Deliverable D7.1 //

Test and Evaluation Plan

Dissemination level	RE
Version	1.1
Due date	30.06.2015
Version date	30.07.2015

This project is co-funded by the European Union

AdaptIVe
Automated Driving Applications and Technologies for Intelligent Vehicles

Document information //

AUTHORS

Carmen Rodarius – TNO

Joris Duflis - TNO

Felix Fahrenkrog – ika

Christian Rösener - ika

András Várhelyi – Lund

Roberto Fernandez – CTAG

Lei Wang - BMW

Patrick Seiniger - BASt

Dehlia Willemsen – TNO

Lex van Rooij – TNO

COORDINATOR

Aria Etemad

Volkswagen Group Research

Hermann-Münch-Str. 1

38440 Wolfsburg

Germany

Phone: +49-5361-896-2334

Email: aria.etemad@volkswagen.de

PROJECT FUNDING

7th Framework Programme

FP7-ICT-2013.6.5: Co-operative mobility

Grant Agreement No. 610428

Large-scale Integrated Project

www.adaptive-ip.eu

LEGAL DISCLAIMER

The information in this document is provided 'as is', and no guarantee or warranty is given that the information is fit for any particular purpose. The above referenced consortium members shall have no liability for damages of any kind including without limitation direct, special, indirect, or consequential damages that may result from the use of these materials subject to any liability which is mandatory due to applicable law.

© 2014 by AdaptIVe Consortium

Revision and history chart //

VERSION	DATE	COMMENT
0.1	04.12.2014	Structure of document
0.2	15.01.2015	First input and update of structure
0.3	03.03.2015	Input from different partners
0.4	18.03.2015	Input from different partners
0.5	19.03.2015	Version for SP7 internal review
0.6	04.04.2015	Version after internal review
1.0	03.06.2015	Considering updates of VSP and preparation of deliverable D7.1
1.1	30.07.2015	Updates according to AdaptIVe reviewer's comments

Table of content //

List of figures //

1 Summary

The general objective of AdaptIVe is to develop and demonstrate new functionalities provided by partially-automated and highly-automated vehicles. These applications cover different speed regimes and driving scenarios and aim at improving safety, energy efficiency, dependability and user-acceptance of automated driving.

The introduction of supervised automated driving is now posing new and specific questions: in particular, the functions embodying automated driving do influence not only a certain defined scenario (for example accidents, near accidents, or safety related situations) but the whole traffic flow. Therefore, the existing evaluation methods are insufficient, and new comprehensive approaches are required.

SP7 "Evaluation" is a horizontal activity within AdaptIVe supporting the vertical subprojects. Its main objective is to develop a common evaluation framework for supervised automated driving applications which is described within this deliverable. This framework addresses two types of the assessment. The first part of the evaluation framework which this report focuses on considers the evaluation of the status quo which consists of the technical, user-related and real-life interaction (in-traffic) evaluation. The second part concentrates on the analysis of the future benefits with respect to safety and environmental aspects, which can be achieved by means of automated driving applications. This will be presented in more detail in the upcoming deliverable D7.3. With the development of each evaluation framework, previous work conducted by earlier projects is considered and included into the procedures where possible.

Starting from an overview on the developed functions and the evaluation activities in previous projects the overall evaluation methodology is described in chapter 2. The evaluation process is split into four assessment types in analogy to the approach of the PReVAL project as well as the interactIVe project. In the technical assessment (chapter 3) the performance of the functions is investigated. The user-related assessment (chapter 4) analyses the interaction between the function and the user as well as the acceptance of the developed functions. The in-traffic assessment (chapter 5) focuses on the effects of automated driving on the surrounding traffic as well as non-users. The impact assessment (chapter 6) determines the potential effects of the function with respect to safety and environmental aspects (e.g. fuel consumption, traffic efficiency). Overall conclusions are presented in the final chapter 7.

For each evaluation, the starting point is the function or system under investigation itself. Based on its description, a classification is performed to determine which evaluation methodologies are most appropriate for the assessment. Within the AdaptIVe sub projects 4 to 6, automation functionalities for close-distance, urban as well as highway scenarios will be developed, respectively. Since a complete evaluation of all of the AdaptIVe functions in all assessments is out of the scope of this project, only selected functions will be evaluated with selected methodologies in order to demonstrate the application of the evaluation framework. Examples of the evaluation procedure are provided with the presentation of each methodology. Within AdaptIVe, two general types of functions are distinguished: event based functions that only operate for a short period of time as well as continuous operating functions which once activated will operate over a longer time period.

For each assessment framework research questions, hypotheses as well as indicators are defined that guide the respective evaluation. The research questions are the first step of the evaluation and provide information on what should be addressed. Based on those research questions, hypotheses to be tested are defined. Testing of the hypotheses is done by using indicators that can be calculated based on signals or be derived from measures logged during the tests. It should be noted, that not all of these research questions, hypotheses and defined indicators might be applicable for all functions or systems. Therefore, for each combination of system and chosen evaluation an appropriate subset needs to be considered.

Many different test tools like balloon cars or real vehicles and test environments like test tracks, public roads or simulators are theoretically available for evaluation. Depending on the function or system under investigation, its development status as well as other requirements like legal boundaries or safety protocols, the most appropriate choice needs to be made for each evaluation. The different considered combinations and possibilities will be described alongside the evaluation frameworks in the respective sections.

2 Evaluation in AdaptIVe

The objective of AdaptIVe is to develop new automated driving applications in order to promote safer and more efficient driving. This deliverable provides a test and evaluation framework for the automated driving applications developed in the project AdaptIVe. Within this framework, test methodologies, performance indicators and test tools needed to assess the AdaptIVe applications with respect to technical performance, user-related effects as well as in-traffic behaviour are defined. Also, first ideas for the impact assessment that considers safety as well as environmental aspects of the AdaptIVe functions are provided. The impact assessment will be described in more detail in the deliverable D7.3

This document describes the whole evaluation process for the three assessments (technical, user-related and in-traffic), as presented in Figure 2.1. At the end of the project, the developed framework will be exemplarily applied to the developed AdaptIVe functions of SP4, 5 and 6.

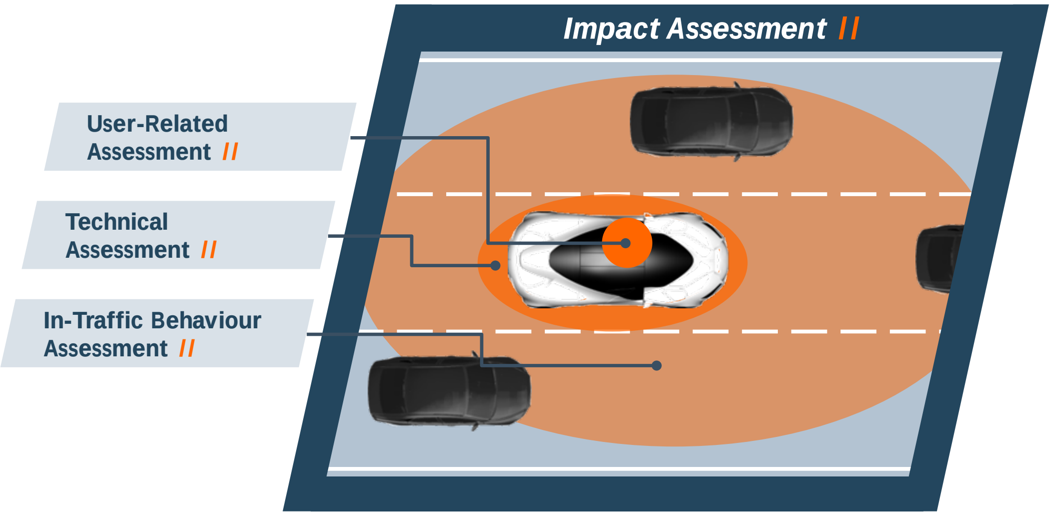


Figure 2.1: Evaluation areas in AdaptIVe

The evaluation framework developed in AdaptIVe builds on the results and experiences from previous European projects, e.g. PReVAL[1], eIMPACT [2], ASSESS [3] and interactIVe [4]. However, these projects dealt mainly with active safety functions or respectively advance driver assistance systems (ADAS). For AdaptIVe, different approaches may be necessary in order to consider also automated driving applications.

2.1 AdaptIVe systems and functions

The project development of the automated driving functions and systems in AdaptIVe are divided into three sub projects (SPs):

Sub project 4: Automation in close-distance scenarios: Addresses manoeuvres at low speed (speeds up to 30 km/h) that are characterised by the presence of close obstacles, such as in parking manoeuvres.

Sub project 5: Automation in urban scenarios: Deals with driving scenarios in urban environments that are characterised by an average speed range of 0 to 70 km/h.

Sub project 6: Automation in highway scenarios: Addresses motorway scenarios (or motorway similar roads) considering velocities up to 130 km/h.

Within these sub-projects, 21 functions are developed. An overview is given in Figure 2.2. In the following the AdaptIVe functions are briefly presented per target area (Highway, urban regions and close distance manoeuvring) as described in[5] and [6].

Sub-project	System Name	Function Name	Demonstrator
4	Construction Site	Construction Site Manoeuvre	Ford
	Parking	Pholova - Park Assistant	
		Automated Parking Garage Pilot	Daimler
5	City Cruise Supervised City Control City Chauffeur	Lane following and speed adaptation	CRF
		Vehicle following in lane	
		Obstacle or VRU on the road	
		Lane change	
		Intersection handling	
		Urban roundabouts handling	
		Traffic light handling	
5	Supervised City Control / Traffic Jam Assist	Lane following and speed adaptation	VCC
		Vehicle following in lane	
		Obstacle or VRU on the road	
		Lane change	
		Intersections handling	
		Urban roundabout handling	
5	Supervised City Control	Lane following and speed adaptation	BMW
		Vehicle following in lane	
		Obstacle or VRU on the road	
6	Highway Automation System	Lane Following	BMW CONTIT VTEC VW
		Lane Change (and overtaking)	
		Stop & Go Driving	
		Speed / time gap adaptation at a motorway entrance ramp	
		Cooperative merging with speed adaptation	VTEC, VW

Sub-project	System Name	Function Name	Demonstrator
		Enter and exit of a motorway	BMW, VW
		Cooperative merging with lane change	VW
		Emergency vehicle on duty	VW

Figure 2.2: AdaptIVe functions, systems and related demonstrator vehicles

The "close-distance manoeuvring" functions concentrate mainly on parking manoeuvres. Two types of parking functions are implemented. The first one aims to perform the actual parking manoeuvre as it is conducted at parking spots or in a garage at home. A particular challenge is on parking into tight parking spaces. The second function is intended to be used in parking garages. The function is designed to drive the vehicle from the entrance of the parking garage to the assigned parking spaces and execute the final parking manoeuvre. The close-distance manoeuvre functions will be integrated into two demonstrator vehicles considering the automation levels from partial up to conditional automation according to the SAE definition [7] [8].

The urban scenario functions enable automated driving in urban areas. Similar to the highway function “automated lane”, vehicle following as well as an automated lane change functions are implemented. A particular challenge for automated driving in urban areas is passing intersections. It needs to be ensured that vehicles’ velocities are adapted appropriately at intersections, traffic lights and roundabouts in order to prevent the vehicle from approaching them at too high speeds. Whether the passing of the intersection itself is handled by the driver or systems depends of the automation level of the systems. A further characteristic of urban traffic is the inconsistency of the traffic flow. This leads to the requirement of appropriated handling for stop and go situations.

The basic highway automation function implements an automated lane following or, if a preceding vehicle is present, automated vehicle following at a certain velocity. The function is supplemented by other (sub-) functions. Examples are the speed and time gap adaptation functionality according to the speed limit that increase or decrease the velocity of the host vehicle respectively the distance to a front vehicle if another vehicle wants to merge in at a motorway entrance ramp. Here the automation function should ensure a smooth and safe merge-in manoeuvre. Another example is the predictive automated driving functionality, which adapts the current velocity according to a new speed limit.

The second basic function for the highway scenario is the automated lane change function. By means of this function overtaking manoeuvres are realized. In addition this function is part of other functionality such as the "highway entry and exit" function. Some functions for the highway automation scenario will utilize V2V communication in order to execute certain manoeuvres. These functions are cooperative filter-in manoeuvres (automated lane change when another vehicle wants to enter the motorway), cooperative lane change on an entrance ramp (automated lane change to an adjacent lane when another vehicle wants to enter the host vehicle's lane) and cooperative response on emergency vehicles on duty (lane change in case an emergency vehicle is approaching from behind).

Next to the presented functions there are also common functions which are developed and used in all three subprojects. These are the stop-and-go and the minimum risk manoeuvre function. The latter defines the vehicle reaction in case of a system failure (SP4) or the system reaches it boundary conditions and the driver does not respond to an overtake request by the function respectively in a transition of control situation (SP4, SP5, SP6). In these situations, measures need to be taken in order to assure a safe vehicle state. Although these functions are developed for all target areas of the project the defined measures for each target area can be different.

2.2 Evaluation in previous projects

The evaluation of the automated functions is an important part of the development process. There are different kinds of evaluations conducted at different stages of the development process. AdaptIVe focuses on the evaluation on functional level at the end of the development process including the technical, user-related, in-traffic and impact assessment of the developed functions. Typically relevant research questions for the evaluation process at this stage are:

- What are appropriate evaluation methods (for automated driving functions)?
- Can existing evaluation methodologies be applied to automated driving functions and if yes to which extend?

In order to address these questions a literature review on existing evaluation methodologies is conducted in a first step. Most of the evaluation activities in the past focused on the evaluation of ADAS functions, whereas only a few projects considered also automated driving functions. Within this literature review different evaluation types have been identified and clustered.

During the actual development of a function a continuous and iterative evaluation of (sub-) functions is conducted. The main objective of the evaluation at this stage is to check, whether the pre-defined requirements are fulfilled by the (sub-) function and whether the defined or specified performance is reached (e.g. SARTRE[9], HAVEit [10] and the German funded research project KONVOI [11], EnergyITS[12]). Of course the tests of fulfillment of requirements or specifications can also be checked at the end of the development process respectively project. The tests for this type of evaluation are mainly conducted on test tracks or in simulation. For these tests typically test cases are defined, which are based on certain use cases of a function or systems under test.

A further evaluation stage is the evaluation towards the end of the development process, where a single function or system is assessed against pre-defined hypotheses. This evaluation stage analyses the outcome of the (whole) function development and is typically applied in research projects (e.g. interactIVe [13]). The function under test is assessed against certain predefined evaluation criteria or/and hypotheses. A general evaluation methodology for this evaluation stage has been introduced by the PReVAL [1] project. The PReVAL approach focuses mainly on

ADAS applications and approach considers three evaluation areas (technical, user-related and safety impact assessment). For each evaluation area the evaluation is done in six steps:

- Step 0: System and function description (this step has been already conducted in the design phase of the functions)
- Step 1: Expected impact and hypotheses
- Step 2: Test scenario definition
- Step 3: Evaluation method selection
- Step 4: Measurement plan
- Step 5: Test execution and analysis

An important question in this evaluation type – in particular in research projects – is often the question concerning the impact of the function on traffic (in terms of safety and environmental aspects). This has been investigated in detail for different ADAS functions in different projects (e.g. TRACE [14], interactIVe [4], eIMPACT [2]), but so far not for automated driving functions. A general framework for the impact assessment was set up with the Nine-Safety Mechanism approach from Draskoczy et al.[15] that describes different areas that can be affected by an intelligent transport systems. In the safety impact assessment often the main focus is on the direct influence of a function on relevant accidents. It is analysed how the accident flow changes with the function. For this purpose different methods have been developed and used in the past. One often used method is the accident resimulation, where a reconstructed accident is simulated taking into account the function under test.

The user-related assessment is the third pillar of the PReVAL evaluation approach. One key aspect in the reported experiment is measuring of the user's acceptance on the developed function or system. Another important aspect in the user-related assessment is measuring the performance of the interaction between the user and the function under tests. Most of the experiments have been carried out in simulators. Some experiments with automated driving applications were also conducted on test tracks with demonstrator vehicles (e.g. SARTRE[9], EnergyITS[12]). Next to the user-related studies in simulator or on test tracks experiments have also be carried out on public roads. However, tests in this environment are only reported for (A)DAS functions (euroFOT [16], TeleFOT [17], etc.). In particular Intelligent Speed Adaptation Systems(ISA) were tested in the field (e.g. PROSPER [18], EVSC [19]).

Another technical evaluation type focuses on the benchmarking of functions and systems. This benchmarking can focus on the evaluation of the performance to make different systems comparable for customers (e.g. Euro NCAP [20]). Another type of benchmarking test is the comparison of the performance of different systems during a competition (e.g. DARPA Grand Challenge [21], DARPA Urban Challenges [22], Grand Cooperative Driving Challenge [23]). For both types within the assessment standardized test cases are analysed. The test cases as well as the rating procedure are defined before the tests independently from the functions or systems under tests. The tests are typically carried out on closed test environments. Simulations are normally not used in this type of evaluation.

Additionally, at the end of the development a safety validation can be conducted to ensure functional safety and a safe use of the end product. Validation of the functional safety evaluates, whether the built-in safety concept is adequate and sufficient in the case of malfunctioning of the system. Safe usage will check whether the driver is aware of the current system state and can be expected to react in an adequate way to system requests or when system limits are reached (this applies for false negative and in particularly false positive behaviour). Standards and guidelines with respect to the evaluation of functional safety (ISO 26262) [24] and a Code of practice for ADAS (RESPONSE 3) [25] already exist. It is currently an ongoing discussion how functional safety for automated driving including perception of the environment can be ensured, since this cannot be fully achieved by tests in the field [26]. The high effort of field tests are shown by previously conducted filed studies as for example euroFOT [16].

2.3 Evaluation approach in AdaptIVe

The evaluation of the AdaptIVe functions is split into four assessment types (analogous to the PReVAL and interactIVe approach):

- Technical assessment
- User-related assessment
- In-traffic assessment
- Impact assessment

In the technical assessment the performance of the functions is investigated. The user-related assessment analyses the interaction between the functions and the user as well as the acceptance of the developed functions. The in-traffic assessment focuses on the effects of automated driving on the surrounding traffic as well as non-users. The impact assessment determines the potential effects of the function with respect to safety and environmental aspects (e.g. fuel consumption, traffic efficiency). The overall approach for the evaluation in AdaptIVe is shown in Figure 2.3.

The initial starting point for the evaluation is a detailed description of the function[1] or system[2] under investigation itself. Based on the description of the function or system a classification is done in order to determine which evaluation methodology for a certain assessment is most appropriate.

[1] A function in the context of the AdaptIVe project is a functionality that performs a certain driving manoeuvre. Examples are the lane following or the lane change functions.

[2] System means in the context of AdaptIVe a bundle of functions that is combined to a automated driving system that can handle different driving manoeuvre (e.g. City Chauffeur).

In the first step, the AdaptIVe functions and systems are classified according to the SAE classification [7] and the automation level they address [8]. The automation level is only one aspect that needs to be taken into account when deciding on the appropriate test method. Another important aspect is the operation time of the function or system that describes how long a function operates while driving, since the operation time is linked to the type of test and the duration of a test. Here, the AdaptIVe functions and systems are divided into two categories:

- Functions that operate only for a short period of time (seconds up to few minutes). Typical examples are automated parking functions and the minimum risk manoeuvre function that defines the vehicle reaction in case of a system failure or if the driver is not responding to a takeover request of the system. These functions are called "**event based**" operating functions in the following.
- Functions that once they are active, can be operated over a longer period of time (minutes up to hours). A typical example for this type of functions is a highway pilot or a motorway automation function. These functions are called "**continuous operating**" functions in the following.

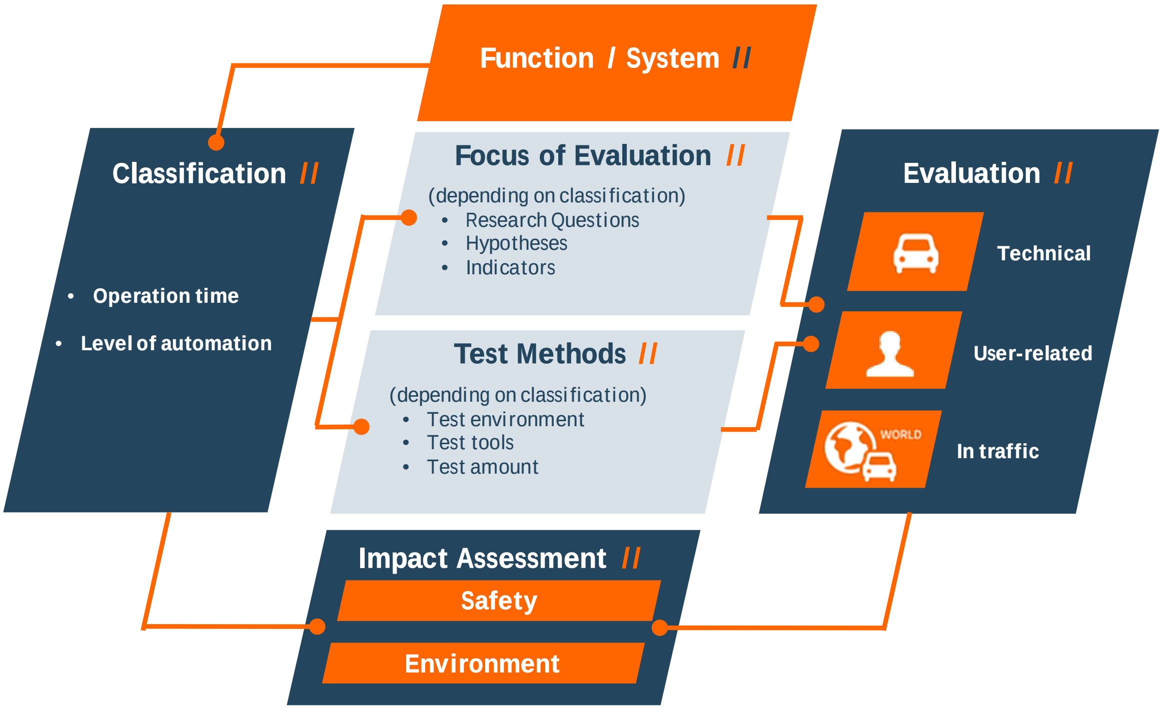


Figure 2.3: Evaluation approach of AdaptIVe

Based on the classification it is decided on the focus of the evaluation and test methods to be applied. The description of the focus of the evaluation includes the definition of research questions that should be addressed by the evaluation. By means of the research questions relevant hypotheses and indicators are derived, similar to the evaluation approach in PReVAL[1]. In AdaptIVe, the research questions, hypotheses and indicators presented in Figure 2.4 have been derived for the technical, user-related and in-traffic assessment. Tables including all research questions, hypotheses and indicators can be found in the corresponding chapters in this document.

Assessment	Research Question	Hypotheses	Indicators
Technical assessment	RQTA(1-35)	HTA (1-45)	ITA(1-38)
User-related assessment	RQUA(1-30)	HUA(1-30)	IUA(1-52)
In-traffic assessment	RQITA(1-6)	HITA(1-9)	IITA(1-8)

Figure 2.4: Research questions, hypotheses and indicators derived in AdaptIVe

With respect to the applied test method depending on the tested function or system it is decided on test environment (e.g. test track, public road, driving simulator) as well as on the required test tools (e.g. balloon cars). Thereby already existing test environments and test tools are tried to be used, as shown in Figure 2.5. One important aspect in this context is also the definition of required test effort respectively amount in order to ensure statistical valid results. This applies in particular for tests on public roads. Here, a trade off between the needed information for the evaluation and the available resources for the evaluation must be found within AdaptIVe. Besides, also other requirements like safety requirements during the tests are taken into account at this point.

Based on the test plan the actual evaluation is conducted in each of the three different assessments. However, as described in the project description of work document, SP7 is not going to test all AdaptIVe functions developed within in the project. The goal of SP7 in AdaptIVe is to show that the developed methodology as described in this document is applicable for different automated driving functions and systems. Therefore the methodologies are applied to selected automated driving functions and systems only during this project.

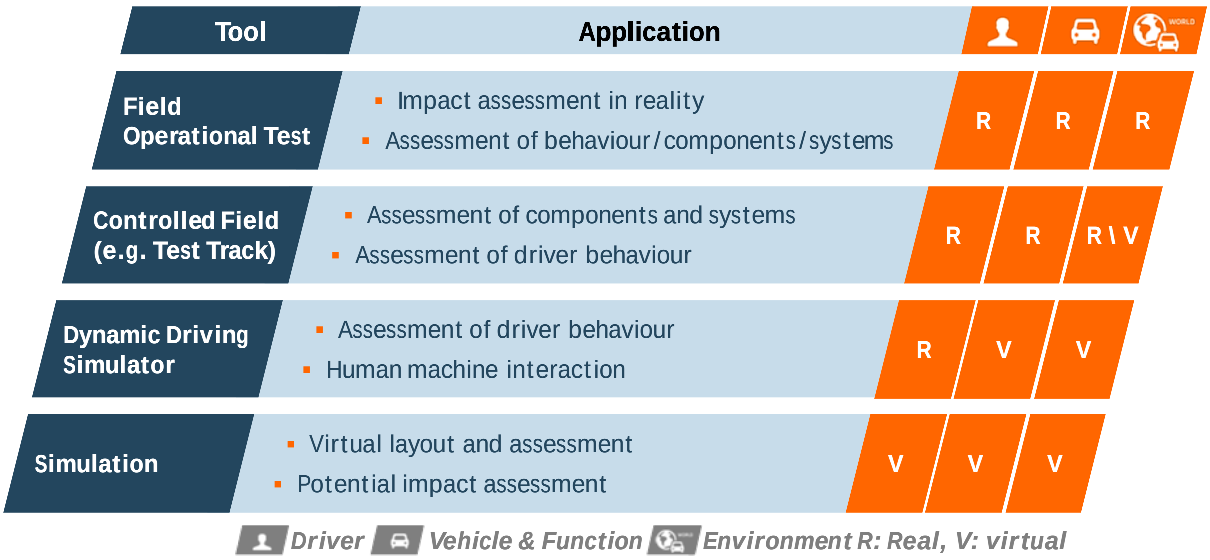


Figure 2.5: Standard evaluation environments and tools according to [27]

The results of the technical, user-related and in-traffic assessment can be used in the last step for the impact assessment. For the impact assessment the classification of the automated driving function must be taken into account, since different systems might require different methods. Similar as for the other assessments the impact assessment focuses on the development of methods.

The developed functions in AdaptIVe are investigated and analysed in different types of situations and scenarios. In order to avoid any confusion, the relevant terms for the evaluation have been defined:

- **Driving Situation**: A driving situation is a specific driving manoeuvre (e.g. a concrete lane change with defined parameters). Thus the driving situation describes in detail a situation that can be simulated and analysed. An example of a driving situation is a lane change at 60.8 km/h with a second vehicle driving at a distance of 10 m behind the host vehicle in the adjacent lane and with a velocity of 65 km/h.
- **Driving Scenario**: A driving scenario is the abstraction and the general description of a driving situation without any specification of the parameters of the driving situation. Thus, it summarises a cluster of homogenous driving situations. Driving scenarios are typically short in time (t < 30 s) and only a few vehicles are involved. An example is lane change to the left lane.
- **Traffic Scenario**: A traffic scenario describes a larger traffic context, which includes different (not pre-defined) driving scenarios. Typically in a traffic scenario a large number of vehicles is analysed over a longer time period. An example of a traffic scenarios could be on a 3 lane high way section of 10 km with 500 vehicles, 2 highway entrances and exits, a speed limit of 130 km/h and investigated time period of 1 h.

Evaluation methodology issues and questions

The objective in SP7 “Evaluation” in AdaptIVe is to develop a complete evaluation methodology for automated driving systems. In order to achieve this objective, different (sub-) questions related to the development of an evaluation methodology need to be investigated and answered. These research questions related to the evaluation sub-project and the evaluation methodology are collected in section 2.3.1 to 2.3.4.

It needs to be distinguished between the general research questions, which are the same for all assessments and the more specific ones, which are only relevant for certain assessments.

Methodology issues and questions deal with the selection of the right method depending on the function and assessment type. Thus, the two main issues and questions are:

- Are different approaches or methods necessary for different types of automated driving functions and systems?

- To what extent can existing evaluation methodologies be also applied to automated driving functions?

The issues and questions related to the evaluation methods and the test tools are presented in the following sub-chapters per assessment type. In the following chapters, some of the issues and questions are already addressed; however some other issues and questions will be addressed during the work in the coming months.

2.3.1 Issues and questions concerning technical assessment

The method issues and questions in the technical assessment in Figure 2.6 focus on two aspects. The first aspect is related to the execution of the tests (test categories, test effort, test cases). The second aspect is related to the evaluation of the test and the test criteria.

Method issues and questions
What are appropriate reference criteria / values for the assessment?
What are adequate performance / validation / verification indicators?
How much test repetitions / test km are required for testing of automated driving functions / systems?
What are appropriate test categories?
Is the human driver an appropriate reference point for the technical assessment? And how is the human behaviour?
How to judge whether a manoeuvre is reasonable or appropriate?
How can (system) safety be assured before going from test track testing to real in-traffic testing?
How to create complex scenarios on test tracks for testing? How to synchronize and create repetitive scenarios? How to automate?
How to define test cases in order to cover all ranges of functional performance with limited time and resources? How to limit the costs of testing?

Figure 2.6: Method issues and questions with respect to the method in the technical assessment

For the test tools, the main issues and questions are related to the logging of data and in particular of Ground-Truth-Data, see Figure 2.7.

Issues and questions concerning test tools
Which tools should be used for the technical assessment?
How to log Ground-Truth-Data (in particular in the field)?
Which signals are needed for the assessment and are they available for the evaluation?
What format and which data will be available for the evaluation (data, video)?

Figure 2.7: Method issues and questions with respect to the tools in the technical assessment

2.3.2 Issues and questions concerning user-related assessment

When selecting evaluation methods and tools for user-related assessment, there are a number of issues and questions to be taken into account. These issues and questions are shown in Figure 2.8 and Figure 2.9 below.

Method issues and questions
How to measure drivers' experience of automated driving?
How to measure driver behaviour in interaction with driving automation?
How to measure long-term changes in driver behaviour in interaction with driving automation?
How to measure overreliance on automation?
How to measure complacency?
How to measure situation awareness?
How to measure driver performance?
How to measure if drivers detect automation failures?
How to measure how automation failure is experienced by the driver?
How to measure the driver's strategy to handle automation failure?
How to measure drivers' regaining control
How to measure whether transfer of control is affected by mental workload?
How to measure driver acceptance?
How to measure drivers' willingness to have/to pay?
How to reveal driver opinions about the system?
How to measure how non-users’ behaviour is influenced by interaction with equipped vehicles?
What scenarios to use to study drivers' interaction with well-functioning driving automation?
What scenarios to use to study drivers' reaction to unexpected functioning, such as automation failure?
What scenarios to use to study non-users’ reaction around automated vehicles?
How should the driver sample look like (distribution by age, gender, driving experience, etc.)?

Figure 2.8: Method issues and questions in user-related assessment

Issues and questions concerning test tools
Are the necessary tools for user-related tests (driving simulator, test track) available to carry out the study?
Are target objects (mock-up vehicles, pedestrians, etc.) for user-related tests available?
Is the vehicle allowed to be driven by naive drivers on public roads? How can safe testing be ensured with naive drivers?
Can all relevant scenarios be covered?
Can all necessary data be logged in the test vehicle?

Figure 2.9: Issues and questions concerning the tools in user-related assessment

2.3.3 Issues and questions concerning in-traffic assessment

For in-traffic assessment, a number of issues were identified with respect to the applicability of various known test tools, with a specific focus on virtual in-traffic assessment. The framework that can be used to answer these issues and questions will be described in Chapter 5.

A method issue and question that is important for in-traffic assessment is safety compliance of the vehicle equipped with the (prototype) automated system. With this difficulty in mind, virtual in-traffic assessment with a Monte Carlo simulation approach is adopted as a viable method. Method issues and questions deal with how to simulate the surrounding traffic, the infrastructure and specifically the variability in scenarios that exists in traffic. Figure 2.10 summarises the identified method issues and questions.

Method issues and questions
How can (system) safety be assured before going from test track testing to real in-traffic testing?
How could a full evaluation of in-traffic behaviour look like?
Should automated and non-automated surrounding traffic be taken into account?
How should a (Monte-Carlo) simulation look like for in-traffic assessment?
Which (test track) tests describe best the general/normal system behaviour?
Is a special device needed to perform the in-traffic assessment and what would be the requirements for such device?
What parameters/signals describe in-traffic behaviour best?
What different infrastructure layouts should be taken into account?

Figure 2.10: Method issues and questions concerning the in-traffic assessment

Issues around test tools required for in-traffic assessment deal with the availability and quality of data and models that sufficiently represent reality. Figure 2.11 summarises the identified issues concerning test tools.

Issues and questions concerning test tools
To what extend can simulations be used to evaluate in-traffic behaviour?
Are databases available (or becoming available) on 'normal' in-traffic behaviour to be used in simulations?
How should a Monte-Carlo simulation be set up to provide adequate results?
Which (test track) tests describe best the general/normal system behaviour?
Is a special device needed to perform the in-traffic assessment and what would be the requirements for such a device?

Figure 2.11: Issues and questions concerning the tools in the in-traffic assessment

2.3.4 Issues and questions concerning impact assessment

The general question for the impact assessment is: what are appropriate methods for the impact assessment. It must be considered that automated driving function will be operated over a longer time frame, which will influence the overall traffic in a different way compared to functions that operate only for a couple of seconds as for instance safety related ADAS functions. This needs to be considered in the impact assessment. Hence, the mentioned issues and questions are related to the selection of the situations respectively scenarios for the impact assessment. A further important issue is the validation of the developed methods. The main issues and questions related to the method are presented in Figure 2.12.

Method issues and questions
How does an impact assessment method for automated driving functions / systems look like (w.r.t. safety, traffic flow, fuel consumption and travel time)?
Which functions / systems should be assessed in AdaptIVe to demonstrate the method?
Which penetration rates should be considered for the impact assessment?
How can the results of the impact assessment be validated?
What kind of situations is relevant for the impact assessment?
Does a take-over situation due to system failure or system limits affect traffic safety?
Are critical situations expected by false positives of the function?
Are critical situations expected by overreliance or other changes in driver behaviour with the system?
Which situations (e.g. accidents) are influenced by the system? And which way are the situations influenced?
What's the ratio of false-positive activations to true-positive activations?

Figure 2.12: Issues / questions concerning the method in the impact assessment

The main focus within the impact assessment is on the development of appropriate methods and fewer issues for the test tools in particular for the input data that has been identified, see Figure 2.13.

Issues and questions concerning test tools
Which tools can be used to simulate the effects of automated driving functions?
Which are appropriate input / output variables for simulations?
How to derive baseline and with system data?
Which results and input parameters (change in driver behaviour with system, driver-system-interaction in take over situation) are available for impact assessment? (also from the other assessments)

Figure 2.13: Issues and questions concerning the tools in the impact assessment

2.4 Time plan for the evaluation

In this chapter, the time plan for the evaluation and testing is presented. It is obvious that the time plan also depends on the progress in other SPs – in particular in the VSPs that develop the AdaptIVe systems and functions. Therefore, at the current stage only a rough planning is possible. A more detailed planning of the tests together with VSPs will be done when it is foreseeable that the demonstrator vehicles will be ready.

The time plan for the evaluation is given in Figure 2.14.

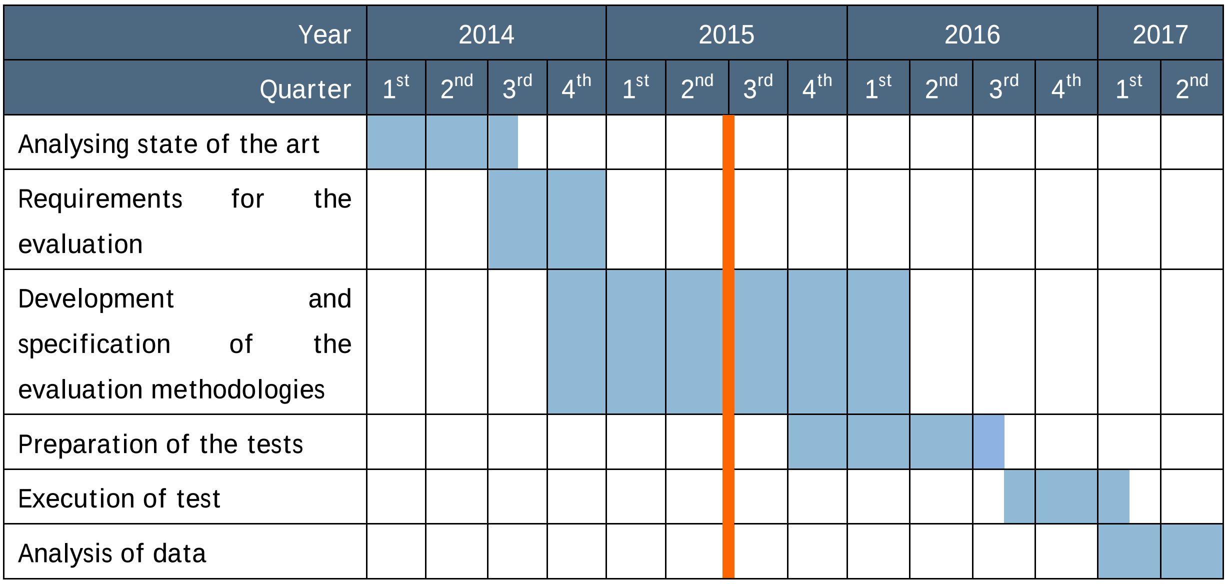


Figure 2.14: Time plan for the evaluation in AdaptIVe (orange line: current time point)

The work in the evaluation sub-project started in January 2014 with analysis of the state of the art with respect to evaluation frameworks for automated driving systems. Based on the state of the art analysis the requirements for the evaluation were defined in an internal report. The outcome of requirements is the basis for this deliverable.

In the 4th quarter of 2014, the definition and specification of the evaluation methodologies was started. The general approaches for the evaluation of an automated driving function are also presented per assessment type in this report. The work in particular the specification and more detailed definition of the evaluation methodology will be continued until 2nd quarter of 2016. In the 4th quarter of 2015, the preparation of the actual tests will start. This task includes the detailed planning of the tests together with the VSPs as well as the selection and (if needed) development of the required tests tools.

The tests will be conducted from the 3rd quarter of 2016 onwards (in case a demonstrator vehicle is available before this time point, also the testing for this vehicle can start earlier). According to DOW the demonstrator vehicle should be available from 30th September 2016 on. This means that the tests will respectively need to be conducted in autumn or winter. It is obvious that not all weather condition will allow testing, for instance a snow cover road. Thus, these weather condition need to be avoided, which implies that the test should be conducted as early as possible. Based on the experience in the previous projects, it is expected that a test of one demonstrator for one assessment will require 2 to 3 weeks. In the remaining time the data will be analysed and results will be documented.

3 Technical Assessment

This chapter describes the evaluation approach for the technical assessment of automated driving functions and system as described in AdaptIVe[5]. In the first section 3.1 the focus of technical evaluation is described. Then, the relevant research questions, hypotheses and indicators are presented. The different testing environments for the technical assessment are discussed in section 3.2. Section 3.3 discusses the requirements for technical assessment with respect to safety, test-tools and test conduction. Finally, in section 3.4 an example of the application of the evaluation methodology for two automated driving functions (event-based and continuously operating) is given.

The objective of the technical assessment methodology developed in AdaptIVe is to set up a general evaluation framework for the evaluation of automated driving functions or systems. The focus within this assessment is on the technical performance of the functions and systems. A major challenge in setting up the evaluation framework is to limit the test effort in order to guarantee an efficient evaluation while ensuring that all important aspects are covered to guarantee a most complete evaluation. Since automated driving systems address the whole driving process, nearly all driving situations need to be taken into account for the evaluation – although not every function addresses all driving situations[3]. It is obvious that detailed analyses for each driving situation might be desirable, however considering the limited resources; it is not feasible, as shown in Figure 3.1.

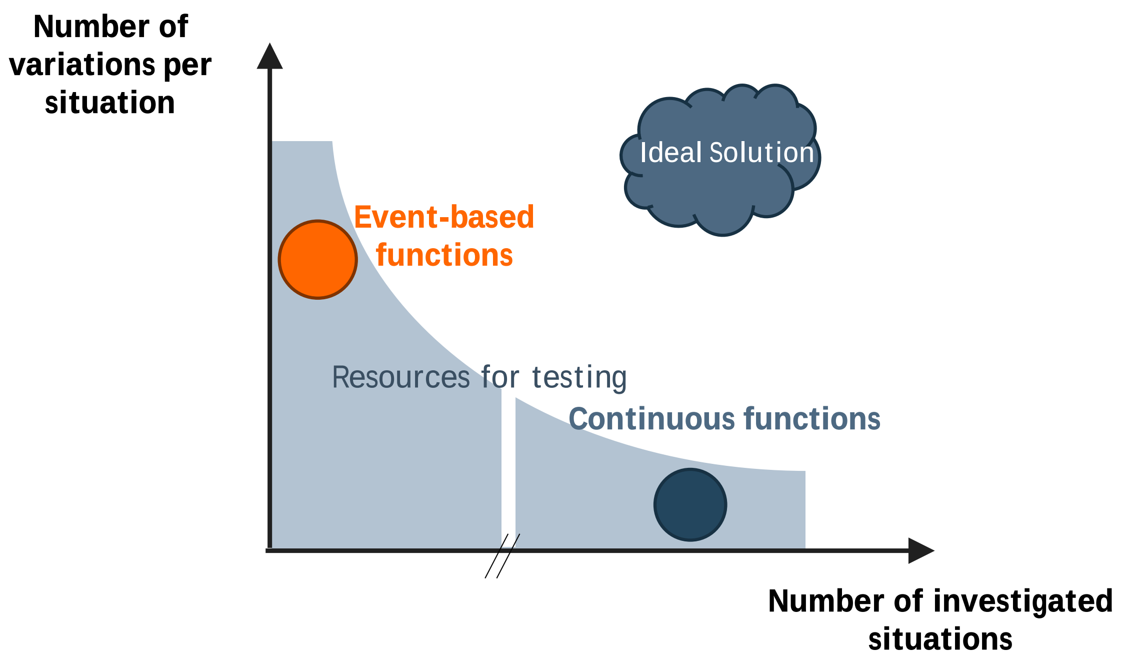


Figure 3.1: Resources for testing of continuous and event-based-functions

[3] In this case the not addressed situations will obviously not be tested for the function.

In the past, for the evaluation of ADAS functions often a use case based approach (e.g. interactIVe [13]) had been used in the technical assessment. This means that first the use cases had been determined based on relevant situations e.g. accidents. Afterwards the test cases had been described based on these use cases. By variation of the test conditions the function performance had been analysed in detail for the test cases. Thus, the test effort is highly depended on the amount of covered use cases. If a function covers nearly all driving situations, this will result in an unfeasible high number of test cases.

For an automated driving function it is distinguished between event-based operation functions and continuously operating functions (as described earlier). For the event-based operating functions a similar approach as for the ADAS function seems to be applicable, since also for these functions typically the focus is on a low number of use cases. Thus, based on the use cases the test cases can be clearly described and the test effort is limited. However depending on the automation level of the tested function a higher test effort compared to ADAS functions might be necessary.

For continuously operating systems, on the other hand, it is difficult to identify certain use cases, since for the system the whole automated drive is already the use case. Within this automated drive different driving situations will occur. These situations can be mapped to use cases of sub-functionalities, like e.g. automated lane change functionality. However, for the technical assessment the focus needs to be first on the whole system, since the user will only experience the whole system and the (sub-) functionalities will only operate as a bundle. Sub-functionalities can also be evaluated within the technical assessment. This will be conducted only in a second step – for AdaptIVe this is depending on the available resources.

Since the use-case based approach seems to be not applicable for continuously operating functions, the focus needs to be shifted. Instead of investigating certain test cases in detail a broader approach will be taken. This means that the objective is to investigate many different driving situations. Considering limited resources this means that the driving situations cannot be analysed on the same level of detail as for the event-based functions.

In the following, the evaluation for both functions types is described in more detail.

Event-based operating functions

For the evaluation of event based operating functions an approach similar to the evaluation of today's ADAS functions, which intervene only for a short time, is used. For ADAS functions the evaluation methodology is well known and already described in former projects (e.g. PReVAL [1], interactIVe [13]). The sequence of the evaluation process is given in Figure 3.2.

According to these projects the first step is the formulation of the research scope by means of the research questions. Therefore, the functions description needs to be analysed in order to decide on which aspect it should be focused during technical assessment. Based on the research question hypotheses are defined which are then analysed during the technical assessment. For this process, adequate performance indicators and evaluation criteria are chosen.

Once the definition of the evaluation requirements is finished the relevant test cases are defined. The basis for the definition of the test cases are normally the function's use cases respectively situations that have been defined as relevant (e.g. certain accident scenarios). The definition of the tests cases should also include a risk assessment in order to ensure safe testing.

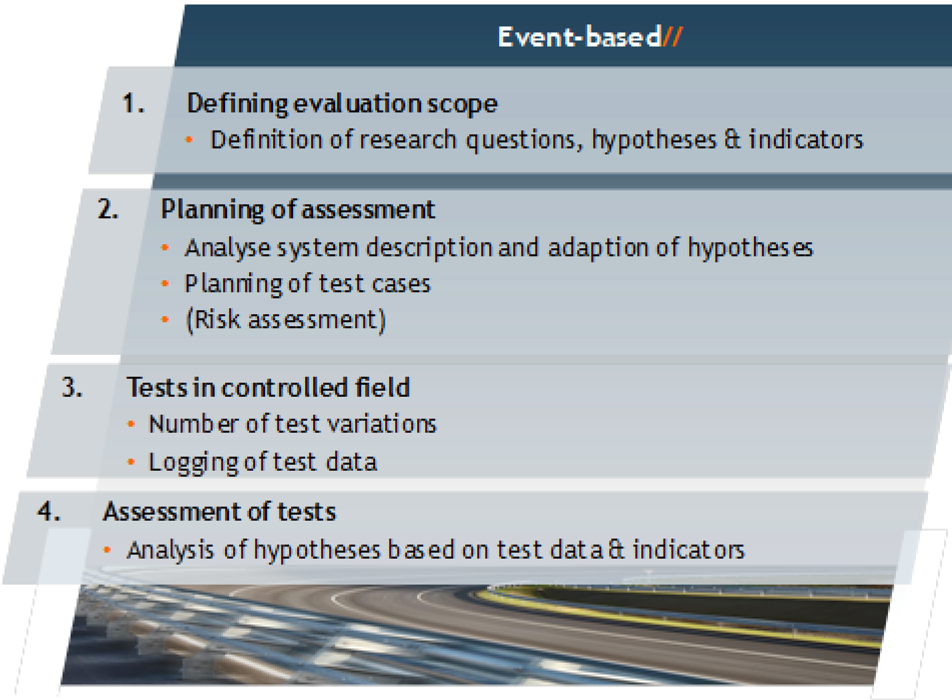


Figure 3.2: Test sequence for event-based operating functions

The actual testing is the second last step of this approach. The tests are typically conducted in a controlled field – mainly test track or closed test garage for parking scenarios. A description of possible test cases that has been derived based on the AdaptIVe use cases can be found in Annex 1. During the testing itself the parameters of the test case (e.g. velocities or relative distances) are varied. However, usually not all parameters are varied due to time and cost restraints. Therefore it needs to be distinguished between varied and fixed parameters. Ideally, each test – a test case with certain parameter set - is repeated several times in order to ensure statistically valid results, as e.g. described in Figure 3.3. Regarding the testing itself it should be avoided to conduct tests in adverse weather conditions[4], which might degrade the performance of the assessed function. Due the project timing (see chapter 2.4) this item can affect the work during evaluation in AdaptIVe.

[4] except the test does not plan tests in adverse weather condition or lighting conditions

The evaluation of the test data is the last step of the methodology. This step includes the calculation of derived measures as well as indicators. Derived measures are signals that cannot be directly obtained during the test. Instead they need to be calculated during the evaluation. A typical example is the Time-To-Collision (TTC), which describes the remaining time to collision in case the current movement of the vehicles is kept constant. Indicators on the other hand are single values that describe the test run in a certain way. Examples are the maximum, minimum or mean values of signals respectively of derived measures.

Based on the indicators the analysis of the hypotheses can be performed. The probability that the defined hypothesis results as true is calculated on the basis of statistical hypothesis testing. For the technical assessment in AdaptIVe, the defined (null) hypotheses are tested against the distribution of the appropriate indicators which are recorded during the test drive, see Figure 3.3.

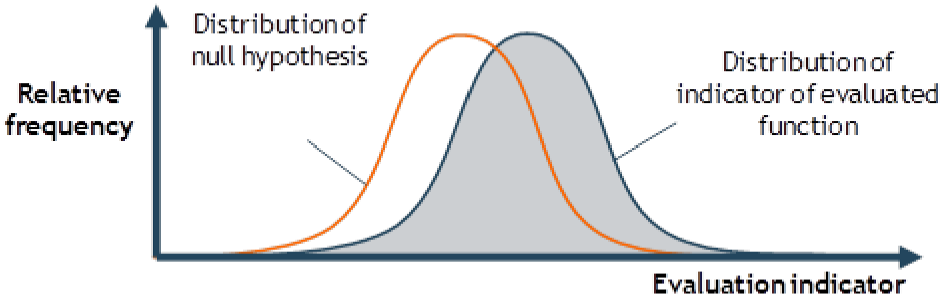


Figure 3.3: Statistical testing of null hypothesis

The hypotheses are tested within the technical assessment for a pre-specified level of statistical significance. The level of statistical significance is attained when a p-value is less than the significance level α. The p-value is the probability of observing an effect given that the null hypothesis is true whereas the significance level is the probability of rejecting the null hypothesis given that it is true [28]. As a matter of good scientific practice, the significance level is chosen before data collection and is usually set to 0.05 (5 %) according to [29]. However, it must also be considered that the AdaptIVe project is a research project. Hence, the developed functions will not be as mature as market ready systems, which mean for these systems that a higher variation within the results can be expected and accepted.

Continuously operating functions

For the continuously operating functions the focus is slightly different in the technical assessment. Since these functions cover different driving situations, also for the assessment a wider scope is necessary. For the continuously operating functions the focus is less on the performance in a certain driving situation but more on overall performance during the whole driving process. Therefore, it seems not be useful to define certain single test cases for these functions. Instead a holistic evaluation approach that covers as many different driving situations as possible is needed. Such an evaluation approach is delivered by a (small) field test, in which the function must be able to handle driving situations that are covered according to the function's specification and occur during the test drive. It will not be sufficient to just cover some driving situations, since it cannot be predicted before the tests under which conditions certain driving situations will occur and how often these driving situations occur. The drawbacks of the field test approach are that it is a rather uncontrolled test set-up and that the effort for a field test is in general quite high with only limited general validity of the results. Therefore, the extent of the field test needs to be limited to a feasible amount.

In order to investigate the performance over the whole driving process adequate indicators are needed. Besides the indicators, also the baseline to which the function behaviour is to be compared needs to be described. For this purpose the basic requirements of automated driving functions and systems needs to be considered. These requirements are:

- safe driving,
- to operate in mixed[5] traffic conditions,
- not affect other traffic in a negative way.

These basic requirements imply that automated driving systems need to operate within the range of normal driving behaviour and should at least be as safe as non-automated driving. Thus, the baseline for the assessment should be the human driver respectively his/her behaviour. Since the driving behaviour of each human driver is different, it can only be described in distributions. These distributions of driver behaviour need to be obtained before the actual assessment is performed. For obtaining these distributions two approaches will be used in AdaptIVe:

- Data of previous field test projects will be used (e.g. filed operational tests like the euroFOT project), since it provides information on the driving behaviour of many different drivers.
- Each test route will be driven several times with and without the system in order to consider specific characteristics of the test region

Next to the distribution of normal driving it can also be analysed whether these boundaries of normal driving are violated during the test with the function. Next to the boundaries related to the normal driving behaviour also legal boundaries must be considered for the evaluation (e.g. speed limits, restriction on passing). However, also the severity of a violation must be taken into account in order to differentiate a slight violation, which might have only a small or even no

[5] Meaning automated and non automated traffic

effect on traffic, from a major violation that can have a serious effect. Additionally, a violation to one side of the defined boundaries (e.g. slower driving as most of the other human drivers) might be less critical compared to a violation to the other side (driving too fast). This needs to be considered for the classification of the violations.

After the more general introduction of the approach for the technical assessment of continuously operating functions the different steps of the approach, as given in Figure 3.4, are described.

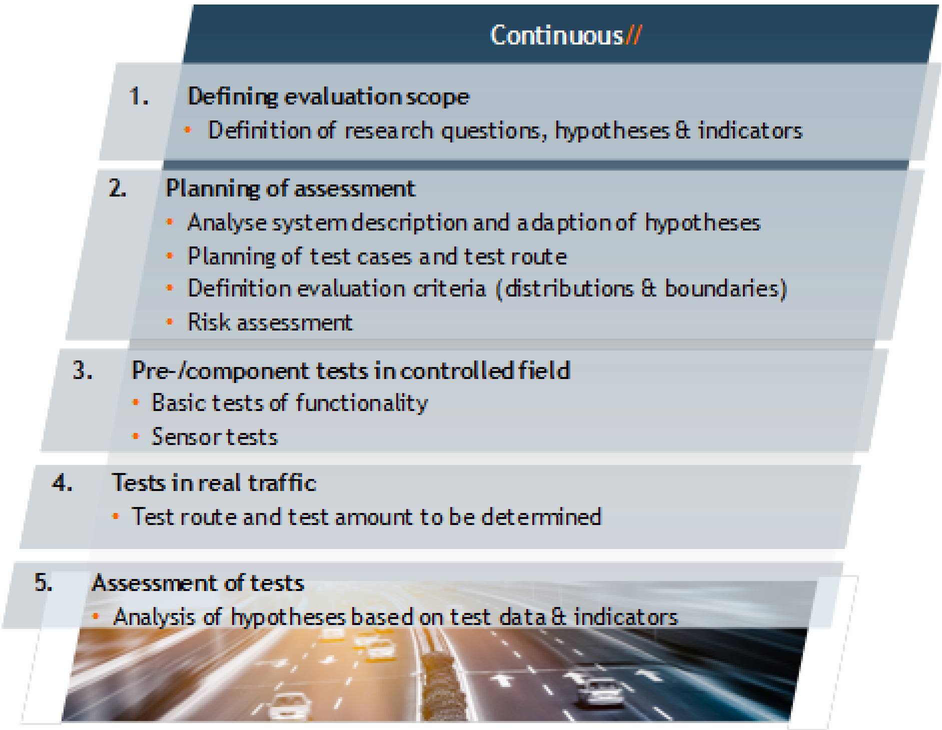


Figure 3.4: Test sequence for continuous operating functions

Analogue to event-based functions, the evaluation approach of continuously operating functions starts with the definition of the research questions and hypotheses (1). This includes also the definition of adequate performance indicators and criteria for the assessment of the hypotheses.

In the next step the actual tests are prepared (2). This includes the definition of the test route and test length. To limit the test effort to a feasible extend, the test route has to be chosen in a way that all relevant driving situations are covered. Therefore, the required test length respectively duration needs to be estimated a priori based on the number of expected relevant driving situations that occur while driving in public traffic. For this estimation, the data and the knowledge gained in previous field operational tests is used. Before the actual test starts a risk assessment is conducted in order to ensure a safe testing.

The actual tests are split into two steps. The first step is the pre-tests (3), in which the basic functionality is checked as well as the accuracy of the used sensors is analysed. These tests are conducted similar to the event-based functions tests on a test track. If the tests have been finished and the function is operating properly, the main test on public roads can start (4).

The test on public roads will take place on the defined test route. The test route will be defined in cooperation between SP7 and the VSPs. During the tests the test route will be driven several times. Before and after the test at least one test drive without the function should be done. It is recommended to conduct test drives at comparable time frames in order to minimize the variation between the test drives. It is also suggested to avoid time frames with heavy traffic (e.g. rush-hour in the morning and evening). Furthermore, tests in adverse weather conditions should be avoided – expect of course for the tests that explicitly foresee tests in certain weather conditions.

After the tests, the data is evaluated analogue to the process described for the event-based function (5). This means that required derived measures and indicators are calculated and that the hypotheses are analysed. Besides to the analysis of the whole test drives also certain driving situations might be of interest. A list of relevant driving manoeuvres is given in Annex 2.

3.1 Focus of technical assessment

This chapter discusses the focus of the technical assessment of automated driving functions and systems. The objective of the AdaptIVe evaluation is to set up a general evaluation framework for automated driving functions and respective systems. Therefore, the focus of the technical assessment is described in a general and generic way. This means that adaptation might be necessary in order to cover also special aspects of certain functions.

As described in the previous chapter the chosen approach follows the standard scientific evaluation procedure by defining relevant research questions in the first step. In this subchapter defined research questions consider different technical aspects, which are presented separately. Based on the research questions the related hypotheses that will be assessed are derived. In order to analyse hypotheses, indicators are required, which are described in the last part of the chapter.

3.1.1 Research questions

This chapter deals with the research questions for the technical assessment. The research questions are the first step of the evaluation and provide information on what should be evaluated in the technical assessment. The research questions are clustered by the evaluation aspects requirements/verification, performance in driving situation, sensor performance, trajectory & planning, safety and function misbehaviour or failure.

The evaluation aspect **requirements/verification** is dealing with research questions related to the basic functionality as well as the fulfilment of the specified requirements of the developed system. Thus, the basic research questions for this aspect are:

ID	Research Question	Function		Addressed level of automation
		Event based	Continuous	
RQTA1	Does the function or system fulfil its functional specification (e.g. speed range, weather conditions)?	x	x	All
RQTA2	Does the function or system inform about its status / the conducted manoeuvre (as described in the technical requirements)?	x	x	All
RQTA3	Is it technically possible to activate / deactivate / override the function or system (at any time)?	x	x	All

Figure 3.5: Research questions for the evaluation aspect “requirements/verification”

Regarding the aspect **performance in driving situation**, research questions related to the driving behaviour and the performance of the automated vehicle in different driving situations is clustered in the following table.

ID	Research Question	Function		Addressed level of automation
		Event based	Continuous	
RQTA4	How much variation occurs for one manoeuvre (considering environmental conditions)?	x		All
RQTA5	Is the function or system performance affected by different environmental conditions	x	x	All
RQTA6	Is driving with the function or system comfortable? Is the driving behaviour of the system / function in line with normal driver behaviour (-> speed, distance behaviour)?		x	All
RQTA7	Does the function or system reduce fuel consumption?		x	All
RQTA8	How much does the distance between the vehicle and relevant obstacles or boundaries vary?		x	All
RQTA9	How accurately the manoeuvre is conducted? (Planning or assumption vs. driven trajectory)	x		Partial
RQTA10	How long does the designated manoeuvre take compared to a human driver?	x	x	Partial
RQTA11	How long does the vehicle drive without driver intervention?		x	Conditional & High
RQTA12	Does the function or system react appropriately to situations, in which the driving lane ends		x	Partial

ID	Research Question	Function		Addressed level of automation
		Event based	Continuous	
RQTA13	Does the function or system react appropriately to situations, which require a slower velocity than the given speed limit?		x	Partial

Figure 3.6: Research questions for the evaluation aspect “performance in driving situation”

The criteria **perception** is dealing with research questions that are relevant for the perception performance of the environment:

ID	Research Question	Function		Addressed level of automation
		Event based	Continuous	
RQTA14	Are all (relevant) static and dynamic objects in sensor range detected?	X	X	All
RQTA15	How big is the difference between sensor measurement and reference measurement?	X	X	All
RQTA16	Which sensor area is covered by the function/system?	X	x	All
RQTA17	What is the range of the V2X communication (if available)?	X	x	All
RQTA18	Can the function or system detect right of way situation / intersection correctly?	(x)	x	Conditional & High

Figure 3.7: Research questions for the evaluation aspect “perception”

The aspect **trajectory & planning** is dealing with all issues related to the planning of the (upcoming) driving manoeuvre:

ID	Research Question	Function		Addressed level of automation
		Event based	Continuous	
RQTA19	Is the conducted manoeuvre reasonable according to the situation? / Are the conducted manoeuvres while driving reasonable according to the situation?	X	x	All
RQTA20	Is driving with the function or system comfortable? Is the trajectory of the vehicle jerk-optimal?	X	x	All
RQTA21	Which acceleration can the function or system apply in case of emergency manoeuvre?	X		Partial
RQTA22	Are the vehicle dynamics of the trajectories in range of average driving profiles / standards (e.g. ISO)		x	Partial
RQTA23	Is the function or system able to determine the		x	Partial

ID	Research Question	Function		Addressed level of automation
		Event based	Continuous	
	optimal trajectory? (e.g. in parking manoeuvre)			
RQTA24	Does the function or system show reactive behaviour within a defined time interval?		x	Conditional & High
RQTA25	Is the current status of the car taken into account? (Fuel level, engine warnings…)	X		Conditional & High

Figure 3.8: Research questions for the evaluation aspect “trajectory & planning”

The evaluation aspect **safety** investigates how safe driving with the automated driving functions / system is. Therefore, in particular critical driving situations will be analysed:

ID	Research Question	Function		Addressed level of automation
		Event based	Continuous	
RQTA26	Do collisions with other objects occur during testing?	x	x	All
RQTA27	Is driving with the function or system safe? What is the minimum distance or time distance to other objects during the manoeuvre?	x	x	All
RQTA28	Is the function or system able to perform emergency manoeuvres?	x	x	All
RQTA29	How often do situations with TTC [6]< TTCcritical (respectively TLC [7]< TLCcritical) occur?		x	All
RQTA30	Does the function or system conduct any emergency manoeuvres (longitudinal acceleration < threshold, evasive manoeuvre, safe stop) during the test?	x	x	All
RQTA31	Is the function or system able to perform emergency manoeuvres?	x	x	All

Figure 3.9: Research questions for the evaluation aspect “safety”

Finally, the criteria **function misbehaviour** is dealing with all issues related to any misbehaviour or misdetection of the function/system that occurs during the tests. Since the in AdaptIVe developed functions are still research functions, it would be unrealistic to expect same maturity of the function as of functions that have been introduced into the market. This aspect needs to be considered during the evaluation.

[6] Time-to-Collision (TTC) describes the time which remains at stationary driving conditions until a collision occurs.
[7] Time-Line-Crossing (TLC) is defined as the time duration available for the driver before any lane boundary crossing.

ID	Research Question	Function		Addressed level of automation
		Event based	Continuous	
RQTA32	Is any misbehaviour of the function or system (false positive and negative) detected during the test?	x	x	All
RQTA33	Does misbehaviour due to false behaviour-decision occur?	x	x	All
RQTA34	Does misbehaviour due to false detection and classification of situations occur?	x	x	All

Figure 3.10: Research questions for the evaluation aspect "function misbehaviour"

3.1.2 Hypotheses

Based on relevant research questions, hypotheses have been defined. The hypotheses are presented in analogy to research questions by the different evaluation aspects (requirements/verification, performance in driving situation, sensor, trajectory & planning, safety and function misbehaviour). Similar to the research questions the hypotheses are formulated in a general way. Therefore, for the hypotheses some adaptation might be required in case any special aspect of a function should be covered. Within the evaluation in AdaptIVe not all hypotheses will be assessed. Instead only relevant hypotheses will be selected.

The evaluation aspect **requirements/verification** is dealing with following hypotheses:

ID	Hypotheses	Reference	Related Research Question
HTA1	The function or system operates only within the defined speed range.	Function or system specification	RQTA1
HTA2	The function or system does not operate in the not defined e.g. weather or lighting conditions.		
HTA3	The function or system operates within the maximum sensor range.		

Figure 3.11: Hypotheses for the evaluation aspect "requirements/verification"

Hypotheses related to **performance in driving situation** are:

ID	Hypotheses	Reference	Related Research Question
HTA4	The standard deviation of the x,y - position for one manoeuvre for multiple tests is below a certain threshold (in all / the defined scenario/s).	Threshold (TBD)	RQTA4
HTA5	The sensor detection range is above a certain threshold in all scenarios.	Threshold (TBD)	RQTA5

ID	Hypotheses	Reference	Related Research Question
HTA6	The longitudinal acceleration is in range of the distribution of human driver behaviour.	Human driver (distribution)	RQTA6
HTA7	The lateral acceleration is in range of the distribution of human driver behaviour.		
HTA8	The difference between current speed and speed limit (v-speed limit) are in range of the distribution of human driver behaviour.		
HTA9	The distance headway is in range of the distribution of human driver behaviour.		
HTA10	The time headway is in range of the distribution of human driver behaviour.		
HTA11	Driving with the function or system reduces the fuel consumption.	Driving without the function or system	RQTA7
HTA12	The standard deviation of the position in lane is below a certain threshold when driving in lane.	Threshold (TBD), distribution	RQTA8
HTA13	The standard deviation of the x,y - position between planned or driven trajectory is below a certain threshold (in all / the defined scenario/s).	Threshold (TBD) or reference manoeuvre	RQTA9
HTA14	The time of the designated manoeuvre is within the distribution of the time of human drivers.	Human driver (distribution)	RQTA10
HTA15	The maximum time of the manoeuvre in automated driving mode is below a certain threshold.	Threshold (TBD)	RQTA11
HTA16	The driver request or lane change is in range of the distribution of human driver behaviour when driving lane ends.	Human driver (distribution)	RQTA12
HTA17	The driven velocity in situations, which require a slower velocity than the given speed limit, is lower than the defined threshold for this situation.	Human driver / Defined speed	RQTA13

Figure 3.12: Hypotheses for the evaluation aspect “performance in driving situation”

All hypotheses, which are relevant for the performance of the perception of the automated driving function, are discussed in the table **perception**.

ID	Hypotheses	Reference	Related Research Question
HTA18	No False negative detections occur during the tests.	No false detection or Threshold (e.g. 99 %) of the detection objects need to be correct	RQTA14
HTA19	No False positive detections occur during the tests.		
HTA20	The difference between sensor measurement (angle to target) and reference measurement (angle to target_ref) is below a certain threshold.	Threshold (Reference Measurement)	RQTA15

ID	Hypotheses	Reference	Related Research Question
HTA21	The achieved sensor coverage and range is higher than X m (in all / the defined scenario/s).	Function or system specification (Sensor coverage)	RQTA16
HTA22	The achieved sensor range of the V2X communication is higher than X m (in all / the defined scenario/s).	Function or system specification (defined sensor range)	RQTA17
HTA23	No false positive detections of right of way situations are occurring during the tests.	No false detection of right of way situation	RQTA18

Figure 3.13: Hypotheses for the evaluation aspect “perception”

The evaluation aspect **trajectory & planning** is covered by the following hypotheses:

ID	Hypotheses	Reference	Related Research Question
HTA24	The conducted manoeuvre decision is in line with evaluator's / driver's expectations (in all / the defined scenario/s).	Questionnaire, Driving Simulator with human driver	RQTA19
HTA25	The jerk in x-direction of the trajectory is below a certain threshold.	Threshold (TBD)	RQTA20
HTA26	The jerk in x-direction of the trajectory is below a certain threshold.		
HTA27	The acceleration in x-direction of the trajectory is below a certain threshold.	Threshold (TBD)	RQTA21
HTA28	The longitudinal acceleration is in range of the distribution of human driver behaviour.	ISO-standards / human driver distribution	RQTA22
HTA29	The lateral acceleration is in range of the distribution of human driver behaviour.		
HTA30	The deviation from the optimal trajectory is below a certain threshold.	Threshold (TBD)	RQTA23
HTA31	The reaction time of the system is below the human driver reaction time in (in all / the defined scenario/s)	Human driver reaction time	RQTA24
HTA32	The trajectory in case the function is well functioning vehicle and the trajectory in cases of an impaired vehicle are the same.	Trajectory (well functioning vehicle)	RQTA25

Figure 3.14: Hypotheses for the evaluation aspect “trajectory & planning”

Hypotheses related to the evaluation aspect **safety** are:

ID	Hypotheses	Reference	Related Research Question
HTA33	The number of collisions is zero.	Vision Zero	RQTA26
HTA34	The distance to objects is below a certain threshold (in all / the defined scenario/s).	Threshold (TBD)	RQTA27
HTA35	The time distance to objects is below a certain threshold (in all / the defined scenario/s).		
HTA36	The max. lateral acceleration is above a certain threshold in critical situations.	Threshold (TBD)	RQTA28
HTA37	The max. longitudinal acceleration is above a certain threshold in critical situations.		
HTA38	The frequency of Situations with TTC < TTCcritical is below a certain threshold.	Human driver	RQTA29
HTA39	The frequency of Situations with TLC < TLCcritical is below a certain threshold.	Human driver	
HTA40	The number of emergency manoeuvres is below a certain Threshold.	Ideal Objective: 0 However, need to be decided on test situations and specification (w.r.t. take-over situation -> any situation, in which the driver takes over but the system should be capable to handle the situation)	RQTA30
HTA41	The number of take-over situations is below a certain Threshold.		
HTA42	The max. lateral acceleration is above a certain threshold in critical situations.	Threshold (TBD)	RQTA31
HTA43	The max. longitudinal acceleration is above a certain threshold in critical situations.		

Figure 3.15: Hypotheses for the evaluation aspect "safety"

Finally, the evaluation aspect **function misbehaviour** is dealing with the following hypotheses:

ID	Hypotheses	Reference	Related Research Question
HTA44	The number of false negative behaviour is below a certain threshold.	Threshold (TBD), e.g. zero	RQTA33 – 35
HTA45	The number of false negative behaviour is below a certain threshold.		

Figure 3.16: Hypotheses for the evaluation aspect "function misbehaviour"

3.1.3 Evaluation indicators

The derived hypotheses are evaluated by using indicators. These indicators are calculated based on signals or derived measure logged during the test. A list of signals that should be logged during the tests - if possible - is provided in Annex 3.

In the following table, all relevant indicators are presented and linked to the hypotheses defined in the previous subchapter. At this point it must be considered that certain functions might require adaptations of the hypotheses and by this also of indicators due to the nature of the function or the used measurement equipment. Therefore, the table presents only a general approach of indicators for the technical assessment.

ID	Indicators	Evaluation Aspects	Related Hypothesis
ITA1	Max. speed, min. speed	requirements/ verification	HTA1
ITA2	Function status at weather conditions		HTA2
ITA3	Maximum distance to target		HTA3
ITA4	Distribution of x,y- position	performance in driving situation	HTA4
ITA5	Sensor detection range, availability of the function or system		HTA5
ITA6	Distribution of different signals (e.g acceleration, velocity, distance to other objects)		HTA6 –HTA10
ITA7	Fuel consumption		HTA11
ITA8	Distribution of position in the lane		HTA12
ITA9	Standard deviation of delta x,y-position		HTA13
ITA10	Time of driving manoeuvre		HTA14
ITA11	Max(time of driving manoeuvre(automated))		HTA15
ITA12	Remaining time to end of lane at driver request or lane change		HTA16
ITA13	Driven speed - threshold (situation depending)		HTA17
ITA14	Number of false negative detections	perception	HTA18
ITA15	Number of false positive detections		HTA19
ITA16	Distance in x-direction to target, distance in y-direction to target, angle to target		HTA20
ITA17	Sensor coverage + sensor range		HTA21
ITA18	Sensor range		HTA22
ITA19	Number of false detections of right of way situations		HTA23
ITA20	Minimum TTC / distance, acceleration, evaluator / drivers' opinions,	trajectory & planning	HTA24
ITA21	Number of jerks in x-direction, jerks in y-direction		HTA25 – HTA26

ID	Indicators	Evaluation Aspects	Related Hypothesis
ITA22	Range of longitudinal acceleration		HTA27
ITA23	Distribution of longitudinal and lateral acceleration		HTA28 – HTA29
ITA24	Deviation from optimal trajectory		HTA30
ITA25	Reaction time of the system		HTA31
ITA26	Deviation of trajectory (x,y-Position)		HTA32
ITA27	Number of collisions	safety	HTA33
ITA28	Distance to objects		HTA34
ITA29	Time distance to objects		HTA35
ITA30	Max. lateral / longitudinal acceleration		HTA36 – HTA37
ITA31	Frequency of Situations with TTC < $TTC_{critical}$		HTA38
ITA32	Frequency of Situations with TLC < $TLC_{critical}$		HTA39
ITA33	Number of emergency manoeuvres		HTA40
ITA34	Number of take-over situations		HTA41
ITA35	Max. lateral / longitudinal acceleration		HTA42 – HTA43
ITA37	Number of false positive behaviour	function misbehaviour	HTA44
ITA38	Number of false negative behaviour		HTA45

Figure 3.17: Indicators for technical assessment in AdaptIVe

3.2 Methods and tools for technical assessment

For technical assessment in AdaptIVe, the test methodology foresees multiple test environments respectively tools. The test environment is chosen based on a classification of the function as presented in

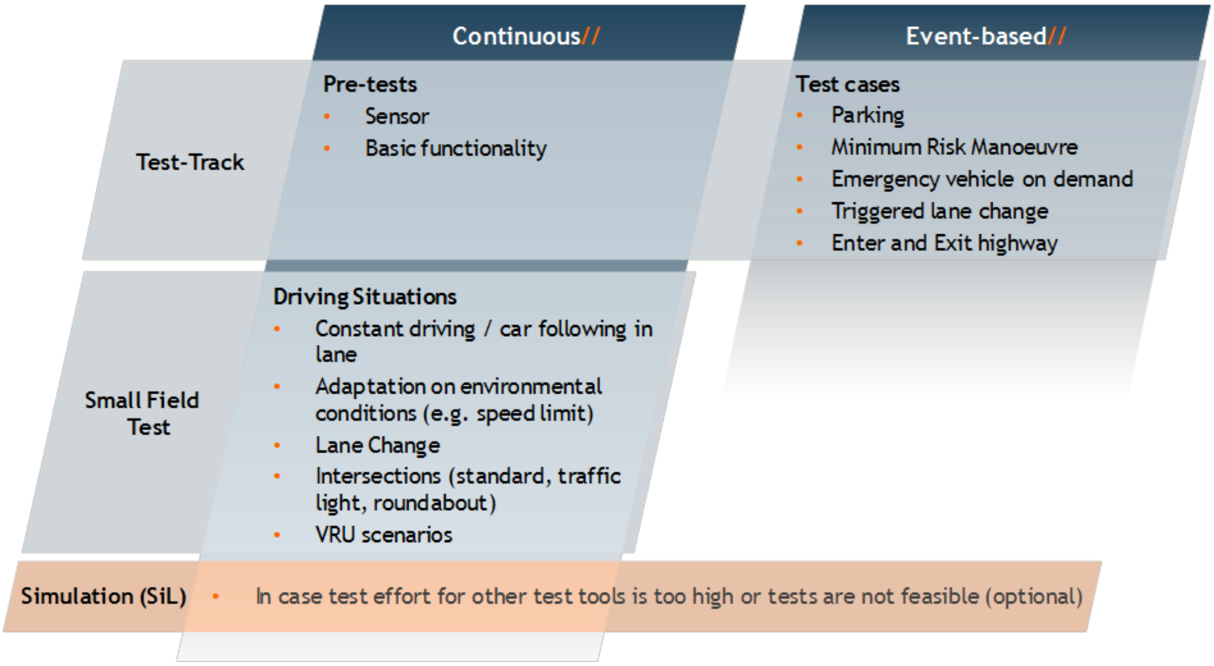


Figure 3.18: Test environments for technical assessment in AdaptIVe

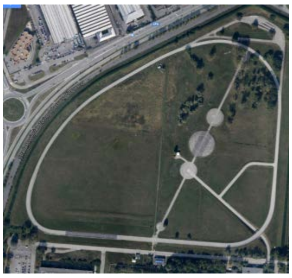
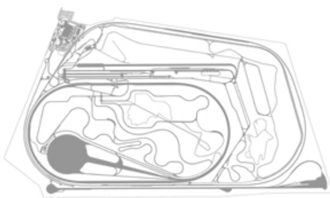
In the following chapter, the different test environments used in the technical assessment are described. The first objective is to test the functions in real conditions. However, simulations might be used in case the required test effort in reality is too high or certain scenarios are too dangerous. Therefore this test environment is also described, although in AdaptIVe it is not foreseen at the moment to conduct test in simulation.

3.2.1 Test-track

For technical assessment tests of the developed functions in reality are indispensable. Tests on a test track (in a controlled environment) are enabling an evaluation of the function in defined situations with variations in the situation. Based on the classification of the functions and systems, the following tests are conducted on a test track:

- Pre-test and sensor tests for continuous operating functions
- Assessment of event-based functions

For the tests on test tracks various test facilities from the AdaptIVe partners are available as presented in Figure 3.19. On which test track the tests will be conducted in the end needs to be decided depending on the assessed demonstrator vehicles and in cooperation with the VSP. For this decision, different aspects must be taken into account (e.g. access to the demonstrator vehicle and test track).

Test tracks	Location	Type	
CRF	Orbassano	Test track	
BMW	Aschheim	Test track	
VW	Ehra	Test track	
Ford	Lommel	Test track	
Volvo	Hällered	Test track	

Test tracks	Location	Type	
ika	Aldenhoven	Test track (including Galileo signals)	
ika	Aachen	Controlled test environment (Parking garage)	

Figure 3.19: Overview on available test track of the AdaptIVe partners

Since a test of all possible scenarios is not feasible due to limited resources, the tests need to be limited to a few relevant test cases. These test cases are derived based on use cases of the developed functions. The test cases for event-based function, which have been defined based on the AdaptIVe use cases, can be found in Annex 1.

3.2.2 Small field test

The methodology for technical assessment in AdaptIVe proposes the evaluation of **continuously operating functions** within a small[8] field test. The test environment itself as well as the required test amount is described in this chapter.

Small field tests in AdaptIVe will be conducted on public roads with real traffic. This is enabling an assessment of the function in many different situations. Before testing in real traffic, a representative test route has to be defined according to specifications and requirements (e.g. road type, length). This will be done in cooperation between SP7 and the responsible persons for the demonstrator vehicles.

When defining the test route for the small field test, the requirements of the function under test must be considered and the infrastructural conditions should be distributed as in the national or European road network. This means that the distribution of e.g. number of lanes and speed limits of the test route should be similar to the distribution of this infrastructure in the European road network and they should be similar to the foreseen environment of the function (e.g. highway pilot). Amendments to the ideal test route design might nevertheless be necessary due to given test conditions. Anyway, lane markings are required on the whole test route. If the small field test foresees a testing on different road types, e.g. urban, rural and motorway, these should be equally distributed in the final test route.

[8] Small in this context means the field test activity is limited to one vehicle (demonstrator vehicle under test) and the test amount is limited to a few weeks.

Due to practicability considerations, the small field tests should be conducted close to places where equipment for maintenance and configuration of the demonstrators is available. If the conduction of field tests is due to serious circumstances not feasible, field tests can be simulated in a controlled environment (test track) or in software in the loop simulation. Finally, the decision for the final test route has to be coordinated between SP7 and the VSPs.

3.2.3 Simulation (SiL)

In case the test effort for a specific function by using the previously stated small field test or test track is too high or the tests are not feasible, the evaluation can also be conducted by using a Software-in-the-Loop (SiL) Simulation in addition.

The following sequence for testing in a virtual environment is foreseen. First, the system description is analysed and the hypotheses are adapted according to the system/function under tests. In the next step the test software is selected. Various tools like e.g. company own developments, PreScan [30], CarMaker [31] and Virtual Test Drive [32] and others are available on the market. At this point the advantages and disadvantages of the different tools are not discussed. Once the test software is selected the preparation of the virtual tests starts. This includes the setup of test case, the parameterisation of the different simulation models (e.g. driver, environment) as well as the integration of the functions respectively systems under test.

Before the test can be conducted in the virtual environment, the simulation tool needs to be verified. For this purpose real world and virtual tests in dedicated test cases need to be conducted and compared. Only in case of a successful verification of the simulation environment the tests can be conducted in the simulation.

The test effort can be calculated similar to tests on a test track and small field tests. Finally for the assessment of the tests the previously defined indicators are calculated and the hypotheses are tested.

3.3 Requirements for technical assessment

In the following chapter, the requirements with respect to safety, test tools and test effort for technical assessment of automated driving functions are discussed. According to the evaluation methodology and used system classification, the requirements are structured in event-based and continuous operating automated driving functions. The requirements are defined in consideration of the SAE Guidelines for Safe On-Road Testing of SAE Level 3,4, and 5 Prototype Automated Driving Systems (ADS) [33].

3.3.1 Safety

Ensuring safety at any time during the test is of major importance for the technical evaluation. This means that testing needs to be safe at any time -meaning any damage of vehicles and persons need to be prevented during the tests. Since the test environments for event-based and continuously operating function are different, also different safety requirements for each function type need to be considered.

Tests for **event-based operating functions** will be mainly conducted in controlled environments, like e.g. test tracks. A controlled environment typically provides the advantages that damages to not involved parties (persons or objects) can be more easily precluded by certain safety measures. During the tests, persons, who are not involved in the tests, need to keep a safety distance[9] to the test vehicle. If the driver is not in the vehicle during the test the safety distance needs to be raised compared to an equal test with the driver in the car. If other objects (vehicles, pedestrians etc.) are involved in the test, crashable dummy objects should be used in order to not cause any damage. Independent of the safety measures for the test environment the driver of the car or any other supervising person[10] must always be capable to regain the control of the vehicle and to switch off the function respectively bring the vehicle (immediately) to a standstill at any time of the tests.

The **continuously operating functions** are evaluated in real traffic. This requires higher safety standards, since such a test involves also other road users. Furthermore, the environment cannot be controlled in the same manner as the test for the event-based functions. Therefore, the safety measures need to focus on the test vehicle, the function and the driver of the test vehicle.

On the vehicle and function side, pre-tests before the actual test in real traffic are required. During the test it should be checked, whether basic behaviour of the function respectively system under test is in line with the specified function behaviour. The conducted tests need to be defined depending on the by the system covered driving situations. During the pre-test also installed safety function (e.g. minimum risk) should be checked, whether they are working properly.

For the tests on public road a trained test driver, who is familiar with the vehicle under test and the tested functions, needs always to be in the driving seat. In the following this driver is called safety driver. The safety-driver must always be capable to switch off the function and regain control at any time during the tests. However, these are only the technical requirements. More

[9] Safety Distance depends on the velocity of the test as well as the foreseen stopping distance of the vehicle and need to be defined in cooperation between SP7 and the VSP

[10] For example for certain (valet) parking functions there might be situations, in which the driver is not in the car.

important is the question when the safety driver should intervene and switch off the function. This question needs to be discussed between persons involved in the tests as well as the persons involved in the function development. In general the safety driver should intervene in case,

- the system's behaviour is not in line with the expected system behaviour in a negative sense,
- a driving situation becomes critical and it is not clear, whether the system can cope with the situation,
- any failure or misbehaviour in one of the components (sensor, computing unit, actuator) or the whole function is detected.

Of course, for the test also legislative aspects have to be taken into account. On the driver side, an appropriate driving license is required. Furthermore, insurance for third party and personal damage must exist. The coverage of the insurance has to be at least as high as requested by the law (e.g. California legislation is requiring for self driving cars $5 million insurance self-insurance or bond). Finally, a road test approval for the test vehicle and the testing region must be available.

3.3.2 Tools

The applied test tools should enable a safe, efficient and accurate testing of all to be evaluated functions in AdaptIVe. According to the evaluation methodology, **event-based operating functions** will be evaluated mainly in a controlled environment or on a test track. The required tools for such tests are:

- **Test track / area**, which needs to be defined in accordance to the tested function or system. Furthermore, it has to be checked whether lane/parking space markings are available and if they are available, what dimensions they have.
- **Test vehicle**, which needs to be equipped with logging equipment.
- **Logging equipment** for logging of all defined signals (see Annex 3). Besides CAN-Signals also video data of the relevant perspectives (front view, rear view, side view) should be logged during the tests
- **Reference measurement system,** e.g. RTK-GPS or appropriate laser scanner
- **Target objects,** which have to be defined depending on test case (other vehicles, balloon cars, road furniture, etc). If possible the used target objects should be crashable. Additionally, they need to be representative of "real world objects" for the sensors and systems on the vehicle under test (VuT).

The **continuous operating functions** are tested on public roads in real driving situations. Therefore the required and used test tools differ:

- **Public Road**, which needs to be defined in accordance to the tested function or system. Here, it has to be checked, whether required information on the road (e.g. digital map data, road markings / traffic signs) are available, and whether the road fulfills the requirements of the functions (e.g. dimensions of the test roads / lanes).
- **Test vehicle** which needs to be equipped with logging equipment
- **Logging equipment** for logging of all defined signals (see Annex 3). Besides to the CAN-Signals also video data of the relevant perspective (front view, rear view, side view) should be logged during the tests

In contrast to the event-based functions, **a reference measurement system** is only optional for the tests on public roads. It is planned to assess the sensor accuracy during the pre-tests, since due to the uncontrolled environment on public roads it might be difficult to obtain highly accurate data of other objects.

3.3.3 Test amount

In this chapter the required test effort is described. Within AdaptIVe, the technical assessment of **event-based functions** is conducted on a test track. The test effort for the evaluation on a test track depends on the number of test parameters, the variations of each test parameter and the number of repetitions of each test configuration. The calculation of the number of test runs is given in Figure 3.20.

	Basis for calculation:	Description:
Number of test runs	$R \cdot \sum_{i=1}^{n} M_i$	R: number of Repetitions; M_i: Number of variations of the test case “i”;
Number of variations of test case	$M_i = \prod_{y=1}^{z} n_y$	n_y: number of variations of test parameter y; max. number of parameters “z”
Number of repetitions	For AdaptIVe: “R” = 5	“R”: repetitions per parameter set

Figure 3.20: Test effort for evaluation on a test-track

In general it should be aimed to test all parameter that can have an influence on a certain driving scenario or test cases. It should also be tried to test parameter in all relevant configuration. This means that variations that occur in the real traffic should be tested. Depending on the driving scenario or the test cases this can lead to a high number of test cases.

In AdaptIVe as for many other project resources related to testing - time as well as money– are limited. Therefore, the number of tests in each test case needs to be limited. For AdaptIVe it is suggested to keep the number of test parameters y, which are varied during the tests, for each test below 5. Furthermore, the number of variations n_y for each test parameter should also be below 5. The number of repetitions of each test configuration is set to five to get statistical significance, depending on the test. This would mean for AdaptIVe that in the worst case 125 test runs have to be tested for one test case. Nevertheless it should be tried for each test case to limit the number of the tests to a minimum.

Regarding **continuous operating functions**, **pre-tests** will also be conducted on test tracks. Purpose of the pre-tests is to ensure that the function is working basically before evaluating the function in real traffic. Since the pre-tests are not the main tests for the technical assessment of the continuously operating function, the test effort needs to be limited even more. Therefore, it is suggested to keep the number of test parameters (y), which is varied during the tests, for each test at max. 2. Furthermore, the number of variations (n_y), for each test parameter, should be at max. 2.

The main test for the continuously operation functions will be performed on public roads in real traffic. It would also for these tests be desirable to have as many test data as possible. This would require an extensive testing on public roads, which is due to the limited resources in the project not feasible. Therefore a trade-off between the test amount that is required for the evaluation and the available resources must be found. This means that, analogue to the event-based test the tests for the continuously operating functions also needs to be limited to a feasible extent.

On the other hand, it needs to be ensured that enough relevant driving situations are detected in order to be able to assess the function under investigation. This means that the test route has to be chosen in a way that all relevant driving situations occur multiple times. Due to the uncontrolled set up of the test this can hardly be guaranteed for each driving situation. Therefore, probability of driving situations occurrence is taken into account. This means that the required test length needs to be estimated a priori based on the number of expected driving manoeuvres that occur while driving in public traffic. The test route is driven several times at the same start conditions (e.g. time) in order to minimize disturbing effects of the traffic.

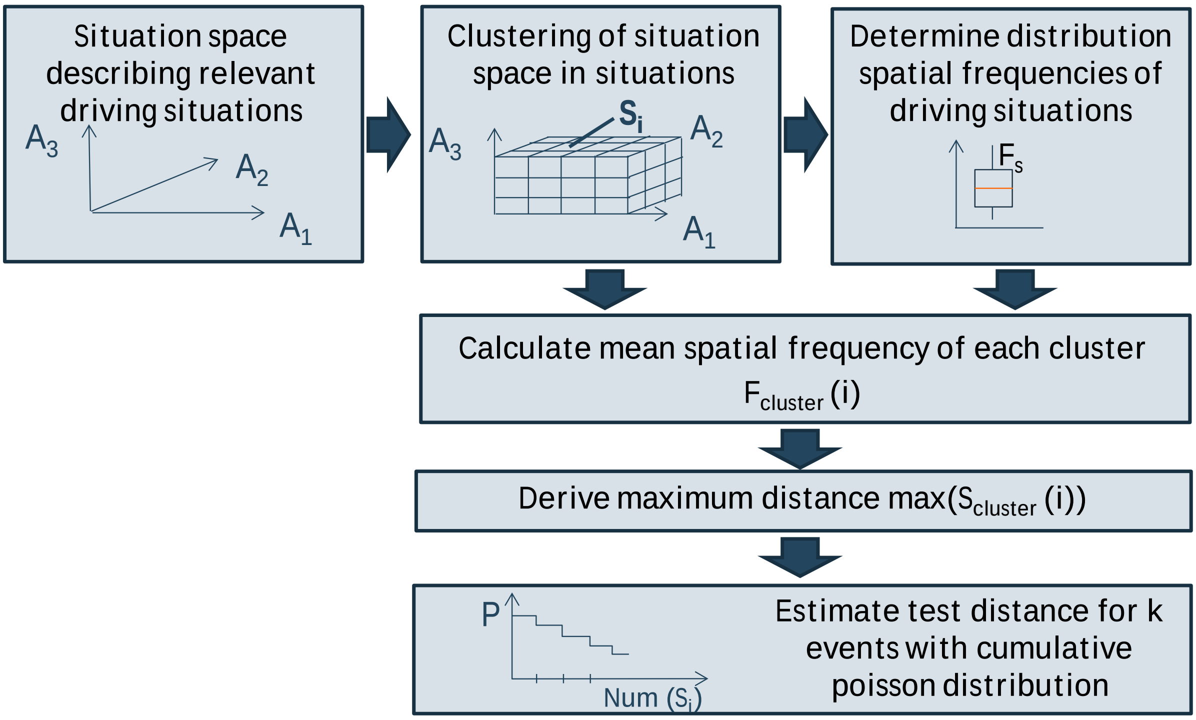


Figure 3.21: Methodology for estimation of test effort

The test effort, particularly the test route length, can be calculated by utilizing data from field operational tests, such as euroFOT [16]. In the following, a methodology is presented to estimate the test distance based on field operational test data, see Figure 3.21. The methodology for estimating the minimal test distance is based on the approach described by Winner et al. in [26].

The driving data of the field operational tests is clustered in relevant driving situations by using a situation space approach. Afterwards, the distribution of spatial frequencies of all relevant driving situations is calculated. The mean of the spatial frequency is chosen to characterize the distribution of spatial frequencies for each driving manoeuvre. The distance necessary for the occurrence of a single driving manoeuvre can be calculated by inverting the mean spatial frequencies.

For calculation of the minimal test distance for the occurrence of k = 5 driving manoeuvres which are necessary to evaluate the function, a cumulative Poisson distribution is assumed in accordance to the approach described in[26]. Based on the mean distance necessary for the occurrence of a single event s_{ref}, the necessary distance is calculated for the occurrence of k events with a probability of $P = 95\,\%$. The basis for the calculation of the minimum distance is given in Figure 3.22.

A detailed description of minimum test distances for relevant driving situations can be found in Annex 24. To ensure the evaluation of automated driving functions in different environments, the minimum test distances are calculated for driving on urban roads, rural roads and on motorways.

	Basis for calculation:	Description:
Test distance	$P = \sum \frac{\lambda^k}{k!} e^{-\lambda}$	P: probability for the occurrence of k driving situations k: desired number of driving situations
Expectancy value	$\lambda = \frac{s_k}{s_{ref}}$	s_k: distance for k driving situations
Test distance for single manoeuvre	$s_{ref} = \max(s_{cluster}(i))$	s_{ref}: mean distance for single driving situation i: Cluster

Figure 3.22: Poisson distribution for calculation of test distance

While most driving situations are detected using data obtained from field operational tests such as euroFOT [16], the data for the driving situation “enter motorway” is obtained by utilizing statistics about the infrastructure of motorways. The minimum test distances for relevant driving situations on motorways are presented in Annex 4. Here, all distances are calculated for the occurrence of k =5, 10, 20, 30 events with a probability of P = 95 % by assuming a cumulative Poisson distribution.

3.4 Example of technical assessment

In this chapter, the within this project developed methodology for technical evaluation will be demonstrated by evaluating exemplary automated driving systems.

3.4.1 Parking function

An automated parking function is chosen as an example of an event based operating function. The exemplary function is able to detect a free parking spot, which can be a parallel, an orthogonal or an angular parking spot and park the car in the detected spot.

According to the testing sequence, suitable research questions and hypotheses have to be defined. Here, it is important to select these based on the focus of evaluation which is classified based on the evaluation aspects. In the following, the exemplary assessment of the “automated parking function” is shown with focus on the evaluation aspect “performance in a driving situation”. Thus, the relevant research questions and hypotheses are presented in Figure 3.23.

ID	Research Question	Hypotheses
RQTA4	How much variation occurs for one manoeuvre (considering environmental conditions)?	The standard deviation of the x,y - position for one manoeuvre for multiple tests is below a certain threshold (in all / the defined scenario/s).
RQTA5	Is the function or system performance affected by different environmental conditions?	The sensor detection range is above a certain threshold in all scenarios.
RQTA9	How accurate the manoeuvre is conducted? (Planning or assumption vs. driven trajectory).	The standard deviation of the x,y - position between planned or driven trajectory is below a certain threshold (in all / the defined scenario/s).
RQTA10	How long does the designated manoeuvre take compared to a human driver?	The time of the designated manoeuvre is within the distribution of the time of human drivers.

Figure 3.23: Research questions and hypotheses for the aspect performance in driving situation

The previously selected hypotheses are evaluated using indicators which are calculated out of the logged signals by the data recording unit over the entire test-drive. Appropriate indicators for the previously stated hypotheses are presented in Figure 3.24.

ID	Indicators
ITA4	Distribution of x,y- position
ITA5	Sensor detection range, availability of the function or system
ITA9	Standard deviation of delta x,y-position
ITA10	Time of driving manoeuvre

Figure 3.24: Indicators for the aspect performance in driving situation

For test preparation, the tested demonstrator has to be equipped with measurement equipment, e.g. reference sensors and a data recording unit. It must be ensured, that all previously defined signals can be recorded during the test drive. The signal list defined in AdaptIVe can be found in Annex 3.

Before carrying out any tests a risk assessment has to be conducted. If any risks are identified, counter measures need to be taken after consultation between SP7 and the responsible persons of the test vehicle. This includes for instance instructions for a safety driver, when to intervene in order to prevent any risky driving situation. For the parking example the parking spot should be outlined by balloon cars and it needs to be ensured that the driver – independent of the fact whether he is in the car or outside – is able to stop the vehicle immediately.

The tests itself are conducted on a test track (controlled environment). Based on the function's use cases to conduct an automated parking manoeuvre, relevant test cases can be selected. For the evaluation of the exemplary automated parking function, the relevant test cases are presented in Figure 3.25. All relevant test cases for the evaluation of event-based functions can be found in Annex 1.

Test-environment	Test cases	ID
Test-track	Parallel parking	E1
	Orthogonal parking	E2
	Angular parking	E3
	Parking with blocking object	E4
	Parking with moving object	E5

Figure 3.25: Test cases for exemplary parking function

During the tests, the parameters of each test case are varied in order to test the function's performance in different driving situations. Nevertheless the test effort needs to be limited to a feasible extent.

As an example, the test case "parallel parking" with the definition of all relevant test parameters is presented in Figure 3.26. For this, 45 tests runs are needed.

After the tests the logged data will be analysed by the SP7 partner responsible for the technical tests (details need to be discussed between the involved partners). In this process, the defined hypotheses will be verified.

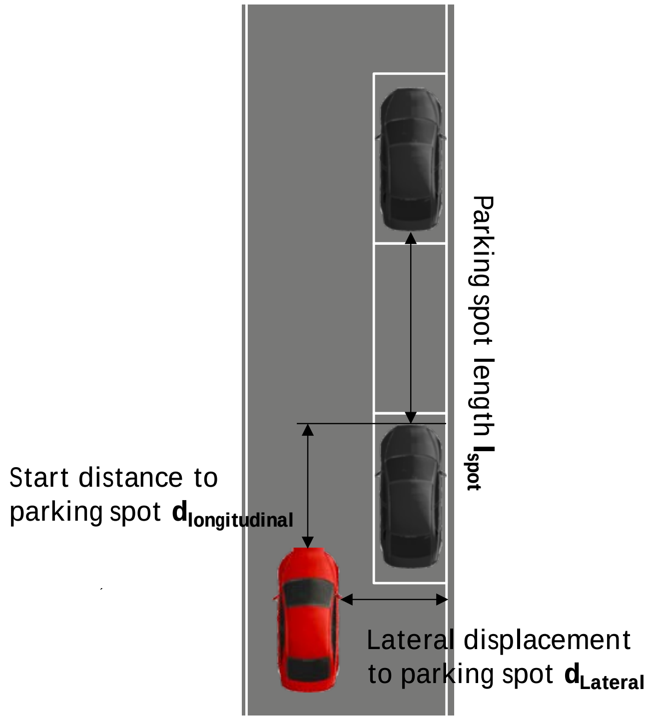
Exemplary test case "parallel parking"	Parameters of test case	
	Start distance to parking spot $d_{longitudinal}$	4.0 m (fixed Parameter)
	Parking spot length l_{spot}	Length of EGO vehicle + 0.8,1.0,1.2 m
	Lateral displacement to parking spot $d_{Lateral}$	0.5, 0.7, 0.9 m
	Number of repetitions per parameter set	R = 5
	Number of test runs	45

Figure 3.26: Exemplary test case "parallel parking"

3.4.2 Highway Automation System

The exemplary system "highway automation system" is a conditional/high automated driving system which is classified as continuously operating system. The system is able to handle different driving situations (car & lane following, automated lane changes, speed adaption) on highways. In the following the evaluation is briefly described.

Of particular importance within this technical assessment is the focus on the safety aspect. Although other aspects might also be relevant for the evaluation and are considered in the final evaluation, the focus in this example is only on this one aspect. The safety related research questions and hypotheses are presented in Figure 3.27.

ID	Research Question	Hypotheses
RQTA26	Do collisions with other objects occur during testing?	The number of collisions is equal to zero.
RQTA27	Is driving with the function or system safe? What is the minimum distance or time distance to other objects during the manoeuvre?	The distance to objects is below a certain threshold (in all / the defined scenario/s).
		The time distance to objects is below a certain threshold (in all / the defined scenario/s).
RQTA28	Is the function or system able to perform	The max. lateral acceleration is above a certain

ID	Research Question	Hypotheses
	emergency manoeuvres?	threshold in critical situations.
		The max. longitudinal acceleration is above a certain threshold in critical situations.
RQTA29	How often do situations with TTC < $TTC_{critical}$ (respectively TLC < $TLC_{critical}$) occur?	The frequency of Situations with TTC < $TTC_{critical}$ (TBD) is below a certain threshold.
		The frequency of Situations with TLC < $TLC_{critical}$ (TBD) is below a certain threshold.
RQTA30	Does the function or system conduct any emergency manoeuvres (ax < threshold, evasive manoeuvre, safe stop) during the test?	The number of false negative behaviour is below a certain threshold.
		The number of false negative behaviour is below a certain threshold.
RQTA31	Is the function or system able to perform emergency manoeuvres?	The max. lateral acceleration is above a certain threshold in critical situations.
		The max. longitudinal acceleration is above a certain threshold in critical situations.

Figure 3.27: Research questions and hypotheses for the aspect safety

The related indicators for the previously stated hypotheses of the evaluation aspect safety are presented in Figure 3.28.

ID	Indicators
ITA27	Number of collisions
ITA28	distance to objects
ITA29	time distance to objects
ITA30	max. lateral / longitudinal acceleration
ITA31	Frequency of Situations with TTC < $TTC_{critical}$
ITA32	Frequency of Situations with TLC < $TLC_{critical}$
ITA33	Number of emergency manoeuvres
ITA34	Number of take-over situations
ITA35	max. lateral / longitudinal acceleration

Figure 3.28: Indicators for the aspect safety

Analogue to the evaluation of event-based functions, the tested demonstrator has to be equipped with measurement equipment. All previously defined signals have to be recorded during the test drive. The signal list defined in AdaptIVe can be found in Annex 3.

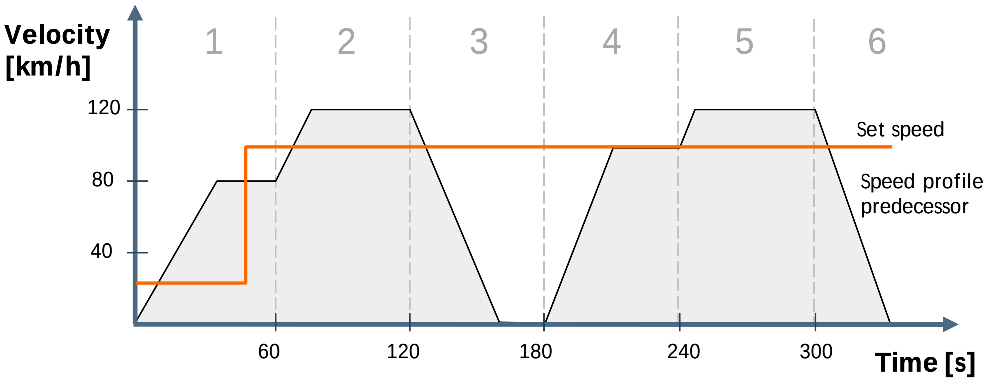
A risk assessment has to be conducted as previously described at the evaluation of event-based functions. If any risks are identified, counter measure need to be taken after consultation of SP7 and the responsible persons of the test vehicle.

The next steps are pre-tests on a test track in a controlled environment. The objectives of the pre-tests are to ensure that the basic functionality of the system is given and the system is operating as specified. Furthermore, the accuracy of the sensors has to be analysed. The test methodology foresees static and dynamic sensor tests, a car following test and tests of minimum risk manoeuvres with and without traffic. All pre-tests are presented in Figure 3.29.

Test-environment	Pre-test	ID
Test-track	Static sensor test	CP1
	Dynamic sensor test I	CP2
	Dynamic sensor test II	CP3
	Basic functionality: car following (see Figure 3.30)	CP4
	Minimum risk manoeuvre	E10
	Minimum risk manoeuvre on motorway	E11
	Minimum risk manoeuvre on motorway with traffic	E12

Figure 3.29: Pre-tests for continuous operating functions

For the car following tests a speed profile for the predecessor vehicle has to be specified. In this example, the speed profile presented in Figure 3.30 is used. The speed profile should include acceleration, braking and constant driving manoeuvres. The preferred speed ("Set speed") of the demonstrator is set to 100 km/h. Thus, the host vehicle has to react in different ways on the predecessor. The evaluation criteria for the pre-test are mainly the distance related indicators like distance headway, time headway and time-to-collision.



Part of test	Part 1		Part 2		Part 3		Part 4		Part 5		Part 6	
Time [s]	0	30	60	80	120	160	180	220	240	260	300	330
Velocity [km/h]	0	80	80	120	120	0	0	100	100	120	120	0

Figure 3.30: Speed profile for test of basic functionality: car following

After successful execution of the pre-tests, the functions are evaluated in the field. For the exemplary system "highway pilot" the following driving situations for small field test are relevant. The detailed description of the relevant driving situations for small field tests can be found in Annex 2.

Pre-tests	Function	Driving situations (ID)		
		Urban	Rural	Motorway
Small-field-test	Following lane	-	-	C5 ($v > 60$ kmh^{-1})
	Stop & go driving	-	-	C6
	Lane change, overtaking manoeuvre	-	-	C10
	Speed and time-gap adaption at motorway entrance ramp	-	-	C22, C23
	Predictive automated driving	-	-	C9
	Danger spot intervention			C8

Figure 3.31: Relevant driving situations for an exemplary highway pilot

The minimum test distance can be derived by utilizing the method to determine the test effort which is described in chapter 3.3.3. The minimum test distance is calculated for all relevant driving situations, whereas the overall testing distance is determined by the driving situation with maximum test distance. In case of the system "highway pilot" the minimum test distance is

300 km, according to the overview of test distances presented in Annex 4. The test distance is exemplarily calculated by assuming the occurrence of k = 10 events with a probability of P = 95 %. If a higher number of driving situations is foreseen, the test effort can be estimated by using the table provided in Annex 4.

In a next step a suitable test route needs to be defined in cooperation between the demonstrator responsible person and SP7. For the definition of the test route different factors, like e.g. required map data available, areas with speed limit as well as the calculated test distance, must be considered. Since for the exemplary evaluation of the system "highway pilot" 3 days of testing are assigned, the test route should at least be 100 km per day long[11]. Thus, a suitable test route could be starting at the Aachen motorway junction, going on the A4 to Kerpen, then taking the A61 to Jackerath and finally driving the A44 back to Aachen. This route is 158 km long and therefore suitable for the evaluation. The overview over the selected test route is presented in Figure 3.32.

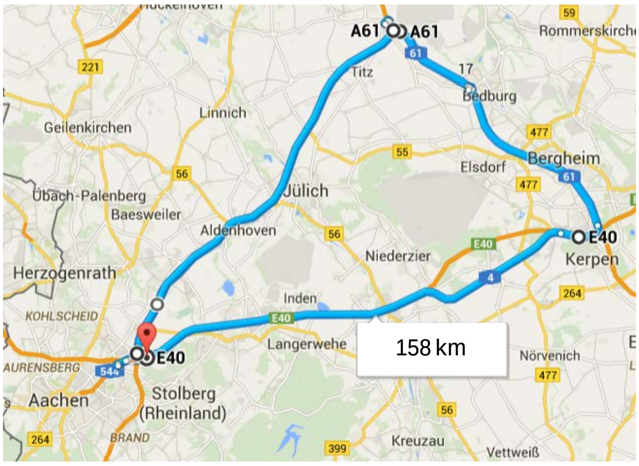
Overview example small field test for "highway pilot"	
Test route	A4 Aachen – Kerpen - A 61 Jackerath – A 44 Aachen
Overview	 Source: Google.maps
Distance	158 km
Motorway entrances	21
Speed limit	9 speed limit changes (no limit, 120, 100, 80 km/h)
Number of lanes	38 % 3 lanes 62 % 2 lanes

Figure 3.32: overview exemplary small field test for "highway pilot"

After the tests the logged data will be analysed by the responsible SP7 member. In this process the defined hypotheses will be verified by means of the logged data.

[11] At this point the detection of five relevant situations for the evaluation is assumed. However more relevant situations are desirable.

4 User-related Assessment

This chapter describes the evaluation approach for the user-related assessment of automated driving functions and systems. In the first section the methodology for evaluation is outlined. Section 4.1 describes focus of a user-related assessment and presents the relevant research questions, hypotheses and indicators. The different methods and tools for user-related assessment are discussed in section 4.2. In section 4.3 an example of the application of the evaluation methodology for an automated driving system is given. Finally, section 4.4 discusses the requirements for testing with respect to safety, test-tools and test conduction.

4.1 Focus of user-related assessment

The user-related assessment of automated driving applications involves a great variety of issues. Stanton and Young presented a number of psychological issues pertinent to vehicle automation that they noted should be considered in empirical evaluation studies [34]. These issues include locus of control, the trust the driver has in the automated system, the situational awareness of the driver, the mental representation that the driver develops of the automated system, the mental and physical workload associated with automation, feedback, driver stress and its implications. The relevance of these and additional issues in more detail are as follows:

- **Behaviour related** issues: People adapt their behaviour as a response to changes in the road-vehicle-user system [35]. Reallocation of attention is an intelligent response to the change[36]. The adaptive process develops dynamically over time, based on operator experiences of interaction with the automated system [37].
- **Understanding of automation** issues: an insufficient and/or erroneous mental model that the driver develops of the automated system may lead to increased risk of user errors [38]; Jenness et al. in a survey among early adopters of in-vehicle technology found that system owners often do not understand the limitations of the systems and manufacturers' warnings [39].
- **Trust and reliance** related issues: over- or under-reliance on automation can have crucial effects on automation outcome. Users rely more on automation they trust more [40]. Over-trust may lead to misuse of automation, leading to failure of the driver to override the system when necessary. Under-trust (i.e. if users fail to rely on automation) may yield abandonment of automation, leading to a lost opportunity of improvements in driving performance [41]. There is a variation among users: younger or older users rely on automation differently [42] Merritt et al. found that user trust in automation was influenced by both implicit and explicit attitudes [43]. They asked participants to complete both a self-report measure of propensity to trust and an Implicit Association Test measuring implicit attitude toward automation and found that explicit propensity to trust and implicit attitude

toward automation did not correlate. They concluded that implicit attitudes have important implications for automation trust and users may not be able to accurately report why they experience a given level of trust. User's implicit attitudes, as well as user mood and emotion may affect their trust in automated systems. Since implicit and explicit processes often dissociate, "implicit preferences may provide predictive power that cannot be obtained via traditional explicit measures"[43]. To understand why users trust or fail to trust automation, measurements of both implicit and explicit predictors are necessary.

- **Locus of control** (i.e. the extent to which removal of control from the driver affects the performance of the vehicle/driver entity). Locus of control refers to the extent to which individuals believe they can control events affecting them. Drivers with an internal locus of control believe their vehicle performance derives primarily from their own actions, while drivers with an external locus of control believe the behaviour of the vehicle is due to the automated system. Stanton and Young mean that some drivers may perceive that they are in overall control of the vehicle when it is in automated mode whereas others may not [34]. According to them, research findings had shown that people with an internal locus of control generally perform better than individuals with an external locus of control which might be attributed to the degree of task engagement for the individual. An internal locus of control may lead drivers to take on an active role, while an external locus of control might lead a driver to assume a passive role with the automated system. Stanton and Young found that the passive drivers failed to intervene when the automated system failed whereas the active drivers took control of the situation [34].
- **Resuming control** of driving is an important issue in automation. Merat, et al. (2014) examined how different methods of transferring control of a highly automated (Level 3) vehicle affected the driver's ability to resume control of driving and found "an overall better performance by drivers when control was transferred after a fixed duration of 6 min, compared to when the automated system disengaged if drivers removed their visual attention away from the road centre" (p 281) [44].
- There is a non-negligible risk that **skill degradation** accompanies automation of the driving tasks due to overreliance, as the reinforcement coming from constant engagement in the driving task becomes absent [45]. [46] refer to Shiff who found that "despite initial manual training, those subjects who had been operating as supervisory controllers of automation in a simulated process control task were slower and more inefficient in bringing the system under control than were subjects who had operated only in a manual mode" (p. 381) [46]. If drivers learn to drive with automated systems initially, without extensive manual experience, appropriate skills may not be developed. Such skills may be important not only for performing a task manually, but also for detecting the need for manual performance [46].

- Automation may affect the **mental workload** of the driver in various ways. It may reduce workload during straightforward driving conditions, but automation monitoring and the need to resume control when attention is directed towards non-driving related tasks may lead to sudden increased **mental workload** [44]**.** Humans are inefficient in monitoring automation [47]; [48]; [49]; [50]. Overreliance on automation contributes to this inefficiency [41]. Banks et al. in a case study concluded that the number of processes conducted by the driver appear to increase as the level of automation increases [51]. In intermediate levels of automation, driver decision-making remains apparent (only at full automation can this decision be removed). The addition of sub-system monitoring increases task loading and hence driver workload as the driver must remain aware of system state and operation. Automation has different effects on users' (younger or older) workload [42].
- **Stress** is a factor that may affect driver workload and safety [52]. Stress and vehicle automation has been studied by [53], who explored (among others) the effects of stress and vehicle automation on driver performance manipulating stress by exposing drivers to a loss of control experience and found that both stress and automation influenced subjective distress, with higher levels of distress under the stressful driving conditions and lower levels of distress under the automated speed control conditions; however, the two factors did not interact. Reimer et al. evaluated the extent to which vehicle-parking-assist systems affected driver stress by using heart rate measurements along with self-reported ratings and found that participants exhibited lower average heart rates and they reported lower stress levels when using the assistive parking system [54].
- **Boredom** in low-task-load environment might lead to distraction [55]. According to Farmer & Sundberg "Boredom is a common emotion, with boredom proneness a predisposition with important individual differences." (p.4) [56]. Stark and Scerbo found significant correlations between boredom proneness, workload, and complacency potential [57], which might indicate that the psychological state of boredom may be a factor that **induces complacency** [58]. Also, Sawin and Scerbo in their vigilance tasks study found association between boredom proneness and vigilance performance [59].
- **Fatigue** may affect driving performance negatively [60] and it reportedly contributes to a significant share of car accidents [61]. Prolonged driving may induce a variety of fatigue symptoms such as drowsiness, boredom, irritation, physical discomfort and daydreaming [62]. [63] found that fatigue induction elicited various subjective fatigue and stress symptoms, raised reported workload, increased heading error, reduced steering activity, and reduced perceptual sensitivity on a secondary detection task. Their results suggest that task-induced fatigue is associated with impaired performance evaluation in "underload" and interventions should be geared towards enhancing driving motivation, rather than reducing attentional demands on the driver. Matthews & Desmond conclude that "passive fatigue"

(associated with tasks requiring monitoring the environment but infrequent response) "may become increasingly common in intelligent vehicle highway systems as control passes from driver to vehicle, and it merits further investigation" (p. 681) [63].

- The **situational awareness** of the driver concerning the driving context and the operational status of the system is of vital relevance. Increased automation may increase the tendency of shifting attention away from the driving task [64]. Endsley & Kiris studied the automation of a navigation task and found that situational awareness is lower under automated conditions than under manual conditions and low situational awareness corresponded with out-of-the-loop performance decrements in decision time following a failure of the system [46]. Based on a review of earlier studies, Stanton et al. conclude that "loss of situational awareness is correlated with poor system performance" and "people who have lost situational awareness may be slower to detect problems with the system they are controlling as well as requiring additional time to diagnose problems and conduct remedial activities when they are finally detected" (p199) [65]. Stanton et al. suggest that "understanding the nature of situational awareness errors can be helpful in deciding upon strategies for developing effective counter-measures" (p. 201) [65]. They discuss various theories of situational awareness and conclude that the "Three-level model" put forward by [66] seems to be the most developed approach, in terms of measures and interventions and it offers a functional model for assessing different degrees of insight in a pragmatic manner [66]. According to Endlsey's definition: "Situational awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and a projection of their status in the near future" (p36) [66].

- **The "out-of-the-loop" performance problem** making the driver handicapped in his/her ability to take over in the event of automation failure is attributed to loss of situational awareness and skill degradation, which leads to declining operator performance [45]. There is no failure free system [67]. If reliability is below 70 %, it is better having no automation at all [68]. Automation failure detection better with varying automation reliability [47]. When exposed to automation failure drivers perform better with a lower level of automation [69]; [70]. Complacency and over-reliance may cause loss of situational awareness leading to errors when automation fails [37].

- **Automation-related complacency** may occur when the operator has to perform both manual tasks and supervise automation. It can be described in terms of an attention allocation strategy where the operator may attended to manual tasks at the expense of the automated task, especially when task load is high [37]. The National Aeronautics and Space Administration Aviation Safety Reporting System (ASRS) coding manual defined complacency as "self-satisfaction which may result in non-vigilance based on an unjustified assumption of satisfactory system state" [71]. "Attention allocation away from the automated task associated with complacency may include not only fixation failures but attention failures as well. In addition to visual attention, automation complacency has been linked to an initial attitude of high trust toward the automation." (pp 389) [37]. Complacency may lead to the consequence that a system malfunction, anomalous condition, or failure is missed or reaction is delayed [37]. Parasuraman et al. reported considerable individual differences in the performance effects associated with automation complacency [47]. Also, Prinzel et al., investigating the relationship between individual differences of complacency potential, boredom proneness and automation-induced complacency, found that personality individual differences are related to whether an individual will succumb to automation induced complacency [58].

- **Automation bias** is reflected in omission errors (the user fails to respond to a critical situation because the automation aid failed to notify him/her) and commission errors (following a wrong recommendation) made by operators when decision aids are imperfect [37]. Commission errors can be "the result of not seeking out confirmatory or disconfirmatory information or discounting other sources of information in the presence of computer-generated cues" [72]. Mosier and Skitka defined automation bias as operators using the decision aid "as a heuristic replacement for vigilant information seeking and processing" (p. 205) [72]. Automation bias can lead to decisions that are not based on a thorough analysis of all available information but that are biased by the advice of decision aid and it can compromise performance considerably in case of automation failures [37]. One reason for automation bias is that users have a tendency to ascribe greater power and authority to automated aids than to other sources of advice [37]. User trust in automated aids as powerful agents with superior capability might make users to overestimate the performance of the aids as they may ascribe to the aid greater authority than to other humans or themselves [37]. Another contributory factor to automation bias is the phenomenon of diffusion of responsibility. When sharing monitoring and decision-making tasks with an automated aid (or other humans in a group) humans may reduce their own effort compared to when they work individually on a given task. The operator may perceive him/herself as less responsible for the outcome and, as a consequence, reduce his/her own effort in monitoring and analysing other available information [73].

- In a review of empirical studies of complacency and bias in human interaction with automated and decision support systems, Parasuraman & Manzey found that **complacency and automation bias** represent different manifestations of overlapping automation-induced phenomena, where attention plays a central role [37]. Further, they found that automation complacency and automation bias occurred in both naive and expert operators and it cannot be prevented by training or instructions. An integrated model of complacency and automation bias put forward by Parasuraman & Manzey implies that these issues result from the dynamic interaction of personal, situational and automation-related characteristics [37].

- **Usability:** Usability of a product is of importance for its success among potential users. There are various definitions of usability, but as Harvey et al. expressed, "consideration of the context of use is essential in defining usability criteria and this will be different for each system under investigation" (pp 563) [74].
- **Acceptance:** Acceptance is a key factor for intended use of new technology in the vehicle [75]. However, as Adell et al. put it "Despite the recognised importance of acceptance there is no established definition of acceptance, and there are almost as many ways to measure acceptance as there are researchers trying to do so" (pp 73) [76]. Adell et al. put forward a proposal for a common definition of acceptance focusing on a system's potential to realise its intended benefits; that is, the incorporation by the driver of the technology into their driving: "Acceptance is the degree to which an individual incorporates the system in his/her driving, or, if the system is not available, intends to use it" (p. 17) [75].

4.1.1 Research questions

The issues, pertinent to vehicle automation were presented above in order to have a background-understanding of the formulated research questions. The research questions listed below, however, don't address all those issues mentioned above, because the aim of the user-related assessment in AdaptIVe is not to explore or verify those issues, but to assess the user-related effects of the developed functions/systems with due consideration to those issues. The research questions below address all levels of automation and both continuous and event based functions. They are presented by the different evaluation aspects (driver behaviour and performance, effects of automation on the driving task, take-over situations and regaining control, trust, opinions and acceptance, issues concerning non-users).

Research questions concerning **driver behaviour and performance** issues are:

ID	Research Question
RQUA1	Does the system give the expected user-related outcome?
RQUA2	Does the driver use the system as intended to be used?
RQUA3	Does the driver use the function/system in all situations for which it is available?
RQUA4	Does the driver stay in the function/system settings suggested by the system?
RQUA5	Does driver behaviour differ when driving with a well-functioning driving automation from driving behaviour without automation?
RQUA6	Are there any long-term changes in driver behaviour when driving with automation?

Figure 4.1: Research questions concerning **driver behaviour and performance**.

Research questions concerning the **effects of automation on the driving task** are:

ID	Research Question
RQUA7	Is situational awareness of the driver influenced by the system?
RQUA8	Is driver stress affected by automation?
RQUA9	Is mental workload of the driver affected by automation?
RQUA10	Does mental workload change after long term use of the system?
RQUA11	Is transfer of control affected by mental workload?
RQUA12	Do drivers engage more in secondary tasks when driving with automation compared to driving without automation?
RQUA13	Do drivers become complacent when driving with automation?
RQUA14	Does the time for the drivers to make decision after a safety critical event differ between manual driving mode and automated driving?
RQUA15	Does driver skill degrade with time using automation?

Figure 4.2: Research questions concerning the **effects of automation on the driving task**.

Research questions concerning **take-over situations and regaining control** are:

ID	Research Question
RQUA16	Is there any change in the drivers' take-over behaviour in long term?
RQUA17	Do drivers detect automation failures?
RQUA18	Do drivers fail to respond to a critical situation because the system failed to notify them?
RQUA19	Do drivers take the right measure to handle automation failure?
RQUA20	Do drivers follow a wrong recommendation instead of vigilant information seeking and processing?
RQUA21	Are drivers confident about the correctness of their decision after a system brake down?
RQUA22	Do drivers with an external locus of control intervene in time when the automated system fails?

Figure 4.3: Research questions concerning **take-over situations and regaining control**.

Research questions concerning the driver's **trust, opinions and acceptance** of the system are:

ID	Research Question
RQUA23	Do drivers have the correct mental representation of the system?
RQUA24	Do drivers have an over- or under-trust in the system?
RQUA25	Do drivers experience automated driving as an improvement in their driving?
RQUA26	What are the drivers' opinions about the system?
RQUA27	Do drivers find the system useful and satisfactory?
RQUA28	Do automation failures influence the driver's attitude towards the system?
RQUA29	What is the level of willingness to have/to pay?

Figure 4.4: Research questions concerning **trust, opinions and acceptance** of the system.

Research questions concerning **non-users** of the system are:

ID	Research Question
RQUA30	Are non-users' behaviour influenced by interaction with equipped vehicles?

Figure 4.5: Research questions concerning **non-users** of the system.

4.1.2 Hypotheses

Based on the research questions described above, the hypotheses are formulated in this chapter. They are presented, analogue to research questions, by the different evaluation aspects (driver behaviour and performance, effects of automation on the driving task, take-over situations and regaining control, trust, opinions and acceptance, issues concerning non-users). During the evaluation in AdaptIVe not all hypotheses will be tested, but - for the individual function/system - relevant hypotheses will be selected.

During the user-related assessment, testing of hypotheses is typically testing if the null hypothesis (i.e. the indicator value for users does not differ from the expected value or from the value when not using the function/system) can be rejected or not. If the probability of a difference between the indicator values is less than or equal to the selected significance level, then the null hypothesis is rejected and the difference is said to be statistically significant. The level of significance is the criterion used for rejecting the null hypothesis. Traditionally, either the 0.05 level (also called the 5 % level) or the 0.01 level (also called the 1 % level) have been used. The lower the significance level, the more the observed mean value must differ from the null hypothesis to be significant. The 0.01 level is more conservative than the 0.05 level. The probability of the difference depends on the number of observations and the variance (standard deviation) of the mean of the observed values. Hence, to use a conservative significance level, a large number of observations (besides low variance in the data) is necessary.

Hypotheses concerning **driver behaviour and performance** issues are:

ID	Hypothesis	Related Research Question
HUA1	The system gives the expected user-related outcome.	RQUA1
HUA2	The driver uses the system as intended to be used.	RQUA2
HUA3	The driver uses the function/system in all situations for which it is available.	RQUA3
HUA4	The driver stays in the function/system settings suggested by the system.	RQUA4
HUA5	Driver behaviour does not differ when driving with a well-functioning driving automation from driving behaviour without automation.	RQUA5
HUA6	There are no long-term changes in driver behaviour when driving with automation.	RQUA6

Figure 4.6: Hypotheses concerning **driver behaviour and performance**.

Hypotheses concerning the **effects of automation on the driving task** are:

ID	Hypothesis	Related Research Question
HUA7	The drivers' situational awareness is not affected by the system.	RQUA7
HUA8	Driver stress is not affected by automation.	RQUA8
HUA9	The mental workload of the driver is not affected by automation.	RQUA9
HUA10	The mental workload does not change after prolonged driving with the system.	RQUA10
HUA11	Transfer of control is not affected by mental workload.	RQUA11
HUA12	The drivers do not engage more in secondary tasks when driving with automation compared to driving without automation.	RQUA12
HUA13	The drivers do not become complacent when driving with automation.	RQUA13

ID	Hypothesis	Related Research Question
HUA14	The time for the drivers to make decision after a safety critical event does not differ between manual driving mode and automated driving.	RQUA14
HUA15	Driving skills don't degrade with time using automation.	RQUA15

Figure 4.7: Hypotheses concerning the **effects of automation on the driving task**.

Hypotheses concerning **take-over situations and regaining control** are:

ID	Hypothesis	Related Research Question
HUA16	There is no change in the drivers' take-over behaviour in long term.	RQUA16
HUA17	The drivers do detect automation failures.	RQUA17
HUA18	The drivers do not fail to respond to a critical situation because the system failed to notify them.	RQUA18
HUA19	The drivers take the appropriate measure after a system brake down.	RQUA19
HUA20	The drivers do not follow a wrong recommendation instead of vigilant information seeking and processing.	RQUA20
HUA21	The drivers are confident about the correctness of their decision after a system brake down.	RQUA21
HUA22	There is no difference in intervention time between drivers with an internal locus of control and those with an external locus of control.	RQUA22

Figure 4.8: Hypotheses concerning **take-over situations and regaining control**.

Hypotheses concerning the driver's **trust, opinions and acceptance** of the system are:

ID	Hypothesis	Related Research Question
HUA23	The drivers have the correct mental representation of the system.	RQUA23
HUA24	The drivers have no over- or under-trust on the system.	RQUA24
HUA25	The drivers experience automated driving as an improvement in their driving.	RQUA25
HUA26	The drivers have their distinct opinion about the system.	RQUA26
HUA27	The drivers find the system useful and satisfactory.	RQUA27

ID	Hypothesis	Related Research Question
HUA28	Automation failures do not influence the drivers' attitude to the system	RQUA28
HUA29	The drivers are interested to have and to pay for the system	RQUA29

Figure 4.9: Hypotheses concerning **trust, opinions and acceptance** of the system.

Hypotheses concerning **non-users** of the system are:

ID	Hypothesis	Related Research Question
HUA30	Non-users' behaviour is not influenced by interaction with equipped vehicles.	RQUA30

Figure 4.10: Hypotheses concerning **non-users** of the system.

4.1.3 Evaluation indicators

Hypothesis testing is done based on indicators. In the following table, all relevant indicators are presented and linked to the above hypotheses (for more details see Annex 5):

ID	Indicators	Evaluation Aspects	Related Hypothesis
IUA1	Position in parking space,	driver behaviour and performance	HUA1
IUA2	Time for parking manoeuvre		HUA1
IUA3	Speed: distribution, mean, stddev.		HUA1,HUA2, HUA5, HUA11, HUA15, HUA30
IUA4	Speed difference		HUA1
IUA5	Properly adapted speed to the situation		HUA1, HUA2, HUA5, HUA15, HUA30
IUA6	Distance forward: distribution, mean, stddev.		HUA1, HUA2, HUA5, HUA15, HUA30
IUA7	Distance back: distribution, mean, stddev.		HUA1
IUA8	Lane position: distribution, mean, stddev		HUA1, HUA2, HUA5, HUA11, HUA15, HUA30
IUA9	Side distance to obstacle		HUA1, HUA2, HUA5, HUA6, HUA30
IUA10	Side distance to VRU		HUA1, HUA2, HUA5, HUA6, HUA30
IUA11	Accepted gap: distribution, mean, stddev.		HUA1, HUA2, HUA5, HUA6, HUA30
IUA12	Safe and lawful lane change		HUA1, HUA2, HUA5, HUA6, HUA15, HUA30

ID	Indicators	Evaluation Aspects	Related Hypothesis
IUA13	Safe and lawful overtaking		HUA1, HUA2, HUA5, HUA6, HUA15, HUA30
IUA14	Safe and lawful merging		HUA1, HUA2, HUA5, HUA6, HUA15, HUA30
IUA15	Safe and lawful passage of intersection		HUA1, HUA2, HUA5, HUA6, HUA30
IUA16	Safe and lawful passage of traffic light		HUA1, HUA2, HUA5, HUA6, HUA30
IUA17	Safe and lawful passage of roundabout		HUA1, HUA2, HUA5, HUA6, HUA30
IUA18	Safe and lawful enter/exit to/from the motorway		HUA1, HUA2, HUA5, HUA6, HUA30
IUA19	The frequency and duration of being in an “unsafe state”		HUA2, HUA5, HUA6, HUA15, HUA30
IUA20	Stopping behaviour		HUA2, HUA5, HUA6, HUA30
IUA21	Yielding behaviour		HUA2, HUA5, HUA6, HUA30
IUA22	Interaction and communication with other road users		HUA2, HUA5, HUA6, HUA30
IUA23	Usage of system in percent of total driving time during relevant situations		HUA3
IUA24	Driving in suggested function/system settings in percent of total time of a certain suggested function/system settings		HUA4
IUA25	SAGAT scores [77]	effects of automation on the driving task	HUA7
IUA26	Short Stress State Questionnaire (SSSQ) scores [78]		HUA8
IUA27	Raw Task Load indeX (RTLX) [79]		HUA9, HUA10, HUA13
IUA28	Percent of driving time the driver being engaged in secondary task		HUA12
IUA29	Task-related Boredom Scale (TBS) [80]		HUA13
IUA30	The probability of detection of automation failure		HUA13
IUA31	Reaction time for detection of automation failure,		HUA13
IUA32	The number of detection errors		HUA13
IUA33	Root-Mean-Squared-Error (RMSE) of secondary task		HUA13
IUA34	The time from a safety critical event arises until the driver takes an action		HUA14

ID	Indicators	Evaluation Aspects	Related Hypothesis
IUA35	Time for the driver to make decision of transfer of control		
IUA36	Number of 1° steering reversals per minute	take-over situations and regaining control	HUA11
IUA37	High Frequency Control of steering (in the 0.3–0.6 Hz band)		HUA11
IUA38	Visual attention measured by eye tracking value of 'Percent Road Centre'		HUA11
IUA39	Driver reaction type in a take-over situation		HUA11
IUA40	Driver reaction time in a take-over situation		HUA16
IUA41	The share of registered automation failures		HUA17
IUA42	The number of driver responses to critical situations related to all situations the system did not notified them		HUA18
IUA43	Driver reaction type to a system brake down		HUA19, HUA20
IUA44	Questionnaire answer on the drivers' confidence about the correctness of their decision after a system brake down		HUA21
IUA45	The time from a safety critical event arises until the driver takes an action		HUA22
IUA46	Questionnaire answer on the drivers' mental representation of the system	understanding, trust, opinions and acceptance of the system	HUA23
IUA47	Scores on the self-report scale of trust [40]		HUA24
IUA48	Questionnaire answer on the drivers' experience if automated driving improved their driving		HUA25
IUA49	Questionnaire answer on the drivers' opinion about the system		HUA26
IUA50	Usefulness and satisfaction scale scores [81]		HUA27
IUA51	Questionnaire answer on the influence of automation failures on the drivers' attitude to the system		HUA28
IUA52	Questionnaire answer on the drivers' interest to have and to pay for the system		HUA29

Figure 4.11: Indicators for user-related assessment in AdaptIVe.

4.2 Methods and tools for user-related assessment

To investigate behaviour related issues when driving with a well-functioning driving automation function/system, logging of driving data and observational studies (either in a driving simulator, on a test track and/or in real traffic) are applicable.

Driver performance

Driver performance can be measured through assessments of drivers' attention to potential hazards (i.e., detection accuracy), accuracy of vehicle control (i.e., variability in lateral position) and variations in mean speed (reflecting the effort to compensate for increased workload). The findings of Matthews & Desmond [63] suggest that loss of on-road driving performance should be assessed through indices of heading error or fine-steering reversals in undemanding conditions.

Logging of driving data either in a driving simulator, on a test track and/or in real traffic can yield indicators, such as:

- driving speed,
- distance to the vehicle ahead,
- lateral position,
- distance to side obstacles,
- accepted gap,
- the frequency and duration of being in an "unsafe state",
- the time from a safety critical event arises until the driver takes an action,
- usage of system in percent of total driving time during relevant situations,
- driving in suggested function/system settings in percent of total time of a certain suggested function/system settings.

Behavioural observations - the Wiener Fahrprobe

To observe driver behaviour and possible changes in them, the in-car observation method (Wiener Fahrprobe), originally developed by Risser & Brandstätter [82] and designed to observe learning drivers can be employed. The method, however, also proved to be useful for studying driver behaviour in real traffic. The observations are carried out by two observers, riding along in the car with the driver, where one of the observers (called the coding observer) studies standardised variables such as speed behaviour, yielding behaviour, lane changes and interaction with other road users. The other observer carries out "free observations" such as conflicts, communication and special events that are hard to predict, let alone to standardise. The method was validated by Risser & Brandstätter [82] when it was shown that there was a correlation between observed risky behaviour and accidents. Other validation work was done by Hjälmdahl and Várhelyi [83] who showed that drivers' speed levels with observers in the car did not differ from their speed levels when driving their own cars. They also demonstrated that it was possible to train observers to perform the observations objectively and reliably. Observations should be carried out both before implementing the system of automation, directly after and after at least 6 months of driving with the system to see if any changes in driver skills occur.

Behavioural observations either in a driving simulator, on a test track and/or in real traffic can yield indicators, such as:

- adaptation of speed to potentially critical situations,
- lane choice, lane change, lane keeping behaviour,
- overtaking behaviour,
- stopping behaviour,
- yielding behaviour,
- behaviour at traffic lights,
- interaction and communication with other road users.

Trust and reliance

Merritt et al. [43] conclude that users do not fully understand why they experience a given level of trust in an automated system. "When asked why they trust or do not trust a system, the users will be capable of describing the effects of only their explicit attitudes — not their implicit ones". Thus, interviews, surveys, and focus groups may provide an incomplete understanding of factors influencing user trust. To better understand why users trust automation, implicit attitudes must be measured using carefully constructed implicit techniques.

Propensity to trust automation can be measured with the six-item scale proposed by used by Merritt et al. [43]. The response options are on a 5-point Likert-type scale ranging from 1 (strongly disagree) to 5 (strongly agree). Propensity to Trust Scale Items are as follows:

1. I usually trust machines until there is a reason not to.
2. For the most part, I distrust machines.
3. In general, I would rely on a machine to assist me.
4. My tendency to trust machines is high.
5. It is easy for me to trust machines to do their job.
6. I am likely to trust a machine even when I have little knowledge about it.

Actual trust in the system in question can be assessed using a six-item self-report scale employed by Merritt [40]. The item responses are on a 5-point Likert-type scale ranging from 1 (strongly disagree) to 5 (strongly agree). Actual Trust Scale Items are as follows:

1. I believe the … system is a competent performer
2. I trust the … system
3. I have confidence in the advice given by the … system

4. I can depend on the … system
5. I can rely on the … system to behave in consistent ways
6. I can rely on the … system to do its best every time I take its advice

Mental representation of the system

Users' understanding of the limitations of the systems can be investigated with help of interviews after have driven with the system. Questions of interest are:

- Can you describe how the … system helps you in car driving?
- Are you aware of any limitations about the … system? If yes, please explain.
- When learning to use the … system, were there things that were especially difficult to learn about the system? If yes, please explain.

Locus of control

There is a great deal of methods to measure locus of control, see e.g. [84]. Having identified drivers with an external locus of control, their tendency to intervene in time when the automated system fails should be tested in laboratory experiments or in a driving simulator.

Transfer of control

To measure driver's ability to resume control of driving, the following variables can be used: mean and minimum values of speed and their Standard deviation, Standard Deviation of Lane Position (SDLP), number of 1° steering reversals per minute, High Frequency Control of steering (in the 0.3–0.6 Hz band) and visual attention measured by eye tracking value of 'Percent Road Centre'. Merat, et al. [44] found that drivers' ability to regain control stabilised after around 40 seconds

Mental workload

Mental workload of the driver can be investigated with the help of the Raw Task Load indeX (RTLX) method proposed by Byers et al. [79]. According to this method, the subjects rate six different workload aspects, namely mental demand, physical demand, time pressure, performance, effort and frustration level. Continuous scales ranging from "very low" (0) to "very high" (100) are used. The difference in workload between driving with the system on compared to off can be calculated for each test driver.

To study if mental workload affects transfer of control to and from the driver, during a driving simulator study, the driver should be engaged with a secondary task.

To investigate if mental workload changes after prolonged driving with the system, observations should be carried out both before implementing the system of automation, directly after and after at least 6 months of driving with the system to see if any changes in driver skills occur.

Stress

To measure impact on driver stress, heart rate as an objective physiological arousal measure can be used [54] (Reimer et al., 2010). There are various heart beat detectors available on the market.

To assess driver stress based on subjective measures, stress state questionnaires can be used. Helton [78] presented validation evidence of a short multidimensional self-report measure of stress state, the Short Stress State Questionnaire (SSSQ) derived from the Dundee Stress State Questionnaire (DSSQ) [63]. Factor analyses differentiated three aspects of subjective stress (similar to the DSSQ): Task Engagement, Distress, and Worry. The SSSQ appeared to be a useful measure of stress state [78]. It consists of 24 items with 5 categories of response, from 1 = strongly disagree to 5 strongly agree, see the items in Figure 4.12. The test drivers should complete the SSSQ before and after test driving.

SSSQ factor	Item
Engagement	I feel alert. I feel active. I am committed to attaining my performance goals. I want to succeed on the task. I am motivated to do the task. I feel confident about my abilities. I expect to perform proficiently on this task. Generally, I feel in control of things.
Distress	I feel dissatisfied. I feel depressed. I feel sad. I feel impatient. I feel annoyed. I feel angry. I feel irritated. I feel grouchy.
Worry	I am trying to figure myself out. I am reflecting about myself. I am daydreaming about myself. I feel self-conscious. I am worried about what other people think of me. I feel concerned about the impression I am making. I thought about how others have done on this task. I thought about how I would feel if I were told how I performed.

Figure 4.12: Factor Structure of the SSSQ task scales, based on Helton [78].

Boredom

To assess perceived boredom experienced by the drivers, the Task-related Boredom Scale (TBS) [80] can be used. The TBS addresses eight factors thought to contribute to feelings of boredom: stress, irritation, relaxation, sleepiness, alertness, concentration, passage of time, and satiation. In addition, respondents are also asked to provide an estimation of their overall feeling of boredom. A total boredom score is calculated by summing all the subscales. The sleepiness, time passage and desire for task to end are reversed scored.

Fatigue

To quantify the progression of driver fatigue, objective and/or subjective measurements can be employed. Objective measures are standard deviation of speed (SDS), standard deviation of the lateral position (SDLP), frequency of extremely large steering wheel movement (SWM) (>N10°), frequency of line crossings and reaction time (RT) [85]. Data to be acquired in simulated driving using a common method to induce fatigue by having subjects perform a demanding secondary task. Ting et al. [85] used a simple RT-test to assess the sustained attention of drivers throughout the driving task: Two red circular images (radius of 25 cm; horizontal angle of 11°–23° left) were randomly displayed on a screen every 2 km. When the visual stimulus appeared (duration of 3.6 s), the subject was required to respond to the stimulus by turning off an identical indicator. The system automatically recorded individual RTs. If no response was made within 3.6 s, a new RT-test was started. The variation in mean RT was employed to assess driver vigilance.

Matthews & Desmond [62]developed a multidimensional measure of subjective fatigue state (SFS). The fatigue scale comprises 24-items, relating to four aspects of fatigue: 1) Visual fatigue, 2) Muscular fatigue; 3) boredom, and 4) Malise, see Figure 4.13. Subjects are required to rate on 0-5 numerical scales the extent to which they experience the 24 items of fatigue symptoms.

Factor (symptom type)	Item
Visual fatigue	Flickering in eyes Feeling of heaviness in the eyes Eyes feel strained Vision is blurred Road appears to 'swim' Unaware of objects off the road
Muscular fatigue	Have a headache Hearing ability reduced Humming in ears
Boredom	Bored Would rather be doing something else Fed up with the task

Factor (symptom type)	Item
	Apathetic Don't care what happens next Don't want to do the task ever again Find the task monotonous Don't want to think about the task
Malaise	Feel ill Feel stomach pains Feel sick or nauseous Feel tired in the whole body Having tremors in the limbs Feel stiff in the legs and arms Unable to straighten up in posture

Figure 4.13: Fatigue factors and items [62].

Situational awareness

Endsley's [66] model of situational awareness is arranged into three hierarchical levels: 1) Perception of the elements in the environment: 2) Comprehension of the current situation: 3) Prediction of future status. Endsley proposes that situational awareness is discussed in terms of mode awareness, spatial awareness and time awareness.

To investigate if automation can have an impact on the out-of-the-loop performance problem and to verify the role of situational awareness in this process, a driving simulator experiment comparing automated driving with manual driving is appropriate. The dependent variables should be: situational awareness, mental workload and time for the drivers to make decision with a simulated system brake down.

Endsley [77] developed the Situation Awareness Global Assessment Technique (SAGAT) for air-fight scenarios. Adapted to the car driving context, SAGAT measures a driver's situational awareness (SA) in the following manner:

- The driver drives the car in the driving simulator in a given scenario using a given automation system in a driver-in-the-loop simulation.
- At some random point in time the simulation is halted and the instrument panel and out-the-window displays are blanked.
- The driver is asked a series of questions in order to determine his/her knowledge of the situation at that exact moment in time. These questions correspond to the driver's SA requirements. The SAGAT queries are programmed on a computer, available at

each driver station, to allow for the rapid input and storage of highly spatial information.

- As it is impossible to query the driver about all of his SA requirements in a given stop, a portion of the SA questions are randomly selected and asked of the driver each time. This random sampling method allows consistency and statistical validity, thus allowing SA scores to be easily compared across trials, drivers, systems and scenarios. Some of the questions in any particular query will pertain to highly important SA information and some of the questions will pertain to more secondary SA information.
- At the completion of the trials, the query answers are evaluated on the basis of what was actually happening in the simulation. This is accomplished by comparing the driver's answers to data collected from the simulation computers. (Where necessary this may be augmented by subjective evaluations from a team of experts) The comparison of the real and perceived situation provides an objective measure of driver SA.
- A composite SAGAT score is then determined for the system under investigation. SAGAT score is stratified into the three zones (immediate, intermediate, and long-range), to provide evaluators with a better picture of the driver's SA. Additionally, individual components contributing to SA can be examined separately to provide more detailed diagnostics to the designer.
- This random sampling process is repeated a number of times for each of several drivers driving with the same system, in order to obtain the number of observations required for statistical significance. SAGAT scores for any system design can then be compared to SAGAT scores for other systems.

Out-of-the-loop performance problem

To investigate the occurrence of the out-of-the-loop performance problem a driving simulator experiment can be employed, where the primary independent variable is manual driving versus automation with a simulated system brake down. The dependent variables are: situational awareness, the decision selected, time for the drivers to make decisions, drivers' confidence about the correctness of decision made and mental workload.

Complacency

Operational definitions of complacency as behaviour need to be based on direct or indirect behavioural indicators of attention allocation. Direct indicators may be derived from eye-tracking analyses or other indicators of monitoring or information-sampling behaviour. Indirect indicators of attention may include assessments of a reallocation of attention resources by means of secondary-task methods. However, attention resources can only be regarded as an indication of complacency or automation bias if the observed effects are compared with some normative model of "optimal attention allocation" in interaction with a given system. Defining appropriate normative models for interaction with given automated systems, however, represents a challenge and needs more research. [37].

Complacency may be influenced by the individual characteristics of the human operator [58]. Singh et al. [86] developed a 20-item scale, the Complacency Potential Rating Scale (CPRS), which measures attitudes toward automation that reflect a potential for developing automation-induced complacency. By factor analysis of the scale, they indicated four complacency-potential related dimensions: trust, confidence, reliance, and safety, which suggest that high scores on these factors are associated with complacency. Although, the CPRS has been shown to be a good indicator of an operator's complacency potential [58] it does not measure factors that may influence the onset of complacency, such as workload, boredom, or cognitive failure. Hence, other measures are also needed to assess automation-induced complacency.

Investigating the relationship between the individual differences of complacency potential, boredom proneness and automation-induced complacency, Prinzel et al. [58] found that operators, performing the monitoring task under variable automation reliability condition did significantly better than those in the constant automation reliability condition, indicating that a constant high automation reliability (87.5 % of malfunctions detected by the automation) impairs an operator's ability to monitor for infrequent automation failures in a multitask environment. Hence in assessing automation-induced complacency, a high automation reliability level (above 87.5 %) should be used with occasional automation failure.

Usability

Usability evaluation is to be made to assess the degree to which a system's human–machine interface (HMI) complies with usability criteria applicable in the specific context of use [87]. Harvey et al. [87] reviewed over 70 usability evaluation methods for In-Vehicle Information Systems (IVIS) and matched each of the selected methods with thirteen usability criteria, clustered in six main factors, i.e. Dual task environment, Range of users, Environmental conditions, Training provision, Frequency of use, Uptake.

To evaluate the users' perceptions of the system under investigation, the subjective method of System Usability Scale (SUS) can be employed. The SUS consist of ten statements, against which participants rate their level of agreement on a 5-point Likert scale with 5 categories of response, from 1 = strongly disagree to 5 strongly agree, [88]:

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.

3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

A single usability score is computed from the ratings, which allows for comparing user opinions across different systems. A total score is calculated by adding the scores of the items. Item 1,3,5,7 and 9 are given a score of scale position minus 1. Items 2, 4, 6, 8 and 10 are given a score of scale position minus 5. The total score is then multiplied by 2.5 to achieve a SUS score that range between 0 (very low usability) and 100 (very high usability).

Acceptance

One of the methods, widely used to assess acceptance of driver assistance systems is the Usefulness and Satisfaction method proposed by van der Laan et al. [81]. According to the method, the subjects assess nine components related to usefulness and satisfaction: "good – bad", "pleasant – unpleasant", "effective – superfluous", "nice – annoying", "likable – irritating", "useful – useless", "assisting – worthless", "desirable – undesirable", "raising alertness – sleep inducing", on a bipolar scale.

More recently, Adell [89] put forward another model, based on the Unified Theory of Acceptance and Use of Technology (UTAUT) used in the area of information technology [90], for analysing acceptance issues of driver assistance systems. Adell [91] undertook a pilot test of her model with promising results. The 17 items for assessing 'behavioural intention', 'performance expectancy', 'effort expectancy' and 'social influence' were adopted from Venkatesh et al. [90], some of them adapted to fit the context of driver assistance systems, see Figure 4.14. Each item is rated on a seven-point scale, ranging from "strongly disagree" (1) to "strongly agree" (7).

Factors	Items
Behavioural intention to use the system	Imagine that the system was on the market and you could get the system in your own car.
	I would intend to use the system in the next 6 months
	I would predict I would use the system in the next 6 months

Factors	Items
	I would plan to use the system in the next 6 months
Performance expectancy	I would find the system useful in my driving
	Using the system enables me to react to the situation more quickly
	Using the system increases my driving performance
	If I use the system, I will decrease my risk of being involved in an accident
Effort expectancy	My interaction with the system would be clear and understandable
	It would be easy for me to become skilful at using the system
	I would find the system easy to use
	Learning to operate the system is easy for me
Social influence	Imagine that the system was on the market and you could get the system in your own car.
	People who influence my behaviour would think that I should use the system
	People who are important to me would think that I should use the system
	The authority would be helpful in the use of the system
	In general, the authority would support the use of the system

Figure 4.14: The adapted UTAUT items to assess driver support systems (from [91]).

Experienced effects

To assess what effects the drivers' experienced when using the system they can be asked to state how they thought different aspects of driving changed when using the system. The drivers can be asked to compare their experiences of using the system to their experience of driving without the system on a bipolar continuous scale from "decreased greatly" to "increased greatly" where "neither" represents the middle point. The following issues are of interest:

- Your safety in traffic
- The risk of getting speeding tickets
- Your travel time
- Your fuel consumption
- Your irritation
- Your stress
- Your enjoyment when driving
- Your feeling of being in the way of others
- Your attention to traffic
- Your image
- Your comfort when driving

Perceived benefits

The drivers can be asked to write down benefits and problems they experienced when using the system. Open questions:

- *"What benefits did you encounter when using the system compared to driving without the system?"*
- *"What differences did you experience when using this system compared to driving without the system? Please mark your estimation with a cross on the scale.*

They can also be asked to state in what extent they think the system would give them benefits or disadvantages if they were to use the system in their everyday driving on some given items: "Do you think the system can give you benefits or disadvantages in your everyday driving?" The answers can be given on a continuous scale from "very large disadvantage" to "very large benefit", with "neither" represented the middle point.

The following issues are of interest:

- Risk of being involved in an accident
- Risk of getting speeding tickets
- Fuel consumption
- Travel time
- Your feeling of being in the way of other drivers
- Image
- Comfort
- Enjoyment when driving
- Other: ……….

Willingness to have and pay

With help of a questionnaire, the subjects should be asked what they would think about having the system in their regular car:

- *"Would the system be mounted in your car for free, would you use it?"*

Answers can be given on a continuous scale from "very bad" to "very good" where "neither" represented the middle point.

Also, how much they would be willing to pay for it:

- *"The system will probably be sold as an optional system in cars. Please indicate at which price you would be willing to buy it."*

Answer can be given by choosing an alternative price-interval.

The methods and tools for investigating the user-related issues described above are presented below per issue. For a summary image see Figure 4.15.

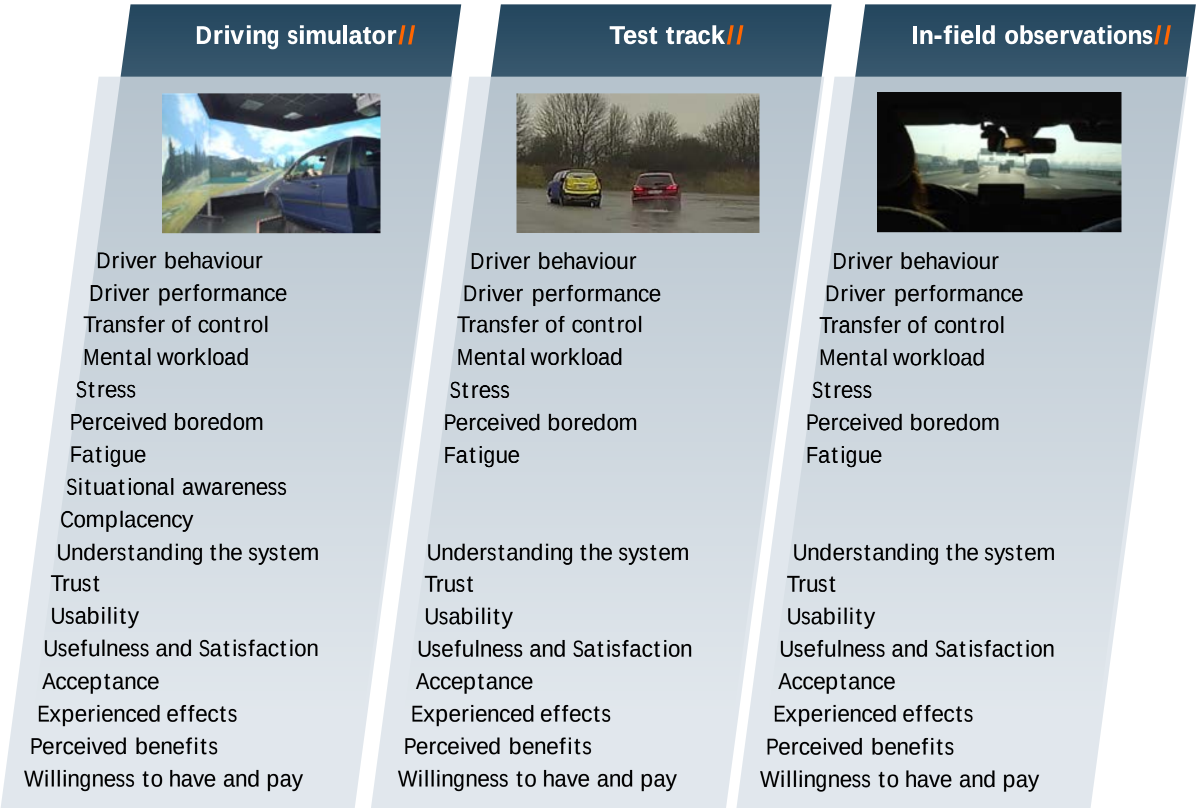


Figure 4.15: Test tools and methods for assessing user-related issues.

4.3 Requirements of the user-related assessment

In the following, the requirements with respect to safety, test tools and test effort for user-related assessment are discussed.

4.3.1 Safety

Ensuring safety at any time during the test is of major importance for the user-related assessment. This means that testing needs to be safe at any time -meaning any damage of vehicles and persons need to be prevented during the tests. The test environments may consists of driving simulator, test track with staged scenarios or driving in real traffic, hence different safety requirements for the different environments need to be considered.

The controlled environment of a driving simulator has the advantage that damages to persons or objects can straightforwardly be provided against.

The controlled environment of a test track also has the advantage that damages to persons or objects can be provided against. During the tests, persons who are not involved in the tests, need to keep a safety distance[12] to the test vehicle. If the driver is not in the vehicle during the test the safety distance needs to be raised compared to an equal test with the driver in the car. If other objects (vehicles, pedestrians, etc.) are involved in the test, crash-able dummy objects should be used in order to not cause any damage. Independent of the safety measures for the test environment the driver of the car or any other supervising person must always be capable to regain the control of the vehicle and to switch off the function respectively bring the vehicle (immediately) to a standstill at any time of the tests.

Tests in real traffic require high safety standards since such tests involve also other road users. Furthermore, the environment cannot be controlled in the same manner as on test tracks. Therefore, the safety measures need to focus on the test vehicle, the function and the driver of the test vehicle.

Before the actual tests in real traffic, pre-tests of the vehicle and the function(s) are required. It should be checked, whether basic behaviour of the function respectively system under test is in line with the specified function behaviour. During the pre-tests also installed safety functions (e.g. minimum risk or safe stop function) should be checked, whether they work properly. During the tests on public roads, a trained co-driver ("safety driver"), who is familiar with the car and the tested functions, must always to be in the front passenger seat. The safety-driver should always be capable to switch off the function and help the test driver regain control at any time of the tests. In general, the safety driver should intervene in case:

- the system's behaviour is not in line with the expected system behaviour in a negative sense,
- a situation becomes critical and it is not clear, whether the system can cope with it,
- any failure or misbehaviour in one of the components (sensor, computing unit, actuator) or the whole function is detected.

A legislative aspect to take into account is that the test drivers should have an appropriate driving license. Furthermore, insurance for third party and personal damage must exist. The coverage of the insurance should be at least as high as requested by the law. Finally, a road test approval for the test vehicle and the testing region must be available.

4.3.2 Tools

The applied test tools should enable a safe, efficient and accurate testing of all to be evaluated functions in AdaptIVe. The required tools for user-related tests are:

[12] Safety Distance depends on the velocity of the test as well as the foreseen stopping distance of the vehicle and need to be defined in cooperation between SP7 and the VSP

- **Driving simulator** which needs to be prepared in accordance to the tested function/system and test scenarios.
- **Test track / area**, which needs to be defined in accordance to the tested function/system.
- **Public Road**, which needs to be defined in accordance to the tested function/system. Here, it has to be checked, whether required information on the road (e.g. digital map data, road markings/traffic signs) are available, and whether the road fulfils the requirements of the functions (e.g. dimensions of the test roads/lanes) and the traffic composition is of relevance for the tested function/system.
- **Test vehicle**, which needs to be equipped with logging equipment.
- **Logging equipment** for logging of all defined signals (see Annex 3). Besides to the CAN-Signals also video data of the relevant perspectives (front view, rear view, side view) should be logged during the tests.

Target objects, which have to be defined depending on test case (other vehicles, balloon cars). As far as possible the used target object should be crashable and representative of the "real life objects" they represent for the sensors used by the functions / systems on the car.

4.4 Example of user-related assessment

The evaluation plan presented below is comprehensive, offered to test all user-related issues of automated driving. It is understood that carrying out all of them is resource and time demanding, hence the set-up of the final evaluation plan will probably confine to the most rewarding ones.

The "user" is represented by a sample of car drivers who take part in the tests. The sample should be representative of the driver population, reflecting the characteristics and needs of the whole population. When electing users for testing, there are a number of issues that should be taken into consideration. Driver characteristics such as age, gender and experience are of importance, where older drivers and novice drivers may have more difficulties in interacting with an in-vehicle technology [87]; but besides user competence, also user personality (e.g. "users scoring high on the personality factor conscientiousness may identify more usability problems because they approach the testing procedure more thoroughly" pp 132), user attitude (openness towards technology), as well as user state (temporary condition of the user) may influence the outcome of usability tests [92].

4.4.1 Assessment plan for a City Chauffeur system

City Chauffeur functions comprise the following driving situations:

- Lane following
- Speed adaptation
- Vehicle following in lane
- Obstacle or VRU on the road
- Lane change
- Intersection handling
- Urban roundabouts handling
- Traffic light handling

To carry out user-related assessment of the City Chauffeur system, either driver simulator experiment, observations of driving on a test track or in real traffic can be employed depending on the focus of the assessment and possible access to facilities and/or permission of naïve drivers to drive the demonstrator vehicle on a test track or on public roads. Observation of driver behaviour in real traffic gives the highest validity of results, while a driver simulator experiment allows for staging situations where also situational awareness and possible complacency can be studied.

Driver Simulator experiment

Participants (20-30) to be tested individually (within-subject design). They should complete two 30-40 min driving simulator sessions on a test route consisting of urban and semi-urban roads. They serve as their own controls (within subject design). The order of driving should be balanced in such a way that every other subject drives first with the system switched off and then with the system switched on. For the following subject the order of driving is reversed. By doing this, the effects of biasing variables, such as getting used to the test route or to the observers and the test situation cannot be eliminated, but such effects can be spread evenly across the situations. Before the test rides the drivers should be informed that the trial is about the system and not about them as drivers and that all data collected will be anonymous. Also, before the use of the system, a brief explanation of the system is to be given to the drivers. The drivers should be instructed to drive as normal as possible and ask for whatever doubts or questions they might have during the test. Each participant should be informed about the system and its limitations and that it might happen that if malfunctions, also that they are required to drive as they usually drive. They should be instructed on the secondary task and given a 10-min practice session to make acquaintance with it. Participants also should be instructed to focus on equally on driving and the secondary task.

Test scenarios to assess user related effects for each of the following functionalities should be staged at least 6 times per test drive in a mixed way:

- Lane following
- Speed adaptation
- Vehicle following in lane

- Obstacle or VRU on the road
- Lane change
- Intersection handling
- Urban roundabouts handling
- Traffic light handling

Driving data as well as driver- and system generated events are to be logged during both riding sessions; Logged data should include:

- driving speed,
- distance to the vehicle ahead,
- lateral position,
- distance to side obstacles,
- steering wheel movement,
- accepted gap,
- the frequency and duration of being in an "unsafe state",
- the time from a safety critical event arises until action is taken,
- usage of system in percent of total driving time during relevant situations,
- driving in suggested function/system settings in percent of total time of a certain suggested function/system settings.

Driver performance is measured through assessments of drivers' attention to potential hazards (i.e., detection accuracy), accuracy of vehicle control (i.e., variability in lateral position) and variations in mean speed (reflecting the effort to compensate for increased workload).

The driver's **ability to resume control** of driving, the following variables to be used: mean and minimum values of speed and their Standard deviation, Standard Deviation of Lane Position (SDLP), number of 1° steering reversals per minute, High Frequency Control of steering (in the 0.3–0.6 Hz band), visual attention measured by eye tracking value of 'Percent Road Centre' and time to stabilised control of these variables) [44].

Mental workload of the driver is assessed with the help of the Raw Task Load indeX (RTLX) method proposed by Byers et al. [79] after both rides.

To measure impact on driver **stress**, heart rate as an objective physiological arousal measure can be used [54] after both rides.

To assess driver's **subjective stress**, the Short Stress State Questionnaire (SSSQ) is to be used [78] after both rides.

To assess **perceived boredom** experienced by the drivers, the Task-related Boredom Scale (TBS) [80] is to be used after both rides.

To quantify the progression of driver **fatigue**, objective measures, such as are standard deviation of speed (SDS), standard deviation of the lateral position (SDLP), frequency of extremely large steering wheel movement (SWM) (>N10°), frequency of line crossings and reaction time (RT) to be used during both rides.

To assess **subjective fatigue** state, the fatigue scale questionnaire [62] to be used after both rides.

To **assess situational awareness** the SAGAT method is to be employed during both rides.

To investigate if automation can have an impact on the **out-of-the-loop performance** problem and to verify the role of situational awareness in this process, the primary independent variable is manual driving versus automation with a simulated system brake down. The dependent variables should be: situational awareness, the decision selected, time for the drivers to make decision with a simulated system brake down, drivers' confidence about the correctness of decision made and mental workload.

To investigate users' **understanding of the limitations of the system**, they are asked to answer the following questions after the second ride:

- Can you describe how the system helps you in car driving?
- Are you aware of any limitations about the system? If yes, please explain.
- When learning to use the system, were there things that were especially difficult to learn about the system? If yes, please explain.

Actual **trust** in the system in question can be assessed using a six-item self-report scale by Merritt [40] after the ride with the system ON.

For the **complacency** study should be carried out during two riding sessions with the system ON and in a multi-task environment [47]. The RTLX and TBS are part of this study as dependent variables as well as system monitoring performance and Secondary task Root-Mean-Squared-Error (RMSE). The performance measures for the system-monitoring task are: (a) the probability of detection of automation failure, (b) reaction time for detection, and (c) the number of detection errors.

To evaluate the users' perceptions of the system, its **usability**, the System Usability Scale (SUS) is to be used [88] after the ride with the system ON.

Usefulness and Satisfaction is assessed by the method proposed by van der Laan et al. [81] after the ride with the system ON.

To assess **acceptance** is the Unified Theory of Acceptance and Use of Technology (UTAUT), proposed by Adell [89] after the ride with the system ON.

The test drivers will be asked to answer questions concerning **experienced effects** of the system, **perceived benefits** with the system and **willingness to have and pay** for the system after the second ride.

Tests on test track

The aim of the tests on test track is to evaluate the effects on driver behaviour, reactions to and acceptance of the system.

Participants (20-30) to be tested individually (within-subject design). They should complete two 30-40 min driving sessions on a test track simulating urban and semi-urban road environments.

Test scenarios to assess user related effects for each of the following functionalities should be staged at least 6 times per test drive in a mixed way:

- Lane following
- Speed adaptation
- Vehicle following in lane
- Obstacle or VRU on the road
- Lane change
- Intersection handling
- Urban roundabouts handling
- Traffic light handling

The test drivers serve as their own controls (within subject design). The order of driving should be balanced in such a way that every other subject drives first with the system switched off and then with the system switched on. For the following subject the order of driving is reversed. By doing this, the effects of biasing variables, such as getting used to the test route or to the observers and the test situation cannot be eliminated, but such effects can be spread evenly across the situations. Before the test rides the drivers should be informed that the trial is about the system and not about them as drivers and that all data collected will be anonymous. Also, before the use of the system, a brief explanation of the system is to be given to the drivers. The drivers should be instructed to drive as normal as possible and ask for whatever doubts or questions they might have during the test. Each participant should be informed about the system and that they are required to drive as they usually drive.

Driving data as well as driver- and system generated events are to be logged during both riding sessions; **Logged data** should include:

- driving speed,
- distance to the vehicle ahead,
- lateral position,
- distance to side obstacles,

- steering wheel movement,
- the frequency and duration of being in an “unsafe state”,
- the time from a safety critical event arises until action is taken,
- usage of system in percent of total driving time during relevant situations,
- driving in suggested function/system settings in percent of total time of a certain suggested function/system settings.

Behavioural observations by the in-car observation method (Wiener Fahrprobe), yields indicators, such as:

- adaptation of speed to potentially critical situations,
- lane choice, lane change, lane keeping behaviour,
- overtaking behaviour,
- stopping behaviour,
- yielding behaviour,
- behaviour at traffic lights,
- interaction and communication with other road users.

Driver performance is measured through assessments of drivers’ attention to potential hazards (i.e., detection accuracy), accuracy of vehicle control (i.e., variability in lateral position) and variations in mean speed (reflecting the effort to compensate for increased workload).

The driver’s **ability to resume control** of driving, the following variables to be used: mean and minimum values of speed and their Standard deviation, Standard Deviation of Lane Position (SDLP), number of 1° steering reversals per minute, High Frequency Control of steering (in the 0.3–0.6 Hz band), visual attention measured by eye tracking value of ‘Percent Road Centre’ and time to stabilised control of these variables [44].

Mental workload of the driver is assessed with the help of the Raw Task Load indeX (RTLX) method proposed by Byers et al. [79] after both rides.

To assess driver’s **subjective stress**, the Short Stress State Questionnaire (SSSQ) is to be used [78] after both rides.

To assess **perceived boredom** experienced by the drivers, the Task-related Boredom Scale (TBS) [80] is to be used after both rides.

To assess **subjective fatigue** state, the fatigue scale questionnaire [62] to be used after both rides.

To investigate users’ **understanding of the limitations of the system**, they are asked to answer the following questions after the second ride:

- Can you describe how the system helps you in car driving?

- Are you aware of any limitations about the system? If yes, please explain.
- When learning to use the system, were there things that were especially difficult to learn about the system? If yes, please explain.

Actual **trust** in the system in question can be assessed using a six-item self-report scale by Merritt [40] after the ride with the system ON.

To evaluate the users' perceptions of the system, its **usability**, the System Usability Scale (SUS) is to be used [88] after the ride with the system ON.

Usefulness and Satisfaction is assessed by the method proposed by van der Laan et al. [81] after the ride with the system ON.

To assess **acceptance** is the Unified Theory of Acceptance and Use of Technology (UTAUT), proposed by Adell [89] after the ride with the system ON.

The test drivers to be asked to answer questions concerning **experienced effects** of the system, **perceived benefits** with the system and **willingness to have and pay** for the system after the second ride.

Tests in real traffic

The aim of the tests in real traffic is to evaluate the effects on driver behaviour, reactions to and acceptance of the system.

Participants (20-30) to be tested individually (within-subject design). They should drive twice along a test route of appr. 40-50 km, consisting of urban and semi-urban roads. They serve as their own controls (within subject design). The order of driving should be balanced in such a way that every other subject drives first with the system switched off and then with the system switched on. For the following subject the order of driving is reversed. By doing this, the effects of biasing variables, such as getting used to the test route or to the observers and the test situation cannot be eliminated, but such effects can be spread evenly across the situations. Every test person is given time to get accommodated to the situation and the car before the real observations are started. Therefore, an additional ten to 15 minutes ride to be done before the test ride. Before the test rides the drivers should be informed that the trial is about the system and not about them as drivers and that all data collected will be anonymous. Also, before the use of the system, a brief explanation of the system is to be given to the drivers. The drivers should be instructed to drive as normal as possible and ask for whatever doubts or questions they might have during the test.

Driving data as well as driver- and system generated events are to be logged during both riding sessions; **Logged data** should include:

- driving speed,
- distance to the vehicle ahead,

- lateral position,
- distance to side obstacles,
- steering wheel movement,
- the frequency and duration of being in an “unsafe state”,
- the time from a safety critical event arises until action is taken,
- usage of system in percent of total driving time during relevant situations,
- driving in suggested function/system settings in percent of total time of a certain suggested function/system settings.

Behavioural observations by the in-car observation method (Wiener Fahrprobe), yields indicators, such as:

- adaptation of speed to potentially critical situations,
- lane choice, lane change, lane keeping behaviour,
- overtaking behaviour,
- stopping behaviour,
- yielding behaviour,
- behaviour at traffic lights,
- interaction and communication with other road users.

Driver performance is measured through assessments of drivers’ attention to potential hazards (i.e., detection accuracy), accuracy of vehicle control (i.e., variability in lateral position) and variations in mean speed (reflecting the effort to compensate for increased workload).

The driver’s **ability to resume control** of driving, the following variables to be used: mean and minimum values of speed and their Standard deviation, Standard Deviation of Lane Position (SDLP), number of 1° steering reversals per minute, High Frequency Control of steering (in the 0.3–0.6 Hz band), visual attention measured by eye tracking value of ‘Percent Road Centre’ and time to stabilised control of these variables [44].

Mental workload of the driver is assessed with the help of the Raw Task Load indeX (RTLX) method proposed by Byers et al. [79]after both rides.

To assess driver’s **subjective stress**, the Short Stress State Questionnaire (SSSQ) is to be used [78] after both rides.

To assess **perceived boredom** experienced by the drivers, the Task-related Boredom Scale (TBS) [80] is to be used after both rides.

To assess **subjective fatigue** state, the fatigue scale questionnaire [62] to be used after both rides.

To investigate users’ **understanding of the limitations of the system**, they are asked to answer the following questions after the second ride:

- Can you describe how the system helps you in car driving?
- Are you aware of any limitations about the system? If yes, please explain.
- When learning to use the system, were there things that were especially difficult to learn about the system? If yes, please explain.

Actual **trust** in the system in question can be assessed using a six-item self-report scale by Merritt [40] after the ride with the system ON.

To evaluate the users' perceptions of the system, its **usability**, the System Usability Scale (SUS) is to be used [88] after the ride with the system ON.

Usefulness and Satisfaction is assessed by the method proposed by van der Laan et al. [81] after the ride with the system ON.

To assess **acceptance** is the Unified Theory of Acceptance and Use of Technology (UTAUT), proposed by Adell [89] after the ride with the system ON.

The test drivers have to be asked to answer questions concerning **experienced effects** of the system, **perceived benefits** with the system and **willingness to have and pay** for the system after the second ride.

5 In-traffic Assessment

While in the previous chapters the assessments of the user-related and technical aspects are described, this chapter discusses the methodology of the in-traffic assessment. The in-traffic assessment is intended to assess the in-traffic performance of automated driving systems in terms of interaction with surrounding traffic and infrastructure, as shown in Figure 2.2. This chapter starts with a global overview of the proposed in-traffic assessment method, which is based on Monte Carlo simulation of driving scenarios focussing on the environment and surrounding traffic. In section 5.1 the focus of the in-traffic assessment methodology is described by defining research questions, hypotheses and indicators. In section 5.2, an overview is provided of the various proposed tools. Section 5.3 details the requirements for the assessment methodology with respect to safety, tools and test conduction. Section 5.4 finalises this chapter by providing an example of the methodology applied to an existing function.

The objective of the in-traffic assessment methodology developed in AdaptIVe is to provide a generic framework for the in-traffic evaluation of automated driving functions in a complete range of traffic situations. For in-traffic assessment, evaluation should extend to a set of scenarios that represent the variation found in normal traffic conditions. In terms of frequency normal driving scenarios are most common, while safety-critical scenarios that also have to be considered in the evaluation are rare and collision scenarios are close to absent, depending on the functionality.

For a comprehensive evaluation in the in-traffic assessment the automated system or function needs to be assessed in all driving situation - not only “standard” situations, but also in all kinds of variations on those situations as well as rare critical situations and accident. The high amount of test kilometres that is required to detect such rare situations can be roughly estimated from the accident data or field data. The 100 car study performed in the USA [93] in which in total 3,220,000 km was driven in predominant urban and suburban, moderate to heavy traffic, resulted in 6.52 crashes per million kilometres and 2810 safety-critical events per million kilometres. The low number of detected accidents indicates already that for a comprehensive evaluation of the in-traffic behaviour of an automated function including all types of driving situations the required test distance needs to be of the order of magnitude of approximately one million kilometres or more.

Currently, in order to perform in-traffic evaluation of an automated function on public roads, a prototype vehicle needs to obtain approval for on-road testing, as a consequence will need to be driven by a professional driver and an vast amount of driving needs to be performed in order to verify the system requirements. The fact that driving is performed by a professional driver in a test-driving context, applies a bias on the naturalistic course of the driving scenarios.

In summary, prototype on-road driving by professional drivers for millions of kilometres is both costly and time consuming. In order to overcome these limitations, a simulation based approach is proposed. A common approach, in other domains, is to represent the variability in real-life traffic in a Monte Carlo sampled set of scenarios from certain data. The in-traffic assessment will be restricted to the use of specific driving scenario datasets. Within the domain of in-traffic evaluation, to the extent of our knowledge, this type of assessment has not been done before, and will develop further with involved partners prior to the performance of real testing.

In Figure 5.1, a global overview of the theoretical method is given. The various components of this method are described as follows:

- First of all, functional requirements are gathered and analysed. The functional requirements can come from a Hazard and Risk Analysis (HARA), where special or critical manoeuvres were selected. Based on this information the working range for the function is defined, which sets the scope for the driving situations to be considered.
- The technical evaluation itself is described in chapter 3. The resulting behaviour of the system in those pre-defined driving situations is input to the Monte Carlo simulations since they help to define and validate the simulation models of vehicle and function.
- The driving scenario dataset is a dataset that contains a series of in-traffic scenario in a descriptive, quantified and parameterised manner.
- The user-related assessment described in chapter 4, provides information on driver behaviour that is required to simulate driver usage of automated driving functions including the variability of this usage.

The Monte Carlo simulation approach is further detailed in Figure 5.2. The variations in the scenario dataset are described in terms of probability density functions and set the scope for the Monte Carlo simulations. A description of the scenario, together with quantified information, allows for modelling the various components to perform the simulations, specifically the driver model and the scenario model. Information on vehicle model and system model should be provided by the vehicle and system supplier and may be validated by means of tests performed during technical evaluation. The methodology for sampling in order to meet requirements will be further detailed in section 5.3.2. Outputs of the Monte Carlo simulations are probability density functions of predefined indicators, for example TTC or driven velocities. These probability density functions allow for assessment of the in-traffic behaviour and performance of the system in terms of interaction with surround infrastructure and traffic, purely in terms of quantified thresholds, not in terms of user-related assessment, which is described in chapter 5.3.2.

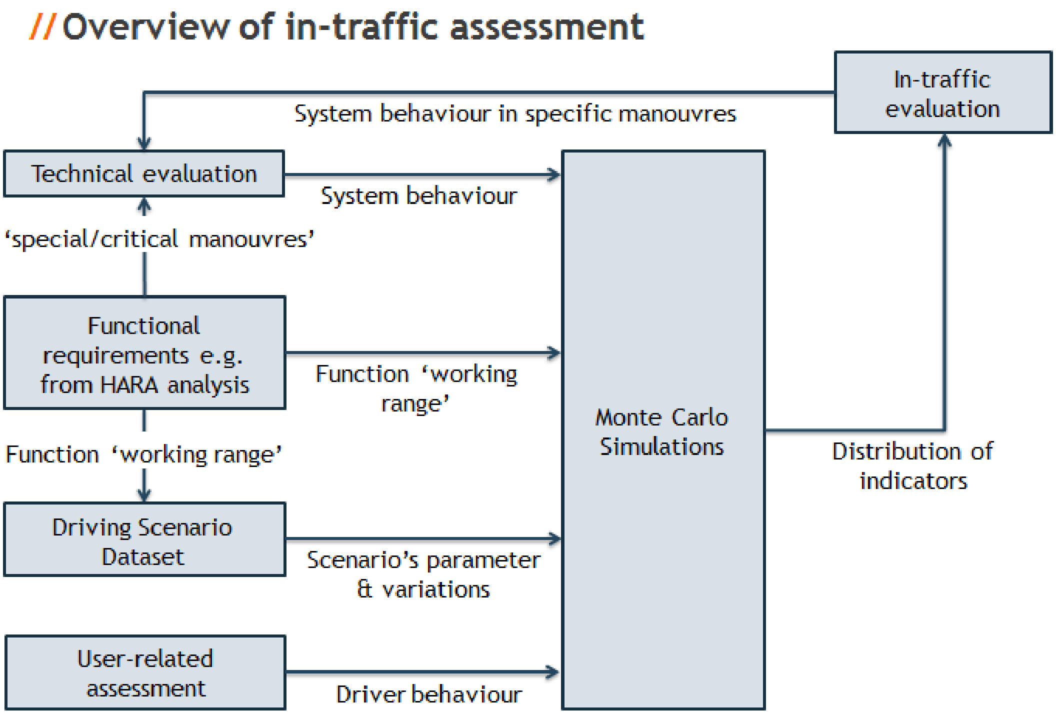


Figure 5.1: Global overview of in-traffic assessment method.

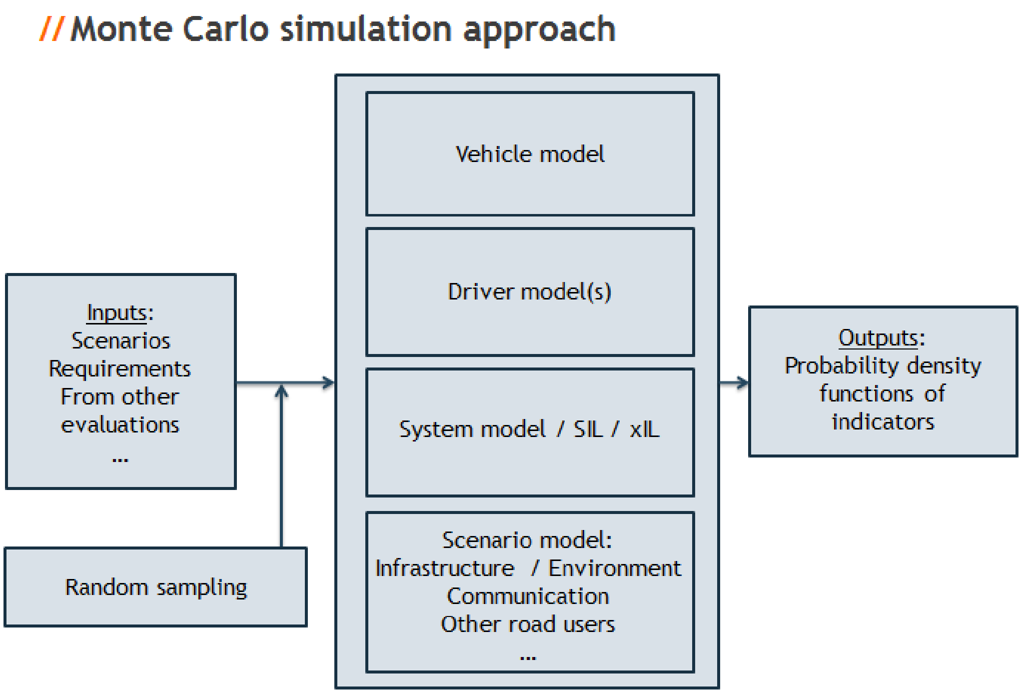


Figure 5.2: Schematic of Monte Carlo simulation approach, detail of block "Monte Carlo Simulations" as shown in Figure 5.1.

Event-based operating functions

Event-based functions in urban and highway driving typically act in scenarios where the impact on surrounding traffic is low. This is due to the fact that event-based functions only operate in a relatively short time (e.g. critical situations or at low speed for instance parking). Since the technical performance of event-based functions will also be analysed in the technical assessment and the safety impact in the safety impact assessment, the scope of the in-traffic assessment will be more on continuously operating functions.

Continuous operating functions

In contrast to event-based operating functions, continuous operating functions need to operate in a large range of different driving scenarios, and therefore, an assessment of the overall in-traffic performance of the driving process is needed. In order to evaluate the in-traffic behaviour and performance of a continuous operating function, Monte Carlo simulation method is suggested. Such a method overcomes limitations such as high costs and time-consuming testing resulting from the significant amount of test driving kilometres needed to cover a large range of different driving scenarios. Using simulation, numerous different scenarios can be tested in a short amount of time.

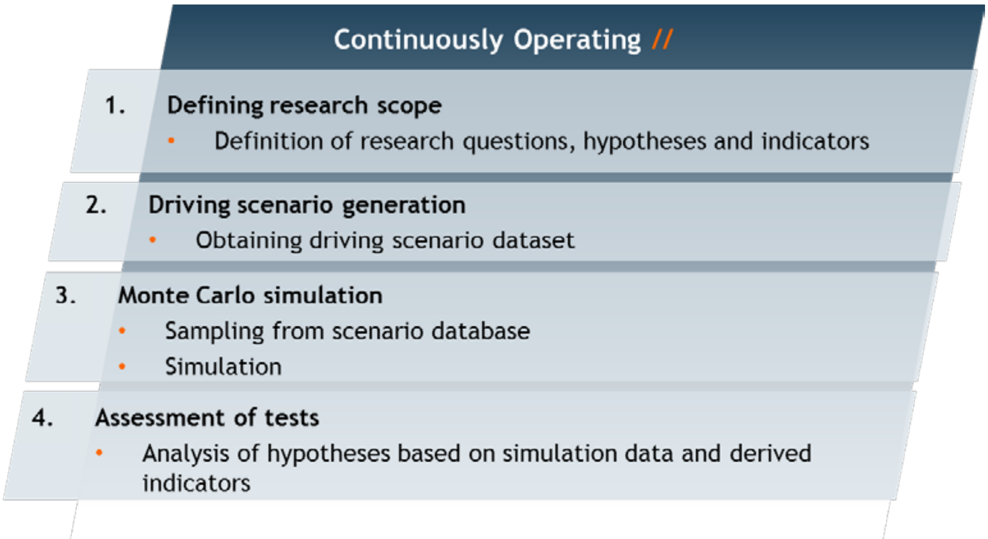


Figure 5.3: Test sequence for continuous operating functions

For the in-traffic assessment of continuously operating functions the different steps of the approach, as given in Figure 5.3, are described. Similar to the technical assessment described in chapter 3, the evaluation of continuous operating functions starts with defining research questions and hypotheses, including adequate indicators. It is also important to define environments for which the continuously operating function is build and should be tested for, e.g. urban, highway etc.

The next step is to derive a driving scenario dataset, typically short in time (t < 30 s), in which the in-traffic behaviour of the continuous operating function will be assessed. The dataset should contain real-life traffic data and will function as input for the Monte Carlo simulation. A dataset can, for example, be obtained by means of a field test. For future purposes, driving scenario datasets can be added to a database, but that is out of the scope of AdaptIVe.

The next step is the Monte Carlo simulation itself which is performed in two steps. From the driving scenario dataset, distributions can be obtained from every variable that is part of the driving scenario, for example, vehicle and systems state, other road user behaviour, system settings and infrastructure. The first step is sampling from these distributions, creating variations in existing real-life traffic scenarios to create new ones. This will be done numerous times and will eventually be the input for the Monte Carlo simulation. The second step is the actual Monte Carlo simulation itself. Different software tools are available on the market to perform in-traffic simulations, such as PreScan [94] or Matlab/Simulink modelling [95]. These can be run multiple times, based on the Monte Carlo sampling with the help of automated scripts. Ideally, not only variations are simulated with the function / system operating, but also without with the vehicle only being driven by a human driver. Next to it the variation of the situation can include depending on the analysed research question also the variation or disturbance of the environment (e.g. communication) in order to investigate how the vehicle and the surrounding traffic react to this disturbance. Such a comparison not only allows for evaluation of the function / system itself, but also for an evaluation of the surrounding traffic. This approach requires the definition of an appropriate driver model(s) that can react on certain traffic situations in a realistic and representative manner. The output of the Monte Carlo simulation will be in the form of distributions of the evaluation indicators defined in section 5.1.3.

After the Monte Carlo simulation the data is evaluated. Based on the output, the evaluation indicators, the analysis of the hypotheses can be done. Using statistical hypothesis testing, the probability that a defined hypothesis is true or false is calculated. The level of statistical significance is specified beforehand. On the basis of the results a conclusion can be made of the in-traffic functioning of the assessed continuous operating function.

5.1 Focus of in-traffic assessment

This chapter discusses which aspects the in-traffic assessment should focus on. AdaptIVe aims at setting up a general evaluation framework for automated driving functions and systems. Therefore, the focus of the in-traffic assessment is described in a general and generic way. This means that adaptations might be necessary in order to cover also special aspects of certain functions.

The first step is to define relevant research questions which cover requirements and reliability of Monte Carlo simulation. Related hypotheses that will be assessed have been derived based on the research questions. In order to analyse hypotheses evaluation indicators are required, which are described in the last part of this chapter.

5.1.1 Research questions

This section discusses the research questions related to in-traffic assessment. The research questions are necessary for the evaluation and should be used as guidance of what should be evaluated in the in-traffic assessment. The research questions are clustered by the evaluation aspects interaction with other traffic participants, interaction at infrastructure environment and communication.

The evaluation aspect **interaction with other traffic participants** handles the research questions related to the basic functionality and fulfilment of the specified requirements for the in-traffic assessment method. The research questions for this aspect are as follows:

ID	Research Question	Function		Adressed level of automation
		Event based	Continuous	
RQITA1	How is the vehicle interaction with other traffic participants?		x	All
RQITA2	How do other traffic participants react on the intervention?		x	All
RQITA3	Are non-user's behaviour influenced by interaction with equipped vehicles?		x	All

Figure 5.4: Research questions for the evaluation aspect interaction with other traffic participants.

The research questions related to **interaction at infrastructure environment** are:

ID	Research Question	Function		Adressed level of automation
		Event based	Continuous	
RQITA4	Can the function handle different infrastructure layouts (e.g. 5 armed crossing, double roundabout, very narrow / British roundabout, splitting roads, etc.)?		x	All

Figure 5.5: Research questions for the evaluation aspect interaction at infrastructure.

Regarding the aspect **communication** with other traffic participants and the infrastructure environment and communication, research questions are:

ID	Research Question	Function		Adressed level of automation
		Event based	Continuous	
RQITA5	Does the function change behaviour if communication fails / changes?		x	All
RQITA6	How does the function handle wrong / inaccurate information?		x	All

Figure 5.6: Research questions for the evaluation aspect communication.

5.1.2 Hypotheses

The hypotheses are defined based on the research questions. The hypotheses are presented analogue to research questions by the different evaluation aspects (method, tool, interaction). The hypotheses are formulated in a general way and it therefore might be required for some adaptation in case a special aspect of a function should be covered. Only relevant hypotheses will be selected since not all hypotheses will be assessed within the evaluation of AdaptIVe.

Hypotheses related to **interaction with other traffic participants** are:

ID	Hypotheses	Reference	Related Research Question
HITA1	The vehicle with the function or systems acts like normal unequipped vehicles	Normal traffic behaviour	RQITA1
HITA2	There is no change in function actions in case of an obvious error	TBD	RQITA1
HITA3	The function complies to the required distributions (of normal driving)	Normal traffic behaviour	RQITA1
HITA4	Other traffic participants do not change behaviour / actions with respect to the equipped vehicle	Normal traffic behaviour	RQITA2
HITA5	Non-user's behaviour is not influenced by interaction with equipped vehicles	Driving without the system	RQITA3

Figure 5.7: Hypotheses for the evaluation aspect interaction with other traffic participants and the infrastructure environment

Hypotheses related to **interaction at infrastructure** are:

ID	Hypotheses	Reference	Related Research Question
HITA6	The function or system can handle most relevant infrastructure layouts; very much like normal drivers do	Expected function behaviour	RQITA4

ID	Hypotheses	Reference	Related Research Question
HITA7	The function or system handles situations that occur less than xxx/h in appropriate manner	Expected function behaviour	RQITA4

Figure 5.8: Hypotheses for the evaluation aspect interaction with other traffic participants and the infrastructure environment

Hypotheses related to **communication** are:

ID	Hypotheses	Reference	Related Research Question
HITA8	The function or system can handle communication loss or changes	TBD	RQITA5
HITA9	The function or system can handle wrong / inaccurate information	Trajectory with correct information	RQITA6

Figure 5.9: Hypotheses for the evaluation aspect interaction with other traffic participants and the infrastructure environment

5.1.3 Evaluation indicators

To evaluate the derived hypotheses, evaluation indicators are used. These indicators are outputs of the Monte Carlo simulation all in the form of **statistical distributions**. In the table below, all relevant evaluation indicators are shown and linked to the hypotheses defined above. The table only presents a general approach of evaluation indicators for the in-traffic assessment. It can be the case that for certain continuously operating functions the hypotheses require some adaptation. By doing this, the evaluation indicators may also require adjustment.

ID	Indicators	Evaluation Aspects	Related Hypothesis
IITA1	Time Headway (THW)	Interaction with other traffic participants, communication	HITA1, HITA3, HITA4, HITA8
IITA2	Time to collision (TTC)	Interaction with other traffic participants, communication	HITA1, HITA3, HITA4, HITA8
IITA3	Acceleration levels	Interaction with other traffic participants, communication	HITA1, HITA3, HITA4, HITA8
IITA4	Speed differences	Interaction with other traffic participants, communication	HITA1, HITA3, HITA4, HITA8
IITA5	Speeds	Interaction with other traffic participants, communication	HITA1, HITA3, HITA4, HITA8
IITA6	Distances to infrastructure elements	Interaction with other traffic participants, communication	HITA1, HITA3, HITA4, HITA8

ID	Indicators	Evaluation Aspects	Related Hypothesis
IITA7	Observed non-users behaviour variables	Interaction with other traffic participants	HITA5
IITA8	Trajectory (x,y-Position)	Communication	HITA9

Figure 5.10: Evaluation indicators to be considered as statistical distributions

5.2 Methods and tools for in-traffic assessment

This chapter describes the tools that are needed for usage of a Monte Carlo simulation approach for in-traffic assessment. The Monte Carlo method will be discussed in more detail with focus on what is needed as generic input for a Monte Carlo in-traffic simulation assessment. By Monte Carlo simulations in the following simulation set ups are meant, where the input parameters to a predefined traffic or driving scenario are varied using Monte Carlo sampling. The simulations themselves will be done using software tools such as PreScan or Matlab/Simulink.

As already shown in Figure 2.5, four general test tools are available for assessment: Field tests on public roads, Tests on test tracks, tests in simulators or simulation. As the first 3 tools are not suitable for an in-traffic assessment due to various reasons, only simulations will be discussed further in the following section.

5.2.1 Monte Carlo simulation

Monte Carlo is a statistical method that relies on repeated random sampling for the derivation of numerical results. This method is widely applied in situations with a high degree of freedom and where it is too inefficient, in terms of time and costs, to test a full factorial combination of input variables. The Monte Carlo method roughly comprises four different steps. Firstly a domain of inputs needs to be defined. The domain of inputs for in-traffic assessment consists of a vehicle model, host vehicle driver behaviour model and scenario parameters. Secondly, from a probability distribution over the domain, inputs are randomly generated (sampling). The third step is the actual computation (simulation). Last but not least is the aggregation of the results, which will be in the form of probability distributions for each evaluation indicator. It should be noted that for system development purposes, logging of the input variables and output indicators is worthwhile to encompass typical areas for system optimisation.

An overview of the Monte Carlo approach for in-traffic assessment is depicted in Figure 5.11. To test continuously operating functions two different types of inputs are necessary, i.e. the variable input which define a scenario (orange) and the continuous operating function (grey) to be tested. The Monte Carlo approach (yellow) itself consists of two parts, which are the Monte Carlo sampling and the actual simulation. The outputs (red) consist of the evaluation indicators defined in section 5.1.3 and their changes in distribution.

A scenario dataset is used to define the parameters of a behaviour model, vehicle model, and scenario model. The behaviour model contains information about how the driver of a vehicle reacts to certain stimuli, which is driver dependent. For example, an older person is more likely to react slower or not at all compared to someone who is younger. The vehicle model is divided into two parts, the vehicle dynamics model and sensor models. Vehicle dynamics says something about the driving performance of a car, e.g. how fast a car can decelerate, which is influenced also by the presence of safety features such as traction control or anti-lock braking system (ABS) etc. Vehicle dynamics are also dependent on external factors, for instance road surface conditions or wind speed. Sensor performance information is stored in the sensor model of the vehicle. The test vehicle can be equipped with multiple sensors such as radar, camera or laser which might or might not work together (sensor fusion). Performance of these sensors can be dependent on several internal factors like sampling rates or amount of laser beams and external factors like disturbance from bad weather conditions.

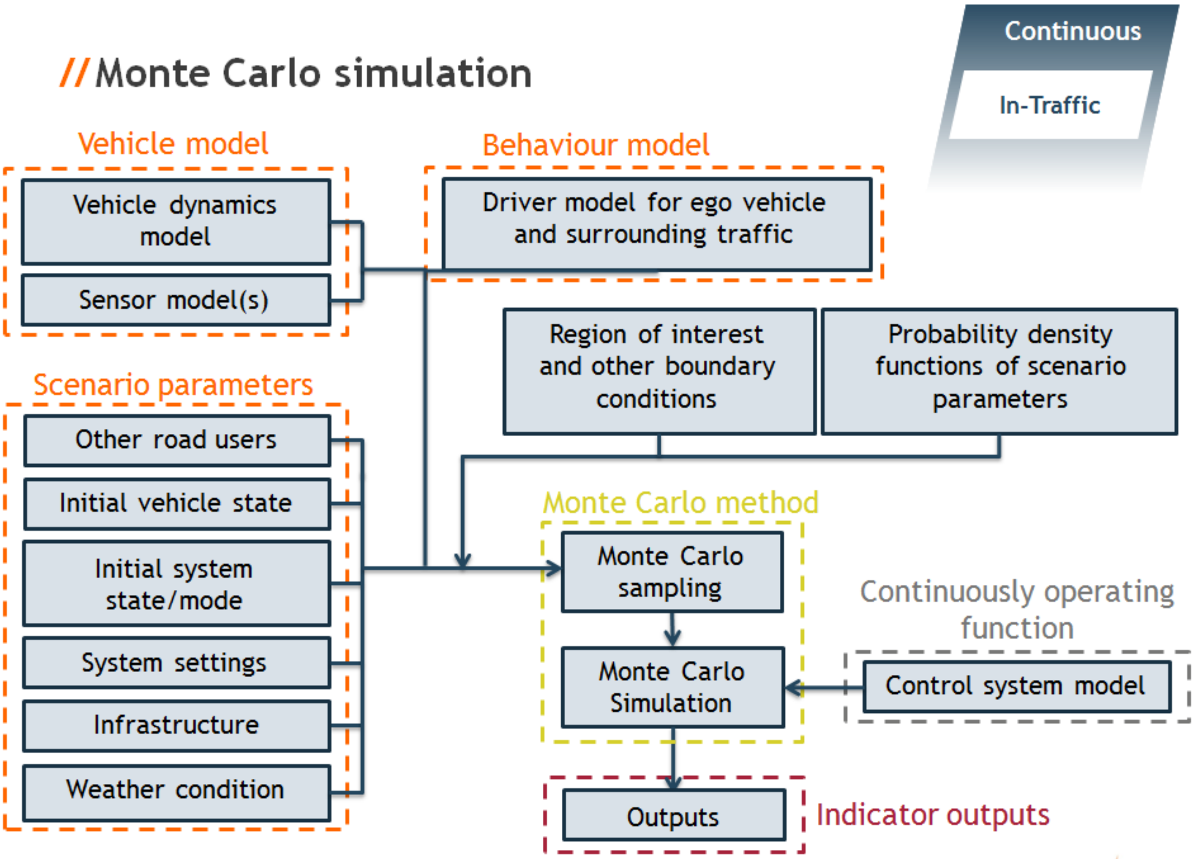


Figure 5.11: Basis concept of Monte Carlo simulation for in-traffic assessment of continuously operating functions

Regarding the scenario parameters, multiple variables are involved. To start with, different vehicle related variables are needed as input. The first is initial vehicle state, such as position, velocity and acceleration. Second is the initial system status or mode, for instance is the system starting or already running. Third and last are the systems settings, e.g. set speed or following distance. Other – external – factors are other road users' presence and behaviour. One can think of variation in following / merging distance or severe sudden braking. Infrastructure also plays a role in terms of road properties (e.g. lines on the road) and road layout. It can be important to test if a continuously operating function still operates for different road curvatures or slopes or if all of a sudden no lane markings are present any longer. The last variable input is the weather condition (like presence of rain, temperature, visibility, etc.) this all plays an important role for vehicle and system performance and therefore for in-traffic behaviour.

These inputs for the simulation can have underlying correlation. This has to be taken into account in the Monte Carlo sampling process. However before sampling, a "region of interests" needs to be defined as not all parameters are of equal interest. To further elaborate the region of interest, consider the amount of rain (millimetre/hour) and road wetness as input parameters for the Monte Carlo sampling. When rainfall is heavy it is highly likely that road wetness will be high as well. Therefore, when sampling in the region of heavy rainfall, the sampling of road wetness should also focus on the more wet than dry road conditions. Safety-critical conditions are potentially more interesting than normal conditions, but do not have a higher priority per definition. Together with defining a region of interest comes choosing the type of Monte Carlo sampling, which will be discussed in section 0.

The discussed inputs are then processed in the simulation. At the end of the simulation for each evaluation indicator a probability density function will be established.

With respect to simulation tools different options do exist, however not all can be evaluated or used simply due to fact that not all tools are available for the AdaptIVe partner, who is conducting the in-traffic assessment. For the assessment within AdaptIVe the following two tools are considered for the in-traffic assessment, although the methodology can also be conducted with other simulation tools:

Test tool option 1: PreScan

PreScan (for more information see [94]) is a physics-based simulation platform that is used in the automotive industry for development of Advanced Driver Assistance Systems (ADAS) that are based on sensor technologies such as radar, laser/lidar, camera and GPS. It allows for design and evaluation of vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication applications as well as autonomous driving applications and can handle model-based controller design (MIL) as well as real-time tests with software-in-the-loop (SIL) and hardware-in-the-loop (HIL) systems.

The tool additionally allows for modelling of traffic and environment including realistic visualization, also for all sensors mounted to the car. When everything has been set up in

PreScan, the tool converts the model into a Matlab/Simulink models. Required test amount depends on how many simulations are needed in order to get a reliable result as well as the required detail of the PreSan scenario.

Test tool option 2: Matlab/Simulink

Matlab/Simulink (for more information see [95]) can be used to both design a continuously operating function as well as for scenario model generation. Typically, the scenario models will be more simplified, containing just the relevant information for evaluation of the function. As such, other vehicles may be represented as objects with relative motion described, instead of a full graphical model with vehicle dynamics features. It can also be used to perform the Monte Carlo sampling of input parameters, the simulation of the continuous operating system and for post processing of the results. This tool is a coding based program and therefore requires more experienced users, but has a higher acceptance over other tools such as PreScan. Overall, Matlab is very versatile and its capabilities are mainly limited by the user and hardware. An advantage is that the calculation is faster since no visualization is involved. This can also be a disadvantage since more experience and skills are required. Also, visualization might be required for example to the presence of camera sensors.

5.3 Requirements of the in-traffic assessment

In the following chapter, the requirements with respect to safety, test tools and test effort for the in-traffic assessment are discussed. Since it was determined that in-traffic assessment was only to be performed for continuously operating automated driving functions, this section will be limited to those. Again, the focus here lies on traffic simulation. In the following subsections, the requirements in terms of safety, test effort and test tools are discussed subsequently.

Test tool	Description		Test effort	Safety	Test tools
Field test	test on public roads with public traffic		Not planned	Not planned	Not planned
Controlled field	test on a closed test track with controlled traffic		Not planned	Not planned	Not planned
Simulation (traffic)	Monte Carlo simulation of specific function (SIL) in specific traffic settings/scenarios		Once programmed the simulation should be automated	No issues as it is simulation	Computer and simulation environment like Matlab/Simulink and e.g. PreScan

Test tool	Description		Test effort	Safety	Test tools
Simulator	test with a simulated vehicle in a simulation environment with simulated traffic		Not planned	Not planned	Not planned

Figure 5.12: Overview of types of test tools and their requirements

5.3.1 Safety

Since simulation was chosen as the only applicable test tool for in-traffic assessment, there are no safety issues concerning damage to parts or injury to persons.

5.3.2 Tools

Scenario dataset:

Data from "real life driving" is required for the in-traffic assessment to build the scenarios in the simulation environment which will not be gathered separately and solely for this task. Ideally, this data could be distracted form a (so far non-existent) "real life driving database" containing information on general driving situations as well as near miss or critical situations or actual accidents. This database could contain data from the following type testing / source:

- Field test data: (Limited) field tests are conducted as part of the technical and user related assessment. This data should be gathered in a manner, that it can also be used for the generation of driving- or traffic scenarios for the in-traffic assessment. Ideally, in addition to data with automated functions also reference data without automated functions should be present.
- Naturalistic driving data: in naturalistic driving data, similarly instrumented vehicles may be used, in this case also without the specific automated system to be assessed. A major difference lies in the fact that these vehicles are typically driven by regular drivers during their regular driving tasks. Within AdaptIVe no such tests will be conducted. If available, data from previous naturalistic driving studies should be consulted.
- In-depth accident data: Databases like GIDAS gather detailed information on general road accidents. This data could also be used to generate traffic or driving scenarios. By varying the input parameters accordingly, an estimate could be done if the automated system could have influenced the situation one way or the other. Also, situations that did lead to an accident in one case will not have led to an accident in many other cases as timing or speeds were just a bit different. Therefore, accidents can be used to generate realistic conditions also for non-accident situations.

In the field of generating scenarios for a real-life database, which is outside the scope of AdaptIVe, naturalistic driving data is seen as a powerful method for generating real-life scenario data. One can choose for adding highly advanced and a complete set of instrumentation on a small number of vehicles or for instrumenting many vehicles with a limited and more cost-effective set of instrumentation. As such, there is a trade-off between the amount of kilometres of data generated versus the amount of information collected in those driven kilometres. Inevitably, large quantities of data are being generated for which innovative methods for filtering, classifying and managing of data are under development [96] [97].

Monte Carlo Assessment

The tools should be capable of performing in-traffic assessment such that the requirements are met. Within aerospace, Monte Carlo assessment is common practice; as such the structure for assessment was derived from [98]. In-traffic assessment tools should therefore be able to:

- Establish compliance with performance limits that need to be specified for the indicators in section 5.1.3.
- Determine any limitations in use of the system for compliance with limits, i.e. document in which situations (involved variable values) the system does not comply.
- Account should be taken of the variation of all the parameters that influence the performance of the system, at the very least those influencing the hypotheses in section 5.1.2.
- Effects of applicable conditions are to be investigated and, if necessary, appropriate limitations derived should be documented.
- Acceptable values for the probabilities of exceeding the limits. These could be determined based on requiring improvement over current probabilities of failure, damage and injury in normal non-automated traffic.
- A programme of scenarios should be completed sufficient to demonstrate the validity of the simulation and support the conclusions of the analysis.
- Individual scenarios should be carried out to demonstrate that errors, which can reasonably be expected to occur, are not hazardous. This includes a strong link to the technical assessment.
- Probability distributions and models of infrastructure and other traffic participants may be used.

In order to achieve the above, Monte Carlo sampling is a proposed method to sample over the variations in scenarios. Monte Carlo is a global term for a number of methods [99] of which a few that are deemed the most applicable will be discussed.

Crude Monte Carlo

The basic crude Monte Carlo (CMC) approach is to take randomly generated samples from the available distribution without further processing. As such, it produces an estimation based on generation of N samples U from a given distribution u:

$$U_1, \dots, U_N \in h$$

These samples are applied in the CMC approach to gain an estimate Ĝ of the original distribution h as follows:

$$\hat{G} = \frac{1}{N}\sum_{k=1}^{N} h(U_k)$$

Importance Sampling

Importance Sampling is a technique that is especially useful for the estimation of rare-event probabilities. Safety-critical events are typical examples of rare event probabilities.

Again, considering a random variable G for a real function with values U and probability density function f:

$$G = E_f U(x) = \int H(x) f(x) dx$$

G(x) is transformed by introducing an estimator distribution g(x) to get:

$$G = \int U(x)\frac{f(x)}{g(x)} g(x) dx = E_g U(x)\frac{f(x)}{g(x)}$$

The estimator of G after sampling is then:

$$G = \frac{1}{N}\sum_{k=1}^{N} U(x) w(x)$$

The function w(x), also known as the likelihood ratio estimator assigns weights to the samples from the probability density f(x) and is computed by the following ratio:

$$w(x) = \frac{f(x)}{g(x)}$$

The function g(x) can then be chosen to apply specific focus on rare-event probabilities. For example, the Chi-Square distribution [100], depicted in Figure 5.13, with accompanying density function depicted in Figure 5.14 may be used as an example of a distribution that amplifies probabilities at low values. Low values are relevant for e.g. THW and TTC. If shape parameter v is chosen as 1, THW values of zero s or close to that are weighted most importantly. In that case crash events are taken into account as well. If shape parameter v is chosen as 3, crash events are neglected but dominant focus is laid on small THW values up to 2 s.

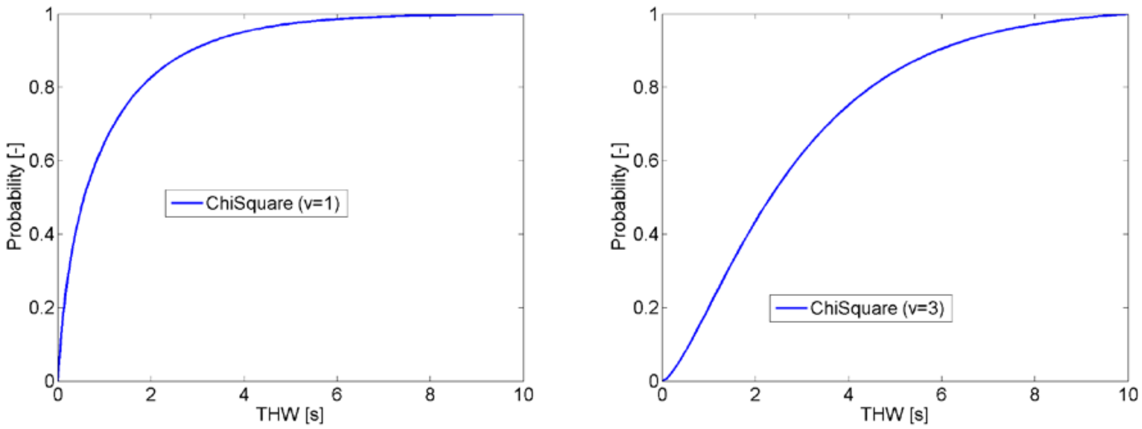


Figure 5.13: Chi-Square (cumulative?) distribution function with different shape parameter v settings: v =1 (left) and v = 3 (right).

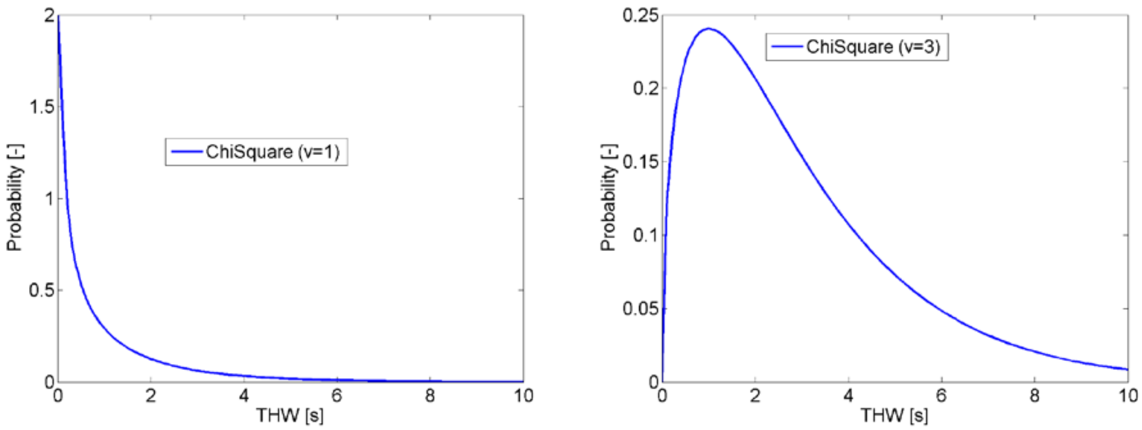


Figure 5.14: Chi-Square density function with different shape parameter v settings: v =1 (left) and v = 3 (right).

In order to provide a qualitative assessment, the accuracy of the estimate to represent the, for example, 95% confidence interval needs to be computed, which is possible for both methods.

5.3.3 Test conduction

This chapter describes the required test effort. In AdaptIVe, the in-traffic assessment of **continuously based functions** is conducted using market available software tools such as PreScan, Matlab/Simulink or comparable.

For the Monte Carlo simulation itself simulation time is difficult to quantify as it mainly depends on integration of the automated function (software, hardware in the loop, …), used sensors and computational hardware for the simulation itself. Where a full traffic simulation of a scenario with three vehicles may take the same time as one scenario would take in real time, a simplified Matlab/Simulink simulation may be orders of magnitude faster. Testing thousands of scenarios would then take more or less the same time than simulating thousands of scenarios. The length of a scenario is to be defined based on the automated function and respective scenario to be evaluated. It should be long enough to cover all important aspects of the situation and the system reaction but not so long that no further information will be gathered. Using parallel computing, the same test would now be able to speed up. Test effort is therefore difficult to determine beforehand, however it is clear that simulating sampled scenarios instead of simulation complete driving sequences, provides a major reduction in simulation time. Setting up the simulation does need preparation by an operator, but once set up; the Monte Carlo sequence should run automatically.

5.4 Example of in-traffic assessment

Since in-traffic assessment methodology is most feasible for continuous functions, an example obtained from a previous in-house investigation is provided in this case an Adaptive Cruise Control (ACC).

An ACC system is designed to perform a car-following task, while balancing between maintaining set speed and the distance to the predecessor. Whether the systems perform this task in a safe manner without the car occupants sustaining discomfort, is vital for the acceptance of such a system.

5.4.1 Real-life dataset

For this example, a dataset is used that consists of 835 minutes of driving. The route consisted of rush-hour commutes on Dutch highways A2, A4, A12, A13, A20 and A67, including few (inter)urban connecting roads. Data was logged based on vehicle CAN data and a Mobileye camera-based sensor. Data for this evaluation was classified according to a safety-critical event selection algorithm, with special focus on indicators for driver distraction [101]. It should be noted, that for further in-traffic evaluation the focus will **not only** be on safety critical events, but on the general driving tasks which might include safety critical events.

A subset of the data of 554 minutes was created, where the driving speed was 54 km/h or more. This excludes low-speed urban traffic and intersections, which are not considered for this ACC example. In Figure 5.15, the number of events in which a certain parameter exceeded a safety-critical threshold is shown

All of these events may be relevant for an ACC system,

Parameter	Threshold	Speed condition	Other condition	Nr.
ax	3.5 m/s^2	> 54 km/h		0
ay	3.5 m/s^2	> 54 km/h		11
TTC	2 s	> 54 km/h	Yaw rate < 6°/s	6
THW	0.6 s	> 54 km/h	Yaw rate < 6°/s	378
Car-following distraction	‘Mode 2’	> 54 km/h		27

Figure 5.15: Example for safety-critical event indicators from 835 minutes of real-life driving data [86] for an ACC application.

One of the events that were classified as safety-critical is shown in Figure 5.16 as an example of the dataset. The host vehicle is approaching the lead vehicle with low THW, with the clear intent to overtake the vehicle. However, the driver is anticipating to a faster vehicle on the left lane which it allows to pass first. The other classified events are all necessarily different and together comprise a set of real-life scenarios. Due to the low amount of kilometres, the specific road sections and the time of day this dataset is not representative however is sufficient as an example.

From this dataset with around 425 car-following related safety-critical events one could perform Monte Carlo sampling, assuming this dataset provides a realistic representation of scenarios in the given setting, i.e. predominant highway car-following related events. However, the data sample of 425 events is too small to perform Monte Carlo sampling if statistical guidelines are followed. Secondly, within these 425 scenarios there most likely is so much variation that the Monte Carlo simulations would provide little to no insight in the influence of a certain parameter on the output response. One obvious way to resolve this is to generate databases with more events by driving with standard vehicles with logging equipment and possibly additional sensor instrumentation. Another means of extending the data and making it more realistic is to extend the dataset with scenarios that are not classified as safety-critical.

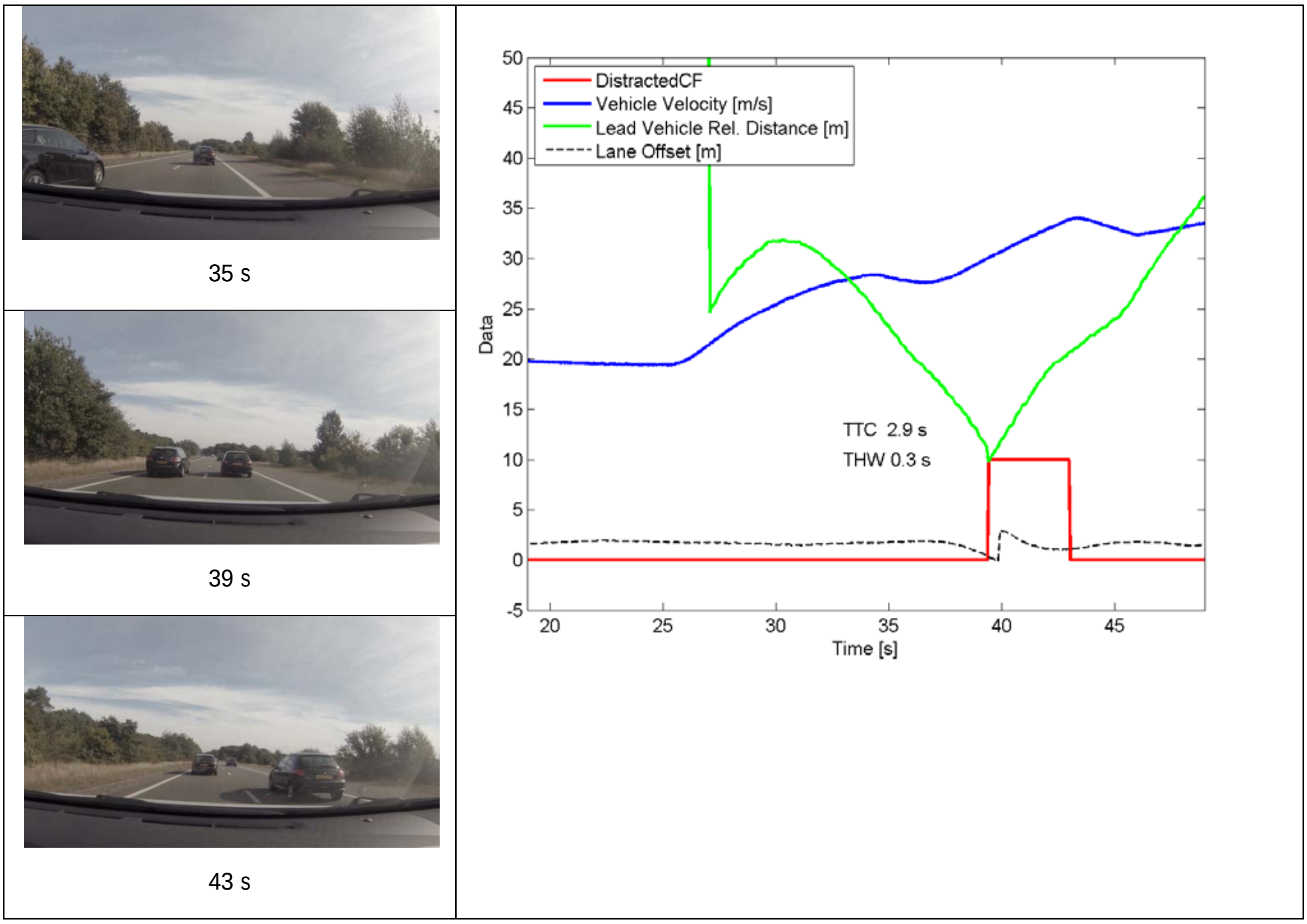


Figure 5.16: Safety-critical event example from 835 minutes of real-life driving data [101].

5.4.2 Parameterised real-life dataset

Another, more statistical approach is to use a dataset to determine typical real-life scenarios and then to apply variations to the scenarios according to statistical distributions of certain parameters. For example, when we consider the example scenario in Figure 5.16, where the host vehicle was approaching the lead vehicle up to a minimum THW of 0.3 s, this scenario can be simulated with a number of variations. From all recorded variables, distribution functions can be derived. In this example, a variation in three variables is assumed:

- Time Headway (THW) is varied according to the recorded distribution shown in Figure 5.17.
- Host vehicle initial velocity is varied according to an assumed normal distribution with 100 km/h mean speed and a standard deviation of 20 km/h.
- Lead vehicle deceleration is varied according to an assumed normal distribution with 3 m/s^2 deceleration and standard deviation of 1 m/s^2.

The probability density functions of Time Headway (THW) and Time-To-Collision (TTC) during the 835 minutes of driving are shown in Figure 5.18. The accompanying cumulative distribution functions are shown in Figure 5.17. It is shown that a THW of around 1.5 s is the most often observed car-following distance. The probability of a critical THW below 0.6 s was approximately 4% of the total driving time. The TTC density is capped at 10 s due to the fact that the sensor processing did not provide TTC values larger than 10 s. It was computed from this distribution that the probability of a critical TTC below 2 s is occurring 2% of the driving time in which the TTC is below 10 s.

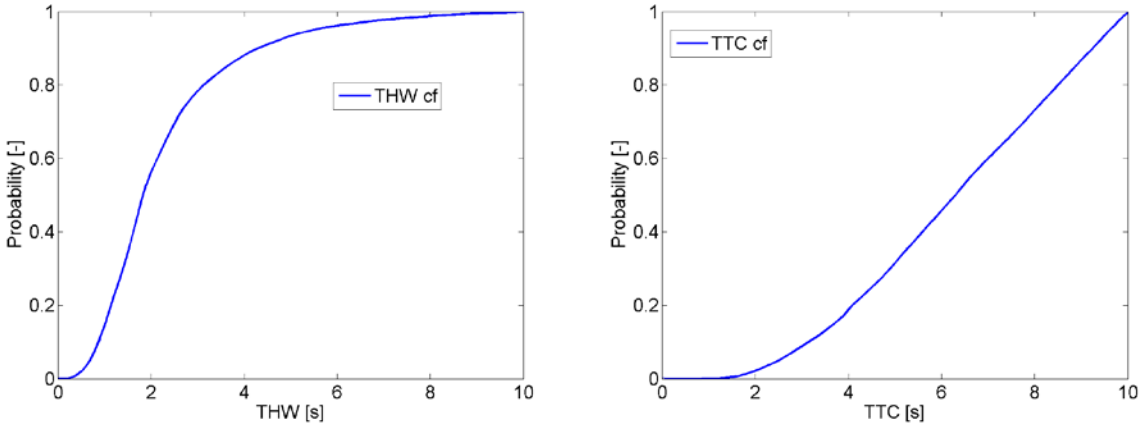


Figure 5.17: THW and TTC cumulative distribution function from 835 minutes of real-life driving data [101].

In-traffic assessment will consist of evaluating the performance of a certain function in the given scenarios. In the example scenario, the paths of the other traffic participants and the initial location and velocity of the host vehicle may be used as input to the scenario. Simulating with vehicle models, driver models and a model of the ACC function will give insight in the performance of the vehicle in this scenario.

Moreover, variations in host vehicle state and behaviour of other traffic participants create a distribution of possible scenarios of this type. Having such a distribution allows for Monte Carlo sampling and simulation of this scenario in all its possible outcomes.

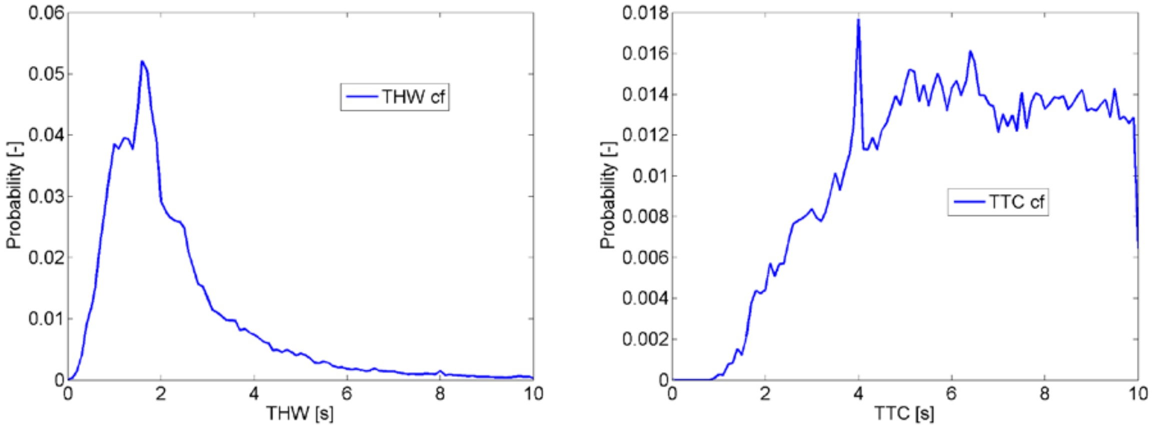


Figure 5.18: THW and TTC probability density function from 835 minutes of real-life driving data [101].

As such, the sampling space consists of three dimensions in which variations occur. Here it is assumed that the three variables are independent. For example, for every possible THW the real-life distribution of host vehicle velocity is the same.

For in-traffic assessment the change of the distribution of the outcomes compared to situations without automated functions is of interest. As such, based on the overview of sampling methods provided in [84], the Weighted Importance Sampling method is proposed for variables that are more infrequent but of considered of major importance like THW for the ACC example. An estimator distribution g(x) is introduced that increases the probability that low THW values are being selected. The function w(x) assigning weights to the samples from the THW probability density f(x) is computed by the following ratio:

$$w(x) = \frac{f(x)}{g(x)}$$

The resulting w(x) or likelihood ratio estimator is shown in Figure 5.19 for shape parameter v=1 and in Figure 5.20 for shape parameter v=3.

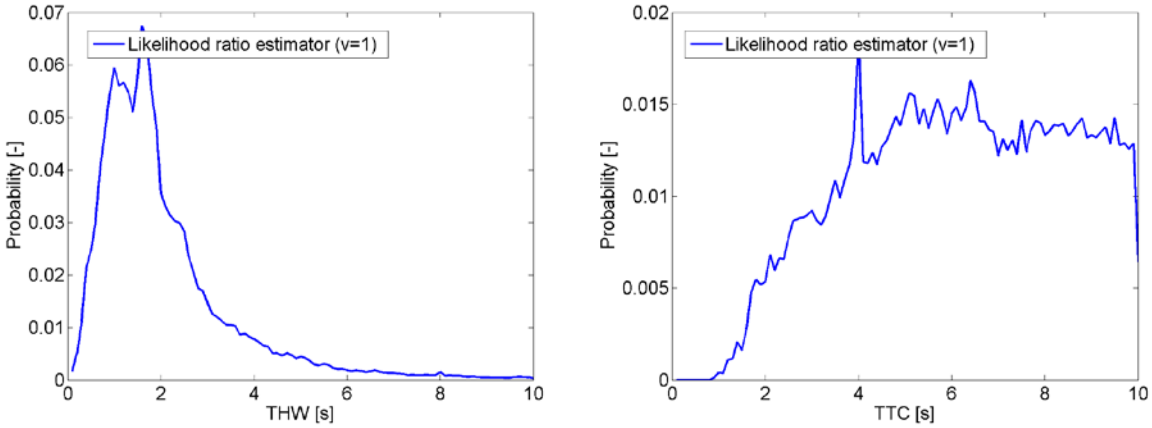


Figure 5.19: Likelihood ratio estimator w(x) for THW and TTC for shape parameter v = 1.

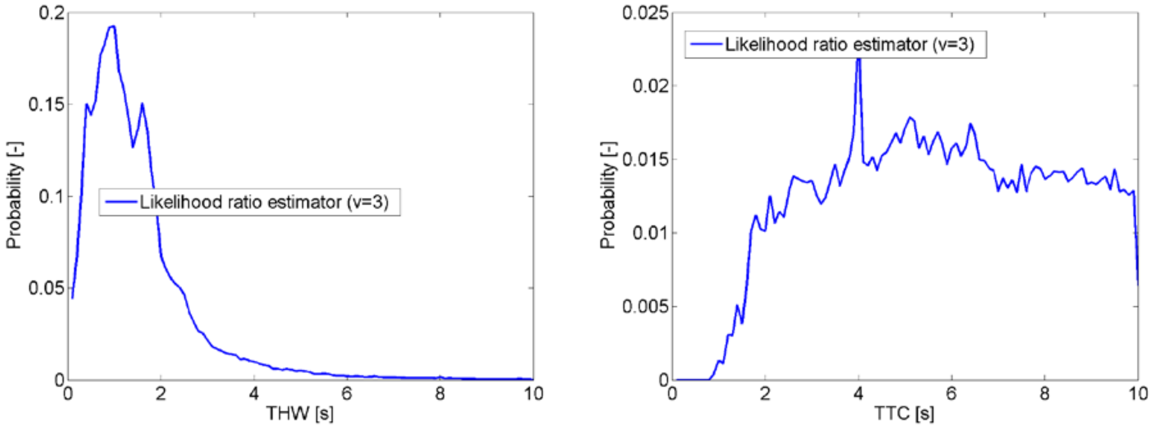


Figure 5.20: Likelihood ratio estimator w(x) for THW and TTC for shape parameter v = 3.

These are essential for the weighted probability density functions and can directly be compared with the original ones in Figure 5.18. It is shown that both THW and TTC have relatively higher probability density for lower values than for higher values. Also, when a Chi-Squared distribution with shape parameter v=3 is used, the THW likelihood ratio estimator provides a substantial probability (of around 0.05) at the first non-zero bin of 0.1 s. As such, more critical events were sampled compared to when the original distribution was used.

Adding to the parameterisation of THW the parameterisation of host vehicle initial velocity and lead vehicle deceleration, a Monte Carlo sampling can be generated over these three examples. This can then be fed into one of the proposed test tools that perform the Monte Carlo simulations, such that the full in-traffic assessment as depicted in Figure 5.1 can be performed.

In order to determine whether the in-traffic assessment method described in this example allows for answering the research questions and test the hypotheses, they are shown again in Figure 5.21.

ID	Research Question	ID	Hypotheses
RQITA14	How is the vehicle interacting with other traffic participants?	HITA01	The vehicle with the function or system acts like normal unequipped vehicles
		HITA02	There is no change in function actions in case of an obvious error.
		HITA03	The function complies to the required distributions (of normal driving)
RQITA15	How do other traffic participants react to the intervention?	HITA04	Other traffic participants do not change behaviour/actions with respect to the equipped vehicle
RQITA16	Are non-user's behaviour influenced by interaction with equipped vehicles?	HITA05	Non-user's behaviour is not influenced by interaction with equipped vehicles.
RQITA17	Can the function or system handle different infrastructure layouts (e.g. 5 armed crossing, double roundabout, very narrow / British roundabout, splitting roads, …)?	HITA06	The function or system can handle most relevant infrastructure layouts; berry much like normal drivers do.
		HITA07	The function or system handles situations that occur less than xxx/h in appropriate manner.
RQITA18	Does the function change behaviour if communication fails/changes?	HITA08	The function or system can handle communication loss or changes.
RQITA19	How does the function handle wrong / inaccurate information?	HITA09	The function or system can handle wrong / inaccurate information.

Figure 5.21: Research questions and hypotheses for the in-traffic assessment of an ACC example

Interaction with other traffic participants was demonstrated by taking into account and varying the motion of the lead vehicle. This scenario was heavily influenced by the traffic participants in the other lane, such that these need to be taken into account as well. Whether and how the other traffic participants react to an intervention from an automated system is obviously not present in the current dataset and therefore should be an implicit part of the driver models of the other traffic participants. The last two scenarios deal with failure modes of the system, which can easily be introduced in the numerical models of the functions.

The indicators are shown in Figure 5.10. This example assumed using variations in THW, speed and acceleration and presented a variation in TTC. Accuracy in trajectory and distance to infrastructure elements are indicators that require a high accuracy world model. Non-user's behaviour variables can be assessed through the use of driver models that need to be developed and validated for the specific scenarios. Since the indicators will be expressed in probability density functions, the in-traffic assessment can be completed with tested hypotheses, such as 'the system can handle situations that occur less than a certain percentage of the time in an appropriate manner.'

6 Impact Assessment

In contrast to the previous assessment (technical, user-related and in-traffic assessment) that analyse the performance of the developed functions or systems the impact assessment analysed the future impact of the developed functions respectively systems on road traffic. For this purpose the impact assessment utilize the results of the previous assessments. Therefore, the impact assessment in AdaptIVe has the objective to develop a methodology to determine the impact on road traffic for the introduction of automated driving applications. The methodology to be developed should be applicable for different types of automated driving applications and should consider different potential effects of the automated driving applications on road traffic. Therefore the assessment has been split into two parts:

- Safety Impact Assessment that will focus on the traffic safety aspects of the automated driving functions,
- Environmental Impact Assessment that will focus on fuel consumption and the traffic flow respectively the travel time.

This deliverable describes the basic ideas for both impact assessments, since the development of the impact assessment is still ongoing. A detailed description of the impact assessment as well as the results of the safety impact assessment will be presented in the deliverable D7.3.

In the impact assessment implemented functions are assessed in different types of situations and scenarios. In order to avoid any confusion, the relevant terms are defined in chapter 2.

6.1 Safety Impact Assessment

Since automated driving functions take over, or at least rigorously co-determine, vehicle control in traffic, traffic safety will either be affected directly (by system intervention) or indirectly (through response of other traffic participants to the system behaviour). Hence, implementation of these functions requires solid proof of their overall safety impact. The objective of the safety impact assessment is to quantify changes - meaning both benefits and risks - in traffic safety induced by automated driving functions. The basic concept will be described, based on which a globally applicable safety impact assessment approach will be developed.

6.1.1 Global requirements of safety impact assessment

The prospective safety assessment demands a methodology that is quantifying, balanced, and realistic.

Quantification refers here to an objective and measurable metric, for example, one safety indicator could be the expected reduction in MAIS2+ injuries that would otherwise have occurred as a measure of safety performance. Hence, all applied indicators must be objective and measurable for the safety impact assessment. Several relevant safety indicators have been listed in the Figure 6.1. Indicators like “controllability estimation”, which is measurable but not objective, will not be included.

Indicator	objective	measurable
Frequency of accidents	x	x
Frequency and distribution of injury severity	x	x
Frequency of critical driving situations	x	x
Number of false positive activations	x	x
Number of false negative activations	x	x
Time in critical situations	x	x

Figure 6.1: Potential safety indicators

As mentioned before, safety effects of automated driving functions theoretically can be positive or negative. Automated longitudinal vehicle control, for example, can initiate a strong deceleration to avoid a rear-end collision with a leading vehicle but induce the risk of a collision with upstream vehicles at the same time. Therefore, the assessment must be **balanced** by including side-effects as well.

Last but not least, the functions must be assessed in a representative traffic environment, which contains equivalent contextual conditions of future customer use cases. The contextual factors include macroscopic traffic state, road condition, weather condition, and microscopic interactions between road users – just to name a few. Furthermore, the common variations of these contextual factors must be taken into account as well.

Conventional methods, as currently used, include hardware-in-the-loop procedures (e.g., for sensor / algorithm testing), testing of technical and human factors (e.g., driving simulator, test track, test rigs), and methods based on real-traffic testing (e.g., controlled studies, field operational tests, observational studies). Each of those tests solely assesses a function on a particular small subset of processes with singular factors influencing the traffic safety. Due to these limitations of conventional methods with respect to the above mentioned requirements, a new method must be developed, as no existing method can meet them all without any limitation. We therefore propose a virtual assessment method. It combines scenario-based stochastic simulation with continuous operation simulation and promise to meet all fundamental requirements adequately.

6.1.2 Methodology of simulative safety impact assessment

The safety impact assessment of automated driving functions can integrate driving scenario simulation and traffic simulation. The general procedure is illustrated in Figure 6.2.

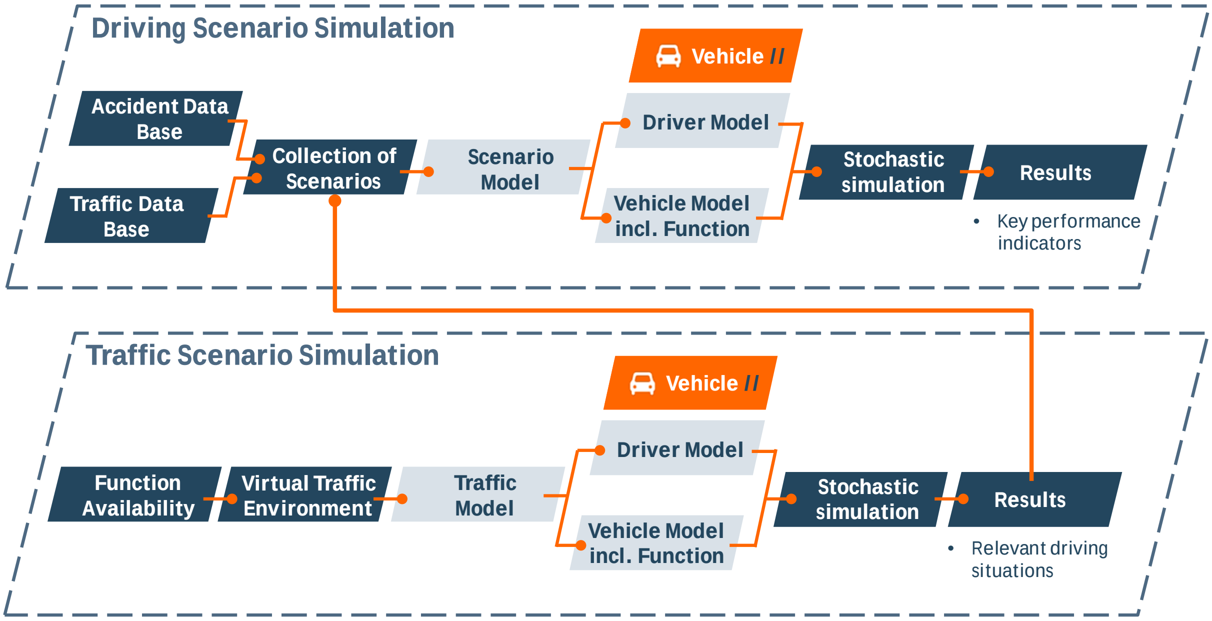


Figure 6.2: General procedure of safety impact assessment of automated driving functions.

The **scenario- simulation** focuses on safety-relevant driving scenarios, which are limited in time and space. These driving scenarios often involve a risky driving situation that can be carried out either by a driver or by a highly automated driving function. Safety performance of human driver and the functions will be determined and compared by simulating the driving scenarios in a replicable way. The first step is identification and specification of relevant driving scenarios, so that they can be modelled sufficiently. For this purpose different data sources can be used. Basically there are two kinds of data sources, (detailed) accident data (e.g., GIDAS database) and traffic data (e.g., data from field operational tests and naturalistic driving studies). Risky driving situations will be extracted and analyzed regarding the underlying interaction process of driver, vehicle and environmental factors. And these processes are crucial for representative and realistic simulation. The abstractions of observed risky driving situations into driver-vehicle-environment-interaction processes are defined as driving scenarios. Furthermore, the driving scenarios will be weighted due to their probability of occurrence.

The driving scenario model contains the representation of the traffic context in a certain driving scenario. The representation includes the specification of road conditions (number of lanes, curvature of the road, speed limit, etc.), macroscopic traffic state (speed of traffic flow, traffic density, homogeneity of traffic in different lanes, etc.) and environmental factors (lighting condition, weather condition, etc.). In each simulation run the initial constellation of the simulated traffic will be realized according to the driving scenario specification. The simulation itself will run autonomously, which means that each traffic participant in the simulation will be controlled by a behavioural model (and if necessary combined with a vehicle model) that works similarly as a human behaves in similar conditions. The driving scenario model, the driver model and the vehicle model can all be parameterized stochastically. Hence, the contextual, inter-individual and intra-individual variations, which influence significantly the simulation result, can be taken into account using stochastic simulation. The results will then give an overall comparison between human safety performance and function safety performance regarding the above-mentioned key indicators.

The continuous operation simulation works with a virtual traffic environment, which is timely and spatially extended. The construction of the virtual traffic environment has the most important objective that the automated functions can unfold their effects in a representative manner. Hence, it should provide a representative variation of traffic context to trigger realistic variation of system response.

In this sense, a section of motorway with two lanes going straight ahead will never be enough even if it is thousands of kilometres long. The simulated part of the traffic system has to be representative regarding the real traffic system. Otherwise, the simulation could be biased. Participants (and their vehicles) in the virtual traffic environment are – as in the scenario-based simulation - controlled by intelligent driver models or directly by the automated driving functions. Critical situations, accidents or generally abnormalities observed during the continuous operation simulation will be registered and analysed. As long as they are caused directly or indirectly by the automated driving functions, the driving situations will be specified as new driving scenarios and added into the scenario collection for the scenario-based simulation.

6.2 Environmental Impact Assessment

The objective of the environmental impact assessment is to describe the change of the traffic with respect to fuel consumption, traffic flow and travel time due to automated driving functions. It can be expected that different user groups will benefit in different manners. Therefore, the environmental impact assessment should also analysis how much different user groups benefit.

Approach for environmental impact assessment

The general approach for the environmental impact assessment is the same for all analysed effects (fuel consumption, traffic flow and travel time) as well as for the analysed functions the same, see Figure 6.3. The taken approach for the environmental impact assessment follows in general the approach of the safety impact assessment. However, the analysed scenarios focus only on the traffic scenario level.

The initial point of the impact assessment is the description of the functions. Based on this description first the relevant driving situations that are addressed by system under investigation are identified. In contrast to the safety impact assessment effects are investigated in the traffic scenario and not in the driving scenario. This means that more vehicles are taken into account and that the considered time frame of a driving situation is bigger. The scenarios that are influenced by the system under investigation are called “relevant scenarios”.

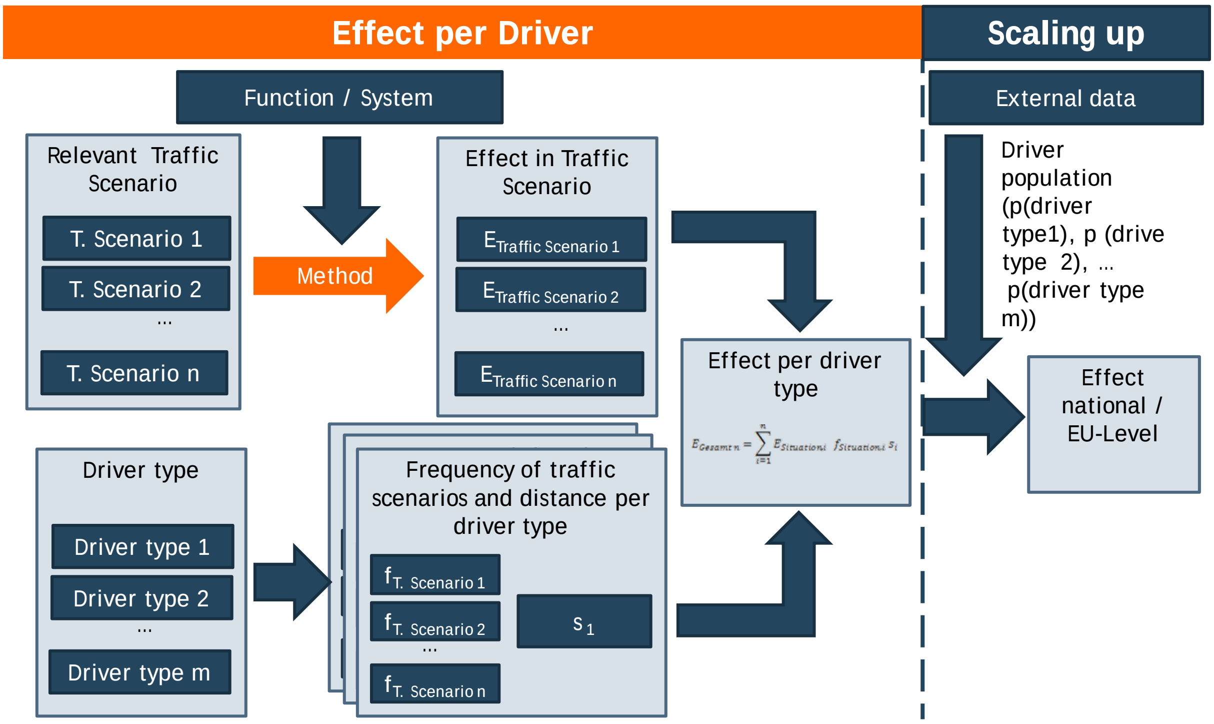


Figure 6.3: Basic concept for the environmental impact assessment.

For each relevant traffic scenario the effect of the automated driving functions is determined by simulating the relevant traffic scenario without and with the automated driving function. Here, different penetrations can be considered. The important question is how the effect in a traffic scenario is obtained. The answer to this question depends on which item (fuel consumption, travel time or traffic flow) should be assessed and which type of function is analysed. The considered types of the function are

- Event-based operating systems,
- Continuous operation systems.

An overview on different used methods for the different items is given in Figure 6.4. For the continuous operation functions mainly the traffic flow simulations will be used. This method has the advantage compared to other methods, like e.g. field operational test or simulator studies, that the required resources are low and the penetration can be varied, which allows also the consideration of mixed traffic conditions. Of course each simulation needs to be adapted to the analysed environment (urban, highway).

For the event-based functions the situation differs. First of all the analysis of the fuel consumption will not be conducted, since due to the short operation time the possible effect is quite low. Hence the focus will be for the travel time and the traffic flow. Since not the travel is considered for the event-based functions, the term travel time is a bit misleading that is why the term “manoeuvre time” is used for the event-based functions. Furthermore for the automated parking functions also the term “traffic flow” is not appropriated. Instead the used parking space will be analysed.

The used method for obtaining the effect of the event-based function in certain driving scenarios and manoeuvres can differ depending on the analysed manoeuvre or function. In principle effects can be obtained by means of simulator studies, simulation, tests on test track or field data. The current approach for the evaluation of a parking function foresees a combination of field observation as well as simulation respectively real world tests.

Research item	Event-based (Parking)	Continuous operating (Highway, Urban)
Fuel consumption	-	Traffic flow simulation
Travel / manoeuvre time	Field data on parking durations without function vs. (simulated / measured) Parking duration with system	Traffic flow simulation
Traffic flow / space	Analyse required needed parking space without (Field data) vs. with (test or simulation) the system	Traffic flow simulation

Figure 6.4: Method for effect in a certain scenario depending on the function type and the analysed item.

The metric that is used to quantify the effect of the function depends also on the investigated item. Depending whether the fuel consumption, the traffic flow or the travel / manoeuvre time is analysed, different indicators will be used. An overview on possible indicators for the evaluation is given in Figure 6.5.

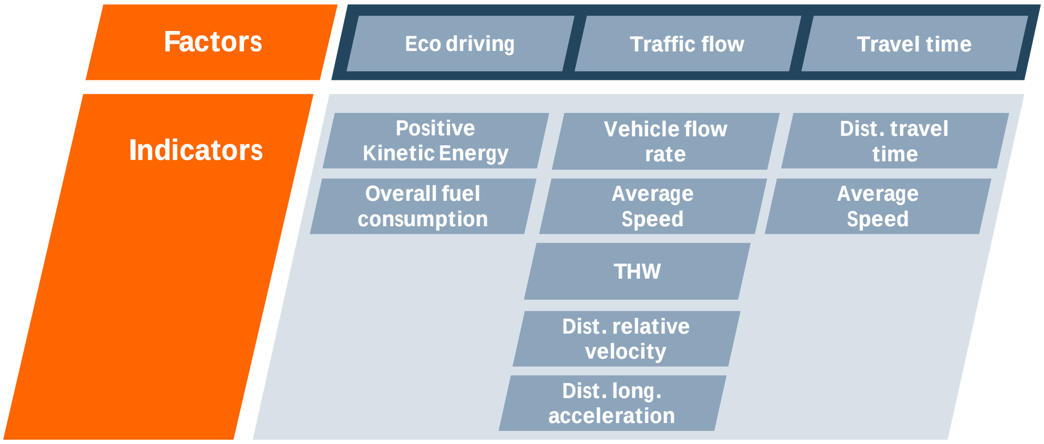


Figure 6.5: Proposed metrics to quantify the effects of automated driving functions.

Next to quantification of the effect per driving scenarios the different driver types are defined. The different drivers will be described based on the travel behaviour. Criteria to characterise the different driver types are:

- km driven per year
- Proportion usage of different road types (urban, rural and motorway)

In the next step for each driver type the (spatial) frequency of the different driving scenarios will be obtained. For this purpose different data source will be used (FOT data, traffic observations, questionnaires, statistical data).

Once the effect in certain driving scenarios, the frequency of the scenario as well as the driven distance per year are obtained, the effect for different driver types can be calculated; see Equation 6.1.

$$E_{Driver\ Type} = \left(\sum_{i=1}^{n} E_{scenario,i} \times f_{scenario,i}\right) \times s_{Driver\ Type} \qquad \text{Equation 6.1}$$

In the last step the single results for each defined driver type are up scaled on national or European level – depending on the available data. Therefore the population of different drivers is taken into account.

As already stated in the beginning of the chapter, the presented approach describes only the basic approach and by this reflects only the current discussion status within the work packages related to impact assessment. The final and detailed methodology will be presented in deliverable D7.3.

7 Conclusion

SP7 “Evaluation” is a horizontal activity supporting the vertical subprojects. Its main objective is to develop a common evaluation framework for supervised automated driving applications which is described within this deliverable. This framework addresses four different assessments for two evaluation stages. The first evaluation stage considers the evaluation of the status quo which consists of the technical, user-related and in-traffic assessment. The second evaluation stage, which will be described in more detail within the upcoming deliverable D7.3, concentrates on the analysis of the future benefits with respect to safety and environmental aspects, which can be achieved by means of automated driving applications

For all four different assessment types, a framework is set up within this deliverable. For each assessment, the starting point is the function or system under investigation itself. Based on its description, a classification will be done to determine which evaluation methodologies are most appropriate for the assessment. Within the AdaptIVe sub projects 4 to 6, automation functionalities for close-distance, urban as well as highway scenarios will be developed. As an evaluation of all of these functions in all assessments types is out of the scope of this project, only selected functions will be evaluated in the different assessments. Examples of the evaluation procedure are provided with the presentation of each methodology. Within AdaptIVe, two general types of functions are distinguished: event based functions that only operate for a short period of time as well as continuous operating functions which once activated will operate over a longer time period.

The respective methodology for each assessment is outlined and the focus of the assessment is presented including the relevant research questions, hypotheses and indicators. The research questions from the first step of the evaluation and provide information on what should be addressed. Based on those research questions, hypotheses to be tested were defined. Testing of the hypotheses will be done by using indicators that can be calculated based on signals or be derived from measures logged during the tests. It should be noted, that not all of these research questions, hypotheses and defined indicators might be applicable for all functions or systems. Therefore, for each combination of system and chosen evaluation an appropriate subset needs to be considered. The different methods and tools for the assessment are discussed and an example of the application of the evaluation methodology for an automated driving system is provided. Finally, also the requirements for testing with respect to safety, test-tools and test conduction are presented.

The framework presented here though extensive might not yet be complete as so far the proof of concept is missing. Only actual evaluation of a system or function according to the specifications set will show if the chosen approach is feasible or still lacks practical aspects and needs refinement. This proof of concept will be provided with the final evaluation in AdaptIVe.

Further work will be conducted in close cooperation with the other SPs to ensure final evaluation of selected automated systems and functions according to a feasible and suitable assessment procedure.

7.1 Outlook

Base on this deliverable the "Evaluation" subproject of AdaptIVe (SP7) will prepare and conduct the tests for the different assessments. In the next step for the different assessments a demonstrator vehicle will be selected and a test plan specific to the chosen demonstrator will be prepared. All test plans will be set up in close cooperation with the VSP's and include a detailed description of the different tests to be conducted. This includes the specification of test parameters; used test tracks respectively test routes, required test tools as well as timing.

Next to the continued planning activities for the tests, SP7 will start to develop the required test tools. In particular the evaluation tool chains – starting from the logging of the data up to the hypothesis testing - for the different assessments will be built up and verified. These tool chains in the different assessments are required in order to ensure an accurate and fast evaluation in the end.

After the preparation of the evaluation activities the actual tests will be conducted according to the time plan provided in chapter 2.4. The test data logged during the tests will form the basis for the evaluation in the different assessments. The evaluation will be carried out as described in this deliverable.

Besides the actual exemplary evaluation of selected AdaptIVe functions and systems, SP7 will analyse in a final step how well the proposed evaluation methodology as described in this document has worked out. Therefore, the experience of the AdaptIVe evaluation will be used to derive "lessons learned", which will be important input for future evaluation activities.

References

[1] J. Scholliers, F. Hendriks, M. L. Aust und e. al., „PReVAL-Project final report and recommondations for future assessments,“ PReVAL Deliverable DL IP D12/D16.4, 2008.

[2] I. Wilmink, W. Janssen, E. Jonkers und e. al., „Impact assessment of intelligent vehicle safety Systems,“ eIMPACT, Deliverable D4, 2008.

[3] C. Rodarius, S. de Hair, P. Seiniger and et.al, "Darf test and assessment protocol for the pre-crash evaltuation of integrated safety systems based on existing, updated and standardized test methods," in *ASSESS, Deliverable D4.2*, ASSESS, Deliverable D4.2, 2011.

[4] F. Fahrenkrog, A. Zlocki, J. Dobberstein und e. al., „Impact Assessment of Developed Applications- Overall interactIVe Assessment,“ interactIVe, Deliverable D7.5, 2014.

[5] G. Alessandretti, M. Aeberhard, C. Hilbert und e. al., „Use Case and Requirements,“ AdaptIVe, Deliverable D1.5, 2014.

[6] A. Saroldi, L. Andreone, F. Visintainer und e. al., „Inital System Specification,“ AdaptIVe, Deliverable D1.6, 2014.

[7] SAE, „Taxonomy and Definitions for Terms Related to On-Road Automated Motor Vehicles,“ SAE document J3016, 2014.

[8] A. Bartels, U. Eberle und A. Knapp, „System Classification and Glossary,“ AdaptIVe Deliverable D2.1, Wolfsburg, 2015.

[9] A. Dávila, „Validation Test Plan and Results,“ SARTRE Projekt, Deliverable D4.1 & D4.2, 2012.

[10] A. Beutner, S. Pettersson, F. Seglö und e. al., „Optimized and validated demonstration vehicles,“ HAVEit, Deliverable D13.1, 2011.

[11] S. Deutschle, „Das KONVOI-Projekt - Entwicklung und Untersuchung des Einsatz von elektrischen LKW-Konvois auf Autobahnen,“ 15. Aachener Kolloquium, Aachen, 2006.

[12] S. Tsugawa, „Final Report on Automated Truck Platoon within Energy ITS Project,“ International Task Force on Vehicle Highway Automation 17th Annual Meeting, Tokyo, 2013.

[13] P. Larsson, I. Esberg, M. van Noort und e. al., „Test and Evaluation Plan,“ interactIVe, Deliverable D7.4, 2012.

[14] V. Karabatsou, M. Pappas, P. Elslande und e. al., „A-priori evaluation of safety functions effectivness - Methodologies,“ Trace, Deliverable D4.1.3, 2006.

[15] M. Draskóczy, O. Carsten und R. Kulmala, „Road Safety Guidelines," Code, Deliverable B5.2, 1998.

[16] C. Kessler, A. Etemad und G. Alessandretti, „Final Report," euroFOT, Deliverable D11.3, 2012.

[17] N.N., „TeleFOT, Field Operaiton Tests of Aftermarket and Nomadic Devices in Vehicles," TeleFOT Fact Scheet, European Commission, Brussels, 2008.

[18] A. Varhelyi, E. Adell, A. Rook und e. al., „Driver behavioural effects from ISA," Deliverable 3.4, PROSPER Consortium, 2005.

[19] O. Carsten und M. Fowkes, „Executive Summary of Project Results," External Vehicle Speed Control, 2000.

[20] N.N., „European New CAr Assessment Programme - Assessment Protocol Safety Assist," Euro NCAP, Version 6, 2013.

[21] N.N., „Report to Congress - DARPA Prize Authority," Defence Advanced Research Projects Agency, Arlington, 2005.

[22] N.N., „Urban Challenge - Technical Evaluation Cirteria," Defence Advance Research Projects Agency, Arlington, 2006.

[23] N.N., „Grand Cooperative Driving Challenge," Information from webpage: http://www.gcdc.net/, (call date) 2014 .

[24] ISO, „ISO 26262 - Road Vehicles - Functional Safety," 2011.

[25] A. Knapp, M. Neumann, M. Brockmann und e. al., „Code of Practice - Design and Evaluation of ADAS," PReVENT Response 3, Delivable, 2009.

[26] H. Winner, „Die Freigabefalle des autonomen Fahrens," 5. Damrstädter Kolloquium, Darmstad, 2011.

[27] L. Eckstein und A. Zlocki, „Safety Potential of ADAS – Combined Methods for an Effective Evaluation," 23rd Enhanced Safety of Vehicles, Seoul, 2013.

[28] S. Schlotzhauer, „S. Elementary Statistics Using JMP," 2007.

[29] R. Craparo, „Significance level: Encyclopedia of Measurement and Statistics," 2007.

[30] F. Hendriks, „Development Tools for Active Safety Systems: PreScan and VeHIL," IEEE, 2010.

[31] B. Schick, „Simulation Methods to Evaluate and Verify Functions, Quality and Safety of Advanced Driver Assistance Systems,“ 2008, IPG Automotive GmbH.

[32] K. v. Neumann-Cosel, „Virtual Test Drive - Simulation umfeldbasierter Fahrzeugfunktion,“ München, 2014.

[33] SAE, „Guidelines for Safe On-Road Testing of SAE Level 3,4, and 5 Prototype Automated Driving Systems (ADS),“ in s *Surface Vehicle Information Report J3018*, SAE, 2015.

[34] N. Stanton und M. Young, „Vehicle Automation and Driving Performance,“ *Ergonomics,* pp. 1014-1028, 1998.

[35] OECD, „Behavioural Adaptations to Changes in the Road Transport System,“ University of Chicago Press, Chicago, USA, 1990.

[36] A. Smiley, „Behavioural Adaptation, Safety and Intelligent Transportation Systems,“ *Transportation Research Record,* pp. 47-51, 2000.

[37] R. Parasuraman und D. Manzey, „Complacency and bias in human use of automation: an attentional integration,“ *Human Factors,* pp. 381-410.

[38] J. Preece, Y. Rogers und H. Sharp, Interaction design: beyond human computer interaction, New York: John Wiley & Sons, 2002.

[39] J. Jenness, N. Lerner, S. Mazor, J. Osberg und B. Tefft, „Use of advanced in-vehicle technology by young and older early adopters: Select results from five technology surveys. (report No. HS 811 004),“ US Department of Transportations, National Highway Traffic Safety Administration, Washington D.C., USA, 2008.

[40] S. Merrit, „Affective process in human-automation interactions,“ *Human Factors,* pp. 356-370, 2011.

[41] R. Parasuraman und V. Riley, „Humans and automation: Use, misuse, disuse, abuse,“ *Human Factors,* pp. 230-253, 1997.

[42] S. McBride, W. Rogers und A. Fisk, „Do younger and older adults differentially depend on an automated system?,“ in s *Paper presented at the Human Factors and Ergonomics Society 54th Annual Meeting*, San Francisco, CA, 2010.

[43] S. Merrit, H. Heimbaugh, J. LaChapell und D. Lee, „I trust it, but I don't know why: Effects of implicit attitudes toward automation on trust in an automated system,“ *Human Factors,* 2012.

[44] N. Merat, A. Jamson, F. Lai, M. Daly und O. Carsten, „Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. Transportation

Research Part F: Traffic Psychology and Behaviour, Volume 27, Part B,“ 2014.

[45] T. Trimble, R. Bishop, J. Morgan und M. Blanco, „Human factors evaluation of level 2 and level 3 automated driving concepts: Past research, state of automation technology and emerging system concepts,“ National Highway Traffic Safety Administration, Washington D.C., 2014.

[46] M. Endsley und E. Kiris, „The out-of–the-loop performance problem and level of control in automation,“ in s *Human Factors, 37*, 1995, pp. 381-394.

[47] R. Parasuraman, R. Molloy und I. Singh, „Performance consequences of automation induced "complacency",“ *The international journal of aviation psychology,* pp. 1-23, 1993.

[48] I. Singh, R. Molloy und R. Parasuraman, „Automation-induced monitoring inefficiency: role of display location,“ *International Journal of Man-Machine Studies,* pp. 17-30, 1997.

[49] T. Sheridan, Humans and Automation: System Design and Research Issues, Santa Monica: Wiley, 2002.

[50] N. Strand, „When Driving Automation Fails. Drivers Experiences and Interactions,“ Chalmers University, Gothenburg, Sweden, 2014.

[51] V. Banks, N. Stanton und C. Harvey, „Sub-systems on the road to vehicle automation: Hands and feet free but not mind free driving,“ in s *Safety Science*, 2014, pp. 505-514.

[52] A. McDonald und G. Davey, „Psychiatric disorders and accidental injury,“ *Clinical Psychological Review,* pp. 105-127, 1996.

[53] G. Funke, G. Matthews, J. Warm und A. Emo, „Vehicle automation: A remedy for driver stress?,“ *Ergonomics,* pp. 1302-1323, 2007.

[54] B. Reimer, B. Mehler und J. Coughlin, „An evaluation of driver reactions to new vehicle parking assist technologies developed to reduce driver stress,“ New England University Transportation Center, Massachusetts Institute of Technology, Cambridge, 2010.

[55] M. Cummings, C. Mastracchio, K. Thornburg und A. Mkrtchyan, „Boredom and distraction in multiple unmanned vehicle supervisory control,“ in s *Interacting with Computers, 25(1)*, 2013, pp. 34-47.

[56] R. Farmer und N. Sundberg, „Boredom proneness - the development and correlates of a new scale,“ in s *Journal of Personality Assessment, 50, 4-17.*, 1986.

[57] J. Stark und M. Scerbo, „The effects of complacency potential and boredom proneness on perceived workload and task performance in an automated environment,“ in s *Human Factors and Ergonomics Society 42nd Annual Meeting*, Chicago, 1998.

[58] L. Prinzel, H. De Vries, F. Freeman und P. Mikulka, „Examination of Automation-Induced Complacency and Individual Difference Variates," NASA/TM-2001-211413, 2001.

[59] D. S. M. Sawin, „Effects of instruction type and boredom proneness in vigilance: Implications for boredom and workload," in s *Human Factors. 37 (4)*, 1995, pp. 752-765.

[60] I. Brown, „Prospects for technological countermeasures against driver fatigue," in s *Accident Analysis & Prevention29, 4*, 1997, p. 525–531.

[61] G. Maycock, „Driver Sleepiness as a Factor in Car and HGV Accidents," *TRL Report,* 1995.

[62] G. Matthews und P. Desmond, „Personality and multiple dimensions of task-induced fatigue: a study of simulated driving," *Personality and Individual Differences,* pp. 443-458, 1998.

[63] G. Matthews und P. Desmond, „Task-induced fatigue states and simulated driving performance," *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology,* pp. 659-686, 2002.

[64] O. Carsten, F. Lai, Y. Barnard, A. Jamson und N. Merat, „Control task substitution in semi-automated driving: Does it matter what aspects are automated?," *Human Factors,* pp. 747-761, 2012.

[65] N. Stanton, P. Chambers und J. Piggott, „Situational awareness and safety," in s *Safety Science 39*, 2001, p. 189–204.

[66] M. Endsley, „Toward a theory of situation awareness in dynamic-systems," 1995.

[67] C. Perrow, Normal accidents: living with high-risk technologies, New York: Basic Books, 1984.

[68] C. Wickens und S. Dixon, „The benefits of imperfect diagnostic automation: a synthesis of the literature," *Theoretical Issues in Ergonomics Science,* pp. 201-212, 2007.

[69] L. Nilsson, „Safety effects of adaptive cruise controls in critical traffic situations," in s *Proceedings of the Second World Congress on Intelligent Transport Systems*, Yokohama, Japan, 1995.

[70] M. Vollrath, S. Schleicher und C. Gelau, „The influence of cruise control and adaptive cruise control on driving behaviour - A driving simulator study," *Accident Analysis and Prevention,* pp. 1134-1139, 2011.

[71] C. Billings, J. Laubner, H. Funkhouser, G. Lyman und E. Huff, „Aviation Safety Reporting System (Technical Report TM-X-3445)," Moffett Field, CA: National Aeronautics and Space Administration, Ames Research Center, 1976.

[72] K. Mosier und L. Skitka, „Human decision-makers and automated decision aids: Made for each other?," in s *Automation and human performance: Theory and application*, Lawrence Erlbaum Associates, 1996, pp. 201-220.

[73] J. Domeinski, R. Wagner, M. Schoebel und D. Manzey, „Human redundancy in automation monitoring: Effects of social loafing and social compensation," in s *Proceedings of the Human Factors and Ergonomics Society 51st Annual Meeting*, Santa Monica, Human Factors and Ergonomics Society, 2007, pp. 587-591.

[74] C. Harvey, N. Stanton, C. Pickering, M. McDonald und P. Zheng, „A usability evaluation toolkit for In-Vehicle Information Systems (IVISs)," *Applied Ergonomics,* pp. 563-574, 2011.

[75] E. Adell, L. Nilsson und A. Várhelyi, „How is Acceptance measured? Overview of Measurement Issues, Methods and Tools," in s *Driver Acceptance of new technology. Theory, Measurement and optimisation*, Ashgate, 2014, pp. 73-89.

[76] E. Adell, A. Várhelyi und L. Nilsson, „The Definition of Acceptance and Acceptability," in s *Driver Acceptance of new technology. Theory, Measurement and optimisation*, Ashgate, 2014, pp. 11-22.

[77] M. Endsley, „Situation Awareness Global Assessment Technique (SAGAT)," in s *Proceedings of the National Aerospace and Electronics Conference*, New York, 1988.

[78] W. Helton, „Validation of a Short Stress State Questionnaire," in s *Proceedings of the Human Factors and Ergonomics society 48th annual meeting*, 2004.

[79] J. Byers, J. Alvah, C. Bittner und S. Hill, „Traditional an raw task load index (TLX) correlations: are paired comparisons necessary?," in s *Annual International Industrial Ergonomics and Safety*, Cincinnati, Ohio, USA, Taylor and Francis, 1989.

[80] M. Scerbo, K. Rettig und C. Bubb-Lewis, „A validation study of a task-related boredom scale," in s *Proceedings of the 2nd Mid-Atlantic Human Factors Conference, (pp. 135-136).*, Washington, D. C., 1994.

[81] J. Van der Laan, A. Heino und D. de Waard, „A simple procedure for the assessment of acceptance of advanced transport telematics," *Transportation Research C,* pp. 1-10, 1997.

[82] R. Risser und C. Brandstätter, „(1985). Die Wiener Fahrprobe. Freie Beobachtung.," KL FACHBUCHR KFV, 21., 1985.

[83] M. Hjälmdahl und A. Várhelyi, „Speed regulation by in-car active accelerator pedal: Effects on driver behaviour," Transportation Research Part F: Traffic Psychology and Behaviour 7.2, 2004.

[84] R. Halpert und R. Hill, „28 Measures of locus of control," 2011. [Online]. Available:

http://teachinternalcontrol.com/uploads/LOC_Measures__1_.pdf.

[85] P. Ting, J. Hwang, J. Doong und M. Jeng, „Driver fatigue and highway driving: A simulator study," *Physiology and Behaviour,* pp. 448-453, 2008.

[86] I. Singh, R. Molloy und R. Parasuraman, „Automation-induced "complacency": Development of the complacency-potential rating scale," *International Journal of Aviation Psychology,* pp. 111-121, 1993.

[87] C. Harvey, N. Stanton, C. Pickering, M. McDonald und P. Zheng, „Context of use a factor in determining the usability of in-vehicle devices," *Theoretical Issues in Ergonomics Science,* pp. 318-338, 2011.

[88] J. Brook, „SUS: a 'quick and dirty' usability scale," in s *P.W. Jordan, B. Thomas, B.A. Weerdmeester, I.L. McClelland: Usability Evaluation in Industry*, London, Taylor and Francis, 1996, pp. 189-194.

[89] E. Adell, „The concept of Acceptance," in s *Proceedings of the 20th ICTCT-workshop*, Valencia, Spain, 2007.

[90] V. Venkatesh, M. Morris, G. Davis und F. Davis, „User acceptance of information technology: Toward a unified view," *MIS Quarterly,* pp. 425-478, 2003.

[91] E. Adell, „Driver experience and acceptance of driver support systems – a case of speed adaptation," Lund University, Sweden, 2009.

[92] J. Sauer, K. Seibel und B. Rüttinger, „The influence of user expertise and prototype fidelity in usability tests," *Applied Ergonomics,* pp. 130-140, 2010.

[93] T. A. Dingus, S. G. Klauer, V. L. Neale und e. al., „The 100-Car Naturalistic Driving Study, Phase II - Results of the 100-Car Field Experiment," DOT HS 810 593, Washington, DC, USA, 2006.

[94] TASS International, „PreScan," TASS International, 2015. [Online]. Available: https://www.tassinternational.com/prescan. [Zugriff am 19 March 2015].

[95] The Mathworks, „Matlab Simulink," The Mathworks, 2015. [Online]. Available: http://uk.mathworks.com/?refresh=true. [Zugriff am 19 March 2015].

[96] Artemis, „Artemis EMC2 - Living Lab 1 Automotive Applications," Artemis, 2015. [Online]. Available: http://www.artemis-emc2.eu/project_overview/ll1_automotive_applications/. [Zugriff am 19 March 2015].

[97] Artemis, „Artemis Crystal - Automotive," Artemis, 2015. [Online]. Available: http://www.crystal-artemis.eu/application-domains/automotive.html. [Zugriff am 19

March 2015].

[98] European Aviation Safety Agency, „DECISION NO. 2003/6/RM on certification specifications, including airworthiness codes and acceptable means of compliance, for all weather operations (« CS-AWO »),“ EASA, Cologne, Germany, 2003.

[99] D. Kroese, T. Taimre und Z. Botev, Handbook of Monte Carlo Methods, New York: John Wiley & Sons, 2011.

[100] R. Saucier, „Computer Generation of Statistical Distributions,“ Army Research Laboratory, Adelphi, MD, USA, 2000.

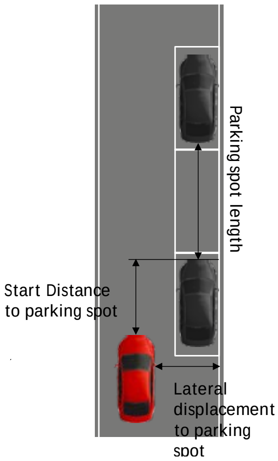
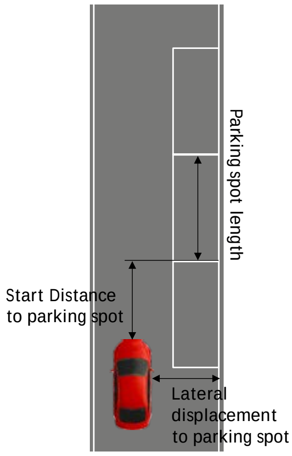
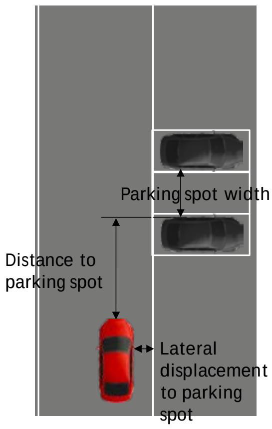
[101] L. v. Rooij, P. Sukumar, L. Kroon, T. Bijlsma und J. Hogema, „Driver Behaviour Monitoring based on Vehicle Sensors for Safety-Critical Event Logging,“ in s *FISTA World Automotive Congress*, Maastricht, the Netherlands, 2014.

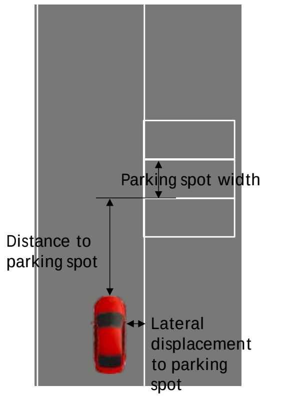
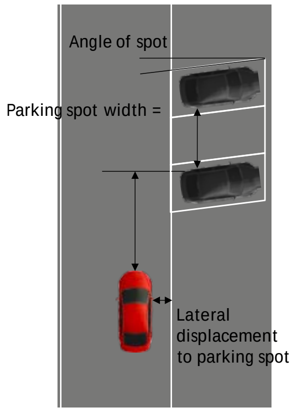
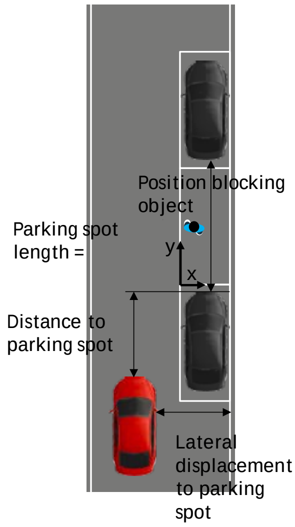
[102] European Commission, „EU Transport in Figures - Statistical Pocketbook 2014,“ European Commission, Brussels, Belgium, 2014.

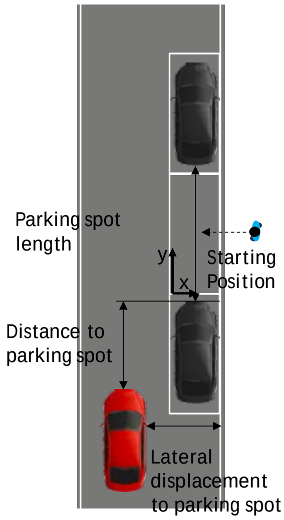
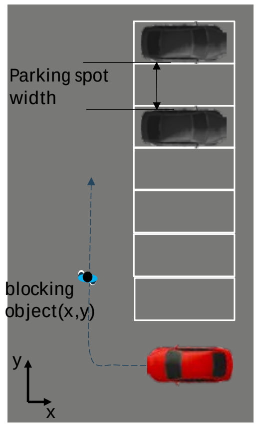
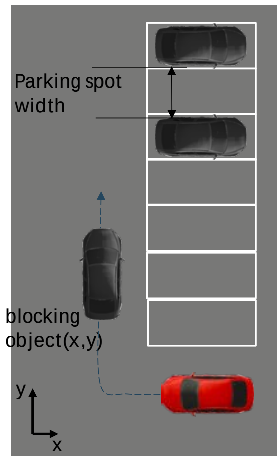
List of abbreviations and acronyms

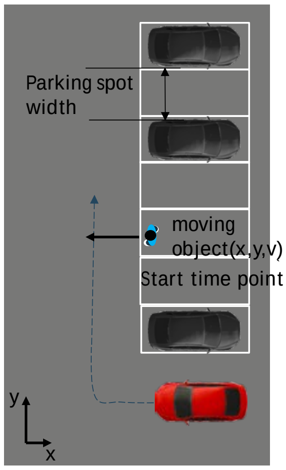
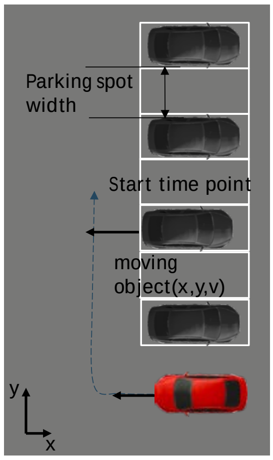
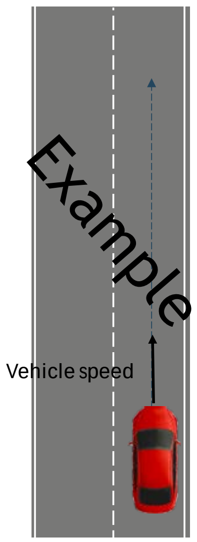
Abbreviation	Meaning
CMC	Crude Monte Carlo
CPRS	Complacency Potential Rating Scale
DSSQ	Dundee Stress State Questionnaire
F_s	Spatial frequency of driving situation
$F_{cluster}$	Spatial frequency of cluster of driving situations
GIDAS	German In-Depth Accident Study
HIL	Hardware-in-the-Loop
MAIS	Maximum Abbreviated Injury Scale
MIL	Model-in-the-Loop
PRC	Percent Road Centre
P	Probability
RTLX	Task Load indeX
RMSE	Root-Mean-Squared-Error
SAE	Society of Automotive Engineers
SAGAT	Situation Awareness Global Assessment Technique
SDLP	Standard Deviation of Lane Position
SiL	Software-in-the-Loop
s_{ref}	Reference distance for a single driving situation
SSSQ	Short Stress State Questionnaire
SUS	System Usability Scale
TBS	Task-related Boredom Scale
THW	Time headway
TLC	Time-Line-Crossing
TTC	Time-To-Collision
UTAUT	Unified Theory of Acceptance and Use of Technology
V2V	Vehicle-to-Vehicle communication
V2X	Vehicle-to-X communication
VuT	Vehicle under test
VSP	Vertical subproject

Annex 1 Test cases for event-based operating functions for technical assessment

ID	Description	Varied Parameters	Fixed Parameters	Equipment	Sketch
E1	**Parallel Parking:** ● Parking into a parallel parking spot which is defined by two vehicles. The ego- vehicle has to detect the free parallel parking spot and has to drive safely into the spot.	● Parking spot length length of EGO vehicle ● Lateral displacement to parking spot	● Start Distance to parking spot	● 2 balloon cars ● Reference sensor (e.g. laser scanner or RTK-GPS)	
E2	**Parallel Parking without objects:** ● Parking into a parallel parking spot which is defined by parking space markings. The ego-vehicle has to detect the parallel parking spot and has to drive safely into the spot.	● Parking spot length length of EGO vehicle ● Lateral displacement to parking spot	● Start Distance to parking spot	● Reference sensor (e.g. laser scanner or RTK-GPS)	
E3	**Orthogonal Parking:** ● Parking into an orthogonal parking spot which is defined by two vehicles. The ego- vehicle has to detect the free parking spot and has to drive safely into the spot.	● Lateral displacement to parking spot	● Parking spot width ● Distance to parking spot	● 2 balloon cars ● Reference sensor (e.g. laser scanner or RTK-GPS)	

ID	Description	Varied Parameters	Fixed Parameters	Equipment	Sketch
E4	**Orthogonal Parking without objects:** • Parking into an orthogonal parking spot which is defined by parking space markings. The ego-vehicle has to detect the parking spot and has to drive safely into the spot.	• Lateral displacement to parking spot	• Parking spot width • Distance to parking spot	• Reference sensor (e.g. laser scanner or RTK-GPS)	
E5	**Angular Parking:** • Parking into an angular parking spot which is defined by two vehicles. The ego- vehicle has to detect the free parking spot and has to drive safely into the spot.	• Lateral displacement to parking spot	• Parking spot width • Distance to parking spot • Angle of parking spot	• 2 balloon cars • Reference sensor (e.g. laser scanner or RTK-GPS)	
E6	**Parking with object blocking:** • Parking into a parallel parking spot which is defined by two vehicles. The ego- vehicle has to detect the free parallel parking spot and has to drive safely into the spot. Furthermore, the vehicle has to detect the blocking object and stop or not to enter the parking spot.	• Lateral displacement to parking spot •Position of blocking object (x,y)	• Parking spot length	• 2 balloon cars • Crashable object (e.g. pedestrian dummy) • Reference sensor (e.g. laser scanner or RTK-GPS)	

ID	Description	Varied Parameters	Fixed Parameters	Equipment	Sketch
E7	**Parking with moving object:** • Parking into a parallel parking spot which is defined by two vehicles. The ego- vehicle has to detect the free parallel parking spot and has to drive safely into the spot. Furthermore, the vehicle has to detect the moving object and stop.	•Lateral displacement to parking spot (0.5, 0.9 m) • Starting Position of object (TBD) • Starting time point of object (TBD)	• Distance to parking spot (TBD) • Parking spot length (length of EGO vehicle + 1.0 m) • Object Speed (y-direction) (5 km/h)	• 2 balloon cars • RTK-GPS Reference sensor (e.g. laser-scanner or RTK-GPS) • Movable Object (pedestrian dummy)	
E8	**Valet Parking:** • Valet parking into an orthogonal parking spot which is defined by two vehicles. The ego- vehicle has to drive to the designated parking spot. Furthermore, the ego-vehicle has to detect the blocking object and stop.	• Position of blocking pedestrian (x,y) (TBD) •Driven Route of the vehicle (TBD)	• Parking spot width (width of ego vehicle + 0.6 m)	• 2 balloon cars • Reference sensor (e.g. laser-scanner or RTK-GPS – if possible) • Pedestrian dummy	
E9	**Valet parking with blocking car:** • Valet parking into an orthogonal parking spot which is defined by two vehicles. The ego- vehicle has to drive to the designated parking spot. Furthermore, the ego-vehicle has to detect the blocking object and stop.	Position of blocking pedestrian (x,y) (TBD) Driven Route of the vehicle (TBD)	Parking spot width (width of ego vehicle + 0.6 m)	• 3 balloon cars • Reference sensor (e.g. laser-scanner or RTK-GPS – if possible)	

ID	Description	Varied Parameters	Fixed Parameters	Equipment	Sketch
E10	**Valet parking with moving pedestrian:** • Valet parking into an orthogonal parking spot which is defined by two vehicles. The ego- vehicle has to drive to the designated parking spot. Furthermore, the ego-vehicle has to detect the moving object and stop.	• Position of pedestrian (x,y) (TBD) • Velocity of pedestrian (v) (TBD) • Driven Route of the vehicle (TBD) • Start time point (TBD)	• Parking spot width (width of ego vehicle + 0.6 m)	• 3 balloon cars (or 2 balloon cars and 1 real vehicle) • Reference sensor (e.g. laser-scanner or RTK-GPS – if possible) • Pedestrian dummy	
E11	**Valet parking with unparking vehicle:** • Valet parking into an orthogonal parking spot which is defined by two vehicles. The ego- vehicle has to drive to the designated parking spot. Furthermore, the ego-vehicle has to detect the moving object and stop.	• Position of moving vehicle (x,y) (TBD) • Velocity of moving vehicle (v) (TBD) • Driven Route of the vehicle (TBD) • Start time point of moving vehicle (TBD)	• Parking spot width (width of ego vehicle + 0.6 m)	• 3 balloon cars (or 2 balloon cars and 1 real vehicle) • Reference sensor (e.g. laser-scanner or RTK-GPS – if possible) • One moving vehicle	
E12	**Minimum Risk Manoeuvre:** Test case will be defined based on definition of the manoeuvres performed by the minimum risk function (information provided by the VSP)	• TBD	• TBD	• Reference sensor (e.g. RTK-GPS)	

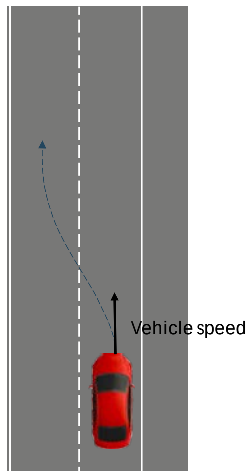
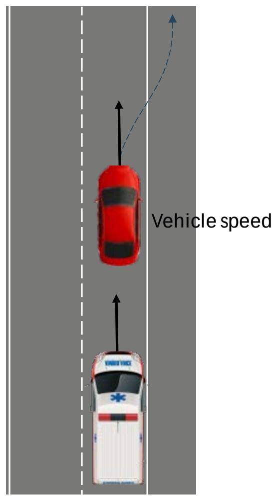
ID	Description	Varied Parameters	Fixed Parameters	Equipment	Sketch
E13	**Manual Triggered Lane Change:** • The ego- vehicle conducts a lane change to the left lane, which initiated by the driver manually.	• Vehicle speed 80 km/h, TBD	-	• Reference sensor (e.g. RTK-GPS)	
E14	**Emergency vehicle on duty:** • The ego- vehicle has to provide an emergency corridor for an approaching emergency vehicle.	• Vehicle speed • V2V communication	-	• Reference sensor (e.g. RTK-GPS)	

Figure 7.1: Overview of test cases for event-based operating functions

Annex 2 Pre-Tests and Driving situations for continuous operating functions

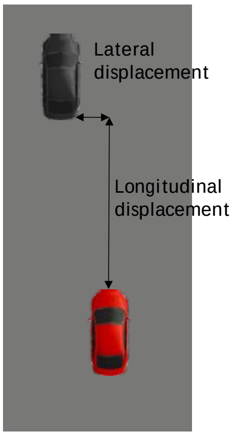
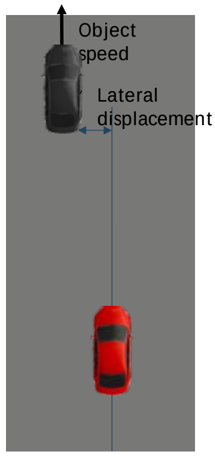
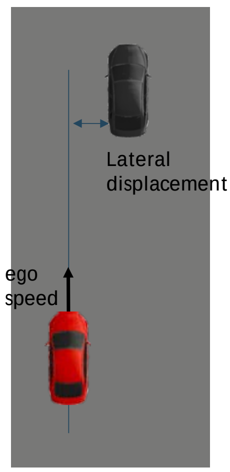
Overview driving situations for continuous operating functions

Driving Manoeuvre	Motorway	Rural	Urban
CP1	X	X	X
CP2	X	X	X
CP3	X	X	X
CP4	X	X	X
C5	X	X	X
C6	X	X	X
C7	X	X	X
C8	X	X	X
C9	X	X	X
C10	X	X	X
C11	X	X	X
C12			X
C13			X

Driving Manoeuvre	Motorway	Rural	Urban
C14			X
C15			X
C16			X
C17			X
C18			X
C19			X
C20	X		
C21	X		
C22	X		
C23	X		
C24	X		
C25	X		

Figure 7.2: overview of driving situations for continuous operating functions

Pre-Tests for continuously operating functions on a test track

ID	Description	Varied Parameters	Fixed Parameters	Equipment	Sketch
CP1	**Static Sensor Test:** • determination of the sensor field of view and sensor accuracy by measuring the position of a static object versus reference sensor.	• Object longitudinal displacement • Object lateral displacement	-	• Reference sensor (e.g. laser-scanner or RTK-GPS) • Sensor target vehicle	
CP2	**Dynamic Sensor Test:** • Acquisition of the sensor field of view and sensor accuracy by measuring the position of a moving object, e.g. car, with a defined displacement to the ego-vehicle.	• Ego-vehicle speed • Object lateral displacement	• Object Speed	• Reference sensor (e.g. laser-scanner or RTK-GPS) • Additional vehicle (real vehicle)	
CP3	**Dynamic Sensor Test:** • Acquisition of the sensor field of view and sensor accuracy by measuring the position of a static object while the ego vehicle is moving	• Object longitudinal displacement • Object lateral displacement	• Test road (including straights and curves)	• Additional vehicle (real vehicle) • Reference sensor (e.g. laser-scanner or RTK-GPS)	

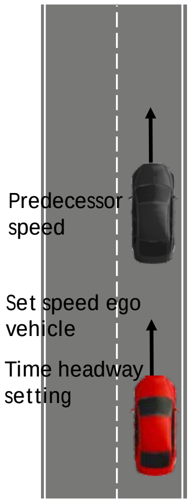
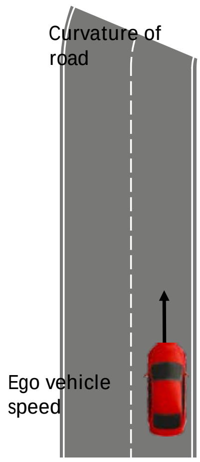
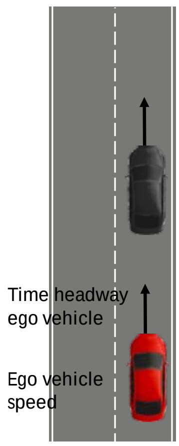
ID	Description	Varied Parameters	Fixed Parameters	Equipment	Sketch
CP4	**Basic Functionality: Car Following** ● The ego- vehicle is following a leading vehicle, which is driving a specified longitudinal trajectory.	● Predecessor speed ● Set speed ego vehicle (if possible) ● Time Headway setting (if possible)	● Test road (including straights and curves)	● Additional real vehicle ● Reference sensor (e.g. laser-scanner or RTK-GPS)	

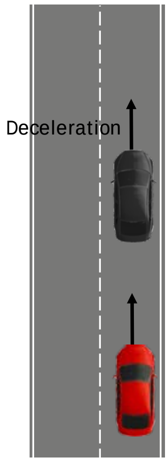
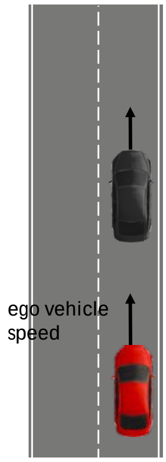
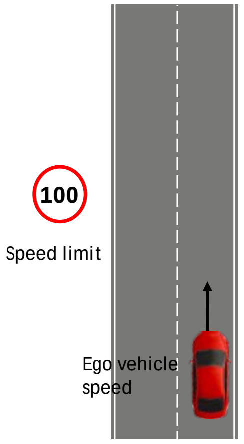
Figure 7.3: Pre-tests for continuous operating functions

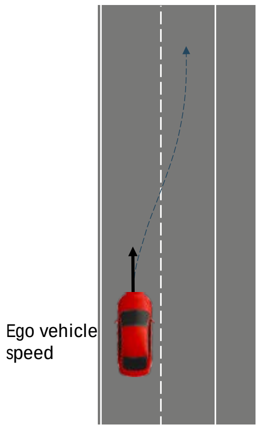
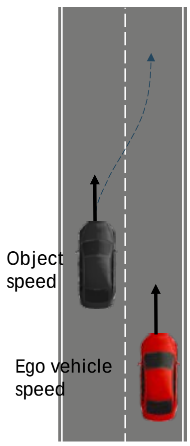
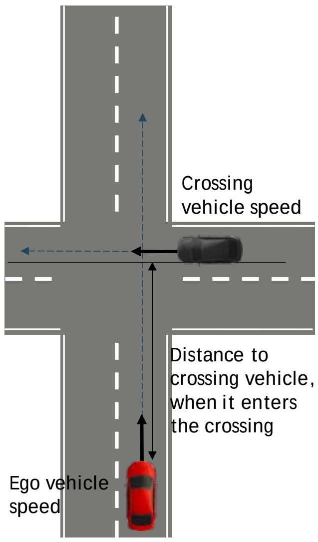
Relevant driving situations for continuously operating functions

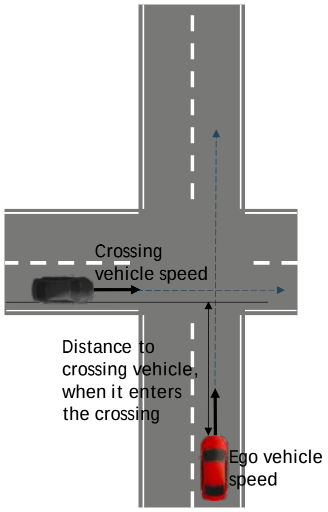
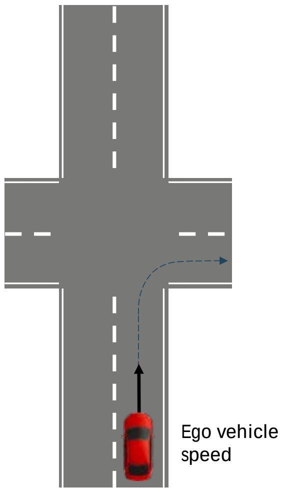
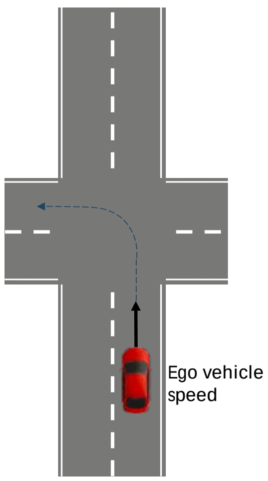
The in the following table mentioned driving situations represent a general overview on possible driving situations. This table also represents only the current status. Additional driving situations might be added over time depending on the demand.

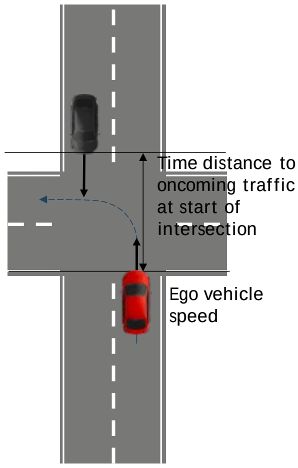
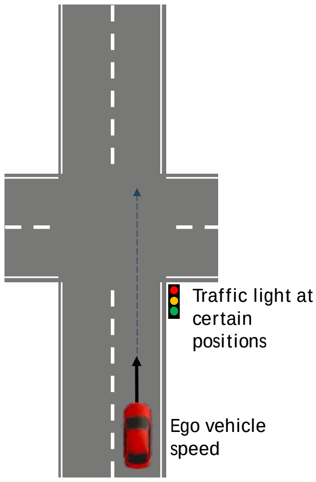
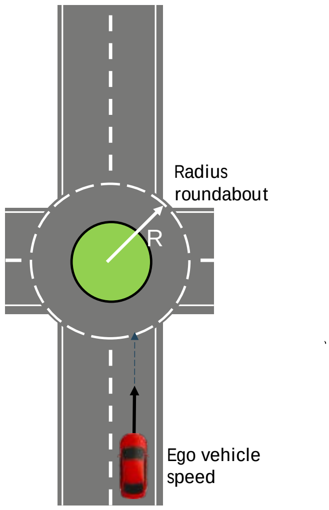
In this context it is important that not all mentioned driving situations are covered by AdaptIVe functions respectively systems. Nevertheless these situations might be relevant for other automated driving function. Therefore, these situations are mentioned in this table. For the evaluation in AdaptIVe only the relevant driving situations will be considered.

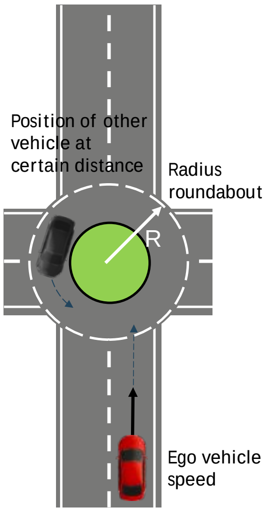
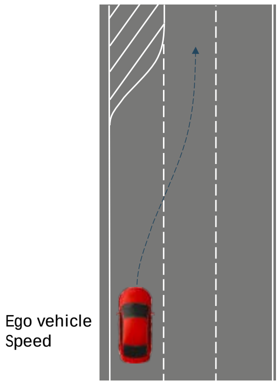
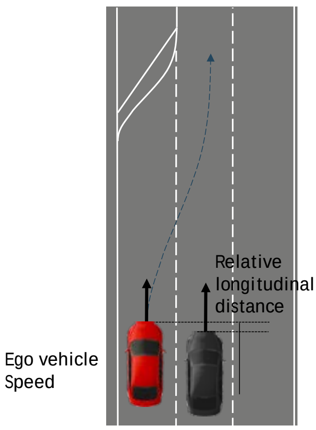
ID	Description	Situation Parameters	Equipment	Sketch
C5	**Constant Driving:** • All situations, in which the test vehicle is driving constant without any predecessor (distance to predecessor > 250 m) • The test is conducted to observe the lateral and longitudinal controlling behaviour of the vehicle while following straight and curved roads.	• Ego vehicle speed • Curvature of road	• Test vehicle with logging equipment • Digital map data • RTK-GPS • Video data	
CP6	**Car Following:** • All situations, in which the ego vehicle is following a leading vehicle.	• Time headway ego vehicle • Ego vehicle speed • V2V communication	• Test vehicle with logging equipment • Digital map data • RTK-GPS (maybe additional reference sensor, like laser-scanner) • Video data	

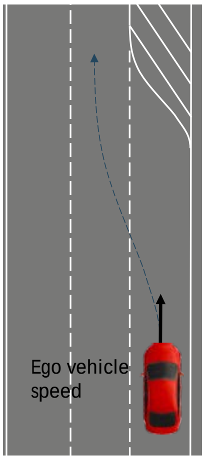
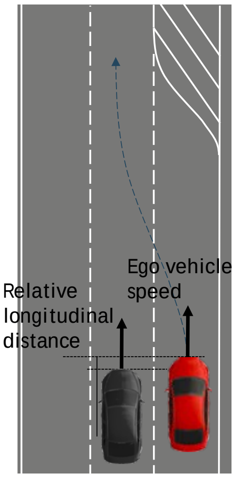
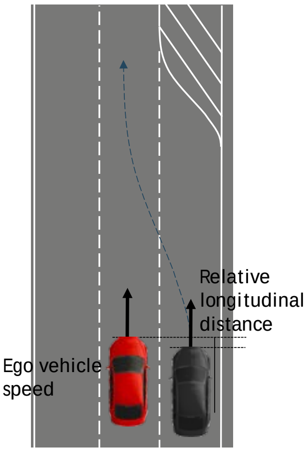
ID	Description	Situation Parameters	Equipment	Sketch
C7	**Car Following with Deceleration:** • All situations, in which the ego vehicle is following a leading vehicle that is slowing down with a deceleration of at least -1 ms-2.	• ego vehicle speed • Time headway ego vehicle (at start of deceleration) • Deceleration • V2V communication	• Test vehicle with logging equipment • Digital map data • RTK-GPS (maybe additional reference sensor, like laser-scanner) • Video data	
C8	**Approaching Object:** • All situations, in which the ego vehicle is approaching in the lane a slower vehicles.	• ego vehicle speed • Velocity of the predecessor vehicle • V2V communication	• Test vehicle with logging equipment • Digital map data • RTK-GPS (maybe additional reference sensor, like laser-scanner) • Video data	
C9	**New Speed Limit:** • All situation, in which the ego vehicle is driving in the lane with a specified velocity and approaching a new speed limit that is not equal to the driven speed .	• Ego vehicle Speed • Speed limit	• Test vehicle with logging equipment • Digital map data • RTK-GPS (maybe additional reference sensor, like laser-scanner) • Video data	

ID	Description	Situation Parameters	Equipment	Sketch
C10	**Lane Change:** ● All situations, in which the ego vehicle conducts a lane change	● Ego vehicle speed ● Direction of lane change (Right Lane, Left Lane)	● Test vehicle with logging equipment ● Digital map data ● RTK-GPS (maybe additional reference sensor, like laser-scanner) ● Video data	
C11	**Cut-In:** ● All situations, in which the ego vehicle is following a lane with a specified velocity while another vehicle is cutting into the ego-lane.	● Ego vehicle speed ● Distance to object, when it enters ego vehicle's lane ● Object speed ● V2V communication	● Test vehicle with logging equipment ● Digital map data ● RTK-GPS (maybe additional reference sensor, like laser-scanner) ● Video data	
C12	**Intersection - Right of Way Situation:** ● All situations, in which the ego vehicle is approaching an intersection with a right of way situation and another vehicle crossing from the right side. ● The ego vehicle has to correctly detect the other traffic participant which has right of way and letting this car pass the intersection. Afterwards, the ego vehicle has to pass the intersection.	● Ego vehicle speed ● Crossing vehicle speed ● (Time) Distance to crossing vehicle, when it enters the crossing ● V2V communication	● Test vehicle with logging equipment ● Digital map data ● RTK-GPS (maybe additional reference sensor, like laser-scanner) ● Video data	

ID	Description	Situation Parameters	Equipment	Sketch
C13	**Intersection - Right of Way Situation:** • All situations, in which the ego vehicle is approaching an intersection with a right of way situation and another vehicle crossing from the left side. • The ego vehicle has to correctly detect the other traffic participant which has right of way and letting this car pass the intersection. Afterwards, the ego vehicle has to pass the intersection.	• Ego vehicle speed • Crossing vehicle speed • (Time) Distance to crossing vehicle, when it enters the crossing • V2V communication	• Test vehicle with logging equipment • Digital map data • RTK-GPS (maybe additional reference sensor, like laser-scanner) • Video data	
C14	**Intersection - Right of Way Situation:** • All situations, in which the ego vehicle is approaching an intersection with a right of way situation and another vehicle crossing from the left side. • The ego vehicle has to correctly detect the other traffic participant which has right of way and letting this car pass the intersection. Afterwards, the ego vehicle has to pass the intersection.	• Ego vehicle speed • Crossing vehicle speed • (Time) Distance to crossing vehicle, when it enters the crossing	• Test vehicle with logging equipment • Digital map data • RTK-GPS (maybe additional reference sensor, like laser-scanner) • Video data	
C15	**Intersection - Left Turn:** • All situations, in which the ego vehicle is approaching an intersection and conducts a left turn without traffic.	• Ego vehicle speed	• Test vehicle with logging equipment • Digital map data • RTK-GPS (maybe additional reference sensor, like laser-scanner) • Video data	

ID	Description	Situation Parameters	Equipment	Sketch
C16	**Intersection - Left Turn with traffic:** ● All situations, in which the ego vehicle is approaches an intersection and conducts a left turn with oncoming traffic.	● ego vehicle speed ● Oncoming traffic speed ● Time distance to oncoming traffic at start of intersection ● V2V communication	● Test vehicle with logging equipment ● Digital map data ● RTK-GPS (maybe additional reference sensor, like laser-scanner) ● Video data	
C17	**Intersection -Traffic Lights:** ● All situations, in which the ego vehicle is approaching an intersection with traffic lights with the intention to cross the intersection. ● It has to correctly detect whether it has to stop or is allowed to cross the intersection.	● Ego vehicle speed ● Traffic light at certain positions	● Test vehicle with logging equipment ● Digital map data ● RTK-GPS (maybe additional reference sensor, like laser-scanner) ● Video data ● Logging of V2I communication	
C18	**Roundabout:** ● All situation, in which the ego vehicle is approaching a roundabout without traffic.	● Ego vehicle speed ● Radius roundabout	● Test vehicle with logging equipment ● Digital map data ● RTK-GPS (maybe additional reference sensor, like laser-scanner) ● Video data	

ID	Description	Situation Parameters	Equipment	Sketch
C19	**Roundabout with traffic:** • All situation, in which the ego vehicle is approaching a roundabout with traffic.	• Ego vehicle speed • Radius roundabout • Position of other vehicle at certain distance (if detectable) • V2V communication	• Test vehicle with logging equipment • Digital map data • RTK-GPS (maybe additional reference sensor, like laser-scanner) • Video data	
C20	**End of Lane:** • All situations, in which the ego-vehicle is approaching the end of a lane of a motorway without traffic. • The ego-vehicle has to decide at which distance to the end of the lane it has to conduct the lane change.	• Ego vehicle speed	• Test vehicle with logging equipment • Digital map data • RTK-GPS (maybe additional reference sensor, like laser-scanner) • Video data	
C21	**End of Lane with traffic:** • All situations, in which the ego-vehicle is approaching the end of a lane of a motorway without traffic. • The ego-vehicle has to decide at which distance to the end of the lane it has to conduct the lane change.	• Vehicle speed • Position of the other vehicle (relative longitudinal distance) • V2V communication	• Test vehicle with logging equipment • Digital map data • RTK-GPS (maybe additional reference sensor, like laser-scanner) • Video data	

ID	Description	Situation Parameters	Equipment	Sketch
C22	**Enter Motorway:** ● All situations, in which the ego-vehicle is approaching the end of a lane of a motorway without traffic. ● The ego-vehicle has to decide at which distance to the end of the lane it has to conduct the lane change.	● Ego vehicle speed	● Test vehicle with logging equipment ● Digital map data ● RTK-GPS (maybe additional reference sensor, like laser-scanner) ● Video data	
C23	**Enter Motorway with traffic:** ● All situations, in which the ego-vehicle is approaching the end of a lane of a motorway without traffic. ● The ego-vehicle has to decide at which distance to the end of the lane it has to conduct the lane change.	● Ego vehicle speed ● V2V communication	● Test vehicle with logging equipment ● Digital map data ● RTK-GPS (maybe additional reference sensor, like laser-scanner) ● Video data	
C24	**Merging on Motorway Entrance:** ● All situations, in which the ego-vehicle is on a motorway and another vehicle which is entering the motorway. ● The ego-vehicle has to consider the traffic vehicle which is entering the motorway. Therefore, the ego vehicle has to decide whether it has to slow down or to accelerate to enable a safe merging of the traffic vehicle.	● Vehicle speed ● Position of the other vehicle at the beginning of the entrance (relative longitudinal distance) ● V2V communication	● Test vehicle with logging equipment ● Digital map data ● RTK-GPS (maybe additional reference sensor, like laser-scanner) ● Video data	

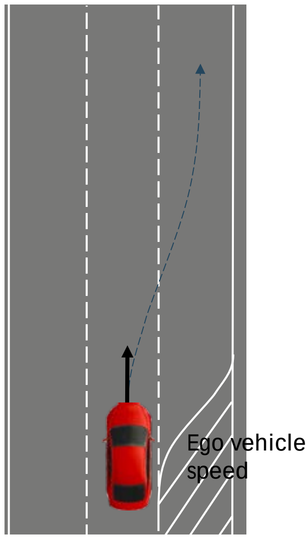
ID	Description	Situation Parameters	Equipment	Sketch
C25	**Exit Motorway:** ● All situations, in which the ego-vehicle is leaving a motorway without traffic. ● The function or system has to decide at which distance to the end of the lane it has to conduct the lane change manoeuvre in a safe manner.	● Vehicle speed	● Test vehicle with logging equipment ● Digital map data ● RTK-GPS (maybe additional reference sensor, like laser-scanner) ● Video data	

Figure 7.4: Overview on driving situations for continuous operating functions

Annex 3 AdaptIVe signal list

No.	Signal	Signal name	Description	Frequency [Hz]	Unit
100	Time (GPS-Time)	time (For targets: time_targetX, where X is the number of the target)	global time stamp with time of day	10	s
101	Date (GPS-Time)	date	date	-	YYYYMMDD
102	Time since start	time_start (For targets: time_start_targetX)	duration of measurement (start at test beginning)	10	s
103	Driven distance	driven_distance	driven distance in the test	10	m
Vehicle state					
200	Vehicle velocity	v (for target vehicle: v_targetX)	driven velocity of the vehicle	10	m/s
201	Longitudinal acceleration (a_x)	acc_long (for target vehicle: acc_long_targetX)	acceleration along the longitudinal axis of the vehicle (measured in COG of vehicle)	10	m/s²
202	Lateral acceleration (a_y)	acc_lat (for target vehicle: acc_lat_targetX)	acceleration along the lateral axis of the vehicle (measured in COG of vehicle)	10	m/s²
203	Yaw Rate	yaw_rate (for target vehicle: yaw_rate_targetX)	yaw rate of the vehicle (measured in COG of vehicle); positive are rotation against clockwise	10	°/s
204	Wheel Speed	v_wheel_fr, v_wheel_fl, v_wheel_rr, v_wheel_rl	wheel speed of all 4 wheels	10	m/s
205	Lateral velocity (v_y)	v_y	velocity along the lateral axis of the vehicle (measured in COG of the vehicle)	10	m/s
206	Lateral position in lane (left side)	dy_lane_l	distance to the left lane boundary measured from the mid of the vehicle	10	m
207	Lateral position in lane (right side)	dy_lane_r	distance to the right lane boundary measured from the mid of the vehicle	10	m
208	Heading angle in lane	heading_lane	heading direction of the vehicle in the lane	10	°
Driver commands					
301	Steering wheel angle	steering_wheel_angle	position of the steering wheel	10	°

No.	Signal	Signal name	Description	Frequency [Hz]	Unit
302	Steering wheel velocity	steering_wheel_v	angular velocity of the steering wheel	10	°/s
303	Steering torque	steering_torque	steering torque	10	Nm
304	Brake pedal position	brake_pedal	position of the brake pedal	10	%
305	Status Brake Light Switch	brake_light_status	status brake light switch	10	0/1
306	Brake pressure	brake_pressure	brake pressure	10	bar
307	Accelerator pedal position	accelerator_pedal	position of the accelerator pedal	10	%
308	Driver activation button	Driver_activation_button	status of driver activation button (generic button or other device to activate the function)	10	-
309	Gear	gear	driven gear	10	-
310	Direction indicator	indicator_l ; indicator_r	status of direction indicator.	10	0/1
311	ACC Set Speed	acc_set_speed	set speed of the adaptive cruise control or the cruise control (if available in the demonstrator)	10	m/s
312	ACC Set Headway	acc_set_headway	time headway setting of the ACC	10	s or System
Detection of all tracked objects (static and moving)					
401	Longitudinal range to object	Object_dx_IDXX	range towards object in longitudinal direction. Measured by vehicle sensors. Referenced to vehicle coordinate system. (signals have to be provided for each detected target)	10	m
402	Lateral range to object	Object_dy_IDXX	range towards object in lateral direction. Measured by vehicle sensors. Referenced to vehicle coordinate system. (signals have to be provided for each detected target)	10	m
403	Longitudinal range rate to object	Object_dvx_IDXX	range rate towards object in longitudinal direction. Measured by vehicle sensors. Referenced to vehicle coordinate system. (signals have to be provided for each detected target)	10	m/s

No.	Signal	Signal name	Description	Frequency [Hz]	Unit
404	Lateral range rate to object	Object_dvy_IDXX	range rate object in lateral direction. Measured by vehicle sensors. Referenced to vehicle coordinate system. (signals have to be provided for each detected target)	10	m /s
405	Angular displacement to object	Object_dpsi_IDXX	angular (heading) displacement of longitudinal axes to object. Referenced to vehicle coordinate system. Measured by the vehicle sensor (signals have to be provided for each detected target)	10	deg
406	Reference Point at object	Object_RefPoint_IDXX	Reference Point at object	10	-
407	ID object	Object_IDXX	ID of object	10	-
408	Longitudinal acceleration of object	Object_ax_IDXX	acceleration of object in longitudinal direction. Measured by vehicle sensors.	10	m/s²
409	Lateral acceleration of object	Object_ay_IDXX	acceleration of object in longitudinal direction. Measured by vehicle sensors.	10	m /s²
410	Object classification	Object_type_IDXX	type of front object	10	-
GPS Position					
501	GPS Position (Latitude, Longitudinal)	GPS_lat, GPS_long (For targets: GPS_lat_targetX, GPS_long_targetX)	position measured by the GPS. If a GPS based system (DGPS RTK-GPS) is used as a reference system, a high accuracy is required.	10	m
502	GPS Position (Latitude, Longitudinal)	GPS_lat_deg, GPS_long_deg (For targets: GPS_lat_deg_targetX, GPS_long_deg_targetX)	position measured by the GPS. If a GPS based system (DGPS RTK-GPS) is used as a reference system, a high accuracy is required.	10	°
503	GPS Altitude	GPS_altitude (For targets: GPS_altitude_targetX)	altitude measured by GPS	10	m
504	GPS Velocity	GPS_vel (For targets: GPS_vel_targetX)	speed of the GPS	10	m/s
505	Heading angle (GPS)	GPS_heading (For targets: GPS_heading_targetX)	track angle of the GPS; with respect to "north line"	10	°
506	Dilution of Precision	GPS_dop (For targets:	quality of GPS Signal	10	-

No.	Signal	Signal name	Description	Frequency [Hz]	Unit
		GPS_dop_targetX)			
507	Number of Satellites	GPS_Satellites	Number of satellites	10	-
Environment information & static objects					
601	Speed limit of current road section	speed_limit	speed limit of current road section	10	km/h
602	Curve Radius of current road section	r_curve	curve Radius of current road section	10	m
603	obstacle (curve/intersection/roadwork/hill)	obstacle_dx	distance to next curve/intersection	10	m
604	Next obstacle classification	obstacle_type	type of next obstacle	10	0/1
605	Distance to next speed limit	speed_limit_dx	distance to next speed limit	10	m
606	Speed limit of next road section/speed limit detected by camera	speed_limit_next	next speed limit	10	km/h
607	Next speed limit source	speed_limit_source	sensor (camera or map) used to determine the next speed limit	10	km/h
608	Status overtaking prohibitions	overtaking_prohibiton_status	Are there any overtaking prohibitions given by lane markings or signs?	10	0/1
609	road or lane curvature	lane_curvature	1 / Road radius respectively 1/ lane radius	10	1/m
Engine data					
701	Fuel consumption	fuel_consumption	current fuel consumption of the engine	10	l/100km
702	Engine Torque	engine_torque	Engine Torque at engine	10	Nm
703	Engine Speed	engine_speed	rotation speed of the engine	10	rpm

No.	Signal	Signal name	Description	Frequency [Hz]	Unit
Function status					
801	Function status	xxx_status	status of each functions (xxx: function name)	10	0 / 1 / 2
802	Function warning status	xxx_warning_status	Describes warning status of the function (xxx: function name, if more than one function is integrated in the demonstrator, each function should have a warning signal)	10	0/1
803	Function warning type	xxx_warning_type	Warning type (only relevant if warnings are issued in different way, to specified by VSP)	10	0/1
804	Function warning level	xxx_warning_level	Different warning levels (only relevant if different warning levels are available)	10	0/1
805	Status ESC	esc_status	status information of the ESC	10	0/1
806	Status ABS	abs_status	status information of the ABS	10	0/1
807	Status Brake assistant	brake_assistant_status	status information about the brake assistant	10	0 / 1
808	Applied brake force	extra_brake_force	extra applied brake force or pressure	10	%
809	Applied steering torque	extra_steering_torque	applied steering torque by function	10	Nm
810	Set speed	Set_speed	Speed, which is set / recommended by the function	10	m/s
Lane information					
901	lane number	lane_number_current	Number of lane, in which the vehicle drives.	10	-
902	lane direction	lane_direction	Driving direction of the lane, in which the vehicle drives	10	-
903	number of lanes	lane_total_number	Number of lanes of the road	10	-
904	Type lane marking (left)	lane_marking_type_l	detected lane type (left)	10	-
905	Type lane marking (right)	lane_marking_type_r	detected lane type (right)	10	-

No.	Signal	Signal name	Description	Frequency [Hz]	Unit
906	Left lane status	lane_status_l	Calculated status of the right lane, indicating possibility to steer into left lane	10	
907	Right lane status	lane_status_r	Calculated status of the left lane availability, indicating possibility to steer into right lane	10	
908	Lane width	lane_width	Road width of the driven road	10	m
Video & V2X Communication					
1001	Video data (TBC)	To be specified	Video showing surrounding traffic	?	-
1002	Video data of driver (TBC)	To be specified	Video showing driver	?	-
1003	Eye movements (TBC)	To be specified	Position and facing direction of driver's eyes	?	-
1004	Head position (TBC)	To be specified	Orientation of driver's head	?	-
1005	Status V2I communication	v2x_status (Comment: I think that should be v2i_status)	Status V2X communication (need only be stored, if the demonstrator uses V2X communication)	10	0/1
1006	V2I communication messages	v2i_messages	Status V2I communication messages (need only be stored, if the demonstrator uses V2I communication)		
1007	Status V2V communication	v2v_status	Status V2V communication (need only be stored, if the demonstrator uses V2V communication)	10	0/1
1008	V2V communication messages	v2v messages	Status V2V communication messages (need only be stored, if the demonstrator uses V2V communication)		
Additional signals (TBD)					
1100	All kinds of driver actions and reactions	TBD	TBD	TBD	TBD

Figure 7.5: AdaptIVe signal list

Annex 4 Test distances for relevant situations (motorway)

It needs to be taken account that the chosen test route might also have an influence on the occurrence of certain events (e.g. new speed limit).

ID	Driving Situation	Cluster	Mean frequency [km^{-1}]	Test distance [km]			
				k = 5	k = 10	k = 20	k = 30
C5	Const	0 km h^{-1} < v < 40 km h^{-1}	0,013914596	760	1220	2090	2925
		40 km h^{-1} < v < 60 km h^{-1}	0,020172132	525	845	1445	2020
		60 km h^{-1} < v < 90 km h^{-1}	0,13987127	80	125	210	295
		90 km h^{-1} < v < 110 km h^{-1}	0,21916875	50	80	135	190
		110 km h^{-1} < v < 130 km h^{-1}	0,247285323	45	70	120	165
		130 km h^{-1} < v	0,194297223	55	90	150	210
C6	Following	0.0 s < THW < 0.9 s	0,13177519	80	130	225	310
		0.9 s < THW < 1.8 s	0,252261564	45	70	120	165
		1.8 s < THW < 2.7 s	0,142820058	75	120	205	285
		2.7 s < THW < 3.6 s	0,073557667	145	235	400	555
		3.6 s < THW < 5.0 s	0,056640449	190	300	515	720
		5.0 < THW	0,333989138	35	55	90	125
C7	Following with deceleration	0.0 s < THW < 0.9 s	0,014772546	715	1150	1970	2755
		0.9 s < THW < 1.8 s	0,016962106	620	1005	1715	2400
		1.8 s < THW < 2.7 s	0,007089554	1485	2395	4100	5740
		2.7 s < THW < 3.6 s	0,003031374	3470	5600	9590	13425
		3.6 s < THW < 5.0 s	0,001879449	-	-	-	-
		5.0 < THW	0,016852257	625	1010	1725	2415
C5	Curve	0 m < R < 10 m	0,000286462	-	-	-	-
		10 m < R < 50 m	0,001136771	-	-	-	-
		50 m < R < 100 m	0,001424293	-	-	-	-
		100 m < R < 250 m	0,014623126	720	1160	1990	2785
		250 m < R < 500 m	0,071650069	150	240	410	570
		1000 m < R < 1500 m	0,378609944	30	45	80	110
C10	Lane change	right -> left	0,160105163	70	110	185	255
		left -> right	0,148375303	75	115	200	275
C8	Critical situation	0.0 s < TTC < 1.0 s	1,03347E-05	-	-	-	-
		1.0 s < TTC < 2.0 s	0,000128219	-	-	-	-
		2.0 s < TTC < 3.0 s	0,000674936	-	-	-	-

ID	Driving Situation	Cluster	Mean frequency [km^{-1}]	Test distance [km]			
				k = 5	k = 10	k = 20	k = 30
		3.0 s < TTC < 4.0 s	0,001747435	-	-	-	-
		5.0 s < TTC < 6.0 s	0	-	-	-	-
		6.0 s < TTC	0	-	-	-	-
	Speeding	0 % < Speeding < 10 %	0,097224537	110	175	300	420
		10 % < Speeding < 20 %	0,055676757	190	305	525	735
		20 % < Speeding < 30 %	0,022478642	470	755	1295	1815
		30 % < Speeding	0,027686854	380	615	1050	1470
C22 - C24	**Motorway Ramp**	Entrance/Exit-ramp	0,25	45	70	120	165
C9	**New speed limit**	No to High	0,633739505	20	30	50	65
		High to Low	0,198180317	55	90	150	210
		Low to High	0,116781858	95	150	250	350

Figure 7.6: Test distances for relevant situations on motorway

Annex 5 Evaluation indicators for user-related assessment

The evaluation indicators related to each individual hypothesis (HUA1-HUA30) are presented below:

HUA1 - The system gives the expected user-related outcome. The relevant indicators to test this hypothesis are specific for the respective systems/functions and they reflect the outcome in the scenario the system is designed for. These indicators for the specific systems/functions are as follows:

Indicators for testing hypothesis HUA1 for the specific AdaptIVe systems/functions are:

System/Function	Indicator
Close-distance scenarios	
Park Assistant	Position in parking space, time for parking manoeuvre
Construction Site Manoeuvre	Speed and side distance at construction site
Automated Parking Garage Pilot	Position in parking space, time for parking manoeuvre
Urban scenarios	
Lane following	Lane position: distribution, mean, stddev.
Speed adaptation	Properly adapted speed to the situation
Vehicle following in lane	Following distance: distribution, mean, stddev.
Obstacle or VRU on the road	speed at and side distance to obstacle or VRU
Lane change	Safe and lawful lane change
Intersection handling	Safe and lawful passage of intersection
Urban roundabouts handling	Safe and lawful passage of roundabout
Traffic light handling	Safe and lawful passage of traffic light
Highway scenarios	
Lane following	Lane position: distribution, mean, stddev.
Lane Change (and overtaking)	Safe and lawful lane change and overtaking
Stop & Go Driving	Following distance: distribution, mean, stddev.
Speed / time gap adaptation at a motorway entry ramp	Speed and accepted gap: distribution, mean, stddev.
Cooperative merging with speed adaptation	Safe and lawful merging; speed difference, accepted gap, distance forward and back: distribution, mean, stddev.
Enter and exit of a motorway	Safe and lawful enter/exit; speed difference, accepted gap, distance forward and back: distribution, mean, stddev.
Cooperative merging with lane change	Safe and lawful merging; speed difference, accepted gap, distance forward and back: distribution, mean, stddev.

Figure 7.7: Indicators for testing hypothesis HUA1 for the specific AdaptIVe systems/functions.

HUA2 - The drivers use the system as intended to be used.

Indicators: driving speed, adaptation of speed to potentially critical situations, the frequency and duration of being in an “unsafe state”, the distance to the vehicle ahead, lane choice, lane change, lane keeping, overtaking, stopping, yielding, behaviour at traffic lights, interaction and communication with other road users.

HUA3 - The drivers use the function/system in all situations for which it is available.

Indicator: usage of system in percent of total driving time during relevant situations.

HUA4 - The drivers stay in the function/system settings suggested by the system during the test drives.

Indicator: driving in suggested function/system settings in percent of total time of a certain suggested function/system settings.

HUA5 - Driver behaviour does not differ when driving with a well-functioning driving automation function/system from driving behaviour without automation.

Indicators: driving speed, adaptation of speed to potentially critical situations, the frequency and duration of being in an “unsafe state”, the distance to the vehicle ahead, lane choice, lane change, lane keeping, overtaking, stopping, yielding, behaviour at traffic lights, interaction and communication with other road users.

HUA6 - There are no long-term changes in driver behaviour when driving with automation.

Indicators: driving speed, adaptation of speed to potentially critical situations, the frequency and duration of being in an “unsafe state”, the distance to the vehicle ahead, lane choice, lane change, lane keeping, overtaking, stopping, yielding, behaviour at traffic lights, interaction and communication with other road users.

HUA7 - The drivers' situational awareness is not affected by the system.

Indicator: SAGAT scores

HUA8 - Driver stress is not affected by automation.

Indicators: heart rate measures and Short Stress State Questionnaire (SSSQ) scores [78].

HUA9 - The mental workload of the driver is not affected by with automation.

Indicators: Subjective rate of the Raw Task Load indeX (RTLX) [79].

HUA10 – The mental workload does not change after prolonged driving with the system.

Indicators: Change in the subjective rate of the Raw Task Load indeX (RTLX [79].

HUA11 - Transfer of control is not affected by mental workload.

Indicator: Time for the driver to make decision of transfer of control, Mean and minimum values of speed and their Standard Deviation, Standard Deviation of Lane Position (SDLP), number of 1° steering reversals per minute, High Frequency Control of steering (in the 0.3–0.6 Hz band) and visual attention measured by eye tracking value of ‘Percent Road Centre’ (PRC)

HUA12 - The drivers do not engage more in secondary tasks when driving with automation compared to driving without automation.

Indicator: Percent of driving time the driver being engaged in secondary task.

HUA13 - The drivers do not become complacent when driving with automation.

Indicators: Task Load indeX (RTLX) [79]., Task-related Boredom Scale (TBS) [80],the probability of detection of automation failure, reaction time for detection of automation failure, and the number of detection errors and Root-Mean-Squared-Error (RMSE) of secondary task

HUA14 - The time for the drivers to make decision after a safety critical event does not differ between manual driving mode and automated driving.

Indicator: The time from a safety critical event arises until the driver takes an action.

HUA15 - Driving skills don’t degrade with time using automation.

Indicators: standard deviation of speed, adaptation of speed to potentially critical situations, the frequency and duration of being in an “unsafe state”, the distance to the vehicle ahead, lane choice, lane change, lane keeping, standard deviation of the lateral position, overtaking.

HUA16 - There is no change in the drivers' take-over behaviour in long term.

Indicators: driver reaction type and reaction time in a take-over situation.

HUA17 - The drivers do detect automation failures.

Indicator: the share of registered automation failures.

HUA18 - The drivers do not fail to respond to a critical situation because the system failed to notify them.

Indicator: The number of driver responses to critical situations related to all situations the system did not notified them.

HUA19 - The drivers take the appropriate measure after a system brake down.

Indicator: driver reaction type to a system brake down.

HUA20 - The drivers do not follow a wrong recommendation instead of vigilant information seeking and processing.

Indicator: driver reaction type after a wrong recommendation in a critical situation.

HUA21 - The drivers are confident about the correctness of their decision after a system brake down.

Indicator: questionnaire answer

HUA22 - There is no difference in intervention time between drivers with an internal locus of control and those with an external locus of control.

Indicator: the time from a safety critical event arises until the driver takes an action.

HUA23 - The drivers have the correct mental representation of the system.

Indicator: questionnaire answer

HUA24 - The drivers have no over- or under-trust on the system.

Indicator: scores on the self-report scale of trust [40].

HUA25 - The drivers experience automated driving as an improvement in their driving.

Indicator: questionnaire answers

HUA26 - The drivers have their distinct opinion about the system.

Indicator: questionnaire answers

HUA27 - The drivers find the system useful and satisfactory.

Indicator: usefulness and satisfaction scale [81].

HUA28 - Automation failures do not influence the drivers' attitude to the system.

Indicator: questionnaire answers

HUA29 - The drivers are interested to have and to pay for the system.

Indicator: questionnaire answers

HUA30 - Non-users' behaviour is not influenced by interaction with equipped vehicles.

Indicators: behavioural indicators of non-users, such as: driving speed, adaptation of speed to potentially critical situations, the frequency and duration of being in an "unsafe state", the distance to the vehicle ahead, lane choice, lane change, lane keeping, overtaking, stopping, yielding, behaviour at traffic lights, interaction and communication with other road users.

Annex 6 User-related Assessment - alternative study designs

The aim of this document is to present alternative evaluation set-ups for user-related assessment in AdaptIVe. The internal report i-7 “Draft Test and Evaluation Plan” of AdaptIVe presents a comprehensive “ideal” evaluation set-up, offering to test all relevant user-related issues of automated driving. It is understood that carrying out all of them is resource and time demanding, hence the set-up of the final evaluation plan will probably confine to the most rewarding ones.

This document aims at presenting alternative set-ups for user-related assessment if “ideal” conditions for assessment cannot be achieved. For details of methods and study design, please refer to the internal report i-7. A more detailed test planning by adaptation of the general approach to the system respectively function under study will start once the relevant function has been selected.

A comprehensive “ideal” evaluation set-up for user-related assessment

The “ideal” set-up for the user related assessment in AdaptIVe is tests in a naturalistic driving environment (real traffic) with naïve (normal) test drivers. Observation of driver behaviour in real traffic gives the highest validity of results, while a driver simulator experiment allows for staging situations where also situational awareness and possible complacency can be studied. For details of methods and study design, please refer to the Internal report i-7.

Tests in a naturalistic driving environment (real traffic) with naïve (normal) test drivers

Participants (20-30) should drive twice along a test route of approximately 40-50 km, consisting of roads relevant for the system/function to be tested (within-subject design). The sample of test drivers should be representative for the driver population (by gender and age). In case of naïve (normal) test drivers are not allowed to drive the test vehicle, a second best alternative is letting employees (with administrative duties in the company) act as test drivers.

In case of relevant equipment (e.g. eye tracking device) is not available to be employed for the methods presented below, the aspect in question can be excluded from the tests.

The following methods to be employed:

1. Driving data as well as driver- and system generated events are logged during both riding sessions.
2. Behavioural observations are carried out by observers in the car riding along during both riding sessions.

3. Driver performance is measured through assessments of drivers' attention to potential hazards (i.e. detection accuracy), accuracy of vehicle control (i.e. variability in lateral position) and variations in mean speed.
4. Measuring the driver's ability to resume control of driving (mean and minimum values of speed and their Standard deviation, Standard Deviation of Lane Position (SDLP), number of 1° steering reversals per minute, High Frequency Control of steering (in the 0.3–0.6 Hz band), visual attention measured by eye tracking value of 'Percent Road Centre' and time to stabilised control of these variables). **Access to Eye tracking device is needed!**
5. Mental workload of the driver is assessed with the help of the Raw Task Load indeX (RTLX) after both rides.
6. The driver's subjective stress is assessed by the Short Stress State Questionnaire (SSSQ) after both rides.
7. The driver's perceived boredom is assessed by the Task-related Boredom Scale (TBS) after both rides.
8. The driver's subjective fatigue state is assessed by the fatigue scale questionnaire after both rides.
9. The driver's understanding of the limitations of the system is assessed by questions after the second ride.
10. Actual trust in the system is assessed using a six-item self-report scale after the ride with the system ON.
11. The driver's perception of the system, its usability is assessed by the System Usability Scale (SUS) after the ride with the system ON.
12. Usefulness and Satisfaction is assessed after the ride with the system ON.
13. The test drivers are asked to answer questions concerning experienced effects of the system, perceived benefits with the system and willingness to have and pay for the system after the second ride.

Downscaled set-ups for user-related assessment

Down-scaled setups are presented below representing different effort levels (from the highest to the lowest i.e. 1) Test driving by naïve (normal) test drivers on a test track with staged scenarios; 2) Familiarising driving by naïve (normal) test drivers on a test track; 3) Demonstration of the function/system for an acceptance study.

SP7 aims to conduct (in case the ideal set up is not possible) option 1 with a selected set of methods and not all described methods. Option 2 and 3 are considered as fall back solution.

Test driving by naïve (normal) test drivers on a test track with staged scenarios

In case of the demonstrator vehicle cannot be driven on public roads and there is no driving simulator available, some tests may be carried out on a test track. Participants (20-30) should complete two 30-40 min driving sessions on a test track simulating road environments relevant for the system/function to be tested (within-subject design). The sample of test drivers should be representative for the driver population (by gender and age). Test scenarios to assess user-related effects for each of the functionalities under study should be staged in a continuous way for offering continuous interaction in a mixed way.

In case of relevant equipment (e.g. eye tracking device) is not available to be employed for the methods presented below, **the aspect in question can be excluded from the tests**.

The following methods to be employed:

1. Driving data as well as driver- and system generated events are logged during both riding sessions.
2. Behavioural observations are carried out by observers in the car riding along during both riding sessions.
3. Driver performance is measured through assessments of drivers' attention to potential hazards (i.e. detection accuracy), accuracy of vehicle control (i.e. variability in lateral position) and variations in mean speed.
4. Measuring the driver's ability to resume control of driving (mean and minimum values of speed and their Standard deviation, Standard Deviation of Lane Position (SDLP), number of 1° steering reversals per minute, High Frequency Control of steering (in the 0.3–0.6 Hz band), visual attention measured by eye tracking value of 'Percent Road Centre' and time to stabilised control of these variables). **Access to Eye tracking device is needed!**
5. Mental workload of the driver is assessed with the help of the Raw Task Load indeX (RTLX) after both rides.
6. The driver's subjective stress is assessed by the Short Stress State Questionnaire (SSSQ) after both rides.
7. The driver's subjective fatigue state is assessed by the fatigue scale questionnaire after both rides.
8. The driver's understanding of the limitations of the system is assessed by questions after the second ride.
9. Actual trust in the system is assessed using a six-item self-report scale after the ride with the system ON.
10. The driver's perception of the system, its usability is assessed by the System Usability Scale (SUS) after the ride with the system ON.

11. Usefulness and Satisfaction is assessed after the ride with the system ON.
12. The test drivers are asked to answer questions concerning experienced effects of the system, perceived benefits with the system and willingness to have and pay for the system after the second ride.

"Familiarising" drive by naïve (normal) test drivers on a test track

In case of test scenarios, relevant to assess the user related effects cannot be staged, the test drivers can drive the demonstrator vehicle just for orientate themselves about the functions of the system.

The following methods to be employed:

1. The driver's perception of the system, its usability is assessed by the System Usability Scale.
2. Usefulness and Satisfaction is assessed after the ride with the system ON.
3. The test drivers are asked to answer questions concerning experienced effects of the system, perceived benefits with the system and willingness to have and pay for the system after the second ride.

Demonstration of the function/system for an acceptance study

In case of naïve (normal) test drivers are not allowed to drive the demonstrator vehicle, participants from the public are demonstrated the system by a company driver, or if driving is not possible, the function is presented from video and an oral description. The assessment is, then limited to studying perceived advantages, disadvantages, usefulness, trust, acceptance, as well as willingness to have and pay for the system. After the demonstration a focus group session is carried out and an individual questionnaire is employed concerning perceived benefits with the system and willingness to have and pay for the system after the second ride.

Driving simulator experiment with naïve (normal) test drivers

In case of a demonstrator vehicle is not available for user-related tests, as an alternative, tests in a driving simulator can be carried out. However, this option is only considered by SP7 as the last option.

Participants (20-30) should drive twice along a simulated test route of appr. 40-50 km, consisting of roads relevant for the system/function to be tested (within-subject design). The sample of test drivers should be representative for the driver population (by gender and age). Test scenarios to assess user related effects for each of the functionalities under study should be staged at least 6 times per test drive in a mixed way.

In case of relevant equipment (e.g. eye tracking device, heart rate measuring equipment) is not available to be employed for the methods presented below, the aspect in question can be excluded from the tests.

The following methods to be employed:

1. Driving data as well as driver- and system generated events are logged during both riding sessions.
2. Driver performance is measured through assessments of drivers' attention to potential hazards (i.e. detection accuracy), accuracy of vehicle control (i.e. variability in lateral position) and variations in mean speed.
3. Measuring the driver's ability to resume control of driving (mean and minimum values of speed and their Standard deviation, Standard Deviation of Lane Position (SDLP), number of 1° steering reversals per minute, High Frequency Control of steering (in the 0.3–0.6 Hz band), visual attention measured by eye tracking value of 'Percent Road Centre' and time to stabilised control of these variables). **Access to Eye tracking device is needed!**
4. Mental workload of the driver is assessed with the help of the Raw Task Load indeX (RTLX) after both rides.
5. The driver's subjective stress is assessed by heart rate measurements during both rides. **Requires heart rate measuring equipment!**
6. The driver's subjective stress is assessed by the Short Stress State Questionnaire (SSSQ) after both rides.
7. The driver's perceived boredom is assessed by the Task-related Boredom Scale (TBS) after both rides.
8. The driver's fatigue is assessed by variables, such as the standard deviation of speed (SDS), standard deviation of the lateral position (SDLP), frequency of extremely large steering wheel movement (SWM) (>N10°), frequency of line crossings and reaction time (RT) to be used during both rides.
9. The driver's subjective fatigue state is assessed by the fatigue scale questionnaire after both rides.
10. The driver's situational awareness is assessed by the SAGAT method during both rides.
11. The driver's out-of-the-loop performance is measured by a simulated system brake down. The dependent variables are: situational awareness, the decision selected, time for the drivers to make decision with a simulated system brake down, drivers' confidence about the correctness of decision made and mental workload.
12. For the complacency study should be carried out during two riding sessions with the system ON and in a multi-task environment. The RTLX and TBS are part of this study as dependent variables as well as system monitoring performance and Secondary task Root-Mean-

Squared-Error (RMSE). The performance measures for the system-monitoring task are: (a) the probability of detection of automation failure, (b) reaction time for detection, and (c) the number of detection errors.

13. The driver's understanding of the limitations of the system is assessed by questions after the second ride.
14. Actual trust in the system is assessed using a six-item self-report scale after the ride with the system ON.
15. The driver's perception of the system, its usability is assessed by the System Usability Scale (SUS) after the ride with the system ON.
16. Usefulness and Satisfaction is assessed after the ride with the system ON.

The test drivers are asked to answer questions concerning experienced effects of the system, perceived benefits with the system and willingness to have and pay for the system after the second ride.