	Turning without conflict	8.6	0.000	26%	0.48

3.2.6 RQ-T12: What Is the Impact of ADF on the Interaction with Other Road Users in a Defined Driving Scenario?

For analysing the interaction with other road users, the distances and relative measures to the lead vehicle were considered, as these had an influence on the behaviour of the ego-vehicle.

Lane-Bound Scenarios

For the interaction in lane-bound scenarios, the time headway (THW) to the preceding vehicles was analysed. The minimum and mean THW was higher in these scenarios with ADF active compared to baseline. However, the effect was small. As already shown in the earlier RQs, the cases where influenced driving occurs were often those where the traffic was limited overall by (infrastructure) restrictions imposed by the urban environment. Therefore, when influenced driving occurred, differences between ADF and baseline are small.

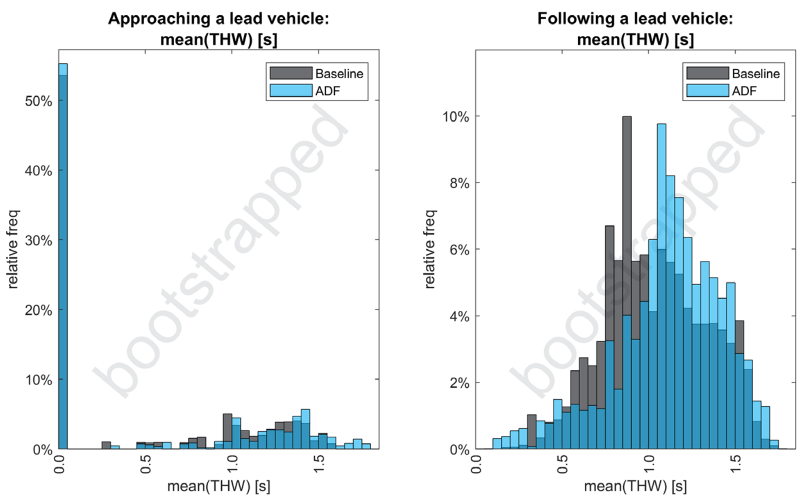


Figure 3.42: Mean THW for lane-bound scenarios.

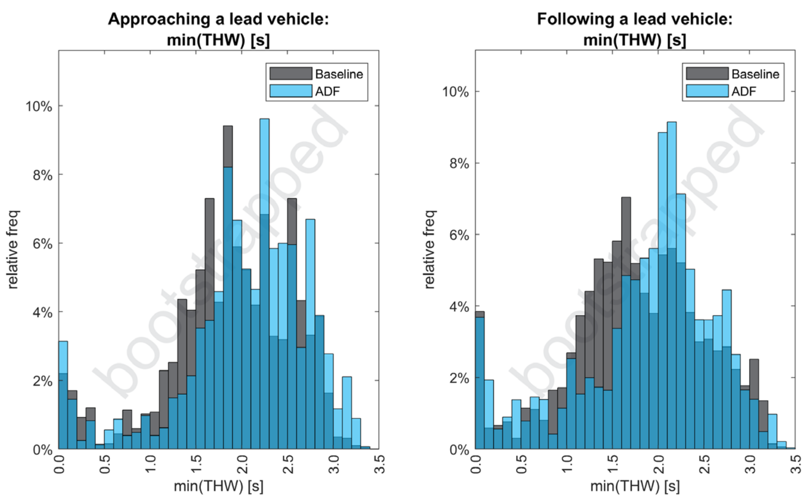


Figure 3.43: Minimum THW for lane-bound scenarios.

Table 3.26: Detailed results for THW indicators for lane-bound scenarios.

Indicator	Scenario	Z	p	Change in %	Effect Size
mean(THW)	Approaching a lead vehicle	5.7	0.000	5%	0.05
	Following a lead vehicle	8.6	0.000	8%	0.27
min(THW)	Approaching a lead vehicle	8.4	0.000	9%	0.24
	Following a lead vehicle	8.2	0.000	6%	0.14

Scenarios Involving Lane Changes

For scenarios involving lane changes, the interaction with other road users was analysed using the duration of these scenarios. Lane changes and thereby the potential interaction with other road users were longer. As already described within RQ-T11 (Section 3.2.5), these interactions were slightly longer. Again, the higher overall speed of other traffic is shown in the relative measures of objects during a cut-in, which were all higher in their mean respective minimum values.

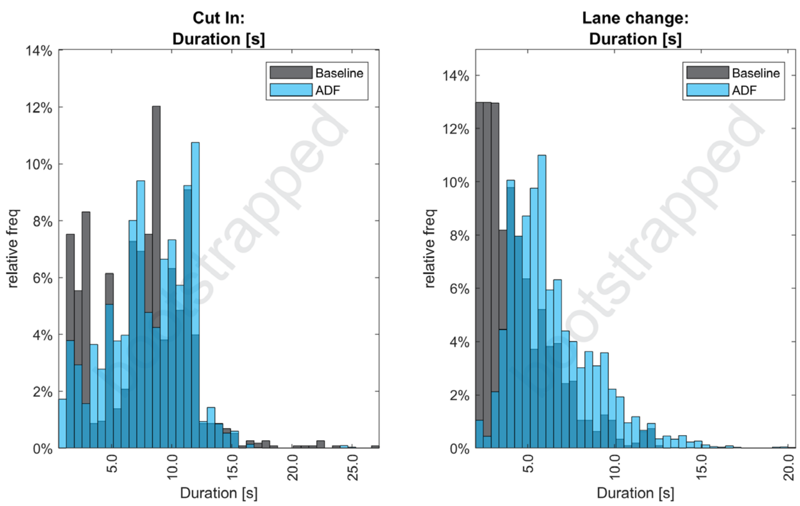


Figure 3.44: Duration of scenarios involving lane changes.

Interaction with other road users during cut-ins happened further away from the ego-vehicle with ADF active compared to baseline.

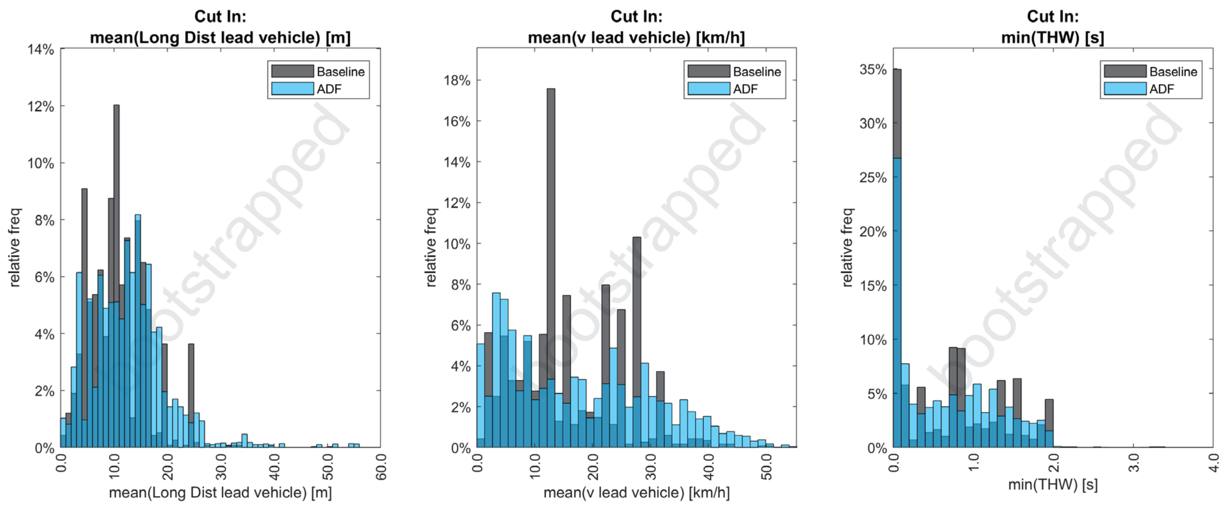


Figure 3.45: Indicators for the interaction in cut-in scenarios.

Table 3.27: Detailed results for the indicators for the interaction in scenarios involving lane changes.

Indicator	Scenario	Z	p	Change in %	Effect Size
mean(Long. distance lead vehicle)	Cut-in	7.9	0.000	22%	0.29
mean(speed lead vehicle)	Cut-in	5.4	0.000	11%	0.14
min(THW)	Cut-in	3.5	0.000	10%	0.08
Duration	Cut-in	5.0	0.000	7%	0.15
	Lane change	8.6	0.000	44%	0.80

Intersections

To analyse interactions at intersections, similar PIs as for driving outside of intersections were analysed. The mean longitudinal distance the ADF kept when travelling through intersections while following a lead vehicle (cf. Figure 3.46) was not different from baseline. For the crossing case, the ADF kept slightly lower distances, but for the turning case, the ADF kept larger distances. Both effects were, however, small.

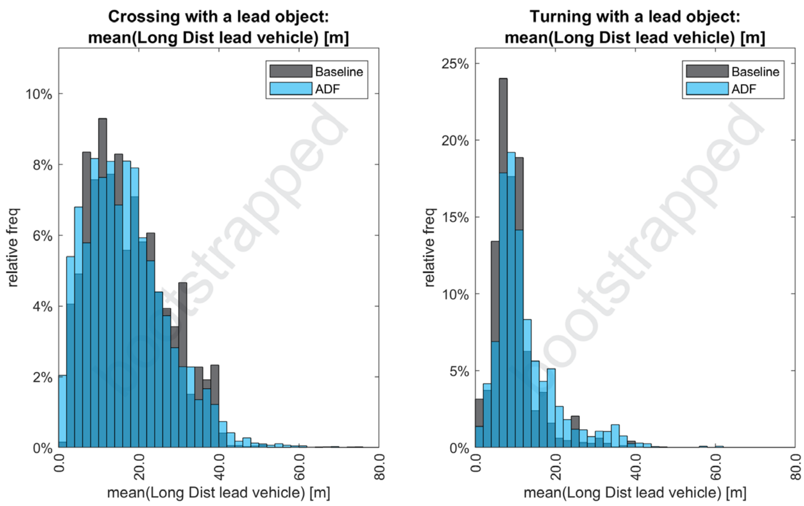


Figure 3.46: Minimum longitudinal distance to the lead vehicle in intersections.

Interactions with laterally moving objects happened before or after the ego-vehicle. Looking at the minimum distances (cf. Figure 3.47), no differences were directly apparent, and the effect was minor (cf. Table 3.28).

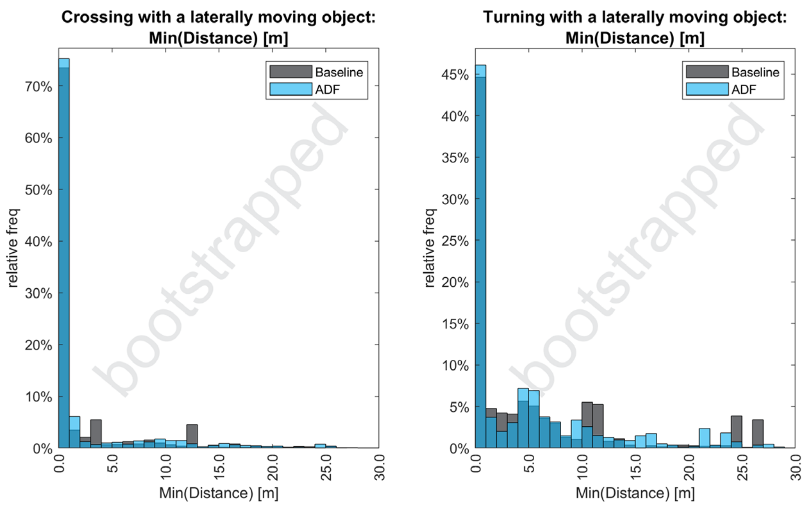


Figure 3.47: Minimum distance to the laterally moving object in intersections.

Table 3.28: Detailed results for the indicators for the interaction in intersections.

Indicator	Scenario	Z	p	Change in %	Effect Size
mean(Long Distance Lead Vehicle)	Crossing with lead object	-6.2	0.000	-4%	-0.07
	Turning with lead object	8.6	0.000	26%	0.34
min(Distance)	Crossing with laterally moving object	1.5	0.128	1%	0.01
	Turning with laterally moving object	0.8	0.452	2%	0.00

3.2.7 RQ-T15: How Does the ADF Influence the Behaviour of Subsequent Vehicles?

Analysing the behaviour of subsequent vehicles proved challenging, as no scenarios were defined which captured this interaction. In comparison to the motorway, an in-depth analysis was also not possible since no videos to the rear were available for all Pilot sites. Therefore, the results in this section are based on the rather noisy signals for the rear vehicles and were only analysed outside of intersections.

Lane-Bound Scenarios

For lane-bound scenarios, the THW to the rear vehicles was mostly smaller for ADF compared to baseline. This is in line with the subsequent analyses in previous RQs, which found that the ADF drove at lower speeds compared to baseline. Thus, assuming higher speed for the overall traffic, this was confirmed by the lower THW values. The only outlier in this consideration was the mean THW for the following scenario. One explanation for this could again be the higher travelled speeds in the baseline resulting in seamlessly flowing along with the traffic, also resulting in the small distances kept in queued driving often seen within urban environments.

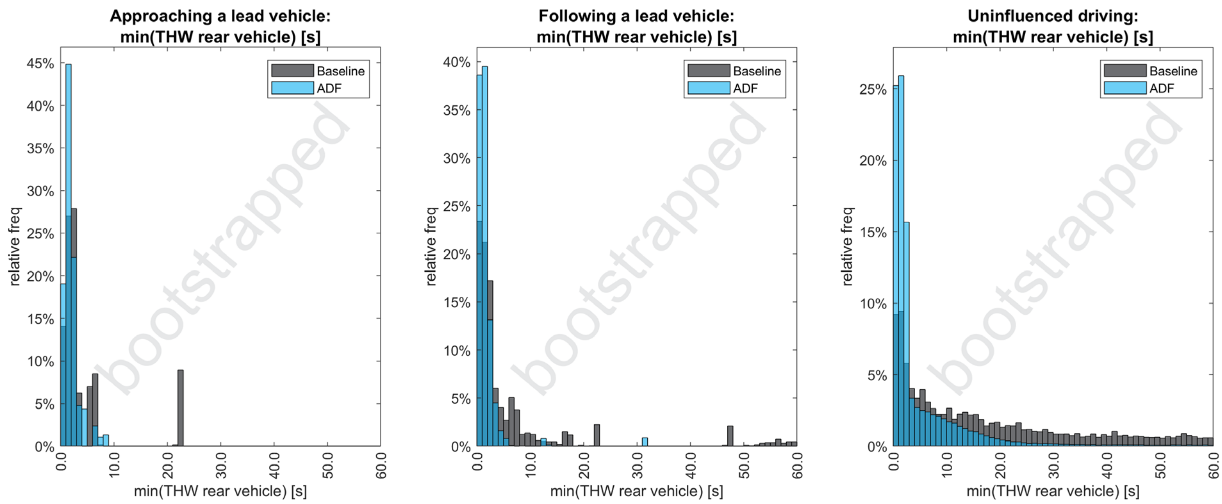


Figure 3.48: Minimum THW of rear vehicles in lane-bound scenarios.

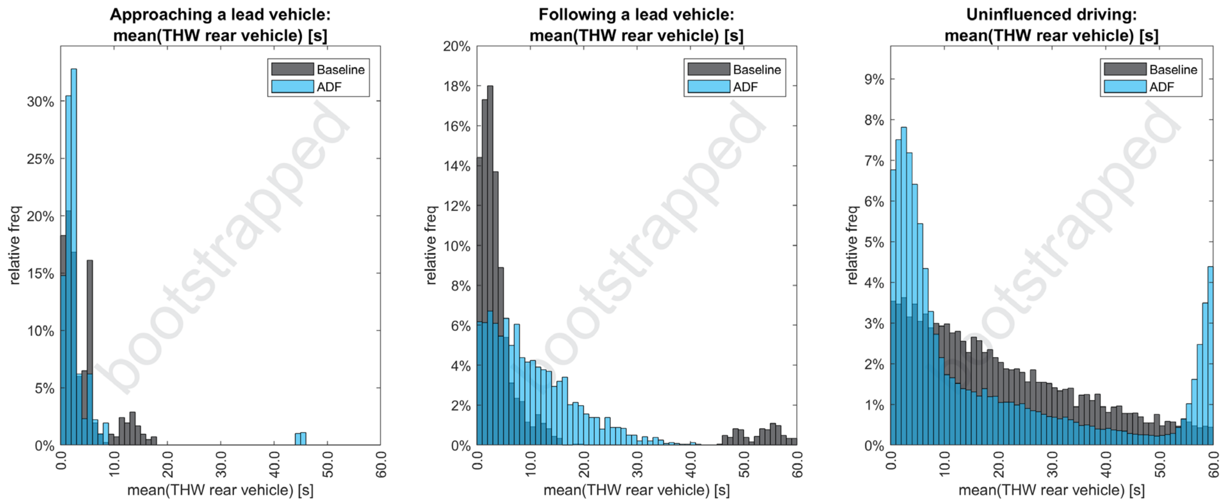


Figure 3.49: Mean THW of rear vehicles in lane-bound scenarios.

Table 3.29: Detailed results for the indicators for subsequent vehicles in lane-bound scenarios.

Indicator	Scenario	Z	p	Change in %	Effect Size
mean(THW rear veh.)	Approaching a lead vehicle	-2.3	0.022	-17%	-0.12
	Following a lead vehicle	5.5	0.000	34%	0.23
	Uninfluenced driving	-7.1	0.000	-9%	-0.10
min(ax rear veh.)	Approaching a lead vehicle	-8.6	0.000	-45%	-0.32
	Following a lead vehicle	-8.6	0.000	-45%	-0.22
	Uninfluenced driving	-8.6	0.000	-51%	-0.14
min(THW rear veh.)	Approaching a lead vehicle	-8.1	0.000	-53%	-0.73
	Following a lead vehicle	-8.6	0.000	-73%	-0.63
	Uninfluenced driving	-8.6	0.000	-69%	-1.20
sd(THW rear veh.)	Approaching a lead vehicle	-1.0	0.314	-6%	-0.02
	Following a lead vehicle	8.4	0.000	80%	0.19
	Uninfluenced driving	8.6	0.000	99%	0.26

Scenarios Involving Lane Changes

For scenarios involving lane changes, the results indicated good predictability of the behaviour of the ADF during lane changes, resulting in higher THWs of the rear vehicle. However, this also resulted in larger decelerations of the rear vehicle to make room for the ego-vehicle early on when ADF was active. That this effect was clearer for the active lane change compared to the cut-in, confirmed this assumption.

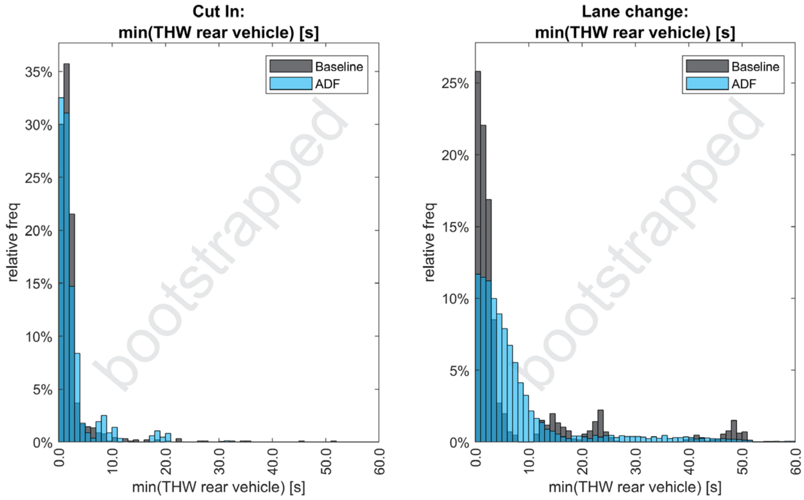


Figure 3.50: Minimum THW for rear vehicles in scenarios involving lane changes.

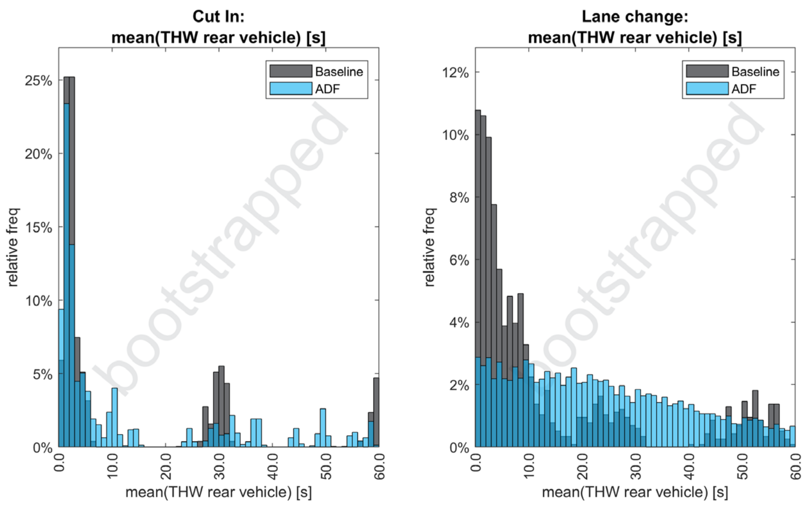


Figure 3.51: Mean THW for rear vehicles in scenarios involving lane changes.

Table 3.30: Detailed results for the indicators for subsequent vehicles in scenarios involving lane changes.

Indicator	Scenario	Z	p	Change in %	Effect Size
mean(THW rear veh.)	Cut-in	0.8	0.428	6%	0.07
	Lane change	8.4	0.000	71%	0.60
min(ax rear veh.)	Cut-in	-3.0	0.003	-14%	-0.10
	Lane change	-8.6	0.000	-68%	-0.28
min(THW rear veh.)	Cut-in	4.4	0.000	13%	0.08
	Lane change	2.9	0.004	15%	0.10
sd(THW rear veh.)	Cut-in	6.8	0.000	75%	0.22

Indicator	Scenario	Z	p	Change in %	Effect Size
	Lane change	8.5	0.000	110%	0.23

3.2.8 RQ-T16: How Does the ADF Influence the Behaviour of Preceding Vehicles?

As already concluded in previous RQs, for many scenarios one of the main influencing factors is the urban environment and not necessarily the difference between ADF and human driver. Therefore, the behaviour of the preceding vehicles was only analysed for lane-bound scenarios or only for the following and approaching scenarios. The assumption is confirmed that following and approaching scenarios happened at lower speeds when ADF was active compared to baseline driving. Table 3.31 shows that the changes were similar in size (as were the effects), so it was concluded that the ADF did not have an influence on the behaviour of the preceding vehicle.

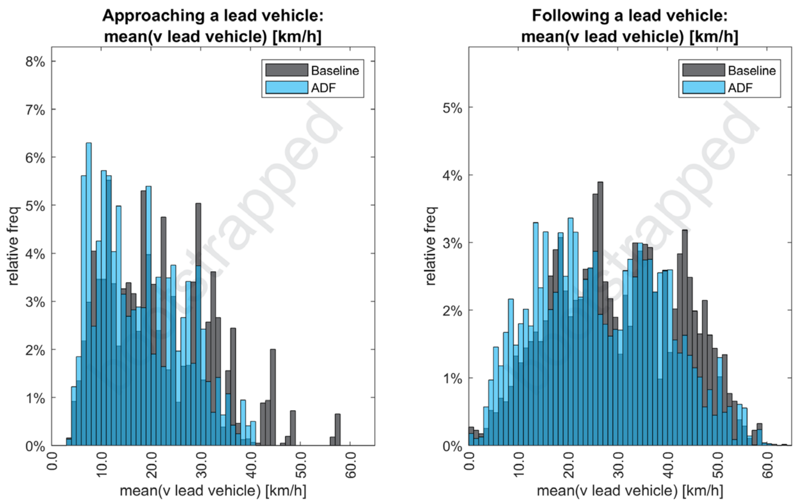


Figure 3.52: Mean driven speed of the preceding vehicle in the approaching and following scenario.

Table 3.31: Detailed results for the mean driven speed of the ego and preceding vehicle in the approaching and following scenario.

Indicator	Scenario	Z	p	Change of mean in %	Effect Size
mean(v)	Approaching a lead vehicle	-8.6	0.000	-12%	-0.4
	Following a lead vehicle	-8.6	0.000	-11%	-0.31
mean(v lead veh.)	Approaching a lead vehicle	-8.6	0.000	-16%	-0.36
	Following a lead vehicle	-8.6	0.000	-11%	-0.26

3.2.9 AIM Mobile Traffic Acquisition Results

This chapter presents and discusses the results of the analysis of ADF and baseline vehicular trajectory data collected at an urban roundabout. First, the results of the driver's own driving behaviour are presented, followed by the interaction behaviour. In Table 3.32 and Table 3.33 summarise the relevant aggregated findings on own driving and interaction behaviour.

Table 3.32: Aggregated findings on own driving behaviour.

RQ	PI	Baseline			ADF		
		entering	circling	exiting	entering	circling	exiting
6 (longitudinal acceleration)	max(a)	2.84	2.25	2.92	4.36	1.39	3.06
	min(a)	-0.49	-1.91	-1.49	-0.75	-1.89	-1.02
	iqr(a)	1.41	1.13	1.55	1.96	0.90	0.91
8 (longitudinal speed)	m(v)	3.47	5.54	6.43	4.25	4.41	5.86
	max(v)	5.36	7.24	8.73	6.28	5.90	8.92
	iqr(v)	1.62	1.89	1.32	1.30	1.29	1.58
13 (traffic flow / journey time)	JT	3.13	1.94	3.12	3.96	3.26	4.51

Table 3.33: Aggregated findings on interaction behaviour.

RQ	PI	car following			merging		cyclist crossing		pedestrian crossing	
		entering	circling	exiting	non-yielding	yielding	non-yielding	yielding	non-yielding	yielding
6 (longitudinal acceleration)	max(a)	1.78	2.96	2.99	2.03	2.28	3.52	3.27	3.62	3.11
	min(a)	-0.94	-2.70	-2.28	-1.12	-1.37	-2.85	-2.74	-2.29	-2.64
	iqr(a)	0.94	1.60	1.67	1.72	1.86	1.94	1.91	1.84	1.77
8 (longitudinal speed)	m(v)	5.84	6.11	7.35	5.29	4.24	7.33	5.04	7.58	4.76
	max(v)	7.59	8.59	10.08	7.56	7.21	10.47	8.81	10.75	8.50
	iqr(v)	1.55	1.73	2.07	2.18	3.00	2.40	3.26	2.68	3.37
12 (road user interaction)	PET or THW	3.26	3.35	2.63	3.58	2.85	2.61	2.58	3.56	2.91
	min(THW)	2.47	1.94	2.03	-	-	-	-	-	-
	min(TTC)	6.06	3.87	6.21	-	-	-	-	-	-
14 (number of incidents / near-crashes)	N/h	139	24	70	37	50	0.24	0.78	0.12	0.33
	N/h PET <1 s / min(TTC) <1 s	1.6	1.8	8.8	0.1	0.1	0.02	0.01	0	0

3.2.9.1 Own Driving Behaviour

In the roundabout, the ADF vehicle was validly detected 28 times by the AIM mobile units. In 27 cases the AV entered the roundabout in the north and exited in the east; in one case it returned to the north. The trajectories and the derived kinematic parameters of these 27 cases were used for further analysis.

To compare the driving behaviour of ADF with manually driven vehicles from the whole data, trajectories of manually driven vehicles as baseline trajectories that matched the conditions of the ADF vehicles were randomly selected. The matching conditions included (i) the same path, i.e., entering in the north and exiting in the east of the roundabout, and (ii) the same environmental and traffic conditions (trajectories were considered only 30 min before and after the presence of the ADF vehicle).

Eventually, 69 MV trajectories were selected for the baseline, which led to an approximate 1:2 relationship of ADF vs. baseline drives. Figure 3.53 shows the trajectories of both. Note the difficulty detecting the ADF vehicle in the upper left of the roundabout (left), which led to trajectory corruptions.

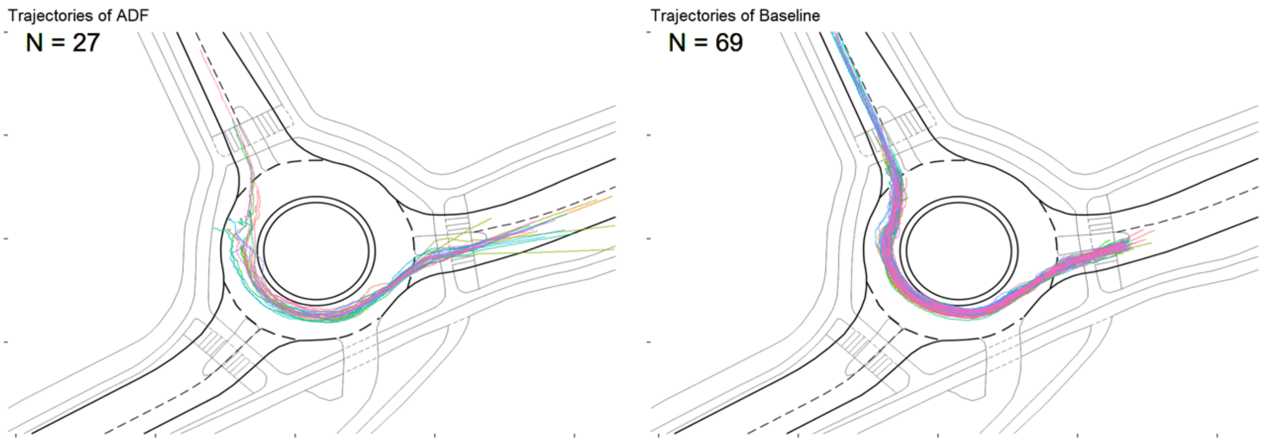


Figure 3.53: ADF trajectories (left) and baseline trajectories (right).

3.2.9.1.1 RQ-T6 / Longitudinal Acceleration

Due to the small sample size (N = 5) of ADF vehicles in the entering phase, the obtained results are highly uncertain. Regardless of the entering phase, the maximum acceleration, the absolute values of the minimum acceleration, and the interquartile ranges (IQR) of the longitudinal acceleration of the ADF vehicle are generally lower compared to baseline.

Considering the maximum acceleration (Figure 3.54, left), a marginal but statistically significant difference was found between ADF and baseline ($F = 6.1$; $df = 1$; $p = .01$; partial $eta^2 = .02$), which is beyond meaningful interpretation. Looking at the minimum acceleration (Figure 3.54, right), no significant differences between ADF and baseline were found, whereas a small but clear effect results from the roundabout phase ($F = 29.4$; $df = 1$; $p = .00$; partial $eta^2 = .10$). The IQR of the longitudinal acceleration was examined (not shown) as an indicator for the smoothness of the passage. With an average difference of 0.4 m/s² the ADF vehicle showed a significantly lower variance than the baseline ($F = 13.1$; $df = 1$; $p = .00$; partial eta² = .05). It is worth noting that in the entering phase–which shows the opposite trend–the subsample size for ADF is only six, since most trajectories were corrupted.

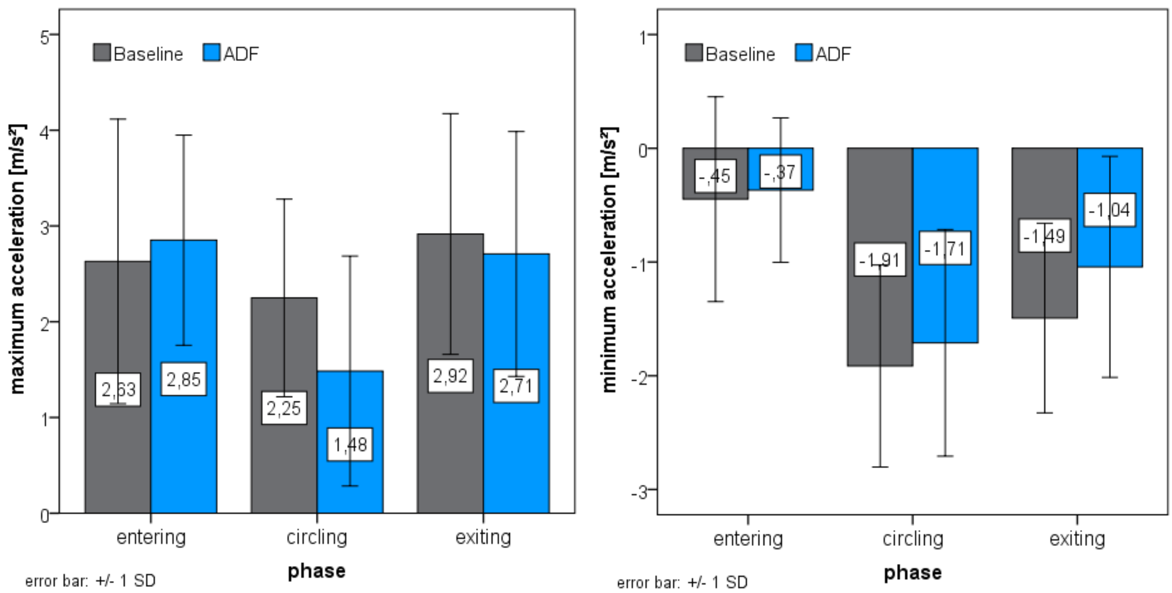


Figure 3.54: Mean and standard deviation of longitudinal acceleration of the baseline (grey) and ADF vehicle (blue): maximum (left) and minimum (right) values.

3.2.9.1.2 RQ-T7-1 / Manoeuvre Precision

To reduce the lateral position noise, the trajectories data of ADF and manually driven vehicles was filtered with an Unscented Kalman Filter (UKF). The method introduced in section 2.5.6.2 was applied to quantify manoeuvre precision at four virtual loops in the roundabout crossed by all ADF and manually driven vehicles, i.e., entering loop 27, circling loop 29 and two exiting loops 30 and 31.

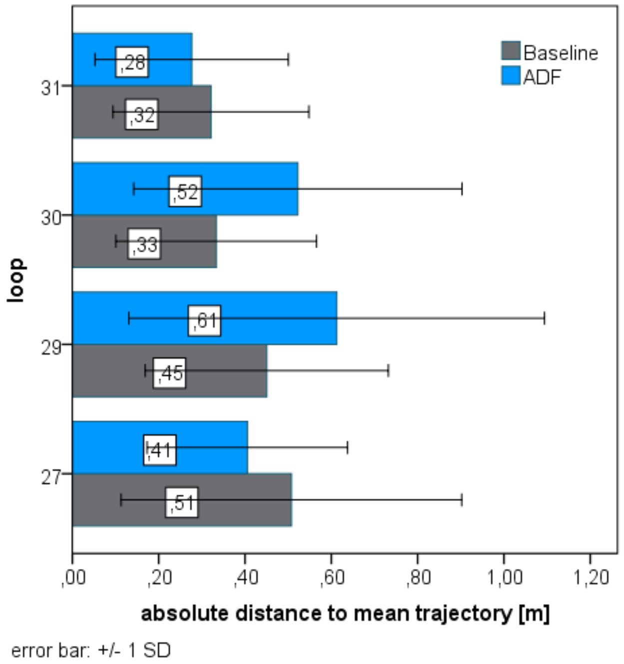


Figure 3.55: Distance to the mean of trajectories intersecting virtual loops 27, 29, 30, 31.

In Figure 3.55 no clear difference between ADF and baseline can be identified. However, using a univariate ANOVA the marginally lower manoeuvre precision of the ADF vehicle reaches statistical significance ($F = 4.8$; $df = 1$; $p = .03$). Considering the occasionally poor detection performance, the extremely small effect size (explained variance: partial $eta^2 = .01$), and the fact that the average difference between ADF and baseline is only 0.09 metres, this outcome should not be overrated.

3.2.9.1.3 RQ-T7-2 / Lane Keeping Performance

Due to the presence of five available ADF trajectories in the entering phase only, the analysis of lane keeping performance focused on the circling and exiting phases, while the latter was separated into exiting (1) and exiting (2). Further, two corrupted ADF trajectories occurred at the beginning of the circling phase, and were thus removed from statistical analysis.

All in all, the ADF vehicle shows a slightly lower lane keeping performance than the baseline, especially in the circling and first part of the exiting phase. In the second part of the exiting phase no significant difference between the baseline and the ADF vehicles occurred.

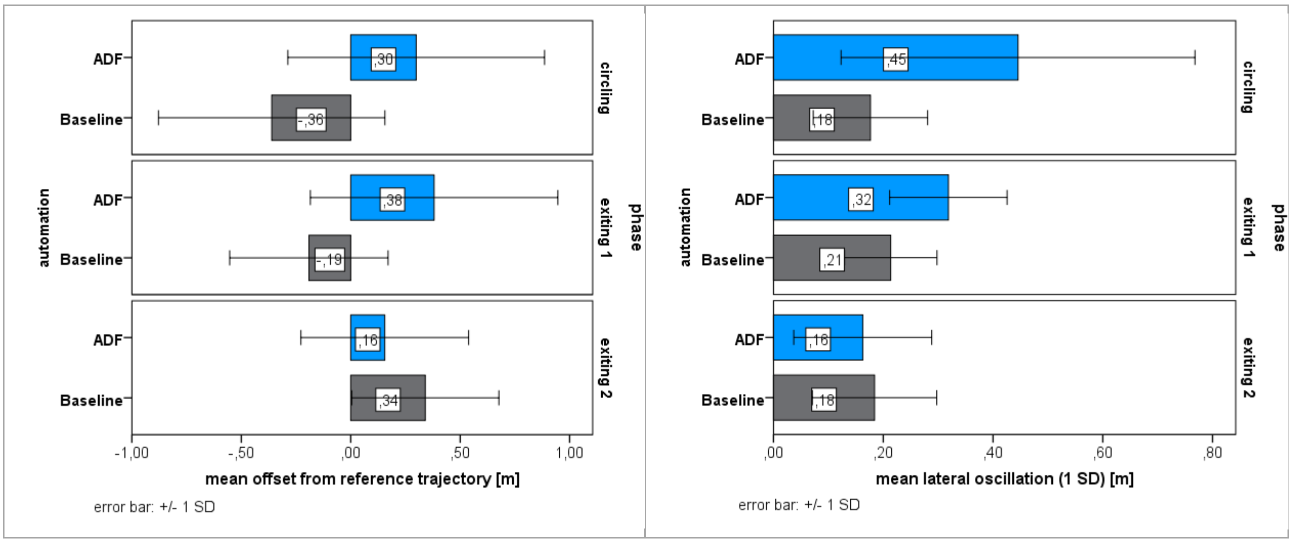


Figure 3.56: Lane keeping performance results of the inference statistical tests; mean offset from the reference trajectory (left), mean lateral oscillation (right).

Assessing the directional stability of each drive, the standard deviation was chosen as the indicator for lateral deviation from the mean (see section 2.5.6.2). A univariate ANOVA revealed that the moderate difference in lane keeping performance between the ADF and baseline vehicles occur in the circling and first exiting phase (Figure 3.56) reached high statistical significance (F = 33.6; df = 1; p = .00) while showing a small effect size (average difference = .12 m; partial eta² = .11). Regarding the offset from the reference trajectory, particularly in the circling phase, manual drivers tended to cut the roundabout (mean drawn to the left) and again in the opposite direction, before exiting on the right. Investigating this mean lateral deviation appeared to show, that manual driving behaviour differed significantly from that with ADF (F = 28.2; df = 1; p = .00; partial eta² = .09). However, the phase in itself had a larger effect on the offset than automation (F = 48.5; df = 1; p = .00; partial eta² = .15).

3.2.9.1.4 RQ-T8 / Driven Velocity

Due to having only five available ADF trajectories in the entering phase the obtained results are highly uncertain. Investigating the average speed (Figure 3.57, left), it is clearly seen that the means for the ADF vehicle are slightly lower in every phase, which leads to a significant result of a univariate ANOVA (F = 21.2; df = 1; p = .00). However, this effect is small (average difference between ADF and baseline vehicles is 0.2 m/s; partial eta² = .08) and the phase has a much larger effect on the average speed than automation (F = 261.2; df = 1; p = .00; partial eta² = .51). Regarding the maximum speed driven, three physically rather unlikely extreme values were removed before the analysis, yielding results similar to the average speeds (Figure 3.57, right): There is a significant but small (-0.8 m/s in average) effect of the ADF vehicle (F = 20.4; df = 1; p = .00; partial eta² = .07), but a much larger one for the phase (F = 421.7; df = 1; p = .00; partial eta² = .62). Finally, assessing the IQR of the speed as a measure of the stability of driving, the differences are small and no clear tendency can be seen (not shown). Consequently, no statistically significant results could be found.

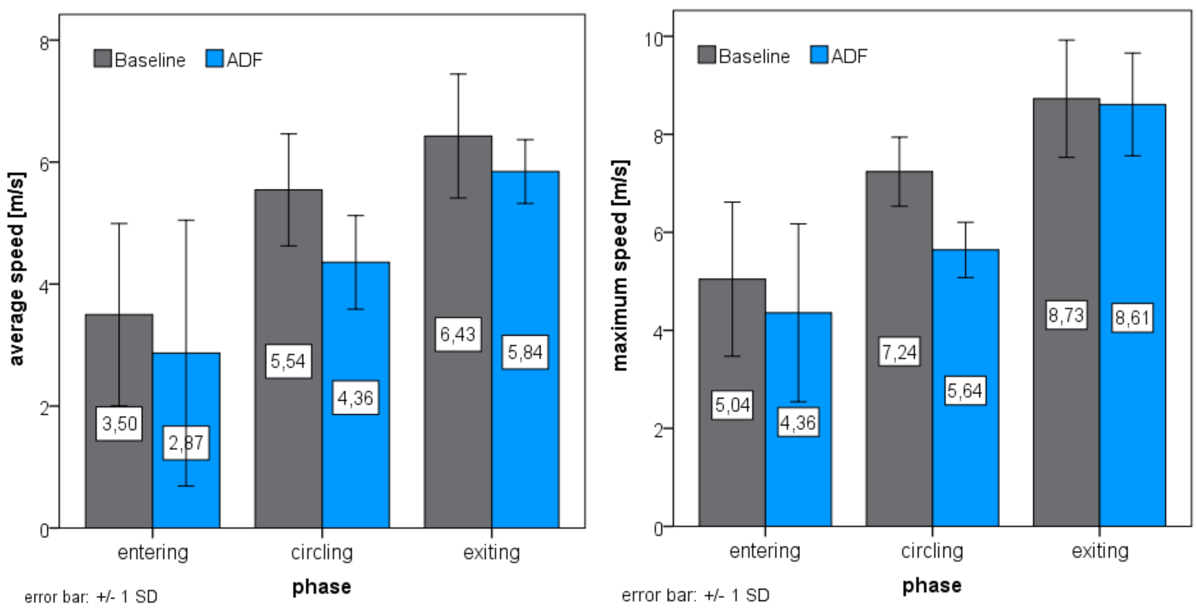


Figure 3.57: Inference statistical values of driven velocity of the baseline (grey) and ADF vehicle (blue): average (left) and maximum (right) values.

3.2.9.2 Interaction with Other Road Users

The 27 valid ADF trajectories going from the north to east exits of the roundabout were considered to measure the interactions with other road users such as VRU and other motorised vehicles. The relevant interaction scenarios considered are car-following, crossing or merging, and are introduced in the AIM methodology (Section 2.5.6). For the driving scenarios car-following, merging and VRU crossing, the relevant thresholds for THW and TTC (car-following) as well as PET (merging / crossing) were set to six seconds in any case. Situations above those thresholds were ignored. It appeared that only one crossing interaction between an ADF vehicle and a VRU occurred and altogether five relevant car-following and six merging situations, which make a sensible analysis and comparison with the baseline impossible; thus, the results are not shown. Instead, the focus was on analysing the normal driving and interaction behaviour of human road users to support maturing the ADF of automated vehicles. Concerning normal behaviour, altogether 80 859 car-following, 30 171 merging and 514 VRU crossing situations were identified. The results obtained regarding car following, VRU behaviour and number of incidents and near-crashes are presented in Annex 5.

3.2.9.2.1 RQ-T13 / Journey Times (JT)

The analysis of JT in the phases entering, circling and exiting required removing some outliers (two in the case of ADF, six in baseline). At an average difference of 1.2 seconds, a univariate ANOVA revealed that the ADF vehicle passed the roundabout significantly slower than the baseline vehicles ($F = 101.8$; $df = 1$; $p = .00$; partial $eta^2 = .30$). Over all three phases the ADF vehicle generates a loss time of approximately 3.5 seconds, which is more than 40% larger than the JT of the baseline vehicles.

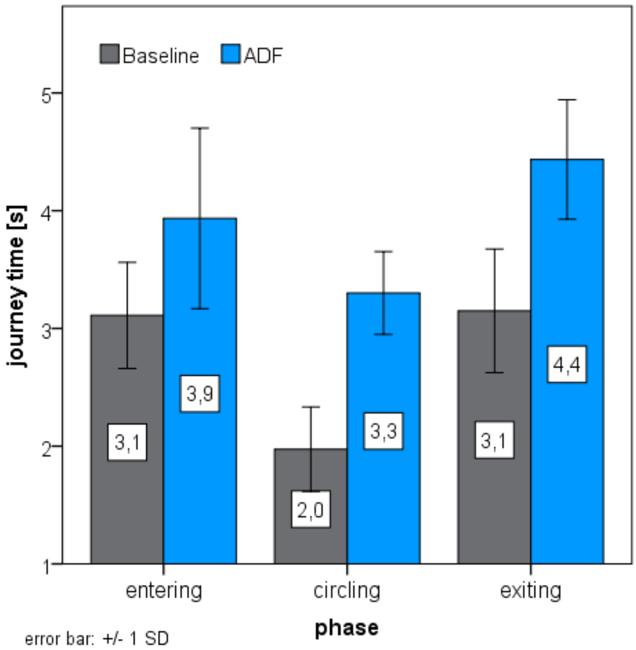


Figure 3.58: Journey time results of the inference statistical analysis.

3.2.9.3 Interpretation and Implications of the Results

Despite significant differences in acceleration between ADF and human drivers, this effect is small and no significant differences in braking occurred. These results imply that the ADF vehicle and human drivers behave similarly in the same situation. Considering that ADF trajectories show a lower variability than the baseline vehicles, it is expected that automated vehicles should be made to accommodate the passenger with smooth driving, whereas a human driver might like to enjoy the acceleration capabilities he or she paid for. Keeping in mind that just one vehicle with ADF and very specific function layout was compared against a set of different human drivers with different skills and different vehicle types, assessing several vehicles with different ADF setups may have yielded a wider variety of driving behaviour, too.

Regarding manoeuvre precision, a very small, but statistically significant effect was found. Considering that the average difference between ADF and baseline was just 0.09 metres – which is within the range of position uncertainty of approximately 0.25 to 0.50 metres – the performance of both can be considered almost identical.

Assessing the lane keeping performance, the ADF occurs to oscillate slightly more than human drivers do. Furthermore, it appeared that human drivers tend to cut the roundabout, while the ADF vehicle sticks closer to a normative trajectory. Again, ADFs are designed for safe, less risky, and comfortable driving and thus should show a less dynamic driving behaviour than human drivers.

Average and maximum speed of the ADF vehicle occur slightly but significantly less than for the human drivers. An explanation for this may be that the ADF has a rather large and stable safety margin, whereas human drivers might engage in more risky driving than the ADF. However, these effects are small–most probably coming from the physical limits that arise from the topology of a roundabout – and the phase has a much larger effect on speed than automation. In conclusion, ADF and human drivers behave rather similarly with regard to speed choice. As a consequence of the lower speeds, the ADF vehicle drives significantly slower through the roundabout than human drivers, showing a longer journey time.

The results obtained may indicate that the variability of different human drivers with their different driving skills using different vehicle types is greater than in case of the automated vehicles, which should show limited acceleration and deceleration performances, and also a bit of predictability of their driving behaviour.

3.2.10 Summary, Urban ADF

As is the case with the motorway ADF, interpretation of the results on system availability and stability needs to be done carefully for the urban ADF. It must be kept in mind that the tested ADFs are still on a prototype level. Furthermore, it is likely that the circumstances of data collection (e.g., prototype ADFs, safety drivers who might prevent sudden TORs through early intervention) impact the results as wells so that a direct conclusion on driving with a market-ready mature ADF is challenging. As an additional factor regarding the urban analysis, one should consider that the urban Pilot sites of this project cover only a small portion of the variety of urban settings and traffic. Just as for the motorway, certain scenarios will probably also not be recordable during the pilots, as safety drivers are always present to intervene. Also, these safety drivers always occupy the driver's seat within the urban Pilot sites, so that driver reactions to misbehaviour of automated driving cannot be thoroughly investigated. Additionally, this also holds true for TORs which cannot be analysed in the urban setting, as most Pilot sites cover the complete Pilot sites, i.e., from starting in the parking lot to ending in the parking lot with no planned TORs in traffic. Additionally, none occurred during the piloting or they were caught by the safety driver quite early.

For the measured impact of ADF on the driving behaviour, the results are not quite as clear as for the motorway. Some basic considerations are as follows:

Automated vehicles adhere to the speed limits, resulting in a mostly overall lower driven speed.

Results on vehicle dynamics are mixed and depend upon the scenario.

The overall conclusion is, that ADF behaviour is similar to that of human driving within the urban environment. Considering the overall reduced speed of the ADF, an increase in safety can be stated. Of course, this will also lead to slightly higher travel times, but the differences here are marginal, so that the overall effect should be minimal. Further, another big advantage of the ADF is increased attentiveness and the fact that the ADF is always alert and has fast reaction times. This can lead to an overall increase in safety by the introduction of ADFs in the urban environment. Since incidents were not within the scope of the analysis within L3Pilot, only assumptions can be made on this topic.

For lane-bound scenarios, a difference between influenced and uninfluenced driving can be observed. In influenced scenarios, the effects are rather small, which can be explained by the ADF having a better sensing of subtle changes in the dynamics of other vehicles and therefore an advantage when reacting to other vehicles. For uninfluenced driving, larger effects were observed. This holds true especially for the dynamics and driven speeds of vehicles.

For intersections however, for most RQs no overall effect can be stated. What can be said is that intersections and the handling of those scenarios are mainly influenced by the infrastructure and other traffic participants. Consequently, effects are often mixed depending on the PI and the scenario. For most intersection scenarios it can be said, that travelling through intersections takes longer when using the ADF.

An additional point which influences the results for the urban ADF is that the data was bootstrapped (cf. Section 2.5.1). The bootstrapping method also added some noise to the data (to prevent identifiability). In principle, this makes it more challenging to find significant effects. It may also exaggerate the extreme values (minimum or maximum) visible in the histograms. However, the amount of noise added was so small that these potential drawbacks cannot have influenced our conclusions. A detailed discussion of the bootstrapping process is given in Annex 4.

The following pages show two summaries of all the results for the urban ADF, one of reported effect sizes and one of changes in relation to baseline values. PIs for which no significant effect could be found are set to zero in both summaries. Negative values show a decrease of the PI (e.g., lower speed) and positive values an increase (e.g., larger distances). With the colour coding it becomes easily visible whether the direction and size of effects are similar across scenarios or vary between them.

Table 3.34: Overview of changes to indicators for urban RQs.

Research question	PI	Approaching lead vehicle	Crossing with laterally moving object	Crossing with lead object	Crossing without conflict	Cut in	Following a lead vehicle	Lane change	Turning with laterally moving object	Turning with lead object	Turning without conflict	Uninfluenced driving
RQ-T6	min(ax)	0%	0%	0%	4%		19%		0%	-26%	39%	43%
	max(ax)	-13%	0%	19%	3%		-19%		-12%	-13%	2%	-34%
	mean(ax)	-11%	9%	38%	9%		0%		-13%	-16%	28%	467%
	sd(ax)	0%	0%	20%	1%		-5%		-16%	-8%	-17%	-23%
	max(abs(ay))	0%	-11%	14%	13%		-5%		-14%	-16%	-29%	-25%
	mean(ay)	0%	-87%	43%	144%		31%		1699%	208%	110%	295%
	sd(ay)	4%	0%	15%	12%		4%		-13%	-5%	0%	-20%
RQ-T7	sd(Pos in Lane)	41%					-20%					18%
	mean(Pos in Lane)	-18%					-20%					-2%
RQ-T8	mean(v)	-12%	11%	-6%	-15%	13%	-11%	-8%	-3%	-10%	-16%	-22%
	max(v)	-10%	9%	-2%	-13%	13%	-12%	-4%	-5%	-9%	-12%	-20%
	sd(v)	11%	9%	53%	34%	22%	-18%	41%	-11%	0%	10%	-21%
RQ-T11	Frequency	-22%	123%	-4%		38%	95%	-26%	-3%	6%		24%
	Duration	0%	-15%	18%	12%	7%	-37%	44%	12%	19%	26%	-44%
RQ-T12	mean(THW)	5%					8%					
	min(THW)	9%				10%	6%					
	Duration					7%		44%				

Table 3.35: Overview of effects for indicators used in urban RQs.

Research question	PI	Approaching lead vehicle	Crossing with laterally moving object	Crossing with lead object	Crossing without conflict	Cut in	Following a lead vehicle	Lane change	Turning with laterally moving object	Turning with lead object	Turning without conflict	Uninfluenced driving
RQ-T6	min(ax)	0	0	0	0.01		0.14		0	-0.09	0.24	0.53
	max(ax)	-0.1	0	0.21	0.04		-0.28		-0.39	-0.25	0.04	-0.65
	mean(ax)	-0.08	0.05	0.16	0.05		0		-0.21	-0.14	0.17	0.23
	sd(ax)	0	0	0.22	0.02		-0.06		-0.58	-0.18	-0.53	-0.4
	max(abs(ay))	0	-0.21	0.19	0.2		-0.09		-0.56	-0.34	-0.93	-0.5
	mean(ay)	0	-0.45	0.09	0.26		0.11		0.28	0.32	0.89	0.68
	sd(ay)	0.06	0	0.18	0.17		0.06		-0.45	-0.09	0	-0.4
RQ-T7	sd(Pos in Lane)	0.09					-0.05					0.06
	mean(Pos in Lane)	-0.2					-0.23					-0.02
RQ-T8	mean(v)	-0.4	0.26	-0.16	-0.56	0.29	-0.31	-0.26	-0.1	-0.21	-0.63	-0.82
	max(v)	-0.38	0.29	-0.07	-0.58	0.35	-0.36	-0.12	-0.17	-0.2	-0.56	-0.91
	sd(v)	0.16	0.12	0.37	0.21	0.26	-0.22	0.4	-0.2	0	0.13	-0.24
RQ-T11	Frequency	-0.2	0.43	-0.13		-0.45	0.37	-0.43	-0.04	0.05		0.59
	Duration	0	-0.36	0.25	0.21	0.15	-0.57	0.8	0.38	0.39	0.48	-0.6
RQ-T12	mean(THW)	0.05					0.27					
	min(THW)	0.25				0.08	0.14					
	Duration					0.15		0.8				

3.3 Parking

For parking, the results are based on effect sizes and percental changes of PI attributes calculated for the five individual studies. Therefore, all graphs show the single data points, the average across all studies and the range between studies. The tables give the number of studies providing results for one specific PI (N studies), the proportion of those studies with significant differences between ADF and baseline and the proportions of studies reporting a significant increase or decrease of the PI.

3.3.1 RQ-T6: What Is the Impact of ADF on Vehicle Dynamics in Defined Driving Situations?

For this RQ, PIs reflecting lateral and longitudinal acceleration of the vehicle are analysed. As can be seen in Figure 3.59, there is a significant reduction of lateral dynamics (measured via maximum absolute lateral acceleration, max(abs(ay))) and standard deviation of lateral acceleration (sd(ay))) for all studies. For longitudinal accelerations, results on the directions of effects are mixed. For instance, two studies report a significant decrease of maximum deceleration (min(ax)), one reports no change and another reports a significant increase. In a similar way, also results on maximum longitudinal acceleration (max(ax)) are mixed. Four out of five studies report a significant reduction of the variation of longitudinal acceleration (sd(ax)).

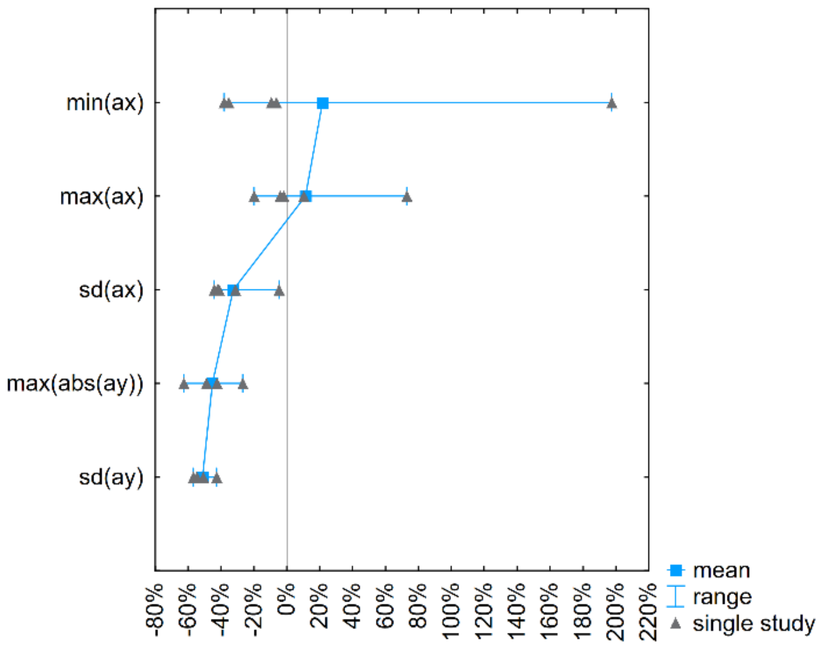


Figure 3.59: Results for indicators assessing vehicle dynamics.

Table 3.36: Summary of results on indicators assessing vehicle dynamics while parking.

Performance indicator	N studies	% studies with significant result	% studies with significant enhancement	% studies with significant reduction
min(ax)	5	100%	20%	80%
max(ax)	5	40%	20%	20%
sd(ax)	5	80%	0%	80%
max(abs(ay))	4	100%	0%	100%
sd(ay)	4	100%	0%	100%

3.3.2 RQ-T7: What Is the Impact of ADF on the Accuracy of Driving?

To analyse the accuracy of parking, the duration of parking manoeuvres (m(duration) and the number of stops (N(stops)) within one manoeuvre are analysed. All studies report that parking needs significantly more time with ADF and that the single manoeuvres include more stops than in manual parking. The duration of manoeuvers increased between 50% and 200%, the number of stops between 35% and 500%.

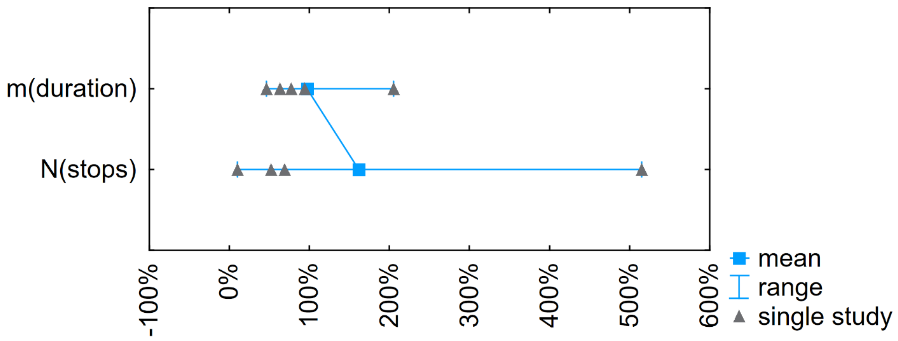


Figure 3.60: Results of indicators assessing the accuracy of parking.

Table 3.37: Summary of results on indicators assessing accuracy of driving while parking.

Performance indicator	N studies	% studies with significant increase	% studies with significant reduction
m(Duration)	5	100%	0%
n(Stops)	4	100%	0%

3.3.3 RQ-T8: What Is the Impact of ADF on the Driven Speed?

All studies report that speed while parking with the ADF decreases significantly. The reduction of mean speed (m(v)) varies between 42% and 66% and for maximum speed (max(v)) between 25% and 50%. Furthermore, the variation of speed (sd(v)) is significantly smaller.

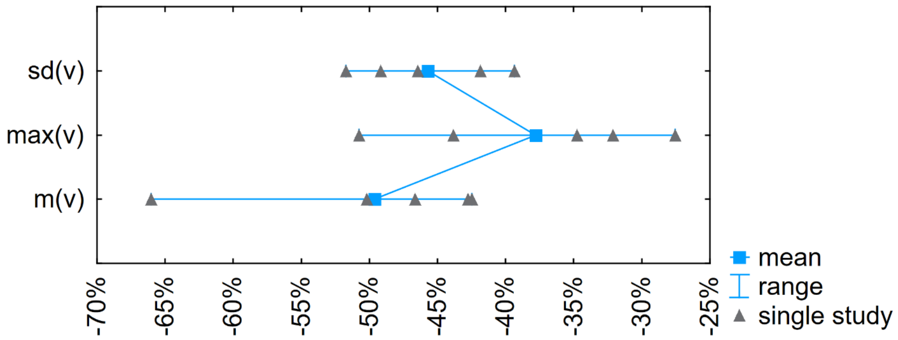


Figure 3.61: Results of indicators assessing the speed during parking.

Table 3.38: Summary of results on indicators assessing speed while parking.

Performance Indicator	N studies	% studies with significant result	% studies with significant enhancement	% studies with significant reduction
m(v)	5	100%	0%	100%
max(v)	5	100%	0%	100%
sd(v)	5	100%	0%	100%

3.3.4 Summary Parking ADF

Over all studies, parking with a parking chauffeur takes significantly longer than manual parking. This is due to lower speed and more stops within a parking manoeuvre. Furthermore, lateral dynamics are reduced while parking with the ADF. It seems that this is also a direct consequence of the reduced speed while parking with the ADF.

4 User and Acceptance Evaluation

Following the results of the Technical and Traffic Evaluation, this chapter presents the results of the User and Acceptance evaluations performed within L3Pilot. It discusses the results for the motorway, urban and parking ADF. As described in the previous sections, a series of questions were administered to investigate the users' evaluation and their acceptance of the system which they had experienced in the Pilot site testing.

In chapter 2, we discussed each ADF in detail. We tested two motorway ADFs, the Traffic Jam Motorway and the Motorway ADF. The Traffic Jam Motorway ADF could be activated on motorways up to 60km/h, whereas Motorway ADF could be activated on motorways up to 130km/h. Traffic Jam Motorway ADF also requires the presence of a lead vehicle, but the Motorway ADF may either follow a lead vehicle or keep to below the speed limit. Both motorway ADF can execute a lane change. As mentioned previously, the findings on motorway ADF reported in this section were not separated into Traffic Jam Motorway ADF or Motorway ADF but are an amalgam of the findings. When Urban ADF is activated the vehicle follows the lane, executes left and right turns, decelerates, accelerates, and can execute lane changes within cities. The vehicle recognises on-coming traffic and VRUs. Finally, when Parking ADF is activated, the vehicle is capable of completing manoeuvring into and out of garages and driveways.

We investigated different research questions by using the questionnaire, e.g.: Are drivers willing to use an ADF? What is the user acceptance of the ADF? What is the impact of ADF on driver state and on driver awareness? What are drivers' expectations regarding system features? What is drivers' secondary task engagement during ADF use? How do drivers respond when they are required to retake control? And what is the impact of ADF use on motion sickness?

4.1 Motorway

4.1.1 RQ-U1: Are Drivers Willing to Use an ADF?

Table 4.1: Question administered to investigate drivers' willingness to use an ADF.

Questions Administered
• I would use this system if it was in my car

As shown in Figure 4.1, drivers were generally very positive and gave high ratings for their willingness to use the motorway ADF if it was in their car. Approximately 83% of professional drivers from the Pilot site agreed or strongly agreed with the statement, while 95% of ordinary drivers from the Pilot site, and 93% of ordinary drivers from the simulator studies, agreed or strongly agreed. No driver stated that they did not know the answer to this question.

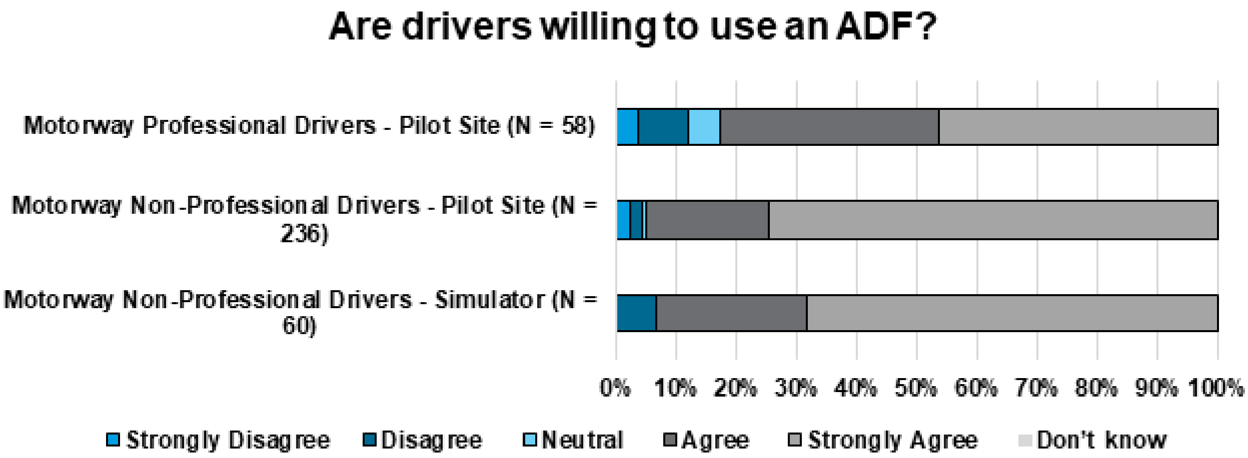


Figure 4.1: Ratings of drivers' willingness to use an ADF for Motorway Professional Drivers, Motorway Ordinary Drivers – Pilot Sites and Motorway Ordinary Drivers – Simulator.

4.1.2 RQ-U3: What Is the User Acceptance of the ADF?

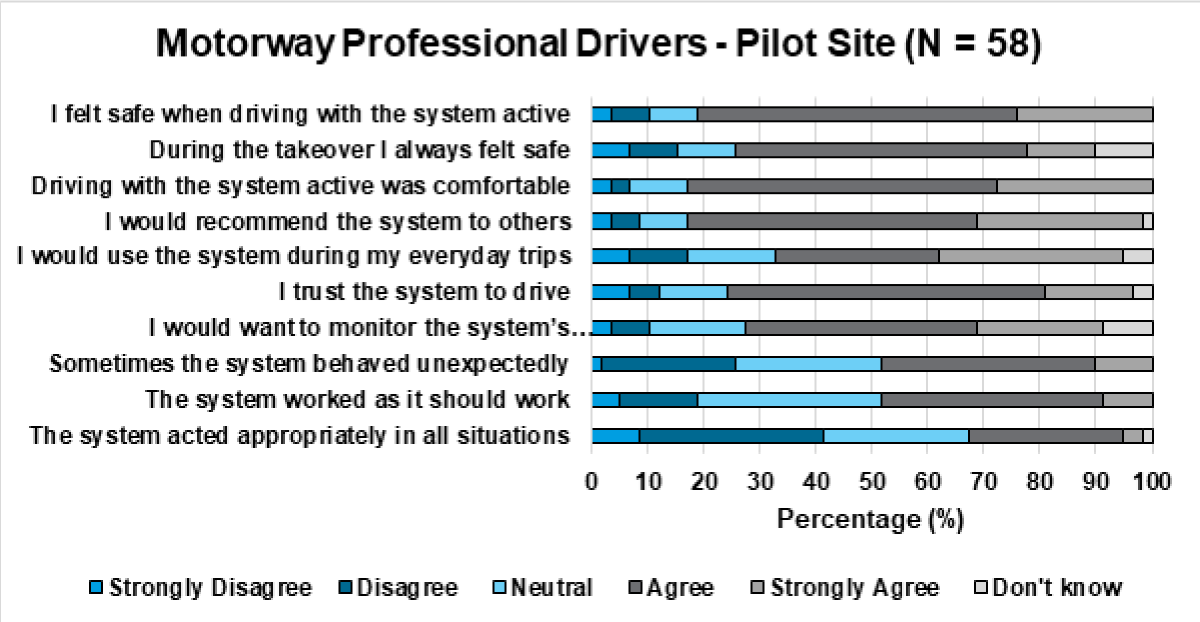
Participants' responses to twelve questions were evaluated to understand their acceptance of the motorway ADF (see Table 4.2).

Table 4.2: Sub-research questions and questions administered to understand user acceptance of the ADF.

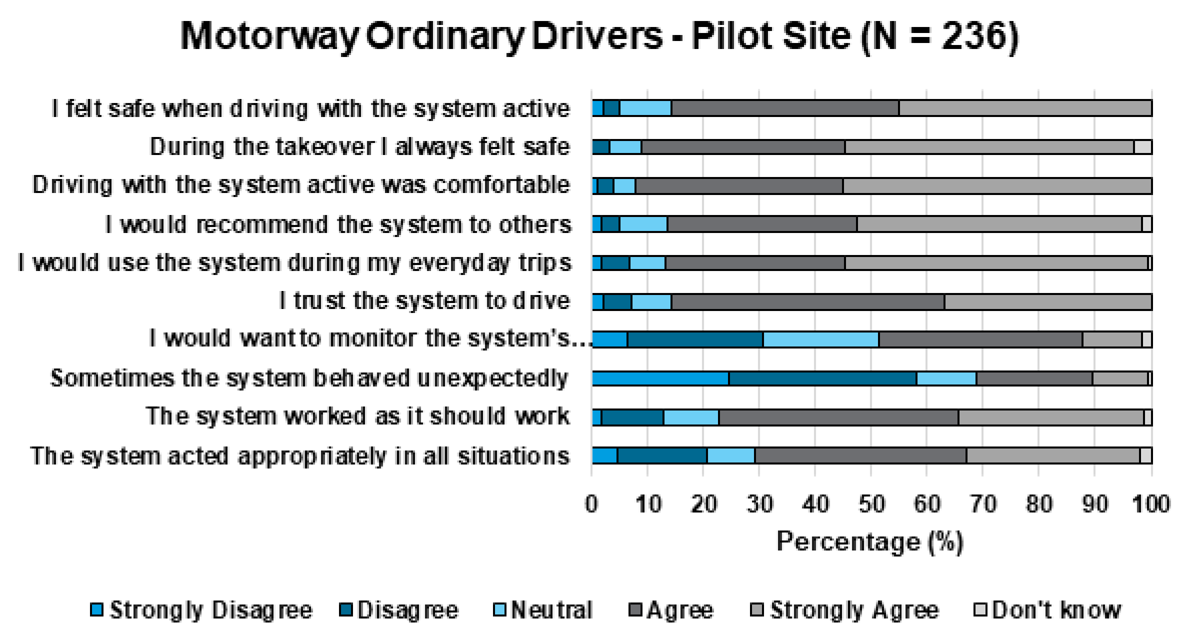
Sub-Research Questions	Questions Administered
• What is the **perceived safety** of the ADF?	• I felt safe when driving with the system active • During the take-over I always felt safe
• What is the **perceived comfort** of the ADF?	• Driving with the system active was comfortable • Rating of each vehicle behaviour
• What is the **perceived usefulness** of the ADF?	• I think the tested system was useful/useless • I would recommend the system to others. • I would use the system during my everyday trips.
• What is the **perceived trust** of the ADF?	• I trust the system to drive • I would want to monitor the system's performance
• How does user **acceptance** differ between ADF types? **(System's Performance)**	• Sometimes the system behaved unexpectedly • The system worked as it should work • The system acted appropriately in all situations

In terms of the **perceived safety** of the ADF (see Figure 4.2), most of the drivers agreed or strongly agreed with the statement 'I felt safe when driving with the system active.' 81% of professional drivers, 86% of ordinary drivers in the Pilot site, and 88% of ordinary drivers in simulator studies were in agreement. Drivers were also asked about their perceived safety during take-overs. The results revealed that 64% of professional drivers from the Pilot sites agreed or strongly agreed that they felt safe during the take-over, whereas 88% of ordinary drivers from the Pilot sites agreed with this statement. This seems to show a trend whereby the ordinary drivers were more positive about their experience. No data was collected for this question from the simulator studies.

(a)



(b)



(c)

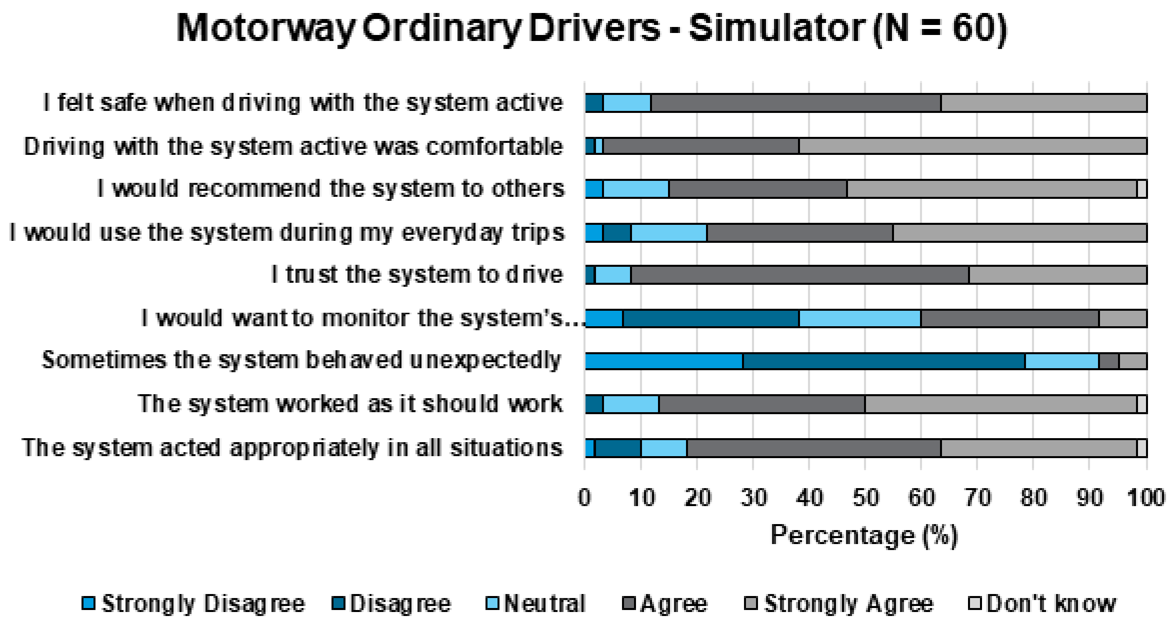
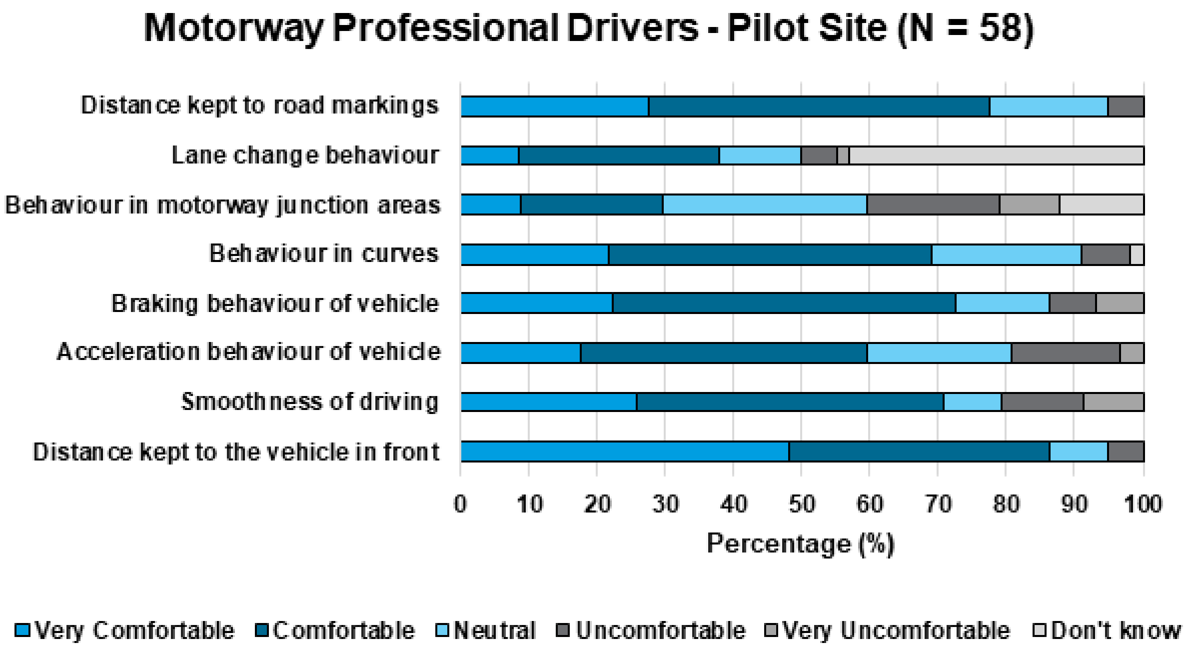


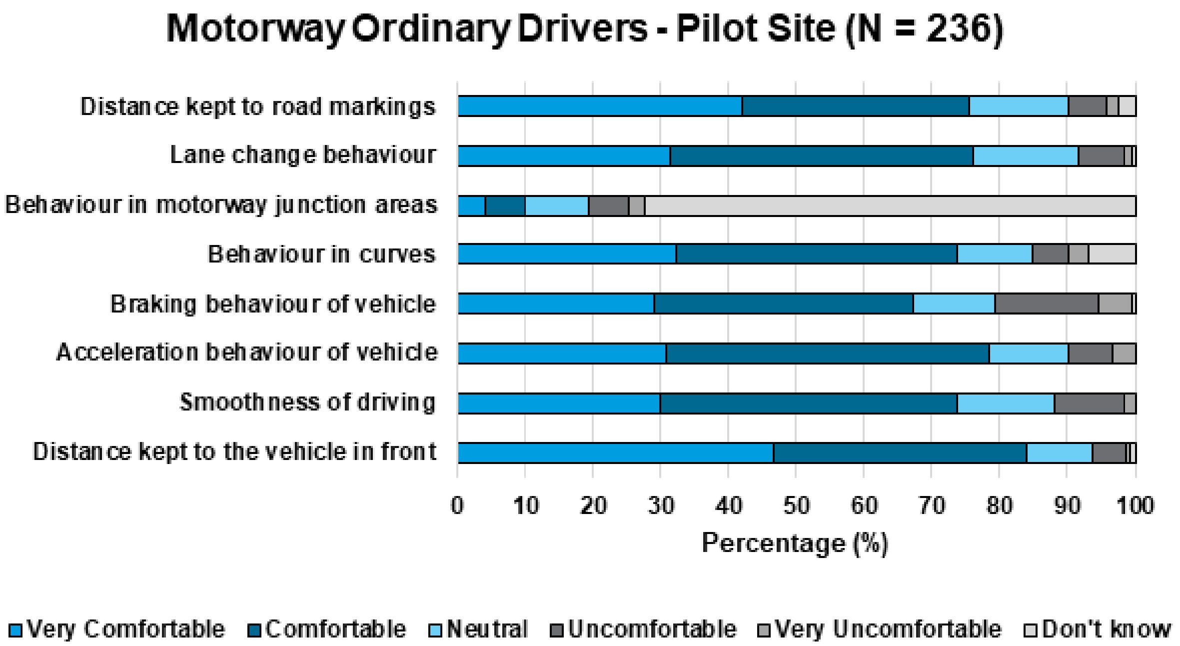
Figure 4.2: Ratings of user acceptance of the ADF for (a) Motorway Professional Drivers (b) Motorway Ordinary Drivers – Pilot Sites and (c) Motorway Ordinary Drivers – Simulator.

For **perceived comfort** of the ADF, most of the drivers agreed that ‘driving with the system active was comfortable’ (see Figure 4.3), whereby 83%, 92%, 97% were in agreement for the respective groups (professional driver in Pilot site studies, ordinary driver in Pilot site studies, and ordinary driver in simulator studies). To provide a deeper understanding of which aspects of the vehicle’s behaviour were more comfortable, drivers were asked to rate eight vehicle behaviours on a six-point scale. For the professional drivers (see Figure 4.3 (a)), the majority (range 60% to 78%) rated most of the behaviours as comfortable or very comfortable. However, behaviour in motorway junction areas (with 30% feeling comfortable and very comfortable, 12% don’t know), and lane change behaviour (with 38% feeling comfortable and very comfortable, 43% don’t know) were not rated as highly in terms of comfort. For the ordinary drivers from the Pilot sites (see Figure 4.3 (b)), the majority (range 67% to 78%) rated most of the vehicle behaviours as comfortable or very comfortable. However, a much smaller number of participants felt comfortable with the behaviour in motorway junction areas (10% feeling comfortable or very comfortable), but this could be because 72% said they didn’t know. Finally, for the ordinary drivers from simulators (see Figure 4.3), the majority (range 50% to 92%) rated vehicle behaviours as comfortable or very comfortable, but again not for the behaviour in motorway junction areas (with 23% feeling comfortable and very comfortable), where the majority (53%) said they didn’t know. Note that the large proportion of drivers answering ‘I don’t know’ was potentially due to them not experiencing the system at that particular area/manoeuvre during the drive. Overall, some of the vehicle behaviours were comfortable, with some having room for improvement.

(a)


Motorway Professional Drivers - Pilot Site (N = 58)
Distance kept to road markings
Lane change behaviour
Behaviour in motorway junction areas
Behaviour in curves
Braking behaviour of vehicle
Acceleration behaviour of vehicle
Smoothness of driving
Distance kept to the vehicle in front
0 10 20 30 40 50 60 70 80 90 100
Percentage (%)
Very Comfortable
Comfortable
Neutral
Uncomfortable
Very Uncomfortable
Don't know

(b)


Motorway Ordinary Drivers - Pilot Site (N = 236)
Distance kept to road markings
Lane change behaviour
Behaviour in motorway junction areas
Behaviour in curves
Braking behaviour of vehicle
Acceleration behaviour of vehicle
Smoothness of driving
Distance kept to the vehicle in front
0 10 20 30 40 50 60 70 80 90 100
Percentage (%)
Very Comfortable
Comfortable
Neutral
Uncomfortable
Very Uncomfortable
Don't know

(c)

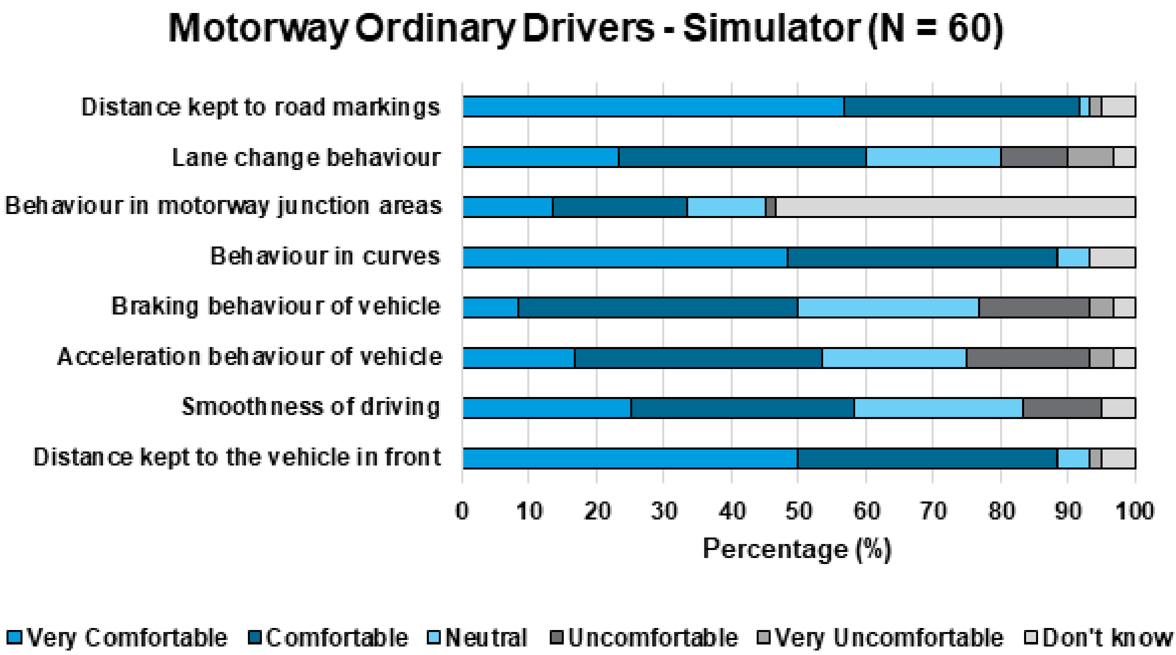


Figure 4.3: Ratings of perceived comfort for each behaviour of the ADF for (a) Motorway Professional Drivers (b) Motorway Ordinary Drivers - Pilot Sites and (c) Motorway Ordinary Drivers – Simulator.

To investigate drivers' **perceived usefulness** of the system (see Figure 4.2), three questions were asked. We asked drivers to rate whether they agreed or not with the statement 'I would recommend the system to others.' Most of drivers agreed with this statement, with 81% of professional drivers in the Pilot sites, 85% ordinary drivers in the Pilot sites, and 83% ordinary drivers in simulator studies agreeing or strongly agreeing. Most of the drivers also agreed with the statement, 'I would use the system during my everyday trips', with 62% professional drivers in Pilot sites, 86% ordinary drivers in Pilot sites, and 78% ordinary drivers in simulator studies agreeing. Thus, a smaller proportion of professional drivers agreed with this statement compared to the other two groups, possibly because professional drivers are more familiar with the system's limitations and the restrictions of a specific prototype. In addition, we also asked drivers to rate 'I think the tested system was useful/useless.' Findings showed that 84%, 94%, and 96% of drivers rated the system as useful, respectively, for each group.

In terms of **perceived trust** in the motorway ADF, and as shown in Figure 4.2, most of the drivers agreed or strongly agreed that they trust the system to drive (72%, 86%, 92% respectively for each group). Interestingly, most of the professional drivers agreed or strongly agreed with the statement 'I would want to monitor the system's performance,' but only 47% (Pilot site) and 40% (simulator) of ordinary drivers agreed with this statement. This, once again, shows the trend that professional drivers may be more cautious with the system.

Drivers were also asked to evaluate the **system's performance**. The results revealed that 48% (professional drivers, Pilot sites), 31% (ordinary drivers, Pilot sites), and 8% (ordinary drivers, simulator) agreed or strongly agreed that sometimes the system behaved unexpectedly, with data collected from the simulators rated more positively. 48%, 76%, and 85% of drivers from each group respectively agreed or strongly agreed that the system worked as it should. Again, the findings revealed that professional drivers who have had extensive experience with the system were less positive than ordinary drivers. Finally, we asked drivers to rate whether 'the system acted appropriately in all situations.' 31%, 68%, and 80% of drivers from each group agreed or strongly agreed with the statement, with professional drivers giving less positive ratings for the system.

4.1.3 RQ-U5: What Is the Impact of ADF on Driver State?

Four questions tapped into understanding drivers' workload or state while interacting with the motorway ADF (see Table 4.3).

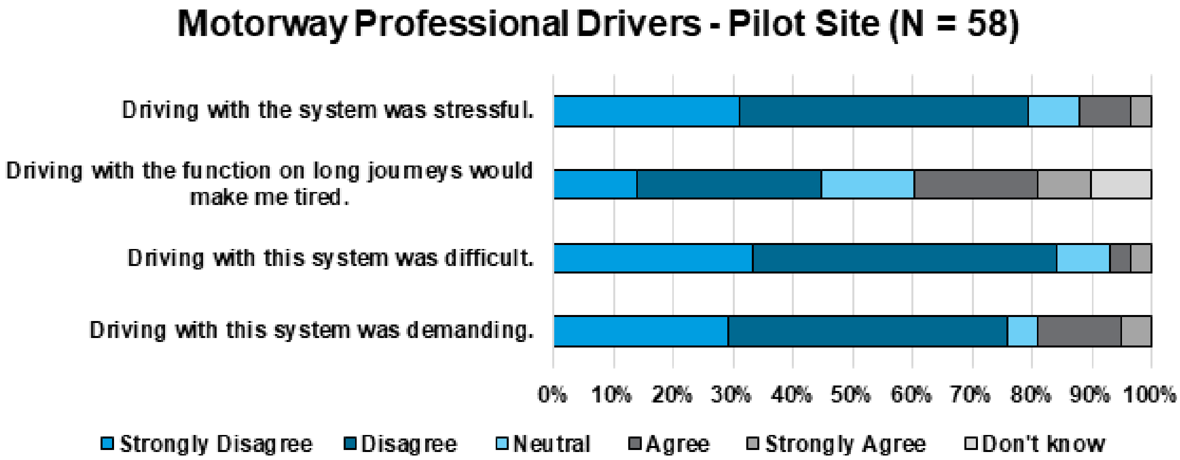
Table 4.3: Sub-research questions and questions administered to understand driver states while using the ADF.

Sub-Research Questions	Questions Administered
• What is the effect of ADF use on drivers' level of stress?	• Driving with the system was stressful
• What is drivers' level of fatigue while using the ADF?	• Driving with the function on long journeys would make me tired
• What is drivers' workload while using the ADF?	• Driving with this system was difficult • Driving with this system was demanding

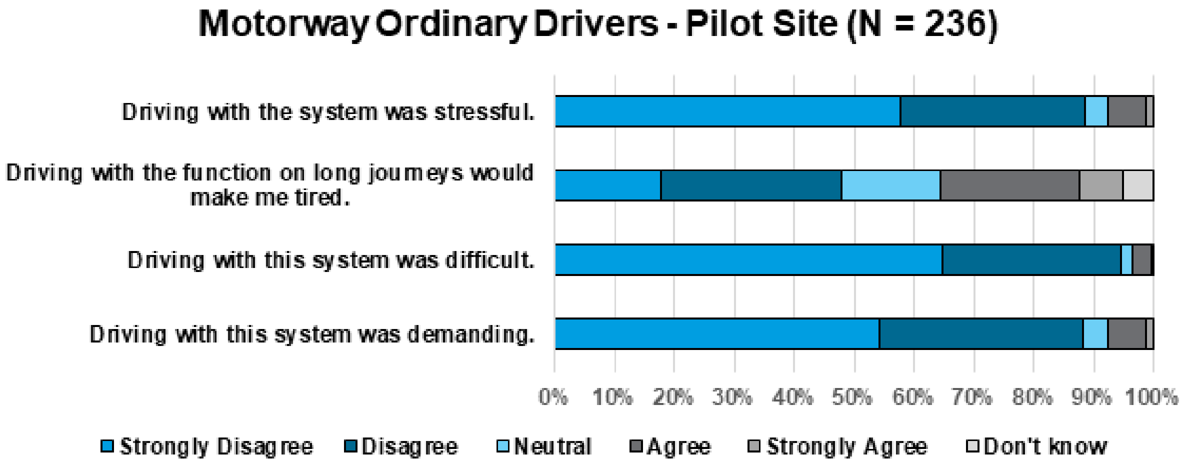
Most of the drivers disagreed or strongly disagreed that driving with the system was stressful. About 95% of ordinary drivers disagreed or strongly disagreed, with a lower percentage of disagreement from professional drivers (79%). Mixed results were obtained in relation to the statement 'Driving with the function on long journeys would make me tired.' For professional drivers from the Pilot sites, 29% agreed or strongly agreed, 18% of them were neutral, and 45% disagreed or strongly disagreed, suggesting that almost half of the professional drivers felt they might get tired when using the system.

Similarly, for ordinary drivers from the Pilot sites, 31% of them agreed or strongly agreed, 18% were neutral, and 48% disagreed or strongly disagreed. On the other hand, most of the ordinary drivers tested in simulator studies agreed or strongly agreed with the statement (60%), 18% remained neutral, and 23% were in one of the disagree categories. This suggests that drivers' experience of tiredness while driving in the simulator is different to the on-road experience, which could be due to the repetition in scenes as well as multiple testing. Most of the drivers disagreed or strongly disagreed that driving with the system was difficult (84% for professional drivers; 94% for ordinary drivers from the Pilot site; 100% for ordinary drivers from simulator studies) or demanding (76% for professional drivers; 88% for ordinary drivers from the Pilot site; 93% for ordinary drivers from simulator studies), suggesting that most drivers found the system easy to use. However, findings also seem to reveal a difference between simulator and on-road testing in system evaluation.

(a)



(b)



(c)

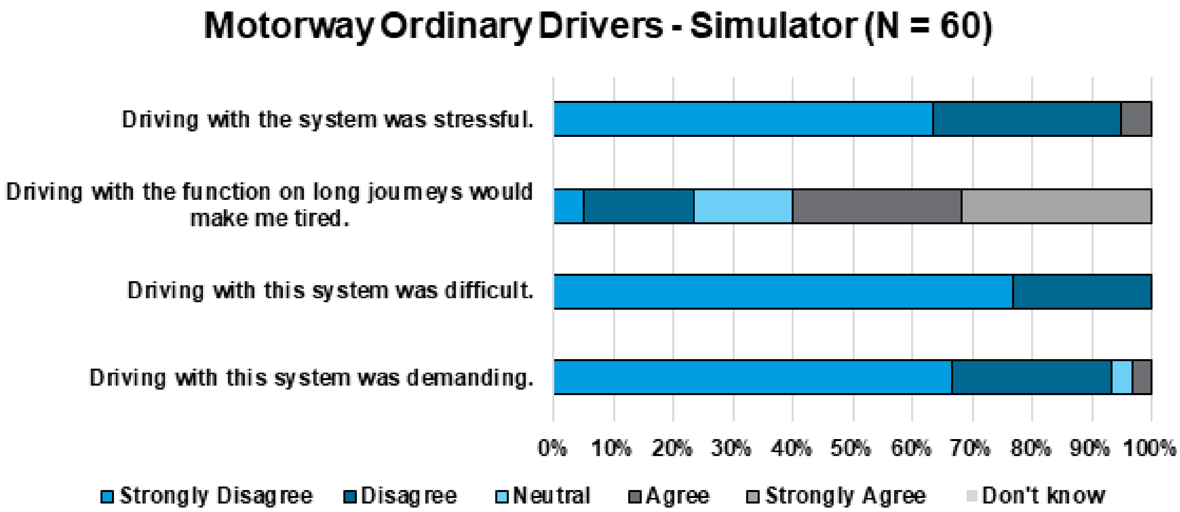


Figure 4.4: Ratings of drivers' workload or state while interacting with the ADF for (a) Motorway Professional Drivers (b) Motorway Ordinary Drivers – Pilot Sites and (c) Motorway Ordinary Drivers – Simulator.

4.1.4 RQU6: What Is the Impact of ADF Use on Driver Awareness?

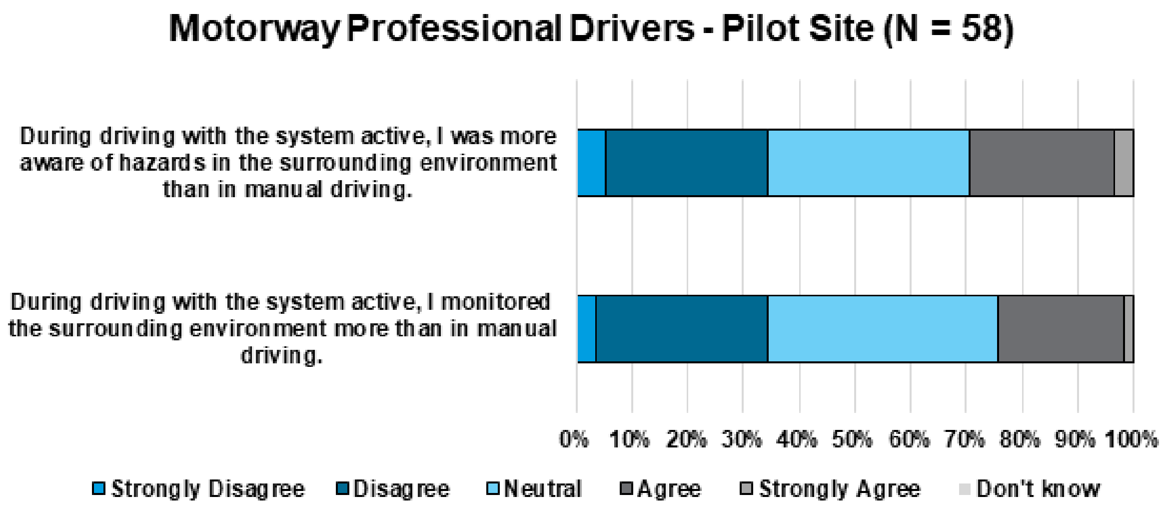
Two items investigated drivers' level of awareness of their environment while using the ADF (see Table 4.4).

Table 4.4: Sub-research questions and questions administered to understand the impact of ADF use on driver awareness.

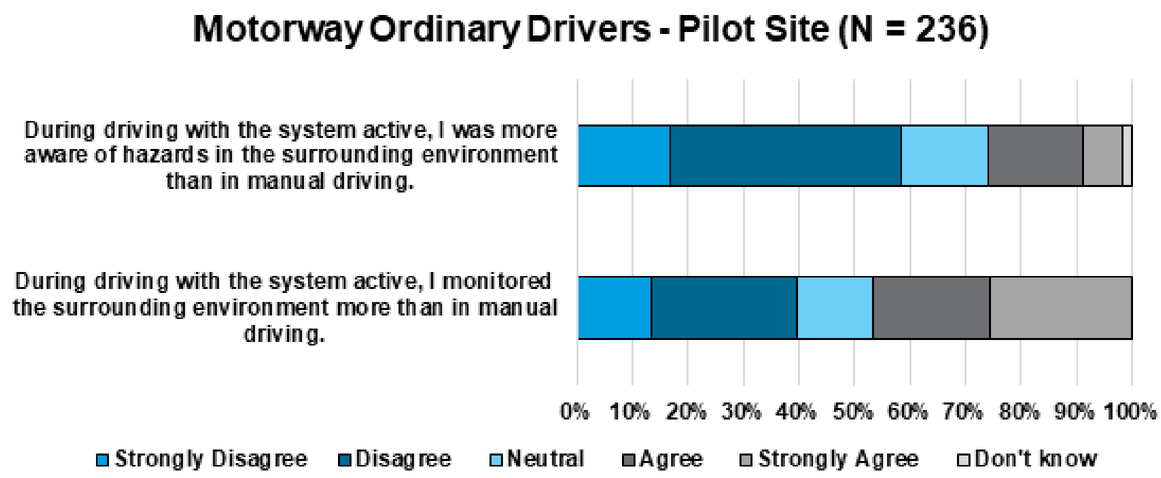
Research Questions	Questions Administered
• What is the effect of ADF use on driver attention to the road/other road users?	• During driving with the system active, I monitored the surrounding environment more than in manual driving
• What is drivers' risk perception while using the ADF?	• During driving with the system active, I was more aware of hazards in the surrounding environment than in manual driving

Findings seem to be mixed in terms of how the "system active" affects **drivers' awareness** and whether they **monitored the surroundings** more than in manual driving. For instance, most of the ordinary drivers from the simulator disagreed or strongly disagreed with these two statements, with 65% and 80% respectively for each statement, suggesting that they did not feel the need to monitor their environment when the automated system was on, perhaps due to simulators being a safe and well-controlled environment, and having gotten used them to after multiple testing. However, in the Pilot sites, 34% of professional drivers disagreed or strongly disagreed with being more aware of hazards in the surroundings, with 24% agreeing or strongly agreeing. Similarly, 34% of professional drivers disagreed or strongly disagreed that they monitored their surroundings more than in manual driving, with 29% agreeing or strongly agreeing with the statement. For ordinary drivers from the Pilot site, 40% disagreed or strongly disagreed that they were more aware of hazards in the surrounding environment, with 47% agreeing or strongly agreeing with the statement. On the other hand, the majority (58%) disagreed or strongly disagreed that they were monitoring the surroundings more than in manual driving, with only 24% agreeing or strongly agreeing with the statement. Overall, this suggests that there was a lot of variance in the attitudes of both professional and ordinary drivers at the Pilot sites.

(a)



(b)



(c)

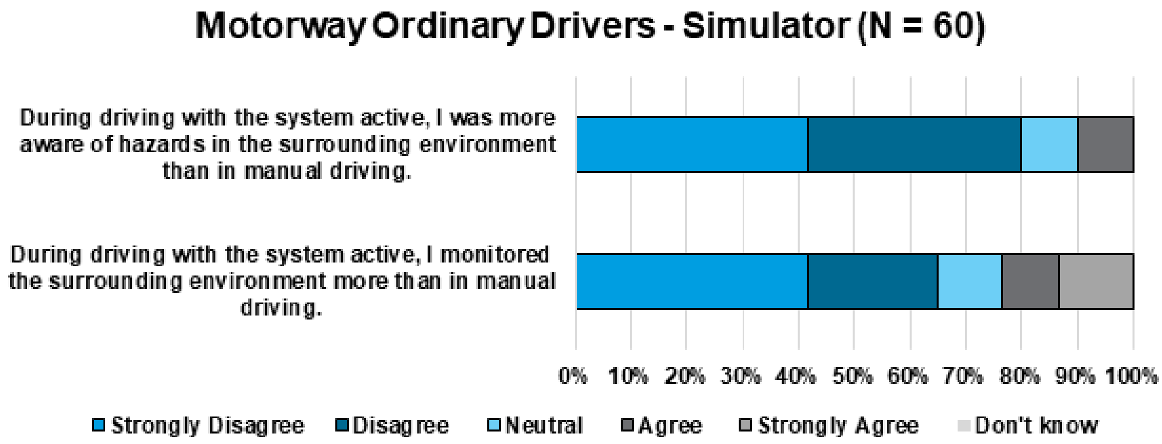


Figure 4.5: Ratings of drivers' workload or state while interacting with the ADF for (a) Motorway Professional Drivers (b) Motorway Ordinary Drivers – Pilot Sites and (c) Motorway Ordinary Drivers – Simulator.

4.1.5 RQ-U4: What Are Drivers' Expectations Regarding System Features?

Three items examined drivers' overall impressions of the Motorway system (see Table 4.5).

Table 4.5: Sub-research questions and questions administered to understand drivers' expectations regarding system features.

Sub-Research Questions	Questions Administered
• What is drivers' overall impression of the system?	• Using the system on motorways was fun • I would make MORE trips if I had the function in my car • I would select destinations further away if I had the function in my car

Interestingly, most of the ordinary drivers agreed or strongly agreed that using the system on motorways was **fun** (86% for Pilot sites and 87% for simulator studies), but only 38% of professional drivers agreed or strongly agreed with the statement, with 35% being neutral and 26% in disagreement. This again could be due to the system being new to ordinary drivers, whereas professional drivers might have experienced the system many times and were more familiar with it. In addition, professional drivers also had a higher responsibility monitoring the system, whereas ordinary drivers could feel more at ease knowing there was a safety driver. Most drivers disagreed or strongly disagreed that they would **make more trips** if they had the function in their car (53% professional drivers, 52% ordinary drivers from the Pilot sites, and 60% ordinary drivers from simulator studies). Most professional drivers disagreed or strongly disagreed that they would select destinations **further away** if they had the function in their car (57%), whereas only 42% of ordinary drivers from the Pilot sites disagreed with the statement. On the other hand, most ordinary drivers from simulator studies agreed or strongly agreed (52%).

(a)

(b)

(c)

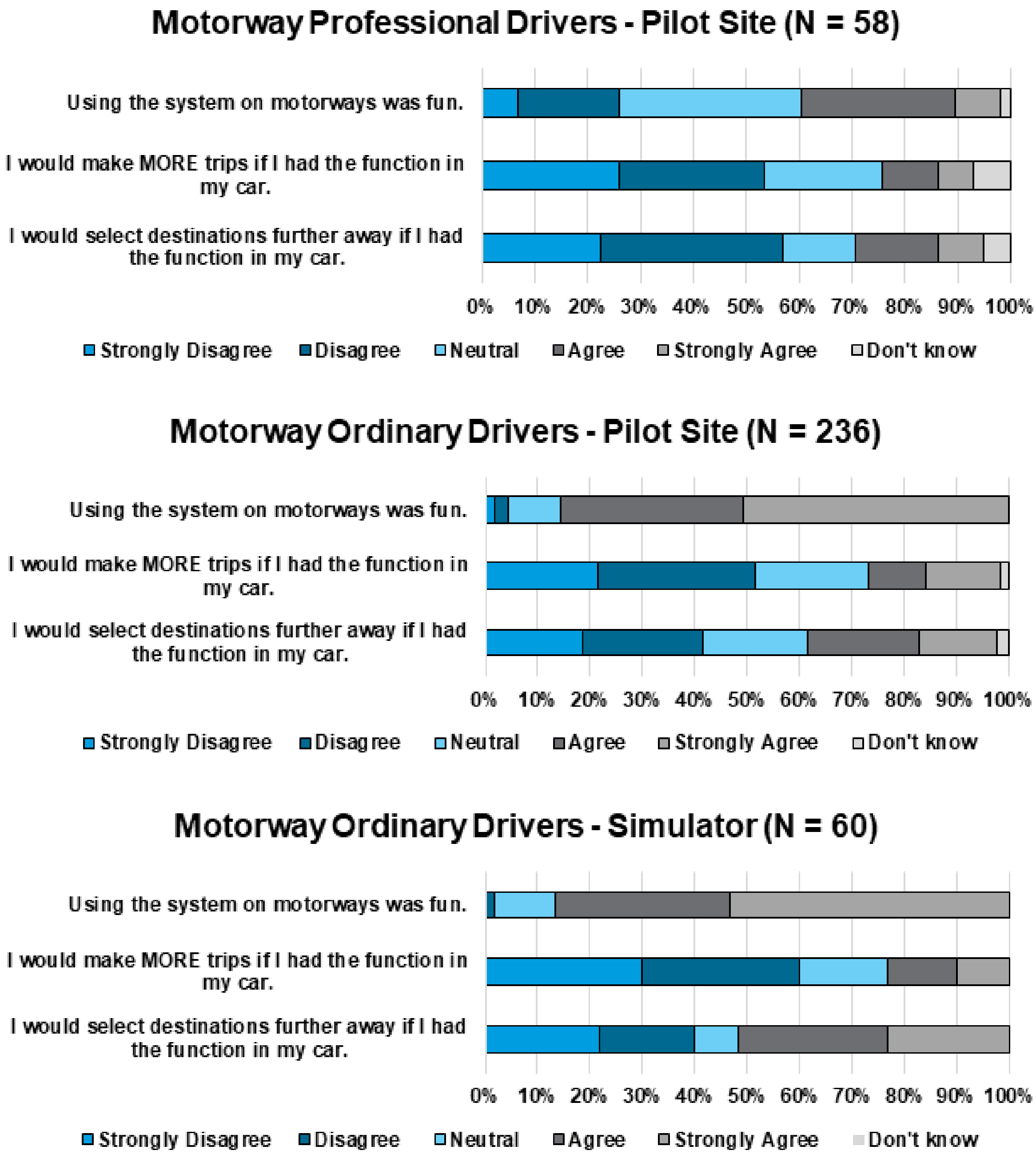


Figure 4.6: Ratings of drivers' overall impression while using the ADF for (a) Motorway Professional Drivers (b) Motorway Ordinary Drivers – Pilot Sites and (c) Motorway Ordinary Drivers – Simulator.

4.1.6 RQ-U9: What Is Drivers' Secondary Task Engagement During ADF Use?

Two sub-questions examined drivers' secondary task engagement during ADF use (see Table 4.6).

Table 4.6: Sub-research questions and questions administered to understand what is drivers' secondary task engagement during ADF use.

Sub-Research Questions	Questions Administered
• What secondary tasks do or would drivers engage in during ADF use?	• I would use the time the system was active to do other activities
• What is the frequency and duration of drivers' secondary task engagement during ADF use?	• Rate how frequent drivers would engage in each activity while the system is active • None • Office/work tasks • Watching movies • Sleeping • Browsing the Internet • Navigation • Social media • Smartphone apps • Personal hygiene/cosmetics • Smoking • Calling • Eating or drinking • Interact with a passenger • Music, radio, audiobooks • Texting

Figure 4.7 reveal that most of the ordinary drivers agreed or strongly agreed that they would use the time the system was active **to do other activities** (82% from Pilot site, 98% from simulator studies), but only 41% of professional drivers agreed with that statement, potentially due to being more familiar with the system's limitations.

We also explored what type of activities drivers liked to engage in and how frequently, when the system was active.

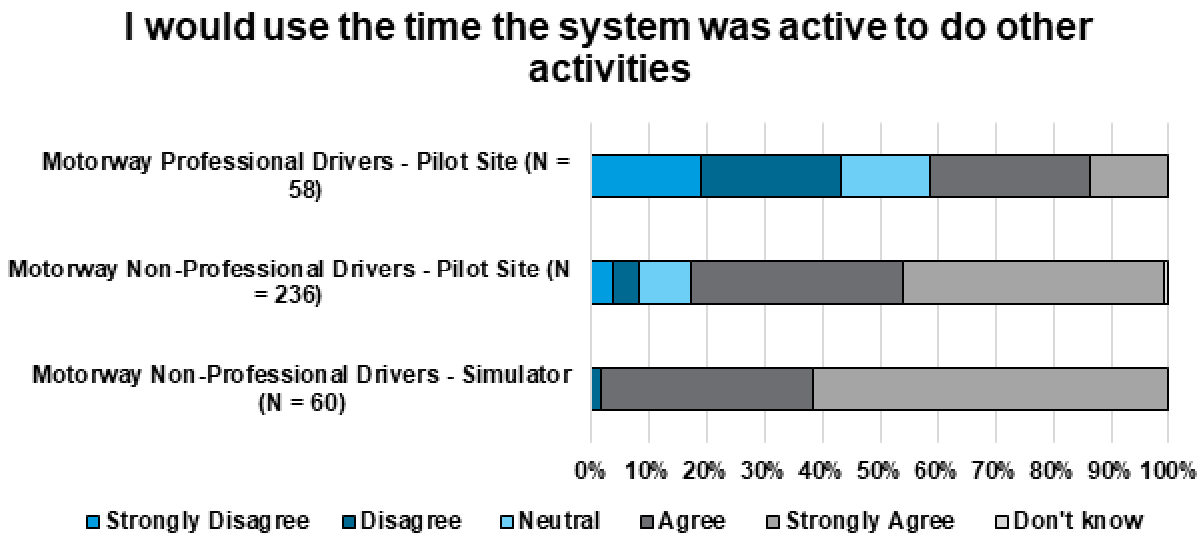
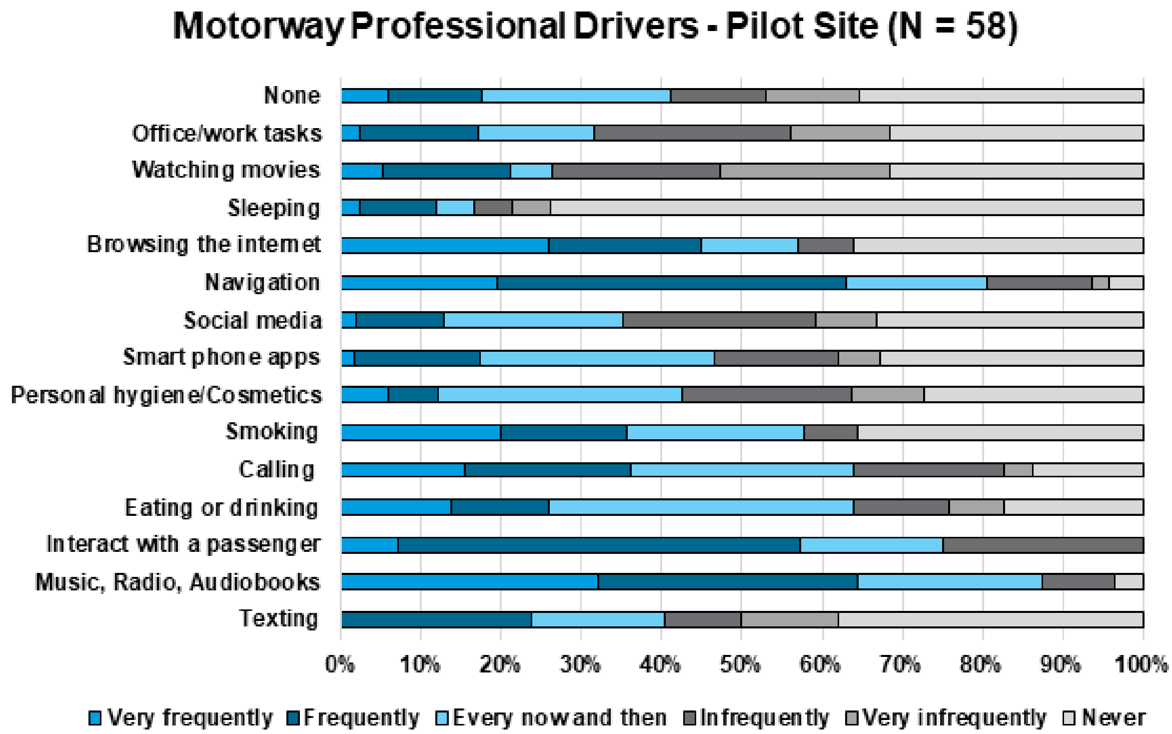


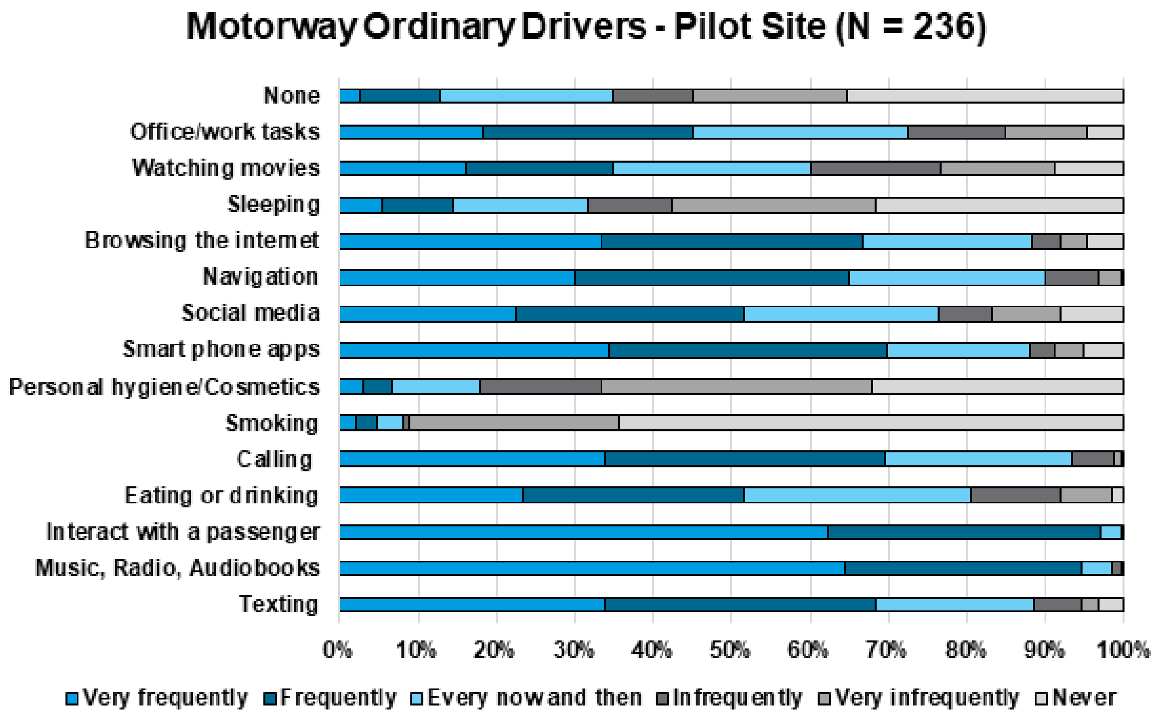
Figure 4.7: Ratings of drivers' willingness to engage in a secondary task while using the ADF for Motorway Professional Drivers, Motorway Ordinary Drivers – Pilot Sites and Motorway Ordinary Drivers – Simulator.

The top three activities that professional drivers would engage in frequently or very frequently were music, radio, audiobook (64%), navigation (63%), and interacting with a passenger (57%); whereas the three activities that they would engage in infrequently, very infrequently, or never were sleeping (83%), watching movies (74%), and office/work tasks (68%). For ordinary drivers tested at the Pilot site, the top three activities that they would engage in frequently or very frequently were interacting with a passenger (97%), music, radio, audiobook (94%), and smartphone apps (70%). The three activities that they would engage in very infrequently, infrequently or never were smoking (92%), personal hygiene/cosmetics (82%), and sleeping (68%). For ordinary drivers tested in the simulator, the top three activities that they would engage in frequently or very frequently were music, radio, audiobook (93%), interacting with a passenger (92%), and texting (78%), whereas the three activities that they would engage in very infrequently, infrequently, or never were smoking (93%), personal hygiene/cosmetics (83%), and sleeping (55%). It is also worth noting that drivers were told that sleeping was not allowed when the system was active. These results show that there was a great degree of similarity across drivers as to the types of activities they would like to engage in while automation is on, with interaction and listening activities selected more often than activities requiring looking away from the road.

(a)


Motorway Professional Drivers - Pilot Site (N = 58)
None
Office/work tasks
Watching movies
Sleeping
Browsing the internet
Navigation
Social media
Smart phone apps
Personal hygiene/Cosmetics
Smoking
Calling
Eating or drinking
Interact with a passenger
Music, Radio, Audiobooks
Texting
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Very frequently
Frequently
Every now and then
Infrequently
Very infrequently
Never

(b)


Motorway Ordinary Drivers - Pilot Site (N = 236)
None
Office/work tasks
Watching movies
Sleeping
Browsing the internet
Navigation
Social media
Smart phone apps
Personal hygiene/Cosmetics
Smoking
Calling
Eating or drinking
Interact with a passenger
Music, Radio, Audiobooks
Texting
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Very frequently
Frequently
Every now and then
Infrequently
Very infrequently
Never

(c)

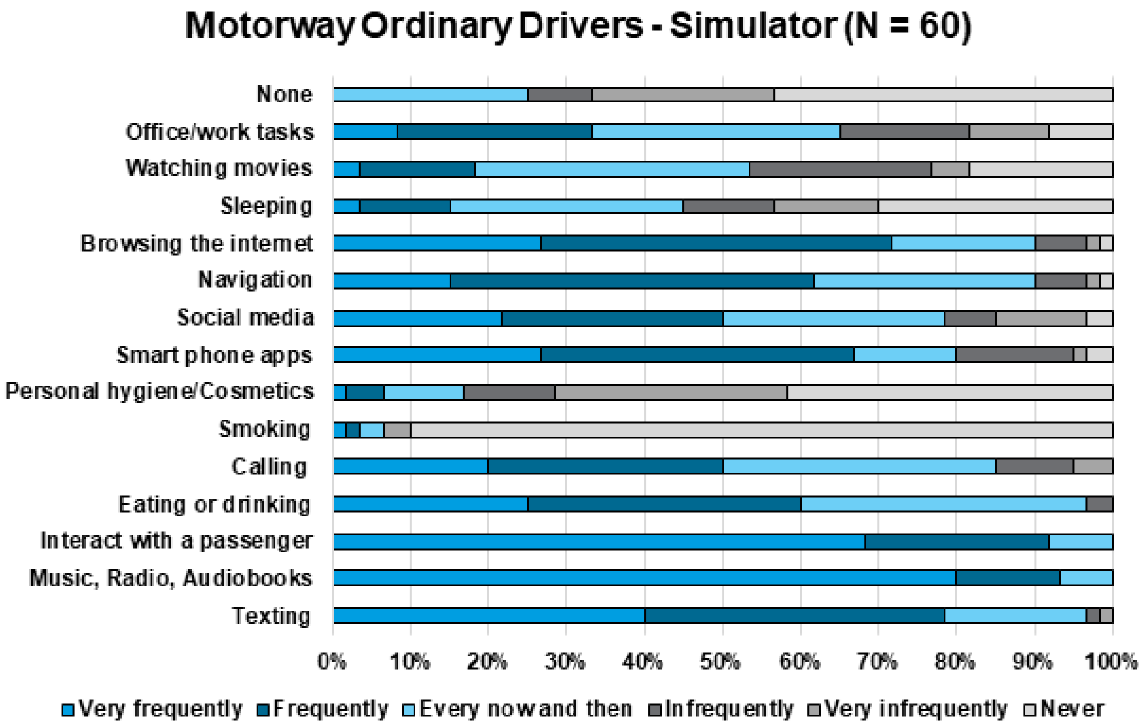


Figure 4.8: Ratings of drivers' willingness to engage in different types of secondary task while using the ADF for (a) Motorway Professional Drivers (b) Motorway Ordinary Drivers – Pilot Sites and (c) Motorway Ordinary Drivers – Simulator.

4.1.7 RQ-U10 - How Do Drivers Respond When They Are Required to Retake Control?

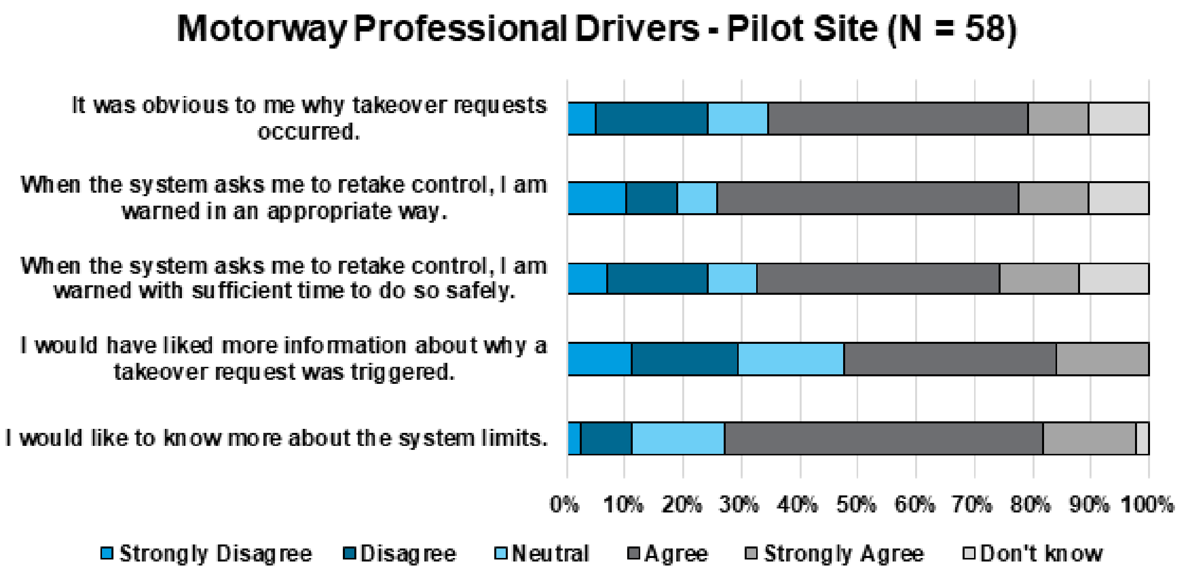
There were six questions examining drivers' responses to planned take-over requests.

Table 4.7: Sub-research questions and questions administered to understand how drivers respond when they are required to retake control.

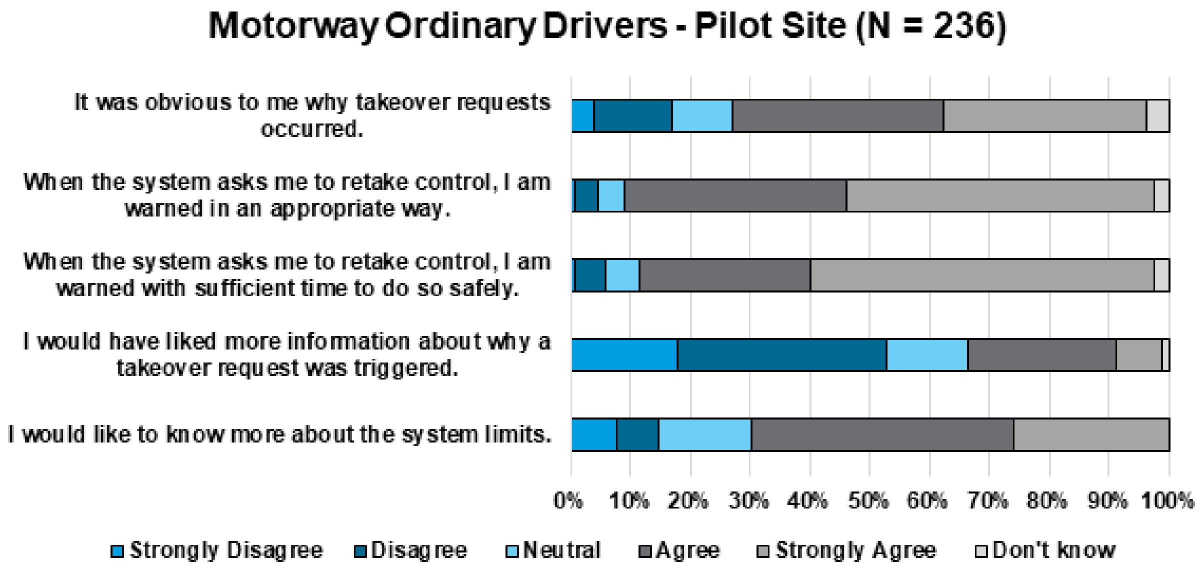
Sub-Research Questions	Questions Administered
• How do drivers respond when they are required to retake control in planned take-overs?	• It was obvious to me why take-over requests occurred • When the system asks me to retake control, I am warned in an appropriate way • When the system asks me to retake control, I am warned with sufficient time to do so safely • How dangerous was the previous take-over situation? • I would have liked more information about why a take-over request was triggered • I would like to know more about the system limits.

Most professional drivers agreed or strongly agreed that it was obvious **why take-over requests occurred** (55%). The proportion was higher for ordinary drivers from the Pilot sites (69%) and was even higher again for ordinary drivers tested in simulators (91%) (see Figure 4.9). This could be due to simulators being a much more controlled environment than Pilot sites. Most of professional drivers agreed or strongly agreed that they were **warned in an appropriate way** when they were asked to retake control by the system (64%), and, once again, the proportion was higher for ordinary drivers from the Pilot sites (89%) and even higher for the ordinary drivers in simulator studies (94%). 55% of the professional drivers agreed or strongly agreed that they were warned with **sufficient time** to safely retake control (55%), and again ordinary drivers from the Pilot sites seemed to be more positive, with 86% of them agreeing with this statement. Most of professional drivers would have liked **more information** on why a take-over request was triggered (52%), but 33% of ordinary drivers from the Pilot sites and 41% of ordinary drivers from the simulator agreed or strongly agreed with the statement. Most of the drivers wanted to **know more about the system limits**: 70% of professional drivers, 70% of ordinary drivers from the Pilot site, and 100% of ordinary drivers from the simulator studies agreed or strongly agreed with this statement.

(a)



(b)



(c)

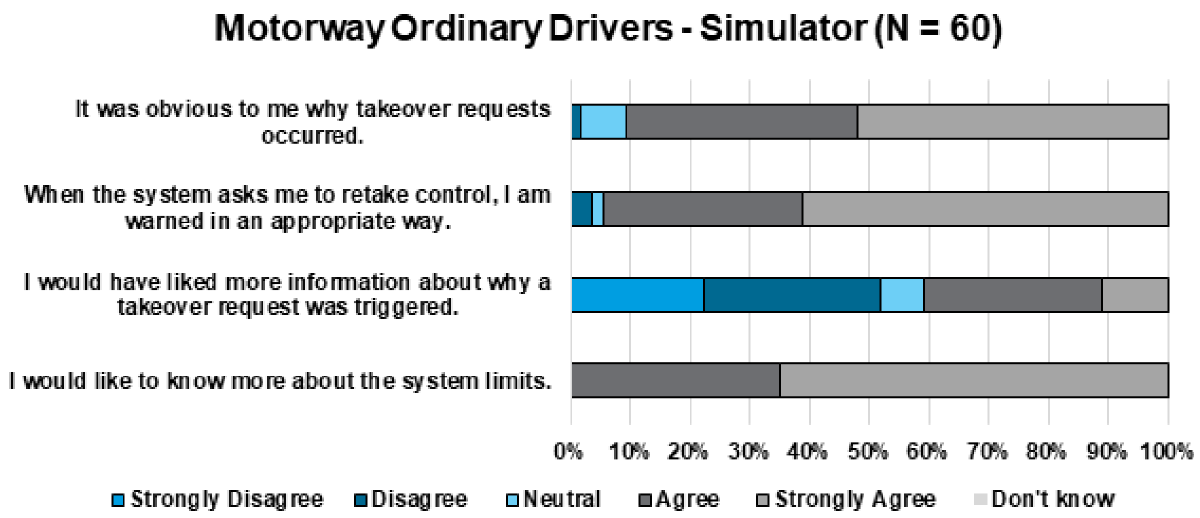


Figure 4.9: Ratings of drivers' experience of take-over control while using the ADF for (a) Motorway Professional Drivers (b) Motorway Ordinary Drivers – Pilot Sites and (c) Motorway Ordinary Drivers – Simulator.

In more than 60% of take-over situations it took less than 4 seconds before drivers reacted to the take-over requests and deactivate the function. The reaction time in 99% of situations was under 10 seconds (see Figure 4.10).

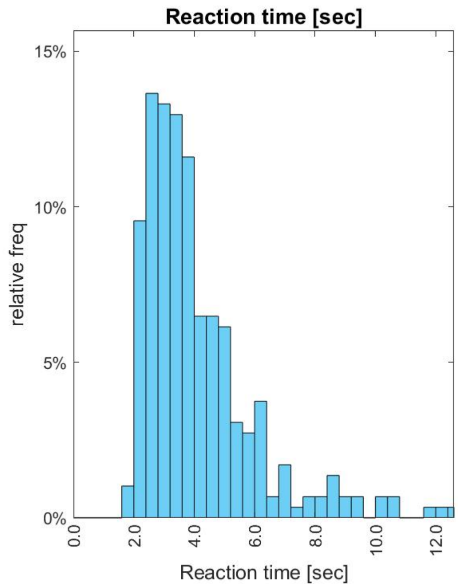


Figure 4.10: Time until the ADF is deactivated after a take-over request.

The impact of take-over requests on driving behaviour and the transition of control from the vehicle to the driver is assessed with the take-over-controllability rating (TOC-rating). The evaluated driving sequence starts with the beginning of the take-over request and ends after the action required to solve the situation leading to the take-over request is terminated.

The results are based on two Pilot sites where non-professional drivers reacted to take-over requests. Take-over requests could happen in three different circumstances:

- Take-over request during stable lane bound-driving, no specific actions need to be taken by the driver after the transition to manual driving.
- Take-over request in a situation where two new lanes merge from the right. The driver can mostly stay in the lane but needs to react to other traffic merging from the right.
- Take-over request before an exit that needs to be taken. The driver needs to change to the exit lane after taking back control.

As can be seen in Figure 4.11, in all evaluated situations the take-over was solved safely and mostly also without driving errors. In the situation where no specific action needed to be taken, nearly 100% of the situations were solved perfectly or with only minor imprecision (rating ≤3). Here, the main weakness was due to visible emotions by the driver that indicated some experience of stress as well as rather small distances laterally. In the situation with the two merging lanes, 96% of situations were rated as being perfect transitions with only minor imprecisions. In this case, the most frequent error type was again small lateral distances. For take-over before the exit, 15% of take-overs were rated as being perfect with minor imprecision, the rest is rated as being safe but with errors. The most frequent error types were missing or too late use of the indicator during 85% of all take-overs, small lateral distances, and rather late initiation of the lane change to the exit lane.

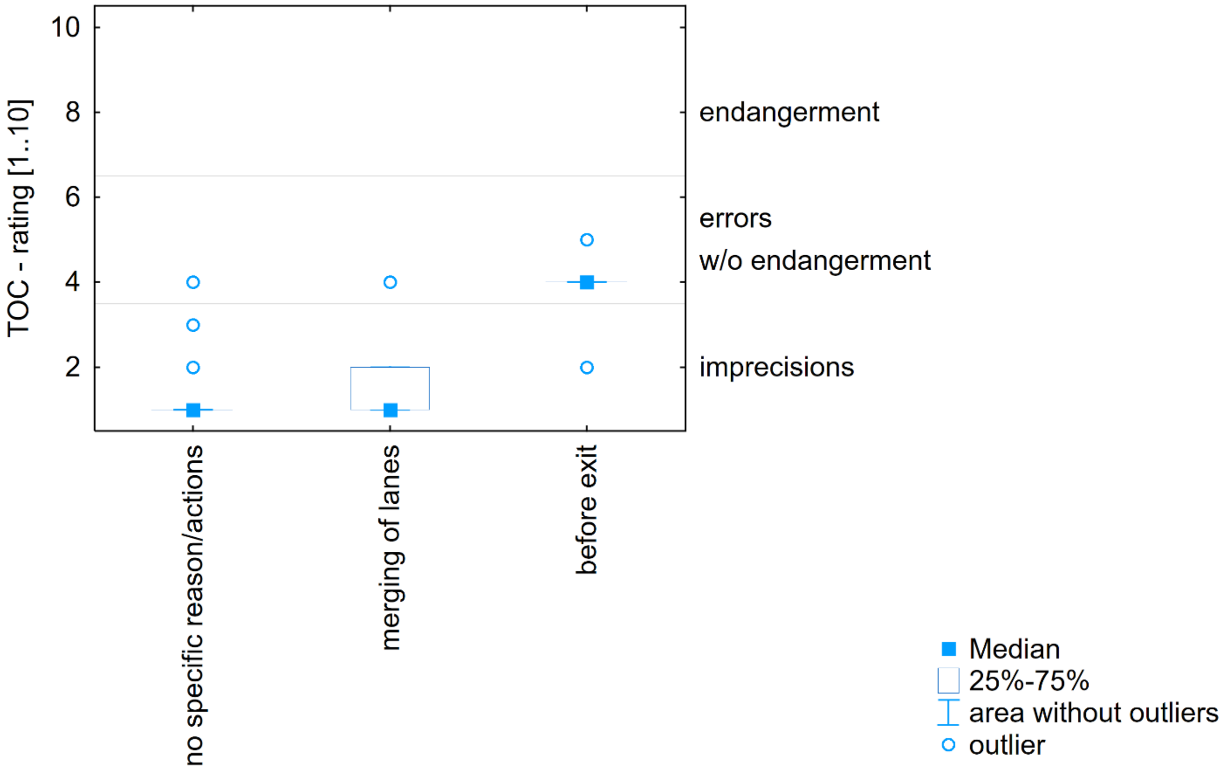


Figure 4.11: Results of TOC-rating in the three different take-over situations.

Results indicate that the overall evaluation of the take-over reaction clearly depends on the complexity of the situation in which the take-over request occurs. In driving environments where the driver can stay in their lane after the transition of control and where no time-critical actions need to be taken, the transition is in more than 99% of analysed situations smooth and without errors (TOC-rating≤3). In situations where specific immediate actions are required, driving errors can occur. Here, it needs to be considered that these are situations where errors occur also in manual driving. For instance, it happens rather frequently that drivers change to an exit lane without using the indicator. For a final evaluation on whether driving errors are more frequent immediately after a take-over request, a comparison is needed of the frequency of errors and the error types between purely manual driving and driving after take-over requests.

4.1.8 RQ-U7: What Is the Impact of ADF Use on Motion Sickness?

Table 4.8: Questions administered to understand the impact of ADF use on motion sickness.

Questions Administered
• Did you experience motion sickness during your test drive with the function active?

100% of the professional drivers answered ‘No’ to the question. 96% of the ordinary drivers from the Pilot sites answered ‘No’ and 4% answered ‘Yes’. No data was collected from simulator studies.

4.1.9 Applications of regression models to motorway ADF user acceptance

Hierarchical Regression Models were conducted to answer two additional RQs:

(1) Which of the User & Acceptance factors predict willingness to use the motorway ADF system, and

(2) Given that professional drivers seem to be less positive in a lot of the ratings compared to ordinary drivers, we also wanted to know whether system familiarity and driver type predict willingness to use the system.

In the following models, we have only included the data collected from Pilot sites (58 professional drivers and 236 ordinary drivers), and not from the driving simulators, because the experience in the simulator could be very different. Hierarchical Regression Models (also known as sequential regression) were used because they allowed the researchers to enter variables in steps or blocks in a predetermined order. This enables us to show if variables of interest (i.e., driver type and system familiarity) explain a statistically significant amount of variance in the Dependent Variable (DV) (i.e., willingness to use) after accounting for all other variables (i.e. User & Acceptance factors), and handling the unbalanced number of participants that was not normally distributed.

Step 1 Independent Variables (see Table 4.9): Factor Analysis grouped items which drivers answered in a similar manner, creating four main factors within the User & Acceptance items. Van Der Laan’s scale is a validated scale that measures acceptance, which consisted of Usefulness and Satisfying constructs. Van Der Laan’s Usefulness and Satisfying scale were added as the fifth and sixth factors. Table 4.9 shows the items loading onto each of the six factors, along with the reliability values for each of the scales. The mean of each of these factors was computed and entered into the first step of the hierarchical regression analysis.

Table 4.9: Items grouped by Factor Analysis and their respective Cronbach’s Alpha.

FACTOR 1: Cronbach’s Alpha 0.914	Workload/Emotion & Expectation
TJM33_33bb	Sometimes the system behaved unexpectedly
TJM33_33m	The system acted appropriately in all situations
TJM33_33k	The system worked as it should work
TJM33_33ii	Driving with this system was demanding
TJM33_33jj	Driving with the system was stressful
TJM33_33c	I felt safe when driving with the system active
TJM33_33u	Using the system on motorway was fun
TJM33_33q	Driving with the system active was comfortable

TJM33_33o	I trust the system to drive
TJM33_33hh	Driving with this system was difficult
FACTOR 2: Cronbach's Alpha 0.771	**Take-over Experience**
TJM33_33beta	When the system asks me to retake control, I am warned in an appropriate way
TJM33_33alpha	When the system asks me to retake control, I am warned with sufficient time to do so safely
TJM33_33z	During the take-over I always felt safe
TJM39_SQ0011	How dangerous was the previous take-over situation?
FACTOR 3: Cronbach's Alpha 0.51	**System Information**
TJM33_33y	I would have liked more information about why a take-over request was triggered
TJM40a40f_40d	I would like to know more about the system limits
FACTOR 4: Cronbach's Alpha 0.5	**System Monitoring**
TJM33_33ll	I would want to monitor the system's performance
TJM33_33n	I would use the time the system was active to do other activities
FACTOR 5: Cronbach's Alpha 0.681	**Van Der Laan's Usefulness**
TJM31_SQ001	Useful-Useless
TJM31_SQ003	Bad-Good
TJM31_SQ005	Effective-Superfluous
TJM31_SQ007	Assisting-Worthless
TJM31_SQ009	Raising alertness-Sleep-inducing
FACTOR 6: Cronbach's Alpha 0.855	**Van Der Laan's Satisfying**
TJM31_SQ002	Pleasant-Unpleasant
TJM31_SQ004	Nice-Annoying
TJM31_SQ006	Irritating-Likeable
TJM31_SQ008	Undesirable-Desirable

Step 2 Independent Variables: In the second step of the Hierarchical Regression, two variables were entered to measure drivers' level of system experience. These were System Familiarity, measured through the item 'Today, you will be operating with the motorway system, how familiar are you with this type of systems you will be using today?' rated from 1 = highly familiar to 5 = highly unfamiliar; and Driver Type (58 Professional Drivers vs 236 Ordinary Drivers).

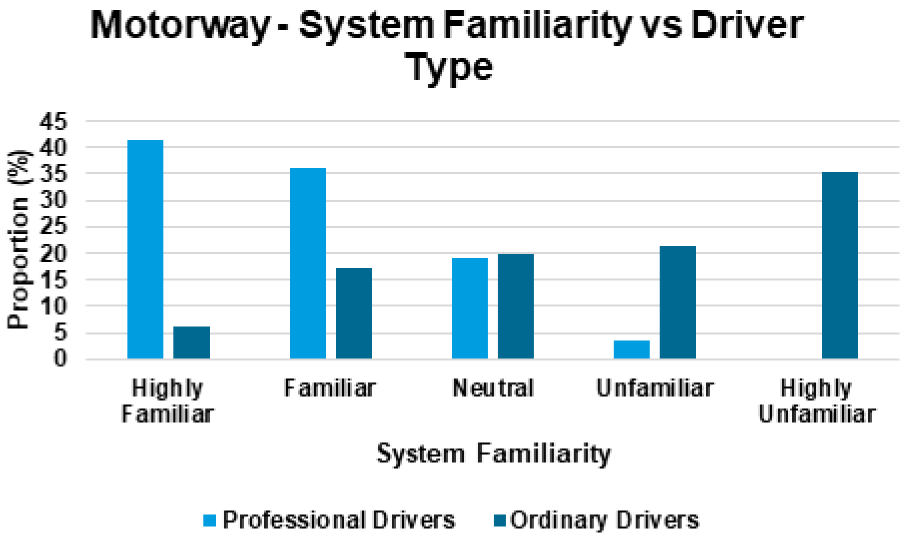


Figure 4.12: Professional and Ordinary Drivers' familiarity with the system.

Figure 4.12 shows the system familiarity of professional drivers and ordinary drivers. It is clear that professional drivers tend to have higher familiarity with the system, whereas ordinary drivers tend to be more unfamiliar with it. Therefore, familiarity was deemed to be a proxy measure of system experience and was included in a separate step to driver type.

A Spearman's correlation was conducted for the factors included in the regression to check for multicollinearity. There was no strong correlation between any of the factors (Coefficient < .07), apart from Van Der Laan's Usefulness & Satisfying (Coefficient = 0.721), which both tap into the underlying construct of 'Acceptance'.

The dependent variable is the mean of three items measuring **Willingness to Use** (Cronbach's alpha = 0.845):

- I would use this system if it was in my car
- I would buy the system
- I would use the system during my everyday trips

Results

- Model 1: Willingness to Use = Six User & Acceptance Factors (R2 = 0.63)
- Model 2: Willingness to Use = Six User & Acceptance Factors + System Familiarity (R2 = 0.635)
- Model 2: Willingness to Use = Six User & Acceptance Factors + Driver Type (R2 = 0.634)

The results of the regression show that the R^2 values for Model 1, 2, and 3, were 0.63, 0.635, and 0.634, respectively. Table 4.10 below shows the coefficients and p values of each factor in each model.

*Table 4.10: Coefficients and p value of each factor in Regression Model 1, 2, and 3 (*p < .05, ** p< .01, ***p < .001).*

Predictor Variables	Model 1	Model 2	Model 3
Workload/Emotion & Expectation	0.334**	0.338**	0.335**
Take-over Experience	0.052	0.063	0.062
System Information	0.221***	0.224***	0.220***
System Monitoring	0.165*	0.167*	0.182*
Van Der Laan's Usefulness	0.254*	0.250*	0.252*
Van Der Laan's Satisfying	0.167	0.176	0.172
System Familiarity	NA	0.070	NA
Driver Type	NA	NA	0.064
R^2	0.63	0.635	0.634
R^2 change	0.63	0.005	0.004

The results revealed that Workload/Emotion & Expectation, System Information, System Monitoring, and Van Der Laan's Usefulness were significant predictors of willingness to use the motorway system. This finding was consistent across three models, whereby the more positive they felt in terms of how the system impacted their workload/emotion and expectation, the higher was their willingness to use. Drivers who reported wanting to know more about the system also had a higher willingness to use. Drivers who would like to engage in secondary tasks and reported being less likely to monitor the system also revealed a higher willingness to use. Finally, the higher the usefulness rating, the higher the willingness to use. In model 2, we did not find system familiarity as a significant predictor of willingness to use, nor did having it as an additional factor significantly improve the variances explained, although R2 did increase slightly, by 0.005. Similarly, we did not find driver type as a significant predictor of willingness to use, nor did having it as an additional factor significantly improve the variances explained (R2 change = 0.004).

4.1.10 Evaluation of Pilot data utilising video coding

To answer some RQs in more detail, video data recorded during the drive was annotated and analysed. This analysis only includes a subset of the ordinary drivers (N=30) who took part in a Wizard-of-Oz study on public roads comparable to motorways. The ADF was mimicked by a hidden second driver, while the participant was under the impression that the vehicle was automated. The video annotations were done based on a codebook with RQ specific features and validated by double annotations to confirm high inter-rater reliability.

4.1.10.1 Reported Trust Related to Behaviour After Hand Over (related to RQ-U3)

To investigate how the trust rated in the pilot-site questionnaire reflects in the driver behaviour, several features were extracted from videos after take-over by the ADF. These features were first glance to the HMI, activation of ADF, taking hands off the steering wheel, taking feet off the pedals in a resting position, and starting a secondary task. There were six phases of ADF available during a drive and the drivers were encouraged to use the system when available. The availability of the ADF is conveyed by visual information on the HMI (display) and an acoustic sound.

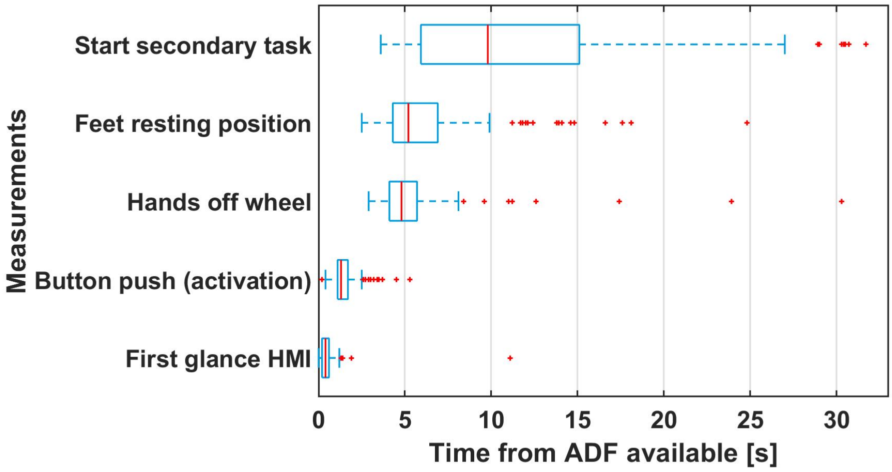


Figure 4.13: Timing of annotated drivers' reactions in the hand over process after ADF becomes available.

The results show that drivers react very fast once the ADF becomes available, looking at the HMI (m=0.5 s) and activating the system on average after only 1.5 s. This is followed by taking the hands off the steering wheel and placing the feet in a resting position (off the pedals), which occurs on average 5.4 s and 6.3 s after the system becomes available, respectively. All these measures showed no significant difference across the six phases where the system was available. However, the start of a secondary task (see Figure 4.13) is spread out, and the timing decreases over the first three ADF phases. Also, not all drivers engage in a secondary task within the 30 seconds after ADF activation.

Besides the hand over process and timing, correlations with the driver's trust rating in the questionnaire were investigated. Among the 30 drivers included in this analysis, only two disagreed with trusting the system and one answered with "neutral". Interestingly, one of the disagreeing drivers did not engage in a secondary task. While the first four measures of occurrence were very homogeneous across all drivers, the start timing of secondary task engagement showed a correlation with the trust rating, meaning that the more they trusted the system, the earlier they engaged in a secondary task when the ADF became available.

4.1.10.2 Drivers' Visual Attention Before and After a Take-Over Request (RQ-U6)

To understand drivers' visual attention before and after a take-over request (TOR), drivers glance locations were coded 30 s before and 30 s after the TOR. The defined areas of interest (i.e., glance locations) were: towards the forward roadway or towards any other off-road area (e.g., secondary task or the instrument cluster). To capture drivers' attention to the forward road, the percent time on forward road (see ISO15007-2) for each driver within different time bins (e.g., 30 s before the TOR) was computed. For comparison, the percent time on forward road was also computed for a corresponding time bin using the manual baseline drive data.

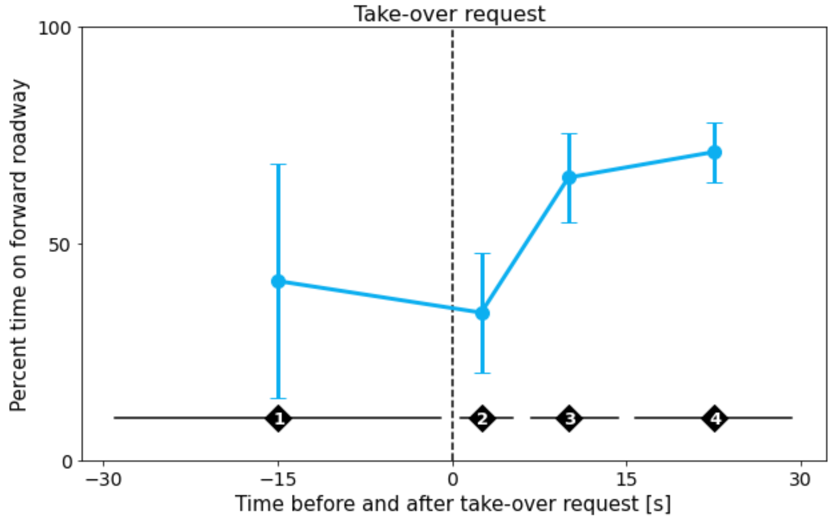


Figure 4.14: Percentage of time that drivers' gaze was directed towards the forward roadway.

Figure 4.14 shows the percentage of time that drivers' gaze was directed towards the forward roadway, within four time intervals, as a function of time anchored at the TOR in L3. Based on how the visual attention to the forward road changed over time, four different intervals were identified during the 30 s before and after the TOR. *Interval 1* captures the visual attention towards the forward roadway during L3 automation leading up to a TOR. *Interval 2* captures the 6 s after the TOR when the drivers deactivate automation. 6 s was chosen because it resembled the time the drivers were given to deactivate automation. While *Interval 4* represents the 15-30 s away from the TOR in which the glance behaviour to the forward roadway has stabilised, *Interval 3* represents the transition from Interval 2 to 4.

A Wilcoxon Signed-Rank Test indicated that the percentage time to the forward road was significantly lower (Mdn = 41%) during L3 automation (Interval 1 in Figure 4.14) compared to the manual 30 s baseline (Mdn = 75%), $Z = -3.975$, $p < 0.001$. Instead of looking at the forward roadway, drivers during L3 automation spent time looking towards secondary task devices (e.g., personal mobile phone) or towards the tablet mounted on top of the centre stack. Shortly after the TOR a noticeable decrease in attention towards the forward roadway can be seen in Figure 4.14 (Interval 2). Instead of looking forward the drivers, in response to the TOR, looked towards the instrument cluster. Then, the visual attention to the road gradually increases again to reach a similar level of attention forward as the manual baseline in Interval 4. A Wilcoxon Signed-Rank Test indicated that the percentage of time on road in Interval 4 (Mdn = 71%) was not statistically significantly different from the 15 s manual baseline, (Mdn = 71%), $Z = -2.67$, $p = 0.79$. A more detailed overview of the results of the glance analysis after take-overs is given by Pipkorn, Dozza and Tivesten, (2021).

4.1.10.3 Secondary Task Engagement Frequency and Duration (RQ-U9)

During the drive with the ADF, the drivers were allowed to engage in secondary tasks. The drivers had two preferences: 1) using their phone (annotated as *texting phone* or *talking phone*) or 2) interacting with a tablet mounted on the centre stack (annotated as *Interact with centre stack*). Only the four longer phases of automation (5-7 minutes) were used for analysis. Figure 4.15 shows the number of engagements across all drivers during the different automation phases. When an action is stopped and restarted it does count again. The feature *Talking/interaction with the passenger* refers to the test leader who was sitting in the passenger seat. When they are getting their phone out, this is considered *Reaching for object* which is a brief activity.

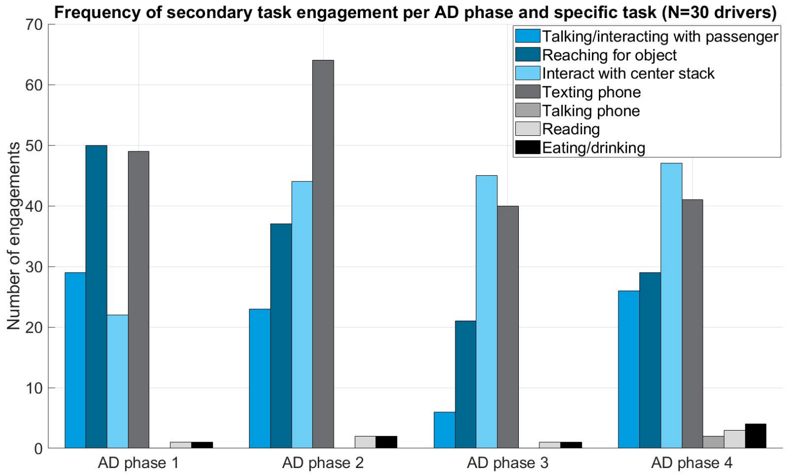


Figure 4.15: Number of engagements in different secondary tasks during 4 different phases of automated driving over all drivers.

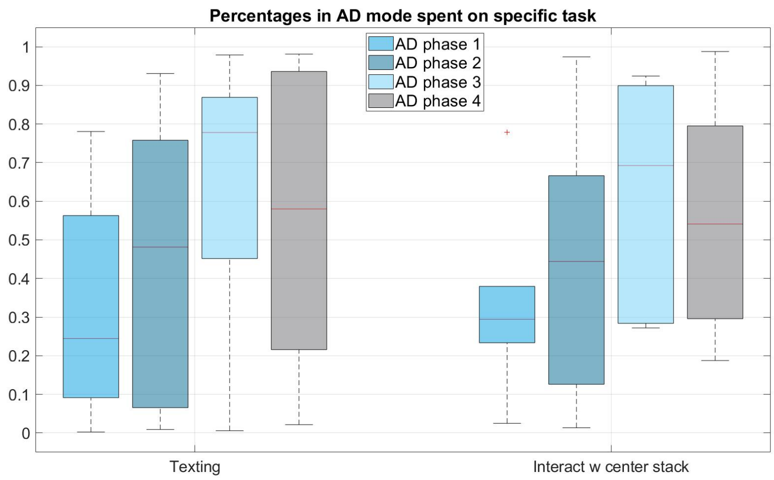


Figure 4.16: Duration of engagement in Texting and Interact with centre stack (using the mounted tablet) across the 4 different phases of automated driving.

The number of engagements is relatively stable over the different phases and indicates that the drivers are comfortable engaging in different secondary tasks. In terms of duration, Figure 4.16 shows the proportion of time during automation spent on *Texting* and *Interact with centre stack.* These two activities were observable for stable time periods and severely drawing attention (especially visual) away from the driving. The amount of time using a phone as well as interacting with a tablet increases over the first three phases on average, indicating that more drivers engage in activities in later exposures. The spread is quite large, implying that there are both drivers engaging not at all or for brief periods and drivers engaging for long parts of the automation duration in one of the two main secondary tasks. Since both are mutually exclusive, drivers could switch between them during a phase of automation. There seems to be no preference between the two activities, as the durations are comparable. The largest proportion is reached in AD phase 3, with 65% time of driving in automation spent with texting on average.

4.1.10.4 Driver Response After Take-over Request (RQ-U10)

The driver response process towards TORs was investigated in detail using video annotations and compared with the results from several supplementary studies (Wizard-of-Oz studies on a test track). These results are reported in detail in D7.2 – L3/L4 Long-term study About user experiences (Metz et al., 2021). All drivers showed quick responses after being prompted to take back control by looking at the HMI, putting the hands on the wheel, glancing at the forward road, deactivating ADF and putting the feet on the pedals. The average take-over time was within 4 seconds after the TOR, and all managed to regain control within 10 seconds. Two drivers needed a second attempt at deactivation, delaying their take-over time.

4.1.11 Summary Findings for Motorway

To evaluate the L3Pilot Motorway system, professional and ordinary drivers were tested at different Pilot sites and in a driving simulator. Generally, drivers were positive about the system, giving high ratings for willingness to use, perceived trust, perceived safety, and system usefulness. Overall, professional drivers were less positive in their ratings (i.e., willingness to use, perceived safety, take-over perceived safety, perceived usefulness, trust, fun) than ordinary drivers. Although all drivers believed that system performance and the expected behaviour of the system could be further improved, the margin for improvement was higher for professional drivers. Most of the drivers did not believe that driving with the system was stressful, difficult, or demanding. However, some thought that using the systems on a long journey would make them tired. This finding seems to be higher for drivers who took part in the simulator study with repetitive testing, which could potentially be due to simulator fatigue. If given the opportunity, most of the ordinary drivers would engage in a secondary task while the system is active, but professional drivers were less inclined to do so. They were also in favour of engaging in secondary tasks which are allowed (i.e., music, audio books, talking to passengers). Most participants agreed that their experience of driving with the system was comfortable, with no reports of motion sickness during the drive. However, future studies should focus on longer engagement in secondary tasks and driving that is more demanding with more lateral and longitudinal movements and environmental change. Vehicle behaviours such as distance kept to other vehicles, driving in curves, braking, acceleration, and smoothness were all rated as comfortable. However, vehicle behaviours in motorway junction areas and during lane changes were deemed to be less comfortable. It is possible that more interactions were required in these situations, possibly leading to the reduced comfort ratings. Future studies should focus on how the system should perform to meet drivers' expectations, and on understanding how the system could improve comfort and acceptance. Most drivers felt safe taking over control from the system, although once again, ordinary drivers provided higher ratings than professional ones.

Overall, professional drivers were less positive than ordinary drivers when evaluating the system, potentially due to system familiarity. However, our regression analysis shows that driver type and system familiarity did not predict willingness to use, suggesting that this difference does not have a significant impact. Willingness to use was predicted by drivers' workload/emotion, system expectation, system information, system monitoring, and Van Der Laan's Usefulness scale. The findings revealed that the more positive drivers were in terms of workload/emotion, system expectation, and usefulness, the higher their willingness to use. Similarly, the more they wanted system information and the more willing they were to engage in a secondary task, the higher their willingness to use the motorway ADF.

It is worth noting that these findings were based on the data collected via different pilots with different study designs. Therefore, findings should be used as a guide to provide the overall impressions of users' evaluation. Pilots were also conducted with the presence of safety drivers, whose role was to deal with any critical situations that arose. This could overestimate users' positive experiences. With continued development and maturation of the system, more studies should be conducted in the future to investigate users' experience, as it predicts willingness to use.

4.2 Urban

The data from the urban ADF was analysed in a similar manner to the Motorway ADF. Because there were only 15 professional drivers who used the urban ADF, their responses were amalgamated with those of ordinary drivers for the analysis presented in this section.

4.2.1 RQU1 – Are Drivers Willing to Use an ADF?

In order to understand whether or not drivers were willing to use the urban ADF, responses to the question "I would use this system if it was in my car" were collated.

Table 4.11: Question administered to investigate drivers' willingness to use an ADF.

Questions Administered
• I would use this system if it was in my car

A majority of participants agreed or strongly agreed that they would **use the system** if it was in their car (76%), although the proportion was generally lower than for the motorway system.

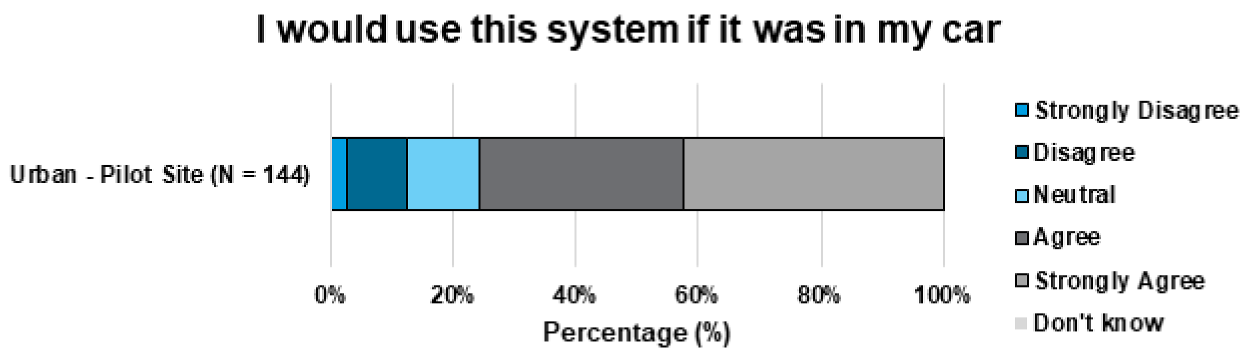


Figure 4.17: Ratings of drivers' willingness to use an ADF.

4.2.2 RQU3 - What Is the User Acceptance of the ADF?

Participants responses to twelve questions were evaluated to understand their acceptance of the Urban ADF (see Table 4.12)

Table 4.12: Sub-research questions and questions administered to understand user accept ance of the ADF.

Sub-Research Questions	Questions Administered
• What is the **perceived safety** of the ADF?	• I felt safe when driving with the system active
• What is the **perceived comfort** of the ADF?	• Driving with the system active was comfortable • Rating of each vehicle behaviour
• What is the **perceived usefulness** of the ADF?	• I think the tested system was Useful/useless • I would recommend the system to others • I would use the system during my everyday trips
• What is the **perceived trust** of the ADF?	• I trust the system to drive • I would want to monitor the system's performance
• How does user **acceptance** differ between ADF types? **(System's Performance)**	• Sometimes the system behaved unexpectedly • The system worked as it should work • The system acted appropriately in all situations

In terms of the **perceived safety** of the Urban ADF (see Figure 4.18), the majority (79%) of participants agreed or strongly agreed with the statement 'I felt safe when driving with the system active'. 81% of the participants agreed or strongly agreed that driving with the system active was **comfortable**. The figure below (Figure 4.19) shows that a majority of the participants felt very comfortable or comfortable with all aspects of the vehicle behaviour. 65% of the participants rated that they would **recommend the system** to others and 69% that they would **use the system** during their everyday trips. 66% **trusted the system** to drive, but 66% would want to **monitor the system's performance**. 49% of the participants agreed or strongly agreed that sometimes the **system behaved unexpectedly**, with 62% rating that the system **worked as it should work**. Finally, 58% agreed or strongly agreed that the **system acted appropriately** in all situations. These results show that, on the whole, user acceptance of the urban system was quite high.

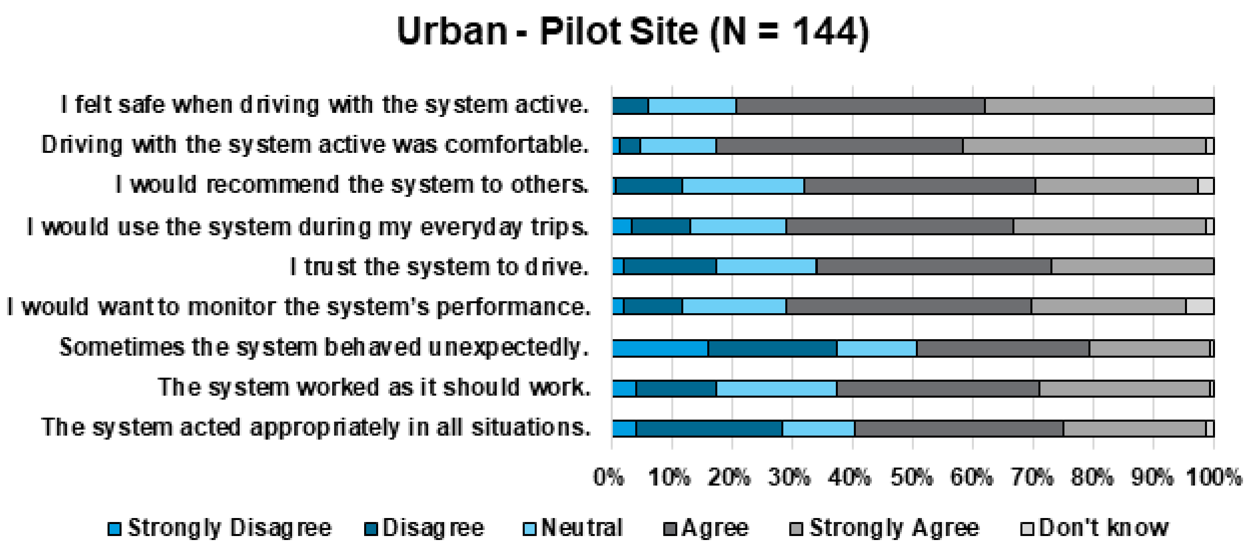


Figure 4.18: Ratings of user acceptance of the ADF.

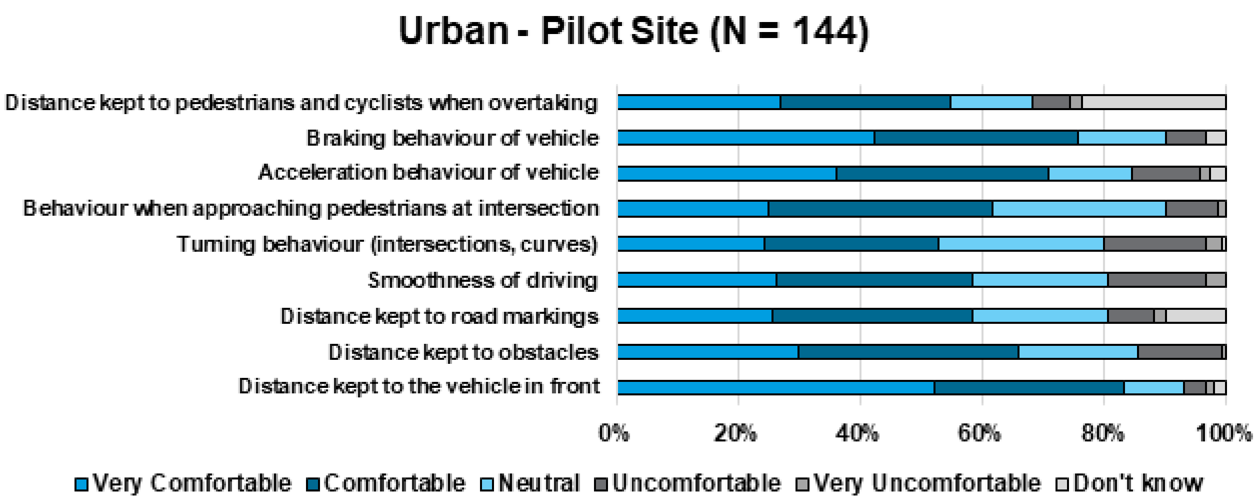


Figure 4.19: Ratings of perceived comfort for each behaviour of the ADF.

4.2.3 RQU5 - What Is the impact of ADF on driver state?

Four questions tapped into understanding drivers' workload or state while interacting with the urban ADF (see Table 4.13).

Table 4.13: Sub-research questions and questions administered to understand driver states while using the ADF.

Sub-Research Questions	Questions Administered
• What is the effect of ADF use on drivers' level of stress?	• Driving with the system was stressful
• What is drivers' level of fatigue while using the ADF?	• Driving with the function on long journeys would make me tired
• What is drivers' workload while using the ADF?	• Driving with this system was difficult • Driving with this system was demanding

Most of the urban participants **disagreed or strongly disagreed** that driving with the system was **demanding** (78%), **difficult** (81%) and **stressful** (71%), indicating that the urban system was deemed to be relatively easy to use. About 43% of them agreed that driving with the function on long journeys would make them **tired**, with 33% disagreeing, suggesting that there was individual variance in how tiring the system was deemed to be.

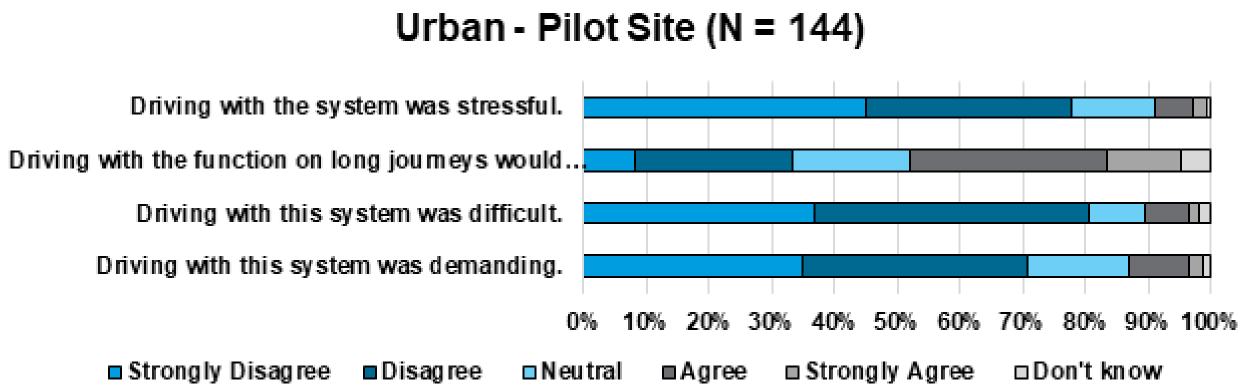


Figure 4.20: Ratings of drivers' workload or state while interacting with the ADF.

4.2.4 RQU6 - What Is the Impact of ADF Use on Driver Awareness?

Two questions investigated drivers' awareness of their environment when using the urban ADF.

Table 4.14: Sub-research questions and questions administered to understand the impact of ADF use on driver awareness.

Research Questions	Questions Administered
• What is the effect of ADF use on driver attention to the road/other road users?	• During driving with the system active, I monitored the surrounding environment more than in manual driving

Research Questions	Questions Administered
• What is drivers' risk perception while using the ADF?	• During driving with the system active, I was more aware of hazards in the surrounding environment than in manual driving

Of the urban participants, 28% agreed or strongly agreed that they were more **aware of the surrounding hazards**, but 48% disagreed or strongly disagreed. A majority of the urban participants disagreed or strongly disagreed that they **monitored the surroundings** more than in manual driving (62%), with 19% of them agreeing or strongly agreeing with the statement. Thus it appears that most of participants (i.e. as passengers) did not feel a need to monitor their environment closely when the urban ADF was engaged.

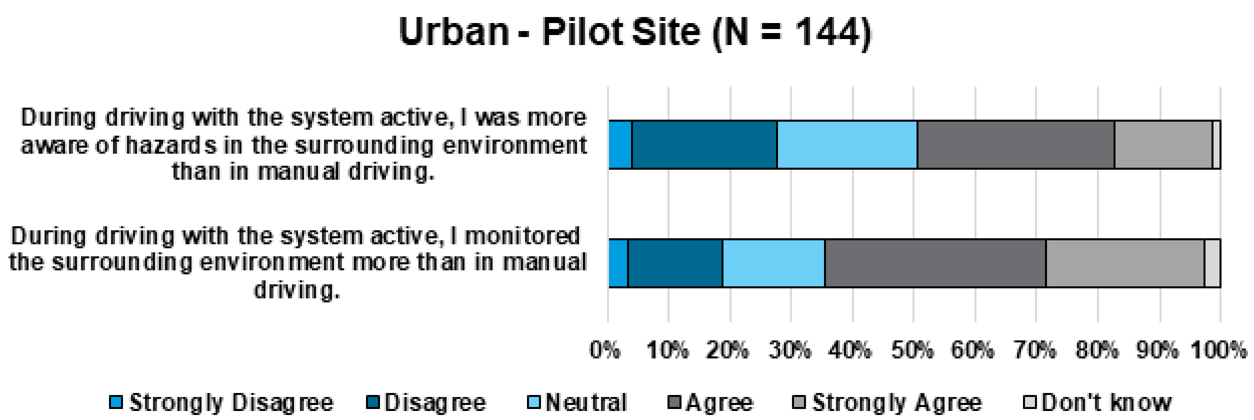


Figure 4.21: Ratings of drivers' level of awareness of their environment while using the ADF.

4.2.5 RQU4 - What Are Drivers' Expectations Regarding System Features?

Two items investigated drivers' expectations about their travel plans with the urban ADF.

Table 4.15: Sub-research questions and questions administered to understand drivers' expectation regarding system features.

Sub-Research Questions	Questions Administered
• What is drivers' overall impression of the system?	• I would make MORE trips if I had the function in my car • I would select destinations further away if I had the function in my car

About 44% of the urban participants agreed or strongly agreed that they would select destinations **further away** if they had the function in their car, with 28% disagreeing or strongly disagreeing with this statement. In addition, 33% of the urban participants agreed or strongly agreed that they would make **more trips** if they had the function in their car, but a larger proportion disagreed or strongly disagreed (44%). This variance suggests that the urban ADF may not have a great impact on travel or trip decisions.

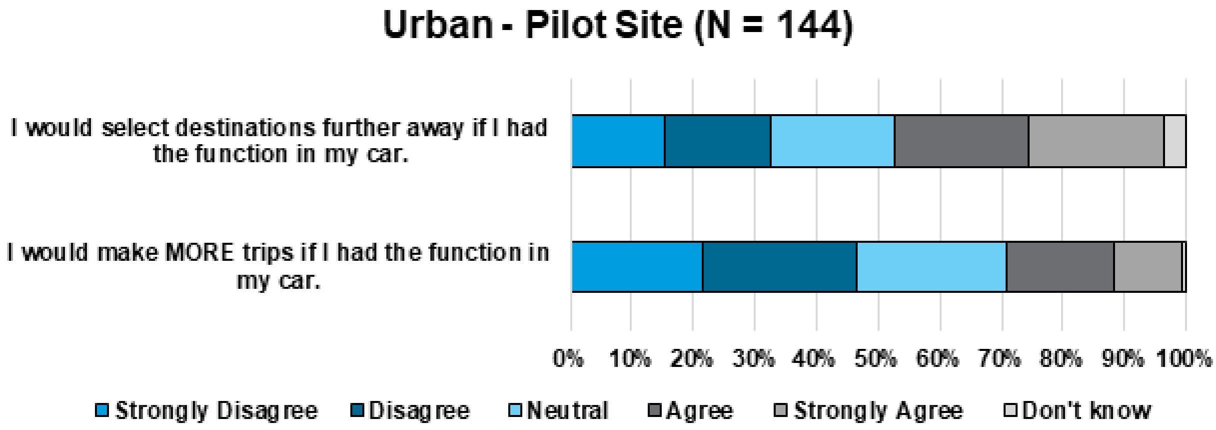


Figure 4.22: Ratings of drivers' overall impression while using the ADF.

4.2.6 RQU9 - What Is Drivers' Secondary Task Engagement During ADF Use?

Two sub-questions addressed drivers' secondary task engagement when the urban ADF was on.

Table 4.16: Sub-research questions and questions administered to understand drivers' secondary task engagement during ADF use.

Sub-Research Questions	Questions Administered
• What secondary tasks do or would drivers engage in during ADF use?	• I would use the time the system was active to do other activities
• What is the frequency and duration of drivers' secondary task engagement during ADF use?	• Rate how frequent drivers would engage in each activity while the system is active • None • Office/work tasks • Watching movies • Sleeping • Browsing the Internet • Navigation • Social media • Smartphone apps • Personal hygiene/cosmetics • Smoking • Calling • Eating or drinking • Interact with a passenger • Music, radio, audiobooks • Texting

A small majority of urban participants agreed or strongly agreed that they would engage in a **secondary task** when the system was active (56%), with 26% disagreeing or strongly disagreeing. The top three activities that urban participants would engage in very frequently or frequently were music, radio, audiobook (91%), interacting with a passenger (87%), and calling (65%); whereas the three activities that they would engage in very infrequently, infrequently or never, were personal hygiene/cosmetics (25%), watching movies (23%), and smartphone apps (22%). Similar to the motorway ADF, it seems that participants were most comfortable engaging in listening tasks or tasks which would not require them to look away from the road.

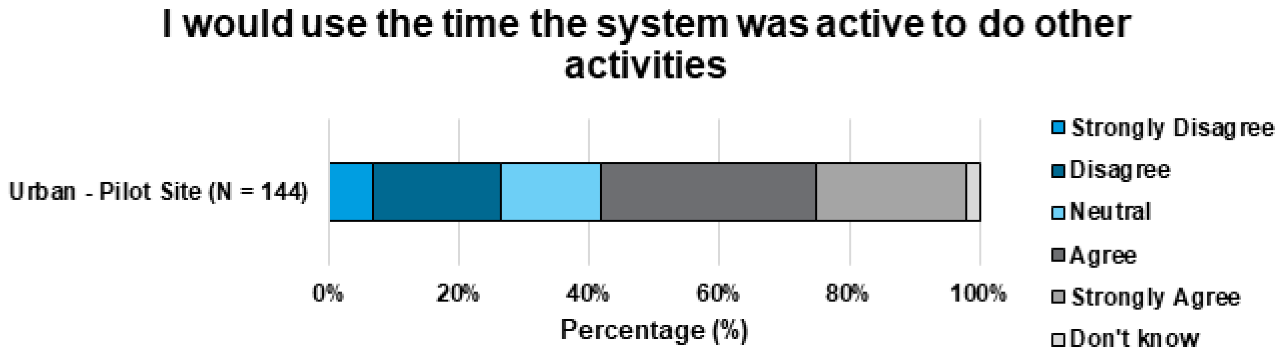


Figure 4.23: Ratings of drivers' willingness to engage in a secondary task while using the ADF.

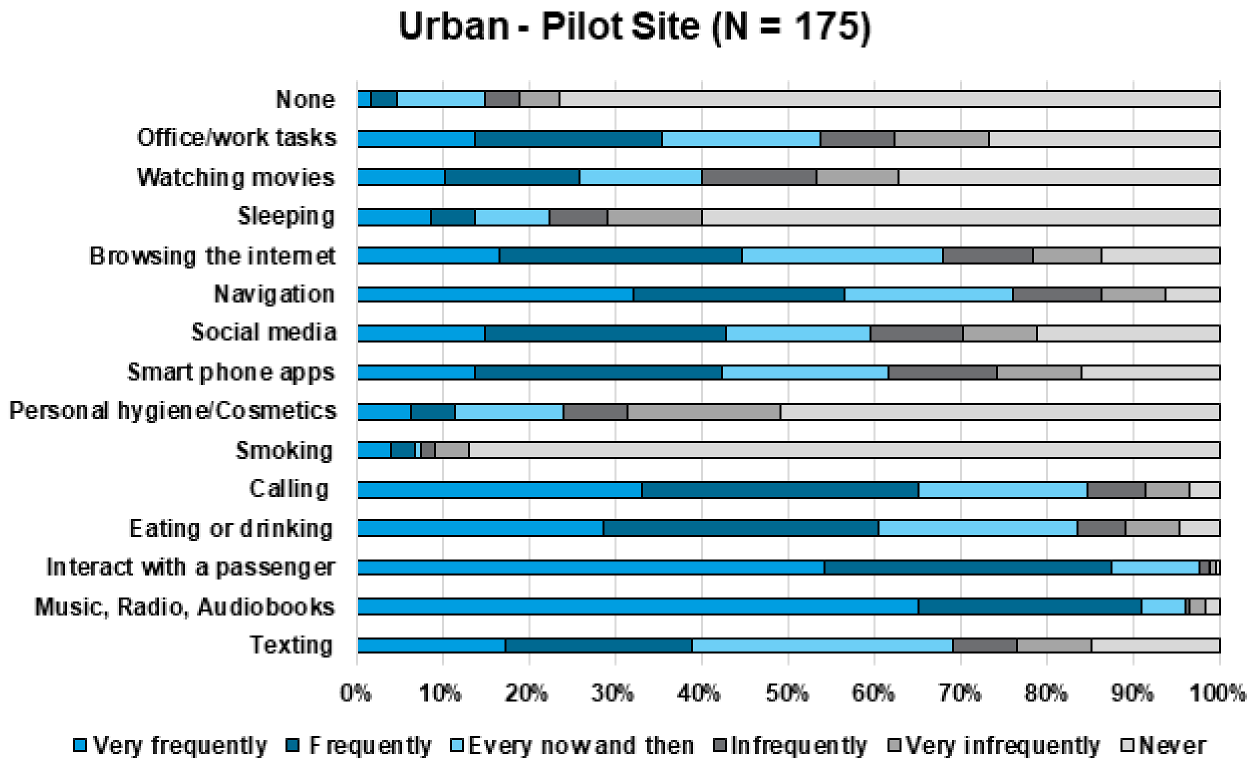


Figure 4.24: Ratings of drivers' willingness to engage in different types of secondary task while using the ADF.

4.2.7 RQU7 – What Is the Impact of ADF Use on Motion Sickness?

Table 4.17: Questions administered to understand the impact of ADF use on motion sickness.

Questions Administered
• Did you experience motion sickness during your test drive with the function active?

Of the 175 participants, 87% reported 'No' motion sickness during the test drive with the function active, with 3% responding 'Yes'. Thus, it appears that the urban ADF did not cause any feelings of illness for a majority of participants.

4.2.8 Applications of regressions models to urban user acceptance

Similar to the Motorway ADF data, Hierarchical Regression Models were conducted to answer two additional RQs:

(1) Which of the User & Acceptance factors predict willingness to use the urban ADF system, and

(2) Whether system familiarity predicts willingness to use the system.

In the following models, we have included data from 144 participants in the Pilot sites studies. As with the previous analysis, the hierarchical regression models enable us to show if variables of interest (i.e., system familiarity) explain a statistically significant amount of variance in the Dependent Variable (DV) (i.e., willingness to use) after accounting for all other variables (i.e., User & Acceptance factors).

Step 1 Independent Variables (see Table 4.18): Due to the lesser number of data items collected, two main factors were grouped in the same way as the Motorway ADF analysis, and Van Der Laan's Usefulness and Satisfying scale were added as the third and fourth factors.

Table 4.18: Items grouped by Factor Analysis and their respective Cronbach's Alpha.

FACTOR 1: Cronbach's Alpha 0.887	Workload/Emotion & Expectation
U33_33m	The system acted appropriately in all situations
U33_33k	The system worked as it should work
U33_33bb	Sometimes the system behaved unexpectedly
U33_33q	Driving with the system active was comfortable
U33_33o	I trust the system to drive
U33_33hh	Driving with this system was difficult
U33_33ii	Driving with this system was demanding
U33_33jj	Driving with the system was stressful
U33_33c	I felt safe when driving with the system active

FACTOR 2: Cronbach's Alpha 0.485	System Monitoring
U33_33ll	I would want to monitor the system's performance
U33_33n	I would use the time the system was active to do other activities
FACTOR 3: Cronbach's Alpha 0.78	**Van Der Laan's Usefulness**
TJM31_SQ001	Useful-Useless
TJM31_SQ003	Bad-Good
TJM31_SQ005	Effective-Superfluous
TJM31_SQ007	Assisting-Worthless
TJM31_SQ009	Raising alertness-Sleep-inducing
FACTOR 4: Cronbach's Alpha 0.839	**Van Der Laan's Satisfying**
TJM31_SQ002	Pleasant-Unpleasant
TJM31_SQ004	Nice-Annoying
TJM31_SQ006	Irritating-Likeable
TJM31_SQ008	Undesirable-Desirable

Step 2 Independent Variables: In the second step we included System Familiarity as an additional item to understand whether knowledge of the system had an impact on willingness to use, above that of the other system experience variables. We asked drivers 'Today, you will be operating with the motorway system. How familiar are you with this type of systems you will be using today? 1 = Highly familiar; 5 = Highly unfamiliar; 6 = Don't know (removed from analysis). Figure 4.25 shows the system familiarity of the urban participants.

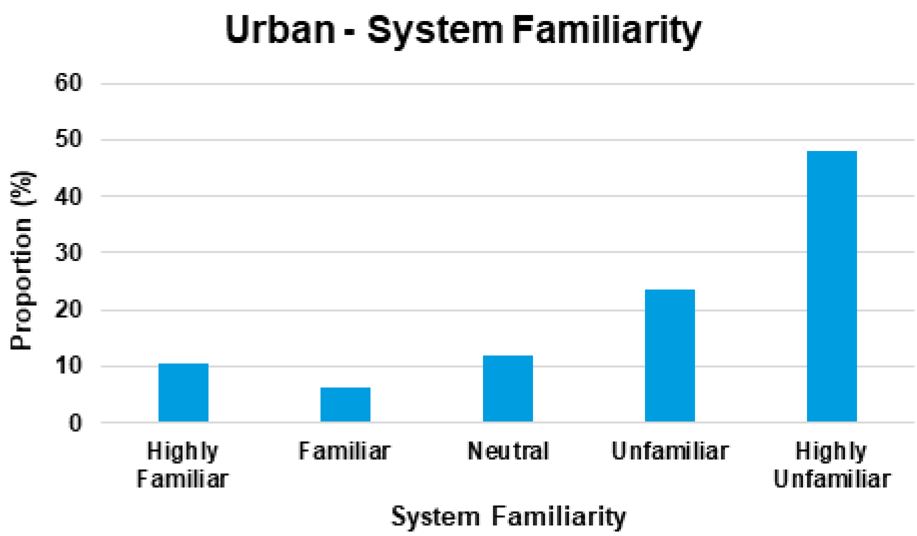


Figure 4.25: Participants' familiarity of the system.

Prior to running the regression analysis, a Spearman correlation was conducted for all six factors to check for multicollinearity. There was no strong correlation between any of the factors (Coefficient < .07), apart from Van Der Laan's Usefulness & Satisfying scale (Coefficient = 0.819), both of which tap into the underlying construct of 'Acceptance'.

The dependent variable, **Willingness to Use Scale** (Cronbach's alpha = 0.909), was calculated as the mean of the following three items:

- I would use this system if it was in my car
- I would buy the system
- I would use the system during my everyday trips

Results

- Model 1: Willingness to Use = Six User & Acceptance Factors (R^2 = 0.604)
- Model 2: Willingness to Use = Six User & Acceptance Factors + System Familiarity (R^2 = 0.625)

The results of the regression show that the R^2 values for Model 1 and Model 2 were 0.604 and 0.625, with Model 2 explaining significantly more variance than model 1. Table 4.19 below shows the coefficients and p values of each factor in each model.

*Table 4.19: Coefficients and p value of each factor in Regression Model 1 and 2 (*p < .05, ** p < .01, ***p < .001).*

Predictor Variables	Model 1	Model 2
Workload/Emotion & Expectation	0.147	0.175*
System Monitoring	0.337***	0.306***
Van Der Laan's Usefulness	0.344***	0.341***
Van Der Laan's Satisfying	0.148	0.163
System Familiarity	NA	0.149**
R^2	0.604	0.625
R^2 change	0.604	0.021*

The results revealed that Workload/Emotion & Expectation, System Monitoring and Van Der Laan's Usefulness were significant predictors of willingness to use the urban system. The more positively participants felt that the system impacted their workload/emotion and expectation, the higher their willingness was to use the system. Drivers who would like to engage in secondary tasks and reported that they were less likely to monitor the system also revealed a higher willingness to use. Finally, the higher the usefulness rating, the higher their willingness to use. In model 2, system familiarity was also a significant predictor of willingness to use, increasing the variance explained by 2.1%.

4.2.9 Summary of Findings for Urban ADF

To evaluate our L3Pilot Urban system, participants (mainly passengers) were tested at different Pilot sites. Similar to the findings for the Motorway ADF, participants were generally positive in their ratings of the system (i.e., willingness to use, perceived safety, usefulness, and trust), but with some margin for improvement for system performance and system expectation. Although participants seem to be quite positive in general, the ratings were not as high as the evaluation of the Motorway system. Most of the participants did not find the system demanding, difficult or stressful to use, but there seems to be individual variance on how tiring participants found the system to be.

Most Urban participants claimed that they monitored the urban system and surrounding environment more than in manual driving. However, this could potentially be due to participants being instructed to monitor their surroundings as they would if they were driving themselves. Slightly more than half of the participants would engage in a secondary task, which was less than the ordinary drivers on the motorway but more than professional drivers on the motorway.

Most of the participants also felt comfortable with the system and reported almost no motion sickness. However, there were a few behaviours that the system could improve on to maximise comfort, such as the distance kept to pedestrians and cyclists when overtaking, and turning behaviour at intersections and curves. Finally, regression results show that unlike the motorway evaluation, system familiarity predicted participants' willingness to use the Urban system. The more familiar they were with the system, the higher was their willingness to use it. The other factors predicting willingness to use the urban system include system monitoring and Van Der Laan's Usefulness scale, whereby higher ratings of system usefulness, and increased willingness to engage in a secondary task, led to increased willingness to use the urban system.

However, most of the participants of the Pilot site studies of the Urban ADF evaluation were passengers. This could again overestimate the positive evaluation of the system. Therefore, the findings should be used as a guide to provide the overall impressions of the system. More future studies should be conducted to investigate users' experience and their evaluation of a more mature system that allows them to be the user seated in the driver's seat. Users' subjective experiences and evaluation of the system are important because they predict willingness to use and general acceptance.

4.3 Parking

For parking, it was not possible to merge the questionnaire data on the level of single responses. Therefore, merging was done on the level of studies on parking ADF. For every study and every analysed questionnaire item, information on the proportion of drivers agreeing or disagreeing with the statement was merged. The combined results are based on answers collected in three different studies on parking. The graphs below show the proportion of drivers agreeing and disagreeing, with every study adding one data point.

4.3.1 RQ-U1: Are Drivers Willing to Use a Parking ADF?

Across all studies, most of drivers agreed or strongly agreed with the statement that they would be willing to use the ADF.

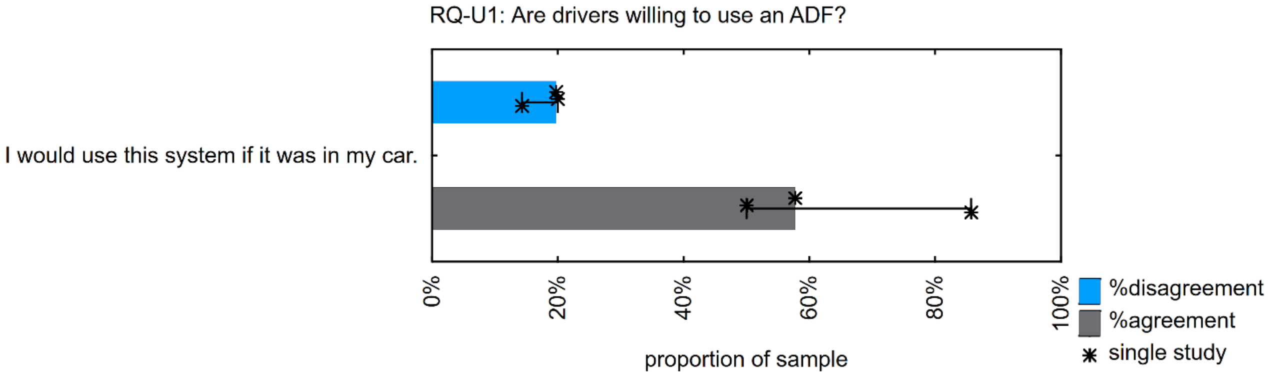


Figure 4.26: Proportion of drivers agreeing and disagreeing with the questionnaire items for RQ-U1.

4.3.2 RQ-U3: What Is the User Acceptance of the ADF?

Across studies, most of drivers stated that they felt safe parking with the parking ADF, trusted the ADF to park, and that they believed the system to be useful. In Figure 4.27 this corresponds to a large proportion of drivers agreeing with the three statements and only a small proportion disagreeing.

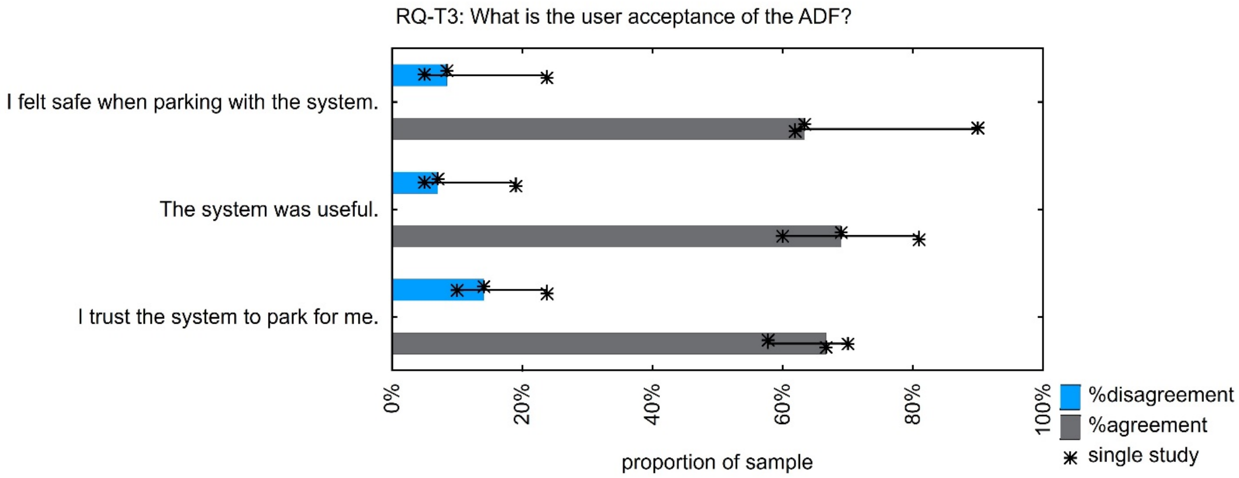


Figure 4.27: Proportion of drivers agreeing and disagreeing with the questionnaire items for RQ-U3.

4.3.3 RQ-U5: What Is the Impact of ADF on Driver State?

In all studies, drivers reported that parking with the ADF was not demanding. Contrary to that, in two studies drivers stated that parking with the ADF was not stressful, while in the third study using the ADF was perceived by most drivers as stressful. Therefore, stress results vary across studies and functions.

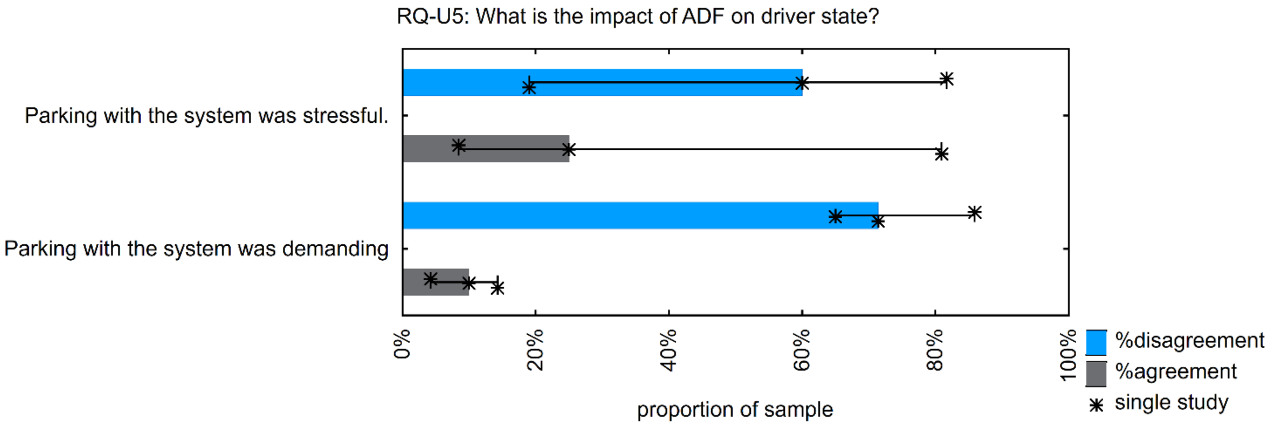


Figure 4.28: Proportion of drivers agreeing and disagreeing with the questionnaire items for RQ-U5.

4.3.4 RQ-U6: What Is the Impact of ADF Use on Driver Awareness?

There were differences between the studies in terms of drivers' perception of how parking with the ADF influenced their awareness of the environment. In one study, a majority of drivers stated that they were more aware of their environment when parking with the ADF. However, in the other studies, drivers tended to disagree with that statement.

It should be considered that the different studies might have used totally different ADFs

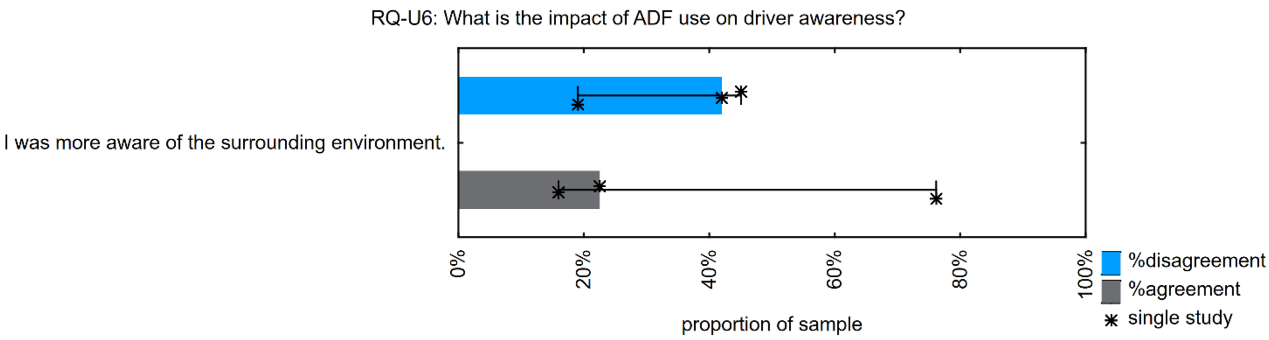


Figure 4.29: Proportion of drivers agreeing and disagreeing with the questionnaire items for RQ-U6.

5 Summary

This deliverable presents the evaluation of the data collected during the Pilots within the European Research project L3Pilot. Two areas are considered, related to Technical and Traffic aspects and User and Acceptance evaluation. Piloting activities were performed for the three types of operative domains: motorway (including traffic jam), urban, and parking. For the RQs formulated for each of these use cases, scenarios and PIs were defined in order to confirm or deny the formulated hypotheses.

Accompanying the evaluation of the Technical & Traffic assessment, which deals with the technical capabilities of the systems and the effects on the behaviour of the piloted vehicle in the traffic context, user related assessments were performed using questionnaires. These questionnaires were handed to the participants of the Pilot studies to capture their expectations and experiences with the Pilot vehicles.

This chapter summarises the main findings detailed in the previous chapters.

5.1 Motorway

For the motorway and traffic jam ADF, the following points are noteworthy (comparisons of ADF vs. manual driving):

- While driving with the ADF, speed is significantly reduced across scenarios.
- While driving with the ADF, the distance kept to the lead vehicle is significantly increased.
- While driving with the ADF, lane keeping is significantly more stable.
- Driving with the ADF leads to a reduction of lane changes of the ADF vehicle and approaching scenarios. More driving time is spent in stable scenarios like car following.
- Drivers who experienced motorway ADF were generally positive about the piloted system, including the take-over experience, but with opportunity for improvement on system performance to match users' expected behaviour, as this would increase willingness to use the system.
- No motion sickness was reported by the studies' participants and drivers agreed that the system was comfortable, especially behaviours in situations where the vehicle was less interactive with other traffic participants.
- The opportunity to engage in secondary tasks and the usefulness of the system increase the willingness to use.
- Professional drivers were revealed to be less positive towards the motorway ADF than ordinary drivers. However, system familiarity and driver type did not affect willingness to use the system.

5.2 Urban

For the urban ADF, the following points are noteworthy:

- Systems are at a lower readiness level compared to motorway systems, so the results need to be considered carefully.
- Urban environments and systems are more complex compared to motorways. This makes it more challenging to evaluate systems and potential effects.
- Due to the complexity of the urban use case requiring a large amount of driving data to cover all facets of urban driving, a clear effect of urban ADFs cannot be stated.
- Overall, ADFs tend to spend more time within intersections given their more cautious approach to conflicts (*not necessarily negative*).
- Results presented in this deliverable are based on bootstrapped data to balance the amounts of data at the different Pilot sites while preserving confidentiality (see Annex 4). This must be considered when viewing and analysing the results presented here.
- Participants were generally positive towards the urban system, but not as positive as with the motorway system. There was opportunity for improvement on system performance to match users' expected behaviour, as this would increase willingness to use the system.
- No motion sickness was reported by the studies' participants; subjects agreed that the system was comfortable, especially behaviours in situations where the vehicle was less interactive with other traffic participants.
- The opportunity to engage in secondary tasks and the usefulness of the system increase the willingness to use.
- Unlike the motorway system, the users familiar with the urban system were willing to use it; moreover, the more familiar they were, the higher their willingness to use.

5.3 Parking

For the parking ADF, the following points are noteworthy:

- Parking with an ADF takes longer and involves more stops.
- While parking with an ADF, speed is lower than during manual driving.
- Drivers state that they are in favour of parking ADFs. They would use the system and considered it to be safe and useful.

5.4 Conclusion

When comparing the different systems (motorway, urban, parking), only a few overall conclusions can be drawn. One point that is common to all is the lower speed driven with the ADFs activated in comparison with baseline driving. Additionally, for urban and motorway cases, driving is more stable, i.e., scenarios such as following a lead vehicle are longer and lane keeping performance is increased. Overall, participants were positive towards the functions, with variations among the functions. In general, participants felt comfortable while experiencing the ADFs. However, for all systems, drivers were also aware of or noticed shortcomings of the tested prototype functions or expected different behaviour of the functions in some cases. However, it can be noted that even with the shortcomings of these, somewhat prototype functions, no motion sickness was encountered by the participants.

Going further into detail, the conclusions for the functions differ, as do the use cases. Therefore, no further overall conclusions can be drawn.

5.5 Discussion, Recommendations, and Implications

The vehicles used in the pilots were equipped with prototype human-machine interfaces and control systems enabling automated driving. These systems are still under development, and their maturity inevitably varied between and within the Pilot sites, should any updates have been required during the prolonged testing schedule at some Pilot sites. The use of “imperfect” prototypes and any unexpected behaviour of the systems may have resulted in unpleasant driving or interaction experiences for users, which may have influenced users’ experience, and thus acceptance of the system, since a development system that is prone to errors is likely to elicit different acceptance ratings, compared to a market-ready system. Therefore, these factors should be borne in mind when considering the results.

The following statements and recommendations can be drawn from the data collected in the Pilots:

- Potential for urban ADF has been shown. However, it will benefit with more and diverse data to be able to show results.
- Pilots were completed either with the presence of safety drivers monitoring the participants’ performance or the vehicles were completely piloted by safety drivers. In either case the role of the safety driver was to deal with any unusual situation. Therefore, participants could potentially have a higher sense of safety as well as a sense of “easiness” which leads to an “overtrust” in the system. Furthermore, the systems would always be overridden if faced with critical situations, which made the evaluation of ADF in such situations impossible.
- We need to be very cautious while comparing systems, driver types, and test and study types. This was mainly due to the nature of the different systems and data being collected via different pilots that consisted of different study designs.

- The reported findings are designed to be used as a guide and to provide overall impressions. More research is needed to understand the implications of these findings for driver interactions with mature systems.
- Users' subjective experiences and their evaluation of the system predict willingness to use and, therefore, it is important to ensure acceptance for those systems and that users have positive experiences in future evaluations.
- Users' ability to engage in secondary tasks was a significant predictor of willingness to use. The most widely reported secondary tasks that users were likely to engage in while the system was active included listening to music, radio, an audiobook, navigation, interacting with a passenger, using smartphone apps, and texting. Therefore, it is important that manufacturers consider these preferences when designing their user interfaces, to make it as comfortable and easy as possible to engage in these tasks.
- User evaluations suggest that there are some system performance elements that could be improved further. These include behaviour at motorway junction areas and lane changing behaviour, the distance kept to obstacles and road markings, the smoothness of driving in the city, turning behaviour in intersections and curves, behaviour when approaching pedestrians at intersections, and the distance kept to pedestrians and cyclists when overtaking in urban environments. The majority of these are situations where interaction with other road users is more likely to occur, and therefore the safety implications may be considered higher.
- In the future, more testing should be conducted, especially to extend the systems' capabilities between ODDs, as well as to investigate users' ability to take-over control in critical situations when the system becomes more mature.
- Studies should be conducted considering long-term exposure to automated driving once the maturity and legal framework for piloting the systems allows such studies. Until then, simulation-based studies should be used complementarily.
- It should also be considered that the SAE L3 ADF are leading to vehicles with much better environment perception (sensors) which can be used besides the L3 ADF for improvement of active safety systems and extend the use of these active safety features way beyond the ODD of the L3 ADF. Therefore, the safety benefit of L3 ADF can be improved significantly at very low additional cost.

6 Lessons Learnt

This chapter contains a number of lessons learned which became obvious to evaluation partners during the various analyses, as well as a resulting recommendation for future research activities involving on-road piloting of automated vehicles. A more detailed view of the lessons learned with regard to data management can be found in Deliverable D5.2 – Guidelines and Lessons Learned (Koskinen et al., 2021).

6.1 Experimental Setup

The FESTA methodology provides a useful guideline which can be applied to piloting studies concerned with functions not yet introduced to the market. The PREPARE phase, and especially the definition of the experimental setup, is of great importance for the relevance of the later design and evaluation. Important considerations to be taken into account are:

- To be able to make a useful interpretation of the data, the prototypes should be described in the greatest possible detail in terms of sensor setup, algorithms, parameters, HMI design, etc., while remembering that the studies are executed with prototypes and not systems intended for production.
- Testing of prototype functions by preference evaluates stable driving behaviour. In contrast, the evaluation of rare or critical events is quite challenging.
- Due to the high complexity of urban environments, the risk of having non-comparable Pilot sites is higher than for motorways. This requires coordination of experimental setups among urban Pilot sites and an extensive data collection to cover all relevant scenarios.
- The lack of regulatory alignment across European countries, such as differences in the permissions granted for the tests, affects the possibilities to achieve consistent results on user testing across Pilot sites. For example, there may be variations across Pilot sites in terms of the types of drivers who are allowed to participate, which raises challenges for comparisons across the project. Harmonising these results across European countries would go a long way towards creating a more productive environment for large-scale testing of automated driving. This is even more the case for urban environments, where the infrastructural differences alone have a much higher influence on the results generated at each Pilot site.
- Pilot sites may vary in terms of participants' specific or unrestricted roles and permitted activities when operating prototype vehicles. For example, since drivers at particular sites were not permitted to engage in non-driving related tasks during the pilot, this may have affected our ability to answer some questions at project level.

6.2 Data Collection and Pre-processing

Apart from the experimental setup, the setup of the data collection has a big influence on what conclusion can be drawn from the data and which efforts are required for a harmonised assessment.

- An early and consistent agreement of data that is to be shared is important and should be adhered to by all partners. Adding data at a later stage often proved to be a bottleneck in the further processing and handling of data as well as in the generation of results.
- When conducting studies across multiple sites, it is essential that any cross-Pilot methods are administered using the same tools and protocols. For example, questionnaires were administered across all Pilot sites using an online tool, which minimises variance, not only between Pilot sites, but also between experimenters. This approach also ensured that the data output could be integrated seamlessly into the common data format and transferred to a consortium-wide CDB.

6.3 Data Analysis

Apart from the pre-processing of data and quality checks, the methods for evaluation data should ensure a harmonised evaluation of the data from the different Pilot sites.

- Usage of a common data format allowed for the creation of one overall toolchain for the analysis of data and generation of results. While it does not solve all the problems that can occur when dealing with data from multiple sources, sensor set-ups, etc., a common format makes the development and integration of the tools substantially easier and allows the easy sharing of tools among evaluation partners.
- A toolchain including data manipulation, computation of PIs, conversion to the common data format, and implementation of the data processing toolchain under the given constrains of treating all data without making single Pilot sites identifiable, was time consuming and sometimes needed repeated revisions of the complete process.
  - If possible, most quality checks of the processed data should already be handled before the upload to the pseudonymised database. This makes it easier to match unexpected values in the output data (PIs) to errors in scripts used or wrong input data.
  - Bugs that become apparent in the merged, pseudonymised data are much more resource consuming, as it is hard to investigate whether all Pilot sites are affected or just a single one.
  - Identified issues, especially those concerning scenario detection, require a complete reprocessing and reupload by the data processing partners. An organised approach to versioning the toolchain and the uploaded data needs to be assured.
- Automatic data annotation and inspection is still a challenge for naturalistic driving data. When verifying data quality and data sequences (e.g., scenario detection by scripts), the role of birds' eye GUIs for delivering object-level information – e.g., tracked objects, lanes, etc., synchronized with a modern video viewer capable of scene pausing/rewind/forward – is significant for the role of the data analysis partner who typically performs the analysis in a mixed manner of manual and automatic inspection.

- Although anonymisation of data origin in the CDB makes data sharing in such a large consortium possible and keeps confidentiality, it also makes error tracing a complex and time-consuming task, in most cases involving all Pilot data processing partners.
- When analysing urban data, much focus has to be put on closely and precisely defining the ODD and the overall capabilities of the function, as small differences in these can lead to big changes in the effects of the functions.
- The variance in the urban environment across different European countries may lead to big differences between different Pilot site setups. It is therefore of utmost importance to clearly define infrastructure elements and the influence they might have on systems. A wider range of urban environments (e.g., spanning more cities, countries, different urban types) could allow for a more encompassing analysis of urban functions.

References

Andreone, L., Borodani, P., Pallaro, N., Tango, F., Bellotti, F., Weber, H., Altpeter, B., Reimer, F., Griffon, T., Sauvaget, J. L., Geonimi, S., Page, Y., Geurineau, T., Willemotte, G., Mäkinen, T. (2021). Deliverable *D6.5 - Pilot Reporting Outcomes,* Deliverable D6.5 of L3Pilot project

Baergman, J., van Nes, N., Christoph, M., Janssen, R., Heijne, V., Carsten, O., Dotzauer, M., Utech, F., Svanberg, E., Cocron, M.P., Forcolin, F., Kovaceva, J., Guyonvarch, L., Hibberd, D., Lotan, T., Winkelbauer, M., Sagberg, F., Stemmler, E., Gellerman, H., Val, C., Quintero, K., Tattegrain, H., Donabauer, M., Pommer, A., Neumann, I., Albert, G., Welsh, R., Fox, C. (2017). *The UDrive dataset and key analysis results. Deliverable 41.1.* Available at: https://erticonetwork.com/wp-content/uploads/2017/12/UDRIVE-D41.1-UDrive-dataset-and-key-analysis-results-with-annotation-codebook.pdf

Bellotti, F., Berta, B., De Gloria, A., Kobeissi, A., Osman, N., Arnold, E., Mozaffari, S., Dianati, M. (2019). *Deliverable D6.2 - Database for Data Collection: Evaluation Format & Common Data Set for Future Researc,.* Deliverable D6.2 of L3Pilot project

Bellotti, F., Berta, R., De Gloria, A., Osman, N. (2020). *Deliverable D6.4 - Data Delivery to Evaluation & Common Data Set for Future Research.*

Benmimoun, M., Fahrenkrog, F., Zlocki, A., & Eckstein, L. (2011, June). *Incident detection based on vehicle CAN-data within the large scale field operational test "euroFOT".* In 22nd Enhanced Safety of Vehicles Conference (ESV 2011), Washington, DC/USA.

Bjorvatn, A., Page, Y., Fahrenkrog, F., Weber, H., Aittoniemi, E., Heum, P., Lehtonen, E., Silla, A., Bärgman, J., Borrack, M., Innamaa, S., Itkonen, T., Malin, F., Pedersen, K., Schuldes, M., Sintonen, H., Streubel, T., Hagleitner, W., Hermitte, T., Hiller, J., Torrao, G., (2021). *Deliverable D7.4 - Impact Evaluation Results,* Deliverable D7.4 of L3Pilot project

Brizzolara, D., Fischer, F., Koskinen, H., Koskinen, S., Heinig, I., Follin, P., Barbier, C., Ojeda, L., Visintainer, F., Fernandez, J., Rämä, P., Rosé, H. (2014). *Report on FOT Operations, DRIVE C2X Deliverable D11.3.* European Commission.

Dingus, T. A., Hankey, J. M., Antin, J. F., Lee, S. E., Eichelberger, L., Stulce, K. E., McGraw, D., Perez, M., Stowe, L. (2015). *Naturalistic driving study: Technical coordination and quality control* (No. SHRP 2 Report S2-S06-RW-1).

Dozza, M., Bärgman, J. & Lee, J. (2013). Chunking: A procedure to improve naturalistic data analysis. Accident Analysis & Prevention. 58. 309–317. 10.1016/j.aap.2012.03.020.

Efron, B., Tibshirani, R. (1986). *Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy.* Statistical Science, 1(1). Available at: https://doi.org/10.1214/ss/1177013815

FOT-Net & CARTRE (2018). *FESTA Handbook. Version 7*. December 2018. Available at: https://connectedautomateddriving.eu/wp-content/uploads/2019/01/FESTA-Handbook-Version-7.pdf (accessed 09 August 2021)

Gettman, D., Head, L. (2003). *Surrogate safety measures from traffic simulation models.* Transportation Research Record, 1840(1), 104-115

Hibberd, D., Louw, T., Aittoniemi, E., Brouwer, R., Dotzauer, M., Fahrenkrog, F., Innamaa, S., Kuisma, S., Merat, N., Metz, B., Nelia, N., Penttinen, M., Guillen, P.P., Rösener, C., Silla, A., Streubel, T., Tango, F., van den Boom, B., Weber, H., Woerle, J., Zerbe, A. (2018). *Deliverable D3.1 - From Research Questions to Logging Requirements,* Deliverable D3.1 of L3Pilot project

Hiller, J., Svanberg, E., Koskinen, S., Bellotti, F., Osman, N. (2019). The L3Pilot Common Data Format—Enabling efficient automated driving data analysis. In *Proceedings of the 26th International Technical Conference on the Enhanced Safety of Vehicles, Eindhoven, The Netherlands* (pp. 10-13)

Hiller, J., Koskinen, S., Berta, R., Osman, N., Nagy, B., Bellotti, F., Rahman, A., Svanberg, E., Weber, H., Arnold, E. H., Dianati, M., De Gloria, A. (2020) '*The L3Pilot Data Management Toolchain for a Level 3 Vehicle Automation Pilot*', Electronics. MDPI AG, 9(5), p. 809. doi: 10.3390/electronics9050809

Innamaa, S., Aittoniemi, E., Bjorvatn, A., Fahrenkrog, F., Gwehenberger, J., Lehtonen, E., Louw, T., Malin, F., Penttinen, M., Schindhelm, R., Silla, A., Weber, H., Borrack, M., Di Lillo, L., Merat, N., Metz, B., Page, Y., Shi, E., Sintonen, H. (2020). *Deliverable D3.4 - Evaluation Plan,* Deliverable D3.4 of L3Pilot project

Jamson, S., Kappe, B., Louw, T. (2014). *Performance Indicators and acceptance analysis plan. ecoDriver Project Deliverable 42.1.* European Commission.

Koskinen, S., Christen, F., Hiller, J., Neila, N., Svanberg, E., Kremer, M. (2021). *Deliverable D5.2 - Guidelines and Lessons Learned on Pilot Tools and Data,* Deliverable D5.2 of L3Pilot project

Kyriakidis, M., Happee, R., de Winter, J. C. (2015). *Public opinion on automated driving: Results of an international questionnaire among 5000 respondents*. Transportation Research Part F: Traffic Psychology and Behaviour, 32, 127-140.

Lai, F., Ströbitzer, E., De Goede, M., Krishnakumar, R., Val, C., Mahmod, M., Malasek, J., Martin, O., Welsh, R., Carsten, O. (2013). *Operation Sites Description and Planning. UDRIVE Deliverable 31.1.* EU FP7 Project UDRIVE Consortium. https://doi.org/10.26323/UDRIVE_D31.1

Louw, T., Merat, N., Metz, B., Wörle, J., Torrao, G., Satu, I. (2019, September). *Assessing user behaviour and acceptance in real-world automated driving: The L3Pilot project approach*. In Proceedings of 8th Transport Research Arena, Helsinki, Finland, 27–30 April 2020.

Madigan, R., Louw, T., Merat, N. (2018). *The effect of varying levels of vehicle automation on drivers' lane changing behaviour*. PloS one, 13(2), e0192190

Mahmud, S. S., Ferreira, L., Hoque, M. S., Tavassoli, A. (2017). *Application of proximal surrogate indicators for safety evaluation: A review of recent developments and research needs*. IATSS research, 41(4), 153-163.

Metz, B., Rösener, C., Louw, T., Aittoniemi, E., Bjorvatn, A., Wörle, J., Weber, H., Torrao, G., Silla, A., Innamaa, S., Malin, F., Lehtonen, E., Fahrenkrog, F., Heum, P., Pedersen, K., Merat, N., Nordhoff, S., Beuster, A., Dotzauer, M., Streubel, T. (2019). *Deliverable D3.3 - Evaluation Methods*, Deliverable D3.3 of L3Pilot project

Metz, B., Wörle, J., Zerbe, A., Schindhelm, R., Frey, A., Bonarens, F., Louw, T., Lee, Y.M., Madigan, R., Merat, N., Pipkorn, L., Streubel, T., Tivesten, E., (2021). *Deliverable D7.2 - L3/L4 Long-Term Study About User Experiences*, Deliverable D3.3 of L3Pilot project

Nagy, B., Hiller, J., Svanberg, E., Kremer, M., Luxen, M., Christen, F., Koskinen, S., Bellotti, F., Osman, N. (2018). *Deliverable D5.1 - Pilot Tools for L3Pilot,* Deliverable D5.1 of L3Pilot project

Neukum, A., Lübbeke, T., Krüger, H. P., Mayser, C., Steinle, J. (2008). *ACC-Stop&Go: Fahrerverhalten an funktionalen Systemgrenzen*. In 5. Workshop Fahrerassistenzsysteme-FAS (pp. 141-150). Karlsruhe.

Nordhoff, S., Beuster, A., Kessel, T., Bjorvatn, A., Innamaa, S., Lehtonen, E., Malin, F., Madigan, R., Lee, Y. M., Merat, N., Louw, T. (2021). *Deliverable D7.1 - Annual Quantitative Survey About User Acceptance towards ADAS and Vehicle Automation.*

Penttinen, M., Rämö, P., Dotzauer, M., Hibberd, D., Innamaa, S., Louw, T., Streubel, T., Metz, B., Woerle, J., Brouwer, R., Rösener, C., Weber, H. (2019). *Deliverable D3.2 - Experimental Procedure,* Deliverable D3.2 of L3Pilot project

Pipkorn, L., Dozza, M., Tivesten, E. (2021). *Drivers glance behaviour when resuming manual control from conditional automation on public road*. Manuscript submitted.

Saurabh V. (2018). CADaS Glossary v.3.7. Available at: http://ec.europa.eu/transport/road_safety/sites/roadsafety/files/cadas_glossary_v_3_7.pdf

van Nes, N., Bärgman, J., Christoph, M., van Schagen, I. (2019). *The potential of naturalistic driving for in-depth understanding of driver behavior: UDRIVE results and beyond*. Safety Science, *119*, 11-20.

List of abbreviations and acronyms

Abbreviation	Meaning
AD	Automated driving
ADAS	Advanced Driver Assistance Systems
ADF	Automated Driving Function
AIM	Application Platform for Intelligent Mobility
API	Application programming interface
AV	Automated Vehicles
BL	Baseline
CDB	Consolidated database
CoP	Code of Practice
DM	Derived Measure
GUI	Graphical user interface
NDRT	Non-driving related task
PI	Performance Indicator
RQ	Research questions
SD	Standard deviation
THW	Time Headway
TOR	Take-over request
TTC	Time-to-Collision
VRU	Vulnerable Road User

Annex 1 Research questions

Table A1.1: Research questions for motorway ADF and evaluated performance indicators

ID	Level 2 RQ	Feasibility	Analysed instances	Analysed Performance indicators
RQ-T1	How reliable is system performance in a given driving and traffic scenario?	Only descriptive analysis	Trip	%time ADF available, duration ADF active
RQ-T2	How often do take-over requests occur?	Only descriptive analysis	Trip	N(TOR)/h
RQ-T3	Does the function initiate a take-over request if required by the boundaries of the ADF?	No	*From the logged data of the system, it could not be evaluated, whether the system reached the ODD boundary to validate whether a TOR was issued*	
RQ-T4	Are there any traffic violations while using the ADF?	No	*No traffic violations could be shared apart from difference to speed limit which is reported in RQ-T8*	
RQ-T5	How do take-over requests affect driving?	Only for selected Pilot sites	Take-over situation	TOC-rating *reported in sections 4.1.7*
RQ-T6	What is the impact of ADF on vehicle dynamics?	Yes	Uninfluenced driving, following, traffic jam, approaching lead vehicle, approaching traffic jam, lane change, cut-in	min(ax), max(ax), sd(ax), max(abs(ay)), sd(ay)
RQ-T7	What is the impact of ADF on the accuracy of driving?	Yes	Uninfluenced driving, following, traffic jam, approaching lead vehicle, approaching traffic jam	sd(lat.Pos.), m(lat.Pos.), sd(v)
RQ-T8	What is the impact of ADF on the driven speed?	Yes	Uninfluenced driving, sd(v) also for following, traffic jam	m(v), max(v)
RQ-T9	What are the impacts of ADF on energy efficiency?	Yes	Trip	m(Energ.Cons.)/100 km
RQ-T10	What is the impact of ADF on the frequency of near-crashes / incidents?	Yes	Following, lane change, cut-in, approaching	N(distance incidents)/scenario, N(dynamic incidents)/scenario
RQ-T11	What is the impact of ADF on the frequency of certain events?	Yes	Trip	N(scenario type)/h, %time(scenario type)

ID	Level 2 RQ	Feasibility	Analysed instances	Analysed Performance indicators
RQ-T12	What is the impact of ADF on the interaction with other road users in a defined driving scenario?	Yes	Following, traffic jam, approaching lead vehicle, cut-in	m(THW), sd(THW), min(THW), min(TTC)
RQ-T14	What is the impact of ADF on the number of near-crashes / incidents with other road users?	Yes, together with RQ-T10		
RQ-T15	How does the ADF influence the behavior of subsequent vehicles?	Only for selected Pilot sites		
RQ-T16	How does the ADF influence the behavior of preceding vehicles?	Yes, together with RQ-T12	Following, approaching lead vehicle trip	diff(v_LeadVeh), sd(v_LeadVeh) N(LaneChange_LeadVeh)/h
RQ-T17	What is the impact of ADF on the number of near-crashes / incidents of other traffic participants?	Only for selected Pilot sites		

Table A1.2: Research questions for urban ADF

ID	Level 2 RQ	Feasibility	Analysed instances	Analysed Performance indicators
RQ-T1	How reliable is system performance in a given driving and traffic scenario?	Only descriptive analysis	Trip	-
RQ-T2	How often do take-over requests occur?	No	Urban experiments did not include TORs	
RQ-T3	Does the function initiate a take-over request if required by the boundaries of the ADF?	No		
RQ-T4	Are there any traffic violations while using the ADF?	No	Only difference to speed limit could be analysed (RQ-T8)	
RQ-T5	How do take-over requests affect driving?	No	Urban experiments did not include TORs	

ID	Level 2 RQ	Feasibility	Analysed instances	Analysed Performance indicators
RQ-T6	What is the impact of ADF on vehicle dynamics?	Yes	Uninfluenced driving, following, approaching lead vehicle, lane change, cut-in, intersections	min(ax), max(ax),m(ax), m(ay) sd(ax), max(abs(ay)), sd(ay)
RQ-T7	What is the impact of ADF on the accuracy of driving?	Yes	Uninfluenced driving, following, approaching lead vehicle	m(Pos. in lane), sd(Pos. in lane
RQ-T8	What is the impact of ADF on the driven speed?	Yes	Uninfluenced driving, following, approaching lead vehicle, lane change, cut-in, intersections	m(v), max(v)
RQ-T9	What are the impacts of ADF on energy efficiency?	No (data not shared)		
RQ-T10	What is the impact of ADF on the frequency of near-crashes / incidents?	No (data not shared)		
RQ-T11	What is the impact of ADF on the frequency of certain events?	Yes	Trip, uninfluenced driving, following, approaching lead vehicle, lane change, cut-in, intersections	N(scenario)/h, scenario duration, m(Dur(v<0,2km/h)), N(v<0,2km/h)/h, N(intersection scenario subtype)/N(intersection scenario type)
RQ-T12	What is the impact of ADF on the interaction with other road users in a defined driving scenario?	Yes	Uninfluenced driving, following, approaching lead vehicle, lane change, cut-in, intersections	m(THW), min(THW) [s], m(long. Dist. Lead veh.), m(v lead veh.), Duration [s], min(Distance) , min(TTCP)
RQ-T14	What is the impact of ADF on the number of near-crashes / incidents with other road users?	No (data not shared)		
RQ-T15	How does the ADF influence the behaviour of subsequent vehicles?	Yes	Uninfluenced driving, following, approaching lead vehicle, lane change, cut-in	m(THW rear veh.), min(THW rear veh.), sd(THW rear veh.), min(ax rear veh.)
RQ-T16	How does the ADF influence the behaviour of preceding vehicles?	Yes	Following, approaching lead vehicle	m(v lead veh), m(v)

ID	Level 2 RQ	Feasibility	Analysed instances	Analysed Performance indicators
RQ-T17	What is the impact of ADF on the number of near-crashes / incidents of other traffic participants?	No (data not shared)		

ID	Level 2 RQ	Feasi-bility	Analysed Performance indicators
RQ-T1	How reliable is system performance in a given driving and traffic scenario?	No	
RQ-T2	How often do take-over requests occur?	No	
RQ-T3	Does the function initiate a take-over request if required by the boundaries of the ADF?	No	
RQ-T5	How do take-over requests affect driving?	No	
RQ-T6	What is the impact of ADF on vehicle dynamics?	Yes	min(ax), max(ax), sd(ax), max(abs(ay)), sd(ay) [m/s2]
RQ-T7	What is the impact of ADF on the accuracy of driving?	Yes	m(duration), N(stops)/manuever
RQ-T8	What is the impact of ADF on the driven speed?	Yes	m(v), max(v), sd(v) [km/h]

Table A1.3: User and acceptance research questions in L3Pilot and the methods used to address them. Vehicle and video data includes TOC rating.

RQ – Level 1	RQ – Level 2	Real world pilot				Other methods		
		Vehicle-based analysis	Video-based analysis	Pilot site questionnaire	Interview/Focus group	Driving Simulator	Wizard-of-Oz	Annual Survey
What is the impact of ADF use on user acceptance & awareness?	Are drivers willing to use an ADF?	X	X	X	X	X		X
	How much are drivers willing to pay for the ADF?			X	X			X
	What is the user acceptance of the ADF?			X	X	X	X	X

RQ – Level 1	RQ – Level 2	Real world pilot				Other methods		
		Vehicle-based analysis	Video-based analysis	Pilot site questionnaire	Interview/Focus group	Driving Simulator	Wizard-of-Oz	Annual Survey
	What is the impact of ADF on driver state?		X		X	X		
	What is the impact of ADF use on driver awareness?		X	X	X	X	X	
	What are drivers' expectations regarding system features?			X	X	X		X
What is the impact of ADF use on user experience?	What is drivers' secondary task engagement during ADF use?		X	X		X	X	
	How do drivers respond when they are required to retake control?	X	X	X	X	X	X	
	How often and under which circumstances do drivers choose to activate/deactivate the ADF?	X	X		X	X		
	What is the impact of ADF use on motion sickness?			X				
	What is the impact of motion sickness on ADF use?			X				

This chapter provides details on how the L3Pilot scenarios' detection scripts (implemented in Data subproject and used and refined in Evaluation subproject) were implemented. Configuration of scenario detection was based on various parameters from the research literature together with empirical knowledge of L3Pilot experts on driving scenarios and events, which was then transformed into conditions for scenario detection.

Annex 2 Driving Scenarios

A2.1. Motorway

For the motorway ADF, a set of mutually exclusive driving scenarios (see A2.1.6 covering the most common interactions on motorways, were defined (see project deliverable D3.3 (Metz et al., 2019) for scenario definitions).

A2.1.1. Lead Object Scenarios

Lead object scenarios are dependent on detecting the lead object in the ego-vehicle lane and then assessing the distance to and speed of the lead object. Depending on the time headway (THW) (the distance to an object divided by the speed of the ego-vehicle), the following definitions are important for understanding the difference between the scenarios:

1. A *Close* object is when the THW is less than 2 seconds between the two objects.
2. An *In-between* object is when THW is between 2 s and 3.5 s.
3. A *Distant* object when THW is more than 3.5 s.
4. No lead object detected.

In addition, the following definitions are set:

1. *Speed tolerance* is set to 1.4 m/s.
2. *Minimum duration* of the scenarios is set to 2 seconds (not applicable to *Approaching a static object*).
3. *Minimum speed* of the ego-vehicle is set to 5.56 m/s (not applicable to *Approaching a lead or static vehicle*).

The definitions stated above are made to reflect actual driving but there is a risk of having long periods not qualified for any scenario, since the *minimum duration* can be quite restrictive when the scenario is switching between *following a lead object* and *approaching a lead object.* If an *approaching a lead object* scenario is between two *following a lead object* scenarios and the distance does not decrease more than 30% from the start until the end of the scenario, then the *approaching a lead vehicle* is replaced by a consecutive *following a lead object* scenario.

Variable definitions:

- THW Close: <2 s
- THW In-between: >= 2 s and < 3.5 s
- THW Distant: 3.5 s
- Speed tolerance: 1.4 m/s
- Minimum duration: 2 s
- Maximum drop out: 0 s

- Minimum speed: 5.56 m/s
- Minimum longitudinal distance decrease: 30%
- Static object speed: 1 m/s

These scenarios are then applying the *Mutually Exclusive Scenario Algorithm* (cf. A2.1.6) described below.

A2.1.1.1. Uninfluenced Driving

Uninfluenced driving is classified if no object is detected, if a detected object is *Distant,* or if a lead object is *In-between* and travelling faster than the ego-vehicle by more than the *Speed tolerance*. The ego-vehicle must travel faster than *Minimum speed* and the consecutive duration of the criteria must be longer than *Minimum duration*.

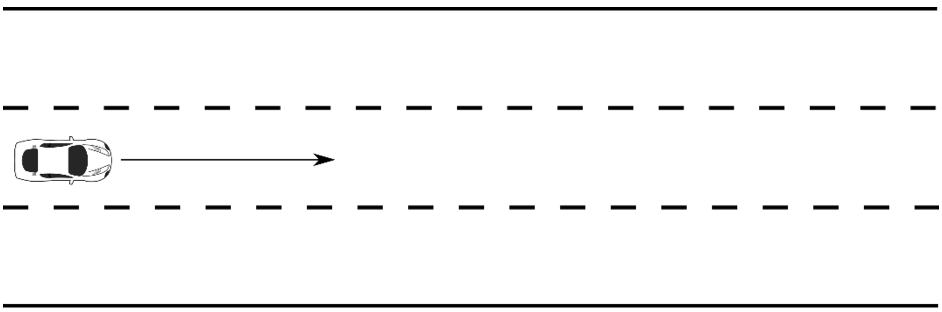
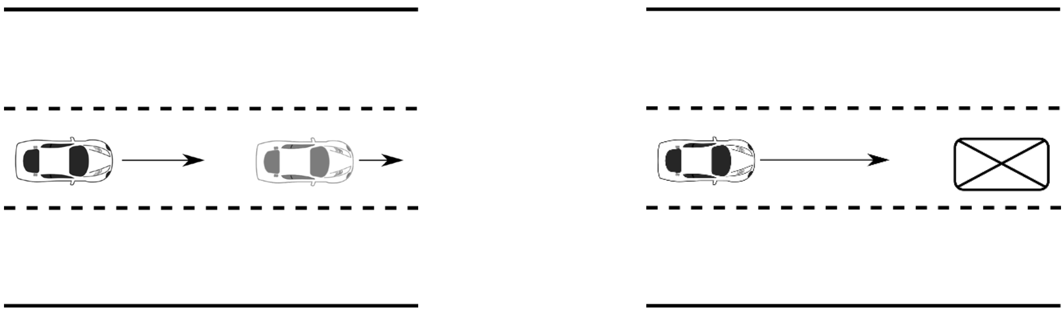


Figure A2.1: Uninfluenced driving scenario.

A2.1.1.2. Approaching a Lead or Static Object

A *Close* or *In-between* lead object is detected and the speed difference between the ego-vehicle and lead object is higher than *Speed tolerance*. There is a specific clause when a *Close* lead object has a lower speed of 1 m/s and then the scenario is classified as *approaching a static object*.



(a) (b)

Figure A2.2: Approaching a (a) lead or (b) static object scenario.

A2.1.1.3. Following a Lead Object

If a *Close* lead object is detected, then the scenario is classified as *following a lead object* as long as the speed difference does not qualify for an approaching scenario, i.e., less than *Speed* *tolerance*. If the relation between the two objects has a THW of *In-between,* then Following a lead vehicle is classified when the speed difference between the two objects are +- *Speed tolerance*.

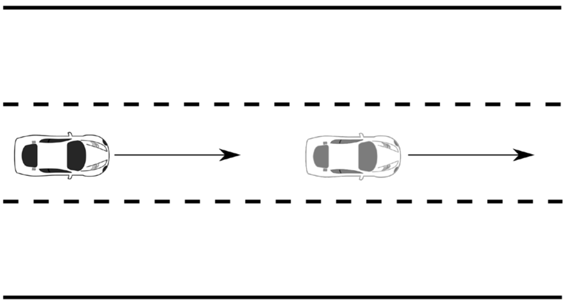


Figure A2.3: Following a lead object scenario.

A2.1.2. Approaching a Traffic Jam

The ego-vehicle is approaching a traffic jam. This scenario is defined and classified as the 20 s before the scenario *driving in a traffic jam*.

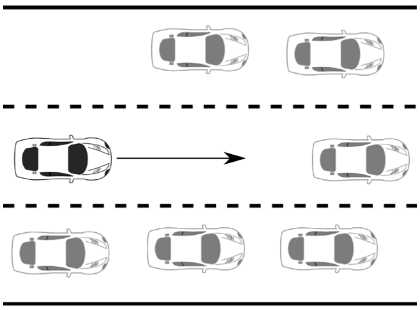


Figure A2.4: Approaching a traffic jam scenario.

A2.1.3. Driving in a traffic jam

The ego-vehicle is travelling in a traffic jam. This is determined by a speed below 60 km/h over a period of at least 180 s.

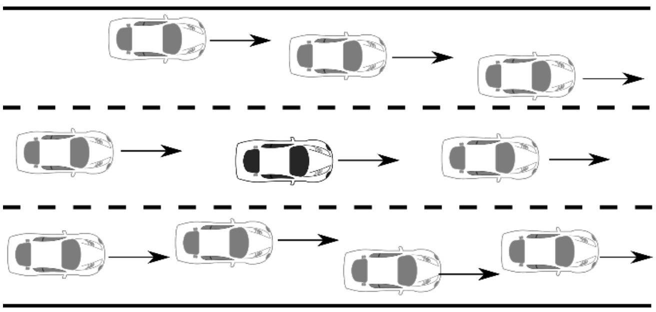


Figure A2.5: Driving in a traffic jam scenario.

A2.1.4. Lane Change

Lane changes of the ego -vehicle are derived from the lateral position of the ego-vehicle with respect to the position of the lane markings. When a left or right marking is crossed, a lane change is detected and its start- and endpoint are determined. The starting point of the lane change is the point at which the car starts moving in the direction of the lane marking before crossing the marking. The end point of the lane change is the point where the car stops moving away from the lane marking after crossing the marking. A maximum window size of 10 s before and after crossing the marking is set to limit the start- and endpoint, respectively. Left and right lane changes are coded separately.

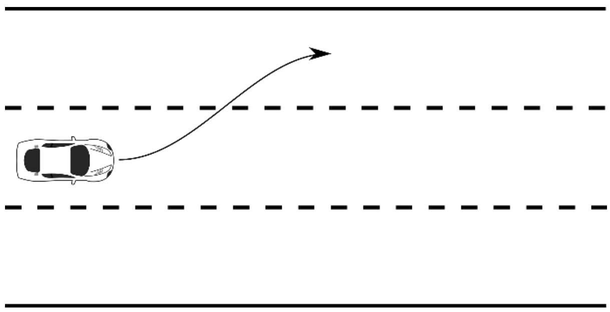


Figure A2.6: Lane change scenario.

A2.1.5. Cut-In

Cut-ins are derived from the data as follows:

For each timestamp at which a change in Lead Vehicle id is detected, the lateral displacement of the object before becoming the newLeadVehicle is checked within a time window (defined by MaxStepsBefore , MaxStepsAfter thresholds) and newLeadVehicle id is classified as cut-in from the left or the right, if all the following criteria hold:

- NewLeadVehicleDistance < ExclusionDistance , where ExclusionDistance = `THW * egoVehicle.Speed`
- NewLeadVehicleVelocity < ObjectSpeedThreshold
- NewLeadVehicleOrigin == {LeftLane, RightLane}
- Lateral displacement (m) between Cut-in start and Cut-in end timepoints > MinLateralMovement. Where:
  - Cut-in start is defined as the moment when newLeadVehicle distance from lane marking is lower than than a threshold (defined by MinDistanceToLaneMarking).
  - Cut-in is considered terminated when the newLeadVehicle is within the ego-vehicle's lane, i.e., the distance of the newLeadVehicle from the lane marking is below a given threshold (defined by the EgoLaneTolerance).
  - Time displacement (in seconds) between Cut-in start and Cut-in end timepoints > MinCutInSamples

Cut-in detection configuration parameters:

- THW – Time gap above which cut-ins are no longer considered (Default: 2 s)

- MaxStepsBefore - Maximum number of steps considered before the new lead vehicle appears (Default: 50)
- MaxStepsAfter - Maximum number of steps considered after the lead vehicle appears (Default: 50)
- MinDistanceToLaneMarkings – Minimum distance the vehicle needs to have to the lane markings in the other lane at start (Default: 1 m)
- EgoLaneTolerance – Minimum distance the lead vehicle needs to have to lane markings in the ego lane (Default: 1 m)
- MinCutInSamples – Minimum number of samples considered for the scenario (Default: 1 s)
- MinLateralMovement – Minimum lateral movement needed by the cutting in object (Default: 1 m)
- ObjectSpeedThreshold – Threshold in relation to the ego velocity from which to exclude objects as cut-ins (default: 2 m/s)

As for all the other scenarios, they also pass through the algorithm for mutually exclusive scenarios.

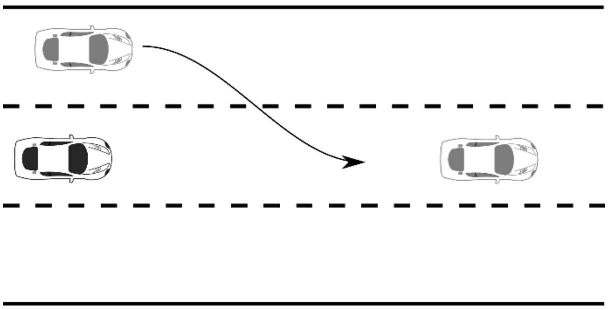


Figure A2.7: Cut-in scenario.

A2.1.6. Mutually Exclusive Scenario Algorithm

Due to signal noise and smoothing, it can occur that the independent scenario extraction pipelines simultaneously report the presence of otherwise mutually exclusive scenarios. To enforce the mutual exclusivity defined within D3.3 (Metz et al., 2019) in a deterministic fashion, an algorithm based on weighted directed acyclic graphs (DAG) was developed.

As input, the algorithm requires the scenario detections for each time step, a list of which transitions from one scenario to another are allowed between time steps, and a minimum duration for each scenario to filter out false positive detections. With these inputs, it can still occur that two simultaneously detected scenarios are deemed valid. In that case, to achieve a deterministic result, a prioritisation must be assigned to each scenario to decide which scenario is chosen over the other. The basic principle of the algorithm is that the nodes represent the time step, the scenario, and the remaining required duration of the scenarios. The validity of transition is indicated through the presence of an edge, and the prioritisation is depicted through weighting of the edges. Finally, the deterministic selection of scenario detections without any overlap of mutually exclusive scenarios is acquired through the shortest path through the graph from start to end.

Each node is specified through the scenario, timestamp of detection, and remaining time required to reach the minimum duration of the scenario. Therefore, as the first step for each positive scenario detection at a given timestamp, corresponding nodes are created for all possible values of the required remaining duration. Then, the edges between nodes that are consecutive in time are set according to the three following rules:

- The two nodes are of the same scenario, consecutive in time, and the required remaining duration decreases by the time step or stays at zero.
- The remaining required duration of the first node is at zero, the transition of the scenario of the first node to the scenario of the second node is valid, and the required remaining duration of the second node equals the respective minimum duration.
- If the two previous rules did not produce any valid edges for a time step, the edges from all the nodes of the first time step to all the nodes of the second time step are set with their remaining required duration at their respective minimum duration.

Finally, a start and end node are added to the graph to be able to compute a path through the network. To incorporate the prioritisation, each edge is weighted by a value according to the prioritisation of the scenario the directed edge points to. The lower the prioritisation of the scenario, the higher the weight of the edge. Now the scenarios for each time step on the shortest path from start to end node are the resulting mutually exclusive scenario detections. Since the *approaching traffic jam* scenario has a fixed duration, a special rule is introduced that allows edges to the corresponding nodes, even if the source node has not yet reached is minimum duration. As a last processing step in the conversion from the nodes of the shortest path to scenario detections, detection sequences shorter than their minimum duration are neglected. Dijkstra’s algorithm is used to find the shortest path.

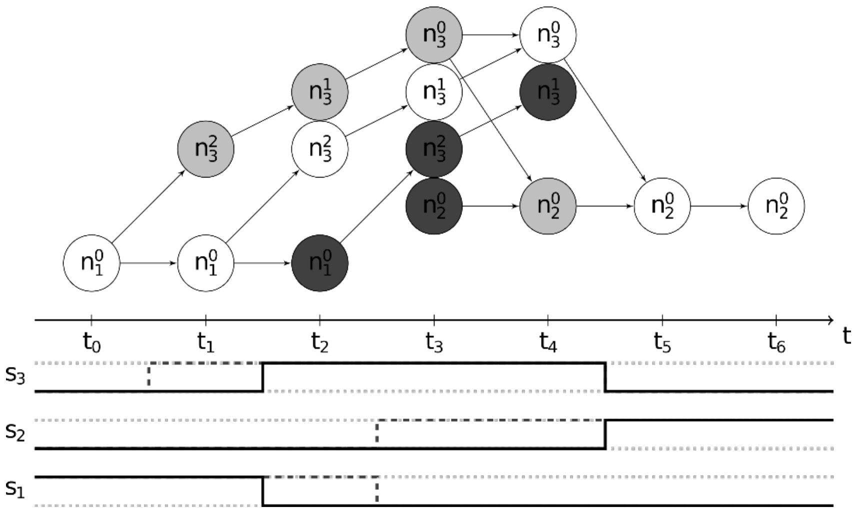


Figure A2.8: An example of a scenario sequence and the DAG.

The figure above (Figure A2.8) illustrates the creation of a DAG. The dotted lines show the original scenario detections from which the nodes and edges are built according to the set rules. The subscript of the node indicates the scenario and the superscript the required remaining minimum duration. The dark grey nodes are not reachable from the start node or the end node, because they are not of sufficient duration or no transition to them was possible. The light grey nodes were not chosen because the scenario of the white nodes was had higher prioritisation (lower weights on the corresponding edges).

A2.2. Urban

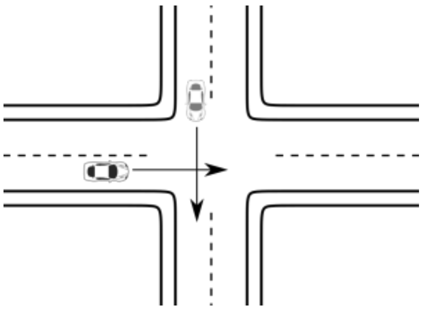
For the urban ADF, the scenarios defined for the motorway were all used, except those regarding traffic jam. Considering the different environment for urban, additional scenarios were defined. These mainly regard intersections, which are not present in the motorway use case. Additionally, scenarios with oncoming traffic are included (see initial definition of scenarios in project deliverable D3.3 – Evaluation Methods (Metz et al., 2019)).

A2.2.1. Intersection Scenarios

For intersections, four different scenarios were defined for the two types of intersection transits (crossing and turning).

A2.2.1.1. Crossing / Turning with Laterally Moving Object

When passing through intersections, the ego-vehicle often has to deal with objects that cross its path. These can be vehicles or VRUs that have the right of way (or take it from the ego-vehicle) in a crossing situation, as well as VRUs who have the right of way at their crossing when the ego-vehicle is turning left or right. In addition, oncoming traffic that needs to pass through when turning is also covered by this scenario.


(a)

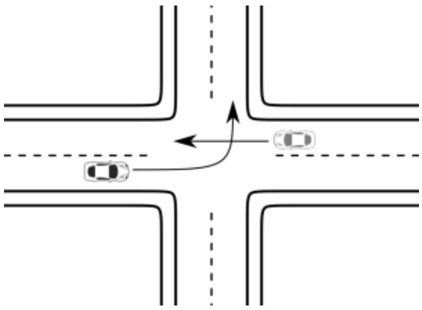

(b)

Figure A2.9: (a) Crossing or (b) turning with laterally moving object scenarios.

A2.2.1.2. Crossing / Turning with Static Object

A final intersection scenario was added where the vehicle passes through an intersection with a static object. This scenario can occur when there is an object blocking the way within the intersection and/or at the desired exit.

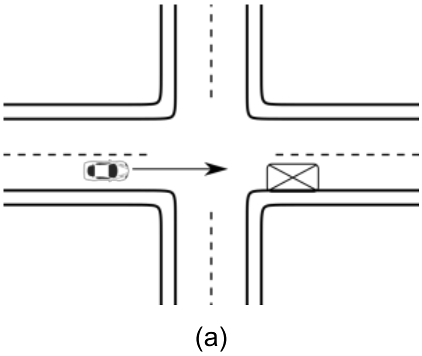


Figure A2.10: (a) Crossing or (b) turning with static object scenarios.

A2.2.1.3. Crossing / Turning with Lead Object

In addition to passing through the intersection without any conflict, the ego-vehicle can also be following a lead object. In the urban use case, this can be a vehicle or any kind of VRU. In this scenario, the ego-vehicle's speed is determined by the object it is following.

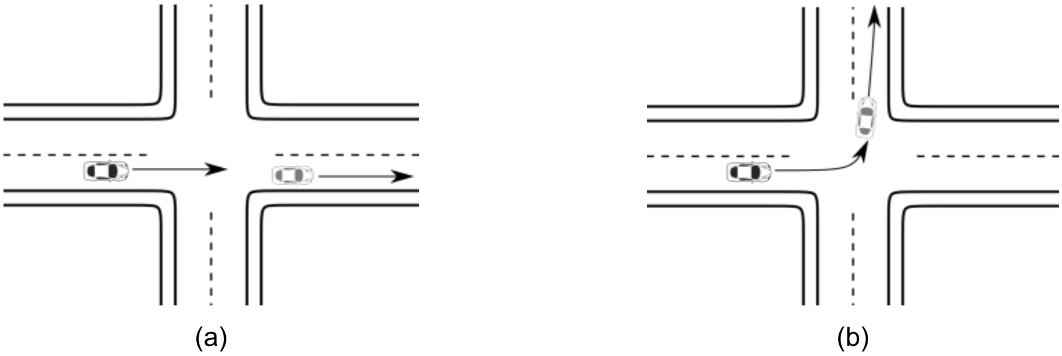


Figure A2.11: (a) Crossing or (b) turning with lead object scenarios.

A2.2.1.4. Crossing / Turning without Conflict

The most straightforward intersection scenario is that of crossing or turning without conflict. In this scenario, the ego-vehicle drives through the intersection without any conflict and without being influenced by other traffic participants in any way.

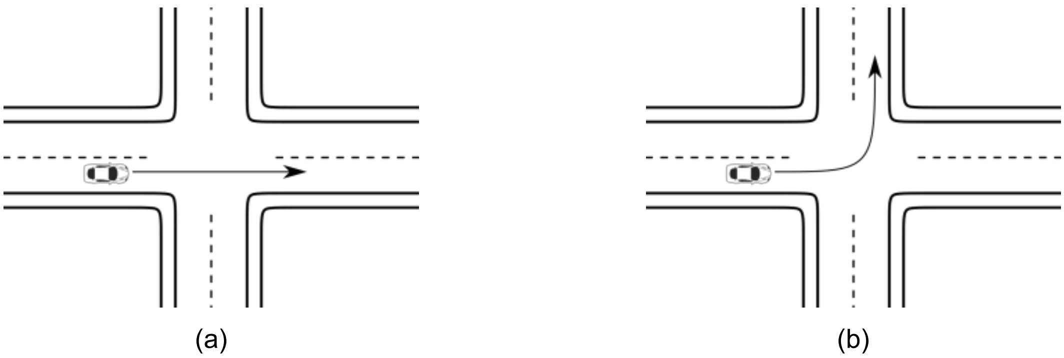


Figure A2.12: (a) Crossing or (b) turning without conflict scenarios.

A2.2.1.5. Calculation of Intersection Scenarios

As specified in the main text (cf. Section 2.5.4.1), scenarios were prioritised in the order in which they appear in this chapter (laterally moving objects, static objects, lead objects, without conflict).

For the implementation of the scenarios, the calculation can therefore not be considered for each scenario separately. Additionally, scenarios are split up after standstills, as these can change the further interaction of the ego-vehicle at a scenario (e.g., when the ego-vehicle has already turned). Within each of these subsections, the following steps are performed until a scenario has been found:

- Check the angle of travel across the intersection and determine turning or crossing from that.
- Check for close encounters of the ego-vehicle trajectory with other trajectories.
  - If yes, check for the intersecting trajectories and calculate the measures for laterally moving objects.
- If no, check if the ego-vehicle has a lead object while travelling through the intersection.
  - A lead object scenario can also occur when the ego-vehicle has a lead object when entering or leaving the intersection.
  - If yes, it was an intersection scenario with a lead object.
- If no, there was no conflict or interaction within the intersection.

A2.2.1.6. Overtaking with Oncoming Traffic (Active / Passive)

Apart from intersections, urban traffic also differs from the motorway use case in that oncoming traffic is not separated by a barrier. This opens up the traffic to the scenario of interacting with oncoming traffic. As an example, this can happen when there is a delivery van parked on the lane of the ego-vehicle which the ego-vehicle needs to overtake in the lane of the oncoming traffic. This can happen in the form of the ego-vehicle doing the overtaking (active) or having to interact with a vehicle that does the overtaking (passive).

Figure A2.13: Overtaking with oncoming object (a) active or (b) passive scenarios.

Annex 3 Intersection Definitions

In the urban evaluation, many of the urban-specific PIs were related to intersections. For all the urban Pilot sites, the intersections had to be defined manually. To ensure that the performance indicator data from all the Pilot sites were merged and analysed in a comparative way, it was important to agree on a common understanding on what an intersection is and how to define it.

The terms junction and intersection are sometimes used interchangeably, but they can have separate definitions. The European road accident database CARE, for example, uses junction as a larger area that can contain multiple intersections (Saurabh, 2018). The intersection is the small intersectional area of two or more roads and a junction is the area 20 metres in each direction from the intersections of the junction. If the distance between two intersections is less than 10 metres, the area between them is also a separate intersection. If the distance is between 10 and 20 metres, it is a “through roadway” area. If the distance is more than 20 metres, the intersections belong to separate junctions.

To determine how we should define the intersection present in the urban Pilot sites of L3Pilot, we had to consider the detected scenarios, derived measures and PIs, and the impact the size and features of the intersection would have on them. If our intersections followed the narrow definition, the time and distance the ego-vehicle travels inside them would be much smaller. This would reduce the number of VRU interactions and turning or crossing scenarios with lead, static or laterally moving objects, since the vehicle would spend less time in an intersection with the relevant objects. This would have an impact on the derived measures calculated from the trajectories of the vehicles in the intersection. It could also make crossing or turning more complicated to analyse in general, as it could create standstills after the ego-vehicle has waited to turn or cross but before entering the intersection.

If we had selected the larger junction definition as our definition of intersections, we would see a reduction in the number of instances of e.g., minimum distance to the lead object in an intersection, as there could be multiple crossings back-to-back from which these values are calculated. Given the limited data from some Pilot sites, a reduction in the number of data points was not a desirable outcome. Furthermore, in a staggered junction there could be both a crossing and a turning manoeuvre taking place inside it. This would break the strict separation of turning and crossing scenarios as designed in the methodology.

For these reasons, we decided to define the intersections as something in between the narrow intersection definition and the broad junction definition used by CARE.

The intersections were drawn by hand on satellite imagery with the help of video data captured by the ego-vehicle. This drawing produced for each intersection a set of coordinates that represent the boundaries of the intersection as a polygon. This allowed for the use of the GPS data captured by the ego-vehicle to determine whether the vehicle was inside the intersection or how close it was to it. The boundaries were set based on road markings, width of the road or lane, curvature of the road or lane, and, as a last resort, by visually comparing them to similar intersections with the above-mentioned features present.

If there were road markings indicating where the vehicles should stop to wait to enter the intersection, the boundary was set to overlap that line. At times there were separate road markings indicating where cyclists should stop, but these were not used in defining the intersection, as the main focus of the exercise was to find when the ego-vehicle was inside an intersection.

The other main type of road marking used to determine the boundaries of an intersection was pedestrian crossings, when they were seen to be part of the intersection system (for example, based on closeness and the presence of traffic lights for pedestrians). The boundary was drawn with the pedestrian crossings within the intersection. Otherwise, the pedestrian crossing would have had to be counted as a separate intersection as was done for pedestrian crossings on line sections of the road. This would have caused problems for the evaluation. For example, there would have been intersections the ego-vehicle never approached, as the next intersection on its path would have been the pedestrian crossing and there would have been no distance to the intersection used by vehicles. Additionally, the time and distance the ego-vehicle spent inside the pedestrian crossing intersection would have been small, which would have made calculating the relevant derived measures more difficult. It was decided that including the pedestrian crossing as part of the intersection and calculating the derived measures separately for vehicle and VRU object classes would produce more meaningful data representing how the ego-vehicle acts at intersections.

Other road markings used were changes in the lane markings, such as switching from dashed to solid. In case no road markings were present at all, or they were inconclusive, other features of the intersection were considered. These included the curvature or width of the road or lane, as they are often used to make more room for vehicles to perform a turning manoeuvre. As a last resort, if no features could be used to select where the intersections starts or they were occluded by buildings, for instance, the intersection was visually compared to similar ones within the same city and the boundaries were selected based on those.

All the intersections were drawn for each Pilot site independently. These were then compared during a workshop to make sure that the selection criteria and resulting boundaries matched those of the other Pilot sites. Additionally, for each intersection the type (X/T/Y intersection, roundabout, etc.) and the priority rule (traffic signs and road markings only, traffic lights with or without partial conflicts, etc.) of the intersection were coded according to the UDRIVE codebook.

Annex 4 Bootstrapping

In the urban Technical and Traffic Evaluation there were three Pilot sites. One of them contributed a large number of driving hours compared to the others. Consequently, also the number of scenario instances was considerably higher for one Pilot site, for many scenarios. This posed a challenge to the data analysis, because simply pooling the data would mean that results from a single Pilot site would dominate the results. This would have compromised the aim to evaluate urban ADFs in general. To balance the contribution of all the urban Pilot sites an additional bootstrapping step was performed before the data was uploaded to the CDB.

In technical terms, the urban Technical and Traffic Evaluation had a class imbalance problem, meaning that the majority class had more data points than the minority class(es). Consequently, the statistical model based on the data might not capture the properties of the minority class and would not generalise to new data well. A straightforward solution to address a class imbalance problem is to either undersample the majority class or oversample the minority classes. Undersampling means that samples smaller than the original data are drawn with replacement to represent the majority class. In oversampling, samples larger than the original data are drawn with replacement to represent the minority class.

In the present case, 50 sampling rounds were performed. In each round, the largest Pilot site was undersampled and for the two smaller sites the sample size was the same as their original data. The samples obtained in each sampling round was pooled and uploaded to the CDB with an extra variable indicating the sampling round. In effect, this created 50 datasets representing the original data in a more balanced way. In other words, the process was similar to using bootstrapping to estimate statistical indicators.

An alternative to the bootstrapping process would have been to create synthetic data based on the minority classes. However, these procedures are potentially complex, and the validity of the data would have needed to be ensured. Another alternative would have been to make the analysis separately for each Pilot site and then weight the results. However, this would have required complex procedures to ensure the confidentiality of the Pilot sites and no data could have been shared.

A4.1. Noise

This bootstrapping step creates another challenge for data confidentiality. The data points at smaller sites are more likely to be sampled in multiple rounds. Consequently, the data from minority sites were more likely to be repeated multiple times and, at least in theory, it would be possible to distinguish larger Pilot sites from the others. To make the identification of data sources less feasible, a small amount of normal noise was added to the variables of interest. This procedure, also called smooth bootstrapping, ensured the uniqueness of data points. The added noise also had the effect of smoothing the distributions.

A4.2. Implications for Statistical Inference

In a typical bootstrapping process, a large number of samples are drawn with replacement from the original data. Based on the samples, the statistical indicators for the original data such as mean or median and their variance can be estimated based on the samples. In short, the bootstrap-based estimation is performed by calculating the statistical indicator for each sample and then deriving the variance estimates based on the distribution of the sample indicators.

Bootstrapped variance estimates are unbiased when the bootstrap sample size is equal to the original data size (Efron & Tibshirani, 1986). Using a different sample size influences the variance estimated. In case of undersampling the variance estimates become larger, and with oversampling smaller.

To illustrate the effect of the under- and oversampling, a simulation was performed. Two datasets were generated, one representing “baseline” and another “treatment”. Both datasets had a variable of interest y. For the baseline, the 100 data points were drawn from a normal distribution N(1, 4). For the treatment, also 100 data points values were drawn from N(1.5, 4). The resulting datasets had M=1.14 and M=1.85 respectively.

Next, both datasets were sampled at different sample sizes. The term *sampling factor* is used to represent that size of the sample relative to the original data size. A sampling factor of one means that the bootstrapped sample size was equal to the original number of data points. Sampling factors of less than one represent undersampling (e.g., 0.2 = 20% of the original data size), and larger than one oversampling (e.g., 2 = 200% of the original data size). In the simulation, 50 sampling rounds were done.

Figure A4.1 shows how the sample means and their 1.96 * standard error for the baseline and treatment groups change with the sampling factor. The standard error of the mean increases when the sampling factor decreases (undersampling) and decreases when the sampling factor increases (oversampling).

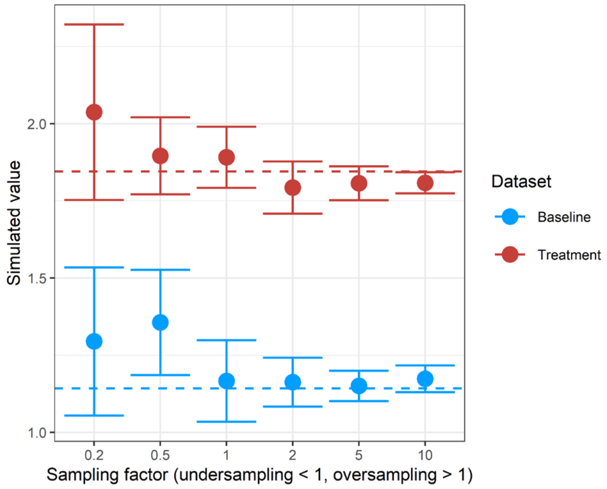


Figure A4.1: Sample means (dots) and their 1.96 standard errors as a function of the sampling factor for simulated baseline and treatment datasets. The true means of the simulated are data shown as dashed lines.

The implications for the statistical inference are clear. Undersampling increases the risk of type II error (false negative), while oversampling increases the risk of type I error (false positive). Consequently, undersampling of the majority site makes statistical inference based on the confidence intervals more conservative as the confidence intervals increase.

Trip PIs and scenario instances were composed of multiple values. Consequently, the bootstrapping procedure preserved the correlations between the different values, e.g., speed and acceleration, within the scenario instances and trips. However, as scenario instances and Trip PIs were treated as independent observations, correlations between different types of scenarios are not preserved.

A4.3. Algorithm

The bootstrapping step was implemented in the data_sampler.py script. For each file type (scenarios and Trip PIs) the following steps were performed:

1. The sample sizes were determined manually for each Pilot site and file type so that data would be in balance.
2. The data files generated by the urban tool chain were read into a data frame.
3. Variables which could not be shared were filtered out.

The processing differed slightly for Trip PIs and scenario instances.

For trips, steps 4 and 5 were performed.

4. A single trip was represented by four rows in the data representing baseline, ADF, System available, and System unavailable conditions (see Bellotti et al., 2020, Chapter 3.1). In the sampling process, these four rows were treated as single observation.

5. Random samples with replacement were drawn among all the trips.

For scenario instances, steps 6 and 7 were performed.

6. A single scenario instance was represented by a single row in the data. Scenario instances were first divided into baseline and ADF conditions (*System_available* and *System_unavailable* were not used).
7. Random samples with replacement were drawn from the baseline and ADF scenarios. Sample sizes for the baseline and ADF could be different.
8. Steps 4 and 5 or 6 and 7 were repeated 50 times to generate 50 datasets.
9. To make it difficult to infer which observations were repeated in the sampling process, a small amount of noise was added to all the measurement variables and to certain index-based variables (*Scenario_Start_Index*, *Number_of_Samples*). For measurement variables, the noise was generated from normal distribution with the mean at zero, and the standard deviation set to be equal to 1% of the standard deviation in the original data. For the index-based variables, a single integer between -3 and 3 was drawn from a uniform distribution.
10. The resulting datasets were stored with new unique *TripID*s generated based on the original ID, sampling round, and Pilot site-specific salt key. Information on the sampling round was preserved.

Annex 5 AIM Results (normal driving behaviour at an urban roundabout)

A5.1. RQ-T12-1 / Car Following

Figure A5.1 shows the results for normal behaviour in car-following regarding the phases entering, circling, and exiting the roundabout. While the minimum THW and minimum TTC values in the circling phase are the lowest, the average THW values in the circling phase are higher than the other two phases. Note that trajectory errors could lead to THW or TTC values close to zero.

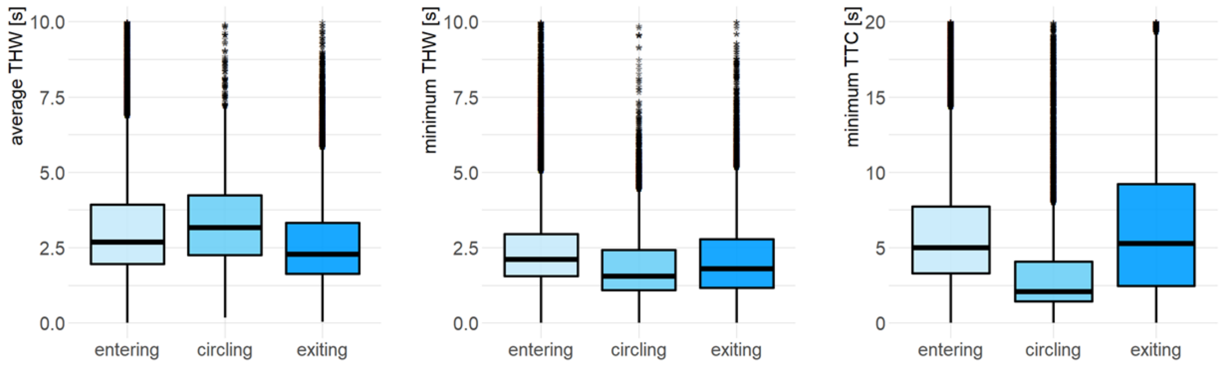


Figure A5.1: Values of average THW (left), minimum THW (middle) and minimum TTC (right) in car-following.

A5.2. RQ-T12-2 / VRU Behaviour

Figure A5.2 shows the results of the interaction of human drivers with VRUs (cyclists: top, pedestrians: bottom) at the eastern arm of the roundabout. Here, the variables PET, T_{Adv} and THW were computed for yielding and non-yielding of the vehicle in the roundabout. While the PET values between human drivers and cyclists are very similar in yielding and non-yielding, the PET is smaller in case of yielding to pedestrians. The T_{Adv} values are larger in case of non-yielding than yielding but differ largely from the values in case of yielding to pedestrians than in case of non-yielding. The THW values are larger in case of non-yielding than in the case of yielding. In case of non-yielding the THW values for cyclists are larger than for pedestrians.

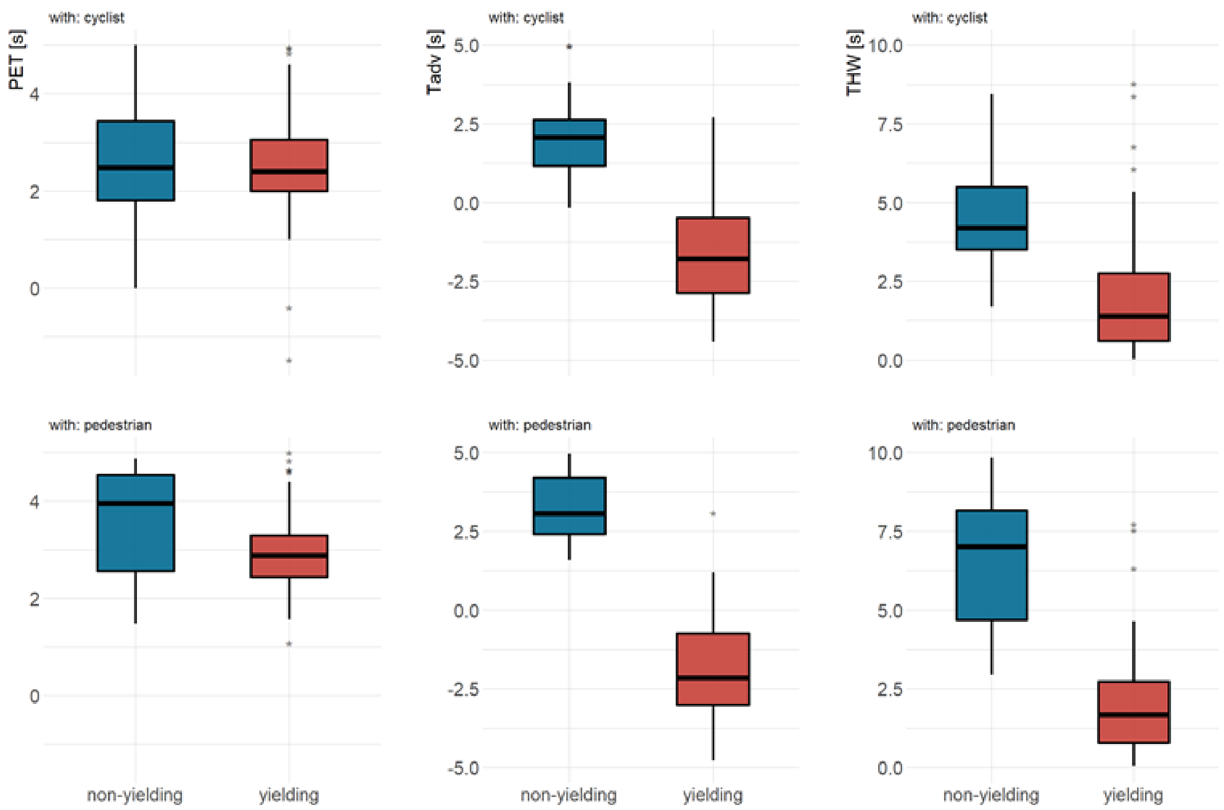


Figure A5.2: Interaction of the vehicle in the roundabout with cyclists (top) and pedestrians (bottom): PET values (left); T_{Adv} values at the moment the vehicle is about to leave the roundabout (middle), positive and negative values respectively indicate the time advantage of the vehicle or VRU; THW values between the VRU and the crossing point, at the moment the vehicle is about to leave the roundabout (right).

A5.3. RQ-T14-1 / Near-Crash Frequency (Vehicles)

Table A5.1 gives the relevant information on near-crash frequencies based on the minimum TTC in car-following scenarios.

Table A5.1: Information on near-crash frequencies in car following scenarios.

	All		Minimum TTC < 2s		Minimum TTC < 1s	
Phase	N	N/h	N	N/h	N	N/h
Entering	48430	139	2259	6	560	2
Circling	8196	24	3230	9	628	2
Exiting	24233	70	4400	13	3046	9
Sum	80859	233	9889	28	4234	13

Table A5.2 gives the relevant information on near-crash frequencies based on the minimum TTC in merging scenarios.

Table A5.2: Information on near-crash frequencies in merging scenarios.

	All		PET < 2s		PET < 1s	
Sub-scenario	N	N/h	N	N/h	N	N/h
Non-yielding	12889	37	1174	3.4	19	0.1
Yielding	17282	50	4276	12.3	19	0.1
Sum	30171	87	5450	15.7	38	0.2

A5.4. RQ-T14-2 / Near-Crash Frequency (VRU)

Table A5.3 gives the relevant information on near-crash frequencies based on the PET. Altogether 514 VRU crossing situations were considered.

Table A5.3: Information on near-crash frequencies between vehicles and VRU.

		All		PET < 2s		PET < 1s	
Interaction with	**Sub-scenario**	N	N/h	N	N/h	N	N/h
Cyclist	Non-yielding	83	< 1	28	< 1	6	< 1
	Yielding	272	1	72	< 1	3	< 1
Pedestrian	Non-yielding	43	< 1	7	< 1	0	0
	Yielding	116	< 1	14	< 1	0	0

Annex 6 AIM Results (normal driving behaviour at an urban intersection)

In the analysis of normal driving behaviour at a signalised urban intersection in Braunschweig, Germany, we focused on kinematic and interaction behaviour of a vehicle (turning left or right, going straight) with oncoming road users (VRU and motorised vehicles). Altogether, 30 days of trajectory data on the relevant scenarios from different months in 2018 and 2019 were analysed. Figure A6.1 shows the relevant scenarios:

Scenario L1: Left-turning following vehicle from west to north interacting with lead vehicle from west to north (Figure A6.1, path 1),

Scenario L2: Left-turning vehicle from west to north interacting with oncoming vehicle from east to west (Figure A6.1, paths 1 and 2),

Scenario L3: Left-turning vehicle from west to north interacting with oncoming bicycle (Figure A6.1, paths 1 and 5),

Scenario R1: Right-turning following vehicle from east to north interacting with lead vehicle from east to north (Figure A6.1, path 3),

Scenario R2: Right-turning vehicle from east to north interacting with bicycle from east to west (Figure A6.1, paths 3 and 4),

Scenario S: Straight driving following vehicle from east to west interacting with lead vehicle from east to west (Figure A6.1, path 2).

For the analysis, the paths of the road users were divided into several relevant phases. For instance, in the case of scenario R2 the four phases "approaching", "turning", "conflict" and "exiting" were considered, while in case of scenario L3, six phases "approaching", "queuing", "conflict" and "exiting/approaching", "conflict" and "exiting" were considered.

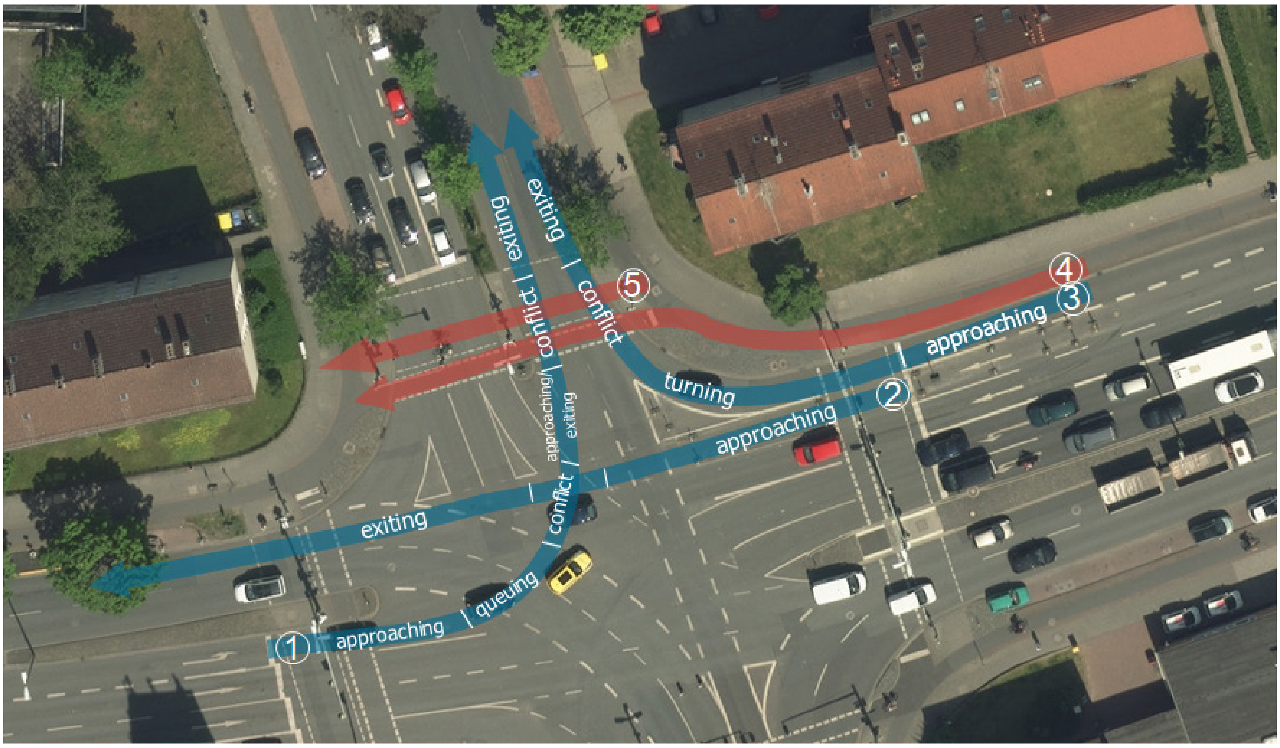


Figure A6.1: Vehicle (blue) and bicycle (red) trajectory paths and phases (picture owned by DLR).

A6.1. Interaction with oncoming road users

For normal driving and interaction behaviour with oncoming road users, the scenarios R2, L2 and L3 and their sub-scenarios "yielding" and "non-yielding" were analysed. “Non-yielding” and “yielding” was taken from the perspective of the turning vehicle. For instance, the situation is categorised as “yielding” if the turning vehicle gives way to an oncoming bicycle. In contrast, the situation is characterised as "non-yielding" if the turning vehicle does not give way to an oncoming bicycle. Table A6.1 summarises the mean values of relevant performance indicators (PI) according to the RQs (see Section 3.8) regarding the vehicle acceleration, driven speed, interaction behaviour, traffic flow / journey times, and the number of incidents and near-crashes.

Table A6.1: Relevant PI in case of the interaction scenarios R2, L2 and L3.

RQ	PI	Unit	Right turning vehicle vs bicycle (R2)		Left turning vehicle vs bicycle (L3)		Left turning vehicle vs oncoming vehicle (L2)	
			Non-yielding	Yielding	Non-yielding	Yielding	Non-yielding	Yielding
6 (longitudinal acceleration)	**max(a)**	m/s^2	2.81	2.95	2.48	2.49	2.44	2.45
	min(a)	m/s^2	-2.41	-3.22	-1.75	-1.87	-1.71	-1.80
	sd(a)	m/s^2	1.07	1.35	0.95	0.93	0.92	0.90
8 (longitudinal speed)	**m(v)**	m/s	5.67	5.07	4.63	3.64	4.69	3.64
	max(v)	m/s	8.96	10.03	10.97	10.58	10.75	10.80
	sd(v)	m/s	1.42	2.26	2.91	2.83	2.81	2.98
12 (interaction behaviour)	**PET**	s	1.57	1.32	2.94	3.08	3.17	2.60
	THW(minTAdv)	s	3.99	4.10	3.97	3.63	2.15	6.02
	Distance(minTAdv)	m	27.68	19.62	22.44	13.27	25.28	10.77
13 (traffic flow / journey time)	**JT**	s	11.90	13.05	15.38	20.58	15.63	20.73
14 (number of incidents / near-crashes)	**N/h**	-	-	-	3	1	24	4
	N/h ***PET < 1s***	-	-	-	0	0	1	0

Figure A6.2 shows the PIs for vehicle acceleration and driven speed in different phases of scenarios R2, L2 and L3.

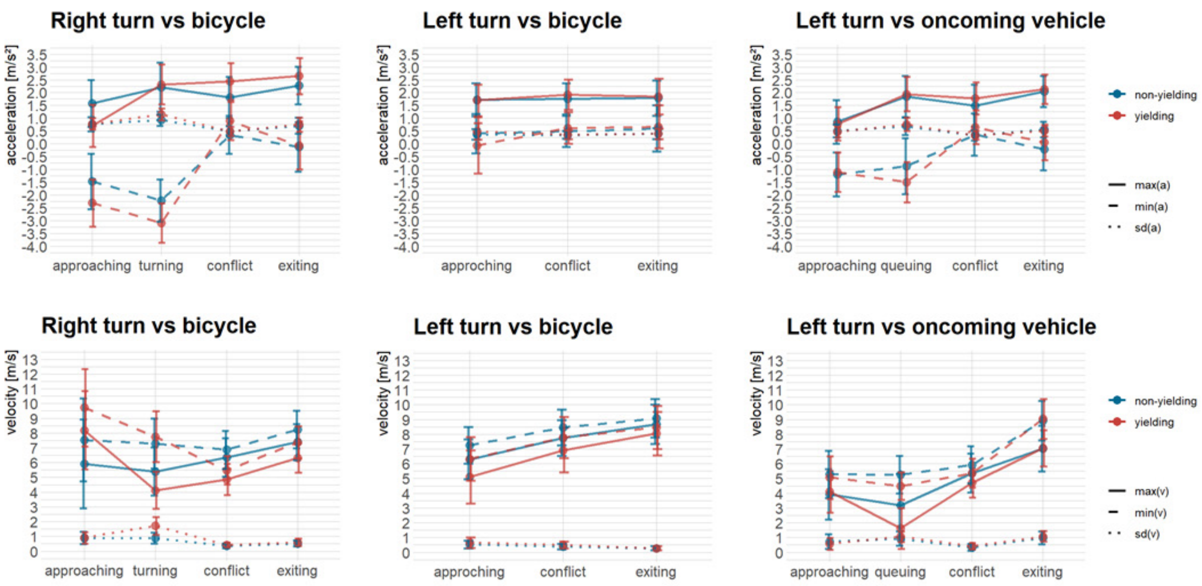


Figure A6.2: Aggregated results on driving behaviour indicators in scenarios R2, L2 and L3.

A6.2. Interaction with leading vehicle (car following)

The normal driving behaviour in car-following in scenarios L1, R1 and S were analysed. Each scenario was divided into two sub-scenarios according to whether the following vehicle has completely stopped in the process or not. Table A6.2 summarises the mean values of relevant PIs for vehicle acceleration, driven speed, interaction behaviour, traffic flow / journey times, and number of incidents and near-crashes.

Table A6.2: Aggregated results on relevant indicators for driving and interaction behaviour in scenarios R1, L1 and S.

RQ	PI	Unit	Right turning following vehicle vs leading vehicle (R1)		Left turning following vehicle vs leading vehicle (L1)		Straight driving following vs leading vehicle (S)	
			Non-stop	Stop	Non-stop	Stop	Non-stop	Stop
6 (longitudinal acceleration)	**max(a)**	m/s²	1.91	2.72	2.27	2.40	1.81	2.58
	min(a)	m/s²	-2.39	-2.67	-1.66	-2.03	-1.99	-2.25
	sd(a)	m/s²	0.99	1.11	0.82	0.89	0.78	1.20
8 (longitudinal speed)	**m(v)**	m/s	6.71	3.23	4.04	2.97	11.85	8.69
	max(v)	m/s	10.16	8.69	10.19	10.39	13.53	12.30
	sd(v)	m/s	1.60	2.56	2.72	3.04	0.93	2.11
12 (interaction behaviour)	**m(THW)**	s	2.39	2.69	6.67	12.16	2.41	4.62
	min(THW)	s	1.77	1.41	1.43	1.45	2.12	3.19
	min(TTC)	s	4.68	3.02	3.22	2.70	8.78	-
13 (traffic flow / journey time)	**JT**	s	9.84	20.13	19.48	23.98	6.95	9.48
14 (number of incidents / near-crashes)	**N/h**	-	75.46	4.44	10.98	11.21	80.9	0.01
	N/h *TTC < 1s*	-	0.57	0.07	0.94	1.1	1.89	0

Figure A6.3 shows the relevant PIs for vehicle acceleration, driven speed, and interaction behaviour in different phases of scenarios R1, L1 and S.

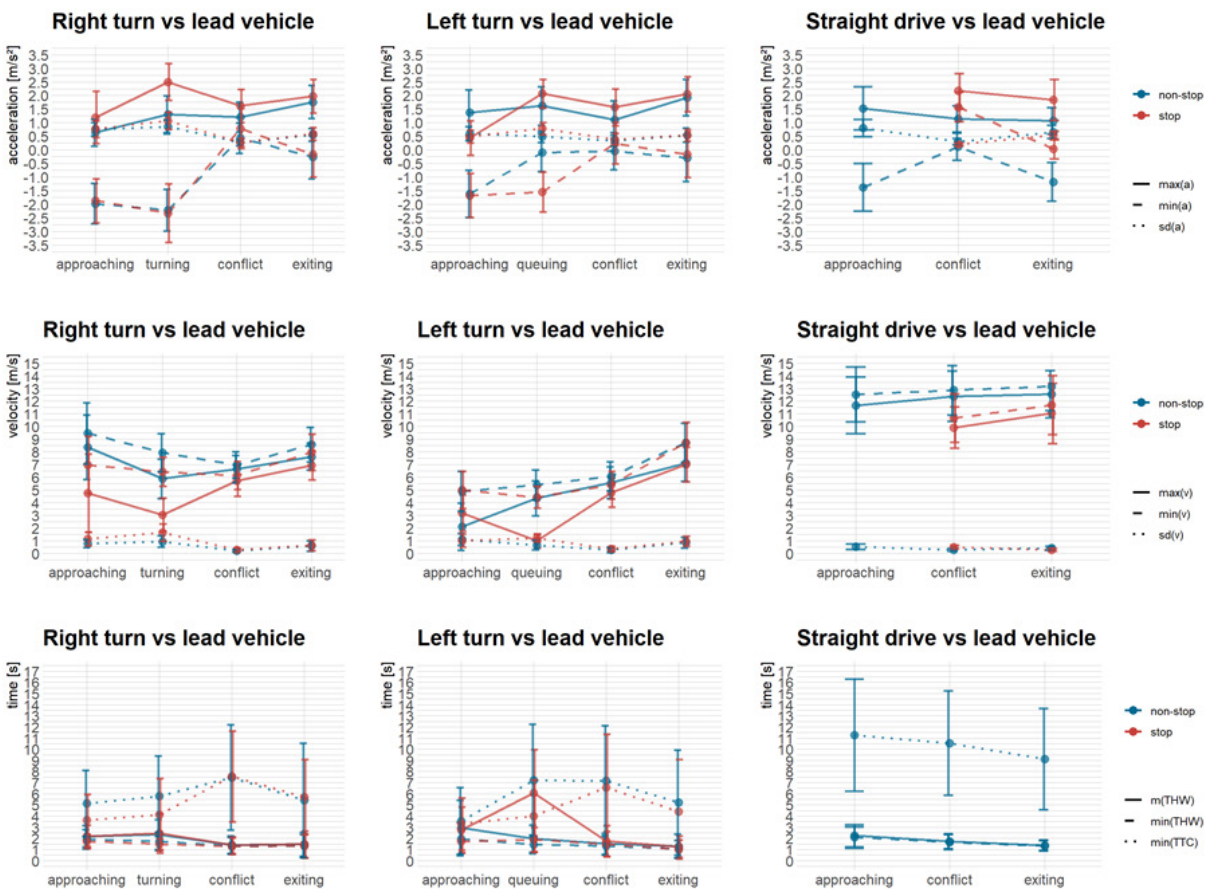


Figure A6.3: Aggregated results on driving behaviour indicators in scenarios R1, L1 and S.