

A Comprehensive Evaluation Approach for Highly Automated Driving

Christian Roesener, Jan Sauerbier

Institut für Kraftfahrzeuge, RWTH Aachen University, Germany

Adrian Zlocki

Forschungsgesellschaft Kraftfahrwesen mbH, Aachen, Germany

Felix Fahrenkrog, Lei Wang

BMW Group, Germany

András Várhelyi

Lund University, Sweden

Erwin de Gelder, Joris Dufils

TNO, Netherlands

Sandra Breunig

German Federal Highway Research Institute – BASt, Germany

Pablo Mejuto

C.T.A.G. - Galician Automotive Technologic Centre, Spain

Fabio Tango, Jessica Lanati

Centro Ricerche Fiat, Italy

Paper Number 17-0259

ABSTRACT

Since the last decade, development efforts by academia and industry for automated driving functions have increased significantly. Also, the European research project AdaptIVe is looking into this topic. Due to the large operation spaces and various complex situations that are covered by these functions, efforts for evaluation increase also significantly. Within AdaptIVe, a comprehensive evaluation approach for automated driving functions ranging from SAE level 2-4 has been developed [1]. The approach splits the evaluation into technical, user-related, in-traffic and impact assessment addressing safety and environmental effects of automated driving. For each evaluation type appropriate test tools and methods are selected e.g. field test for technical assessment, trials on test track and in real traffic for the user-related assessments and simulations for the in-traffic and impact assessment. Next to the assessment type also the characteristics of the function must be considered when deciding for specific test tools. Hence, besides to the level of automation [8] the automated driving functions are classified into continuous and event-based operating functions. Whereas event-based operating functions are only operating for a short period in time (e.g. automated parking), continuous operating functions are, once they are active, operating for longer time periods (e.g. highway automation). Based on the classification the aspects to be evaluated and test methods are selected for all assessment types. The developed methodology has been applied to several automated driving functions developed within AdaptIVe. As an example, for the technical assessment of continuous operating functions it has been assessed whether the driving behavior of the developed functions is similar to human driving behavior and therefore not disturbing human traffic. In the user-related assessment, issues related to driver behavior, understanding of automation, trust, mental workload, resuming control, vigilance, usability and acceptance has been looked at. In this paper the key aspects of the AdaptIVe evaluation methodology for technical, user-related, in-traffic and impact assessment are presented as well as the key results of the application of this methodology on the within AdaptIVe developed automated driving functions.

INTRODUCTION

Automated driving is a vision since the early 20th century. A first step towards this vision was the introduction of ADAS (advanced driver assistance systems) in the last decade of the 20th century. Following the successful introduction of ADAS research on higher automated driving functions is ongoing since many years. These functions were intensively investigated and demonstrated during the DARPA Challenges [2] [3] as well as activities of Google and their so called Google self-driving cars [4] in the US and in Europe by the Berta Benz Drive [5] and the GCDC [6]. This chosen path is continued by the European research project AdaptIVe. Within AdaptIVe, 21 different automated driving functions for different speed ranges and target areas are developed [7].

However, with the increasing complexity of ADAS and automated driving functions, assessment efforts are expected to rise dramatically as stated in [8]. Therefore new assessment methods which are enabling an efficient assessment of these functions have to be designed. Besides new methods and frameworks for assessment, metrics for measuring the performance of automated driving functions have to be identified as well.

AIM

The aim of this paper to present the key aspects and results of the evaluation approach [1] developed within the European research project AdaptIVe, which feature a comprehensive evaluation in the areas of user-related, technical, in-traffic and impact assessment. All considered evaluation areas are presented in figure 1.

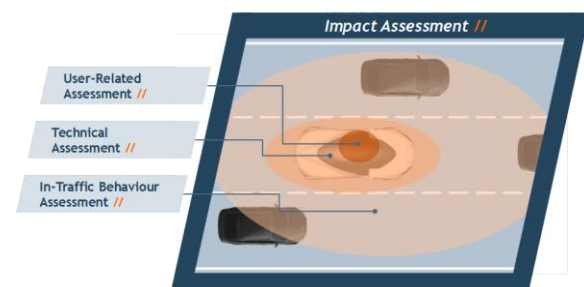


Figure 1. Evaluation areas in AdaptIVe

METHODOLOGY

As described previously, different aspects are analysed in the several evaluation areas. In the technical assessment the performance of the functions is investigated. The user-related

assessment analyses the interaction between the functions and the user, trust, usability, as well as 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.

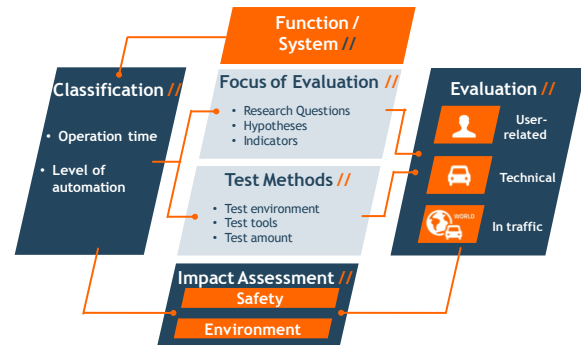


Figure 2. Overall AdaptIVe evaluation approach

The initial starting point for the evaluation is a detailed description of the function or system 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.

In the first step, the AdaptIVe functions and systems are classified according to the SAE classification [9] and the automation level they address. 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:

- **Event-based operating functions:** Functions that operate only for a short period of time (seconds up to few minutes). Typical examples are automated parking functions and minimum risk manoeuvre function
- **Continuous operating functions:** 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.

Based on the classification it is decided on the focus of the evaluation and test methods to be applied. 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 will be used. By using already existing test tools the evaluation approach is enabling an efficient assessment of the developed automated driving functions.

Technical assessment

The objective of the technical assessment is the evaluation of the performance of automated driving functions. While the assessment frameworks developed in previous European projects, e.g. PReVAL [10], eIMPACT [11], interactIVe [12] and others dealt mainly with active safety functions or respectively ADAS, where the assessment focused mainly on testing of the functions in predefined use cases, the approach for continuous automated driving functions has to be different. Because contrary to active safety functions continuous automated driving functions are active for longer time periods and are operating in a huge variety of scenarios. Following these requirements existing assessment approaches have to be extended in order to ensure that the whole situation space which is addressed by the function is covered. A major challenge within this assessment is to limit the test effort to a feasible amount while ensuring that all important aspects are covered. Since automated driving systems address the whole driving process, nearly all driving situations are relevant for this assessment. It might be desirable to test the function's behaviour in a high number of driving situations and in different variations of these situations. Considering the limited resources within the assessment, this is hardly feasible. Therefore, a prioritization of the test approach within the technical assessment is required. As already mentioned previously, the automated driving functions need to be distinguished in:

- Event-based operating functions
- Continuously operating functions

For continuous automated driving functions a so-called "scenario-based assessment" is used for

assessment. Instead of defining single test cases a (small) field test is conducted for assessing the automated driving functions. During the field test the function must be able to handle driving situations that are covered according to the function's specification and occur during the test drive. Afterwards, the driving data is clustered into relevant driving scenarios in which the functions are assessed. The functions are assessed by analysing two aspects:

- Change of frequency of relevant driving scenarios compared to reference behaviour
- Change of performance of automated driving functions in driving scenarios compared to reference performance

In order to investigate the performance in the defined driving scenarios adequate indicators are to be defined. 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 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. 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, data from euroFOT [13] has been used in AdaptIVe.

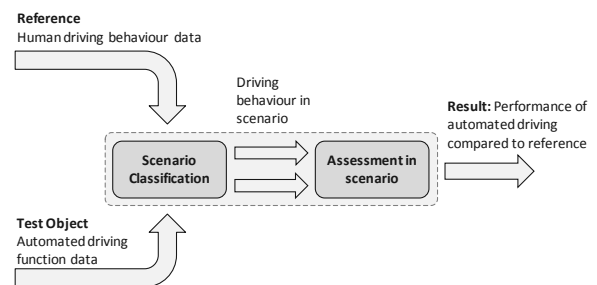


Figure 3. Method for technical assessment

For identification and classification of the defined driving scenarios, feature-based machine learning approaches [14, 15] are used for detection of the defined driving scenarios. The methodology developed is shown in figure 3.

As presented already in the previous section, the assessment is done by analysing the change in frequency of the occurrence of defined scenarios and the change in performance with and without automated driving functions in defined scenarios. Therefore, these scenarios have to be defined before the assessment, see table 1.

Table 1.
Driving scenarios for technical assessment

Scenario	Semantic description
Free driving/ Vehicle following	No predecessor, ego vehicle keeps lane
	Ego vehicle's intention is to keep the lane and is influenced by a predecessor vehicle
Lane change	Ego vehicle's intention is to change to the next lane
Cut-In	Another vehicle intends to merge into the lane of the ego vehicle

In order to ensure that all relevant driving situations occur during the test, the duration of the dedicated small field test is estimated by means of data from previous FOT, such as euroFOT [13]. The reference data of the field operational test is clustered in relevant driving situations by using a situation space approach. Afterwards, the distribution of frequencies of all relevant driving situations is calculated. In accordance to [8] a Poisson distribution for the occurrence of driving situations is assumed and the minimal test length for the occurrence of at least $k = 5$ driving situations is calculated. After classification of the relevant driving scenarios the predefined hypotheses can be tested. 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 probability for the occurrence of a driving situation is given by:

$$P = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{Equation (1)}$$

Where:

- P Probability
- λ Expectation value
- k Number of events

The expectation value can be obtained by:

$$\lambda = \frac{s_k}{s_{ref}} \quad \text{Equation (2)}$$

Where:

- s_{ref} mean distance necessary for the occurrence of a single event
- s_k Distance for k events

For determining whether the behaviour of the automated driving function is within the range of normal driving behaviour, and furthermore to quantify the deviation from normal driving behaviour, an appropriate method has to be identified. Therefore the usage of the quantitative measure 'effect size' is proposed in this approach, which is according to [16] a simple way of quantifying the difference between two groups, that reveals many advantages over the use of tests of statistical significance alone. As depicted in [16], the effect size is a standardized mean difference between two groups and emphasizes the size of the difference rather than confounding this with sample size. The effect size d is calculated in order to estimate the deviation of the behaviour of the automated driving function compared to human driving behaviour by using the following equation:

$$d = \frac{\mu_{experimental} - \mu_{Reference}}{\sqrt{\frac{\sigma_{experimental}^2 + \sigma_{Reference}^2}{2}}} \quad \text{Equation (3)}$$

Where:

- $\mu_{experimental}$ Experimental mean value
- $\sigma_{experimental}$ Experimental standard deviation,
- $\mu_{Reference}$ Reference mean value
- $\sigma_{Reference}$ Reference standard deviation

User-related assessment

User-related assessment was carried out for the "Traffic Jam Assist" system providing automated speed control and lane keeping. If a lead vehicle is present, the automated vehicle adapts its speed in

order to maintain a pre-set time distance to the lead vehicle.

Due to restrictions of driving by naïve drivers in real traffic conditions, assessment activities were limited to driving on a test track by a number of test drivers (employees at Volvo Car Company with administrative duties) to be demonstrated of the system.

Fifteen persons took part in the study, 12 males and 3 females. To collect information trust, usefulness, perceived advantages, disadvantages, acceptance, and willingness to pay for the driver assistance system, after driving on the test track and experiencing the system in action, the participants filled in a questionnaire.

To assess actual trust in the system a six-item self-report scale proposed by Merritt [17] was used. To evaluate the users' perceptions of the system, the System Usability Scale (SUS) [18] was employed. The system's Usefulness and Satisfactoriness was assessed using a modified version of the method proposed by van der Laan et al. [19].

In-traffic assessment

The objective of the in-traffic assessment is to provide a generic framework for assessment of automated driving functions (ADF) in a complete range of traffic situations. For the in-traffic assessment, the set of test cases should resemble the variation found in actual real-life traffic. Automatically, in terms of frequency, normal driving scenarios are most common, while safety-critical scenarios are rare and collision scenarios are close to absent, depending on the functionality.

Estimates of the amount of hours that need to be driven by a vehicle with ADS before it can be regarded as being able to safely handle all scenarios range from one million to billions of hours. If not infeasible, this is at least very costly. Therefore, simulation based assessment should be used. Here, the challenge is to define proper test-scenarios. These test-scenarios can be knowledge-driven or data-driven [21]. A drawback of knowledge-based test-scenarios is that they do not allow to generalize the results to the performance of the system-under-test when operating in traffic, i.e. the test cases may not be valid or representative for real life traffic.

Therefore, a data-driven approach is chosen, which allows generalization of the results.

Within the in-traffic assessment, a way of assessing the in-traffic behavior of automated functions using parameterized scenarios which are extracted from recorded driving data is presented. These parameterized scenarios are used for generating test cases for Monte-Carlo simulations. As real driving data is used, the assessment allows to draw conclusions on how the ADF would perform in real traffic. Since the simulations allow for probabilistic results, there is no need to 'drive' all (billions of) kilometers to draw conclusions.

The process can be roughly divided into three steps. The first step is referring to the collection of the real-life scenarios. A scenario combines the actions of the ego vehicle, the static environment (e.g. infrastructure and weather), and the ongoing activity of the dynamic environment (including the other traffic participants) for a certain period of time. Typically the duration of a scenario is of the order of seconds [21].

The second step concerns the generation of new test cases based on the recorded scenarios. First the scenarios are parameterized. Kernel Density Estimation (KDE) [22] [23] is employed to fit a distribution to model the scenario parameters. The KDE can then be used to generate new test cases.

The final step is the simulation of a generated test case and the post-processing of the resulting data. The combination of generating new test cases and their simulation form the basis of the Monte Carlo Simulation approach.

Safety impact assessment

In Adaptive a virtual assessment approach is applied for the safety impact assessment that combines scenario-based stochastic simulations with continuous operation simulations. The chosen approach is illustrated in figure 4.

The traffic scenario respectively continuous operation simulation works in a virtual traffic environment, which considers many different traffic participants. The virtual traffic environment has the objective to analyze the automated driving functions' behavior in the traffic context considering changes in the frequency of certain driving scenarios. Therefore, the traffic scenario needs to provide a representative variation of

traffic context to trigger realistic variation of system response.

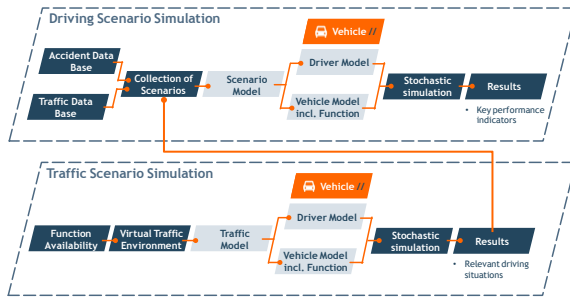


Figure 4. General procedure of safety impact assessment of automated driving functions.

The driving scenario simulation focuses on safety-relevant driving scenarios, which are limited in time and space and represent different conflict types. Safety performance of human driver and the automated driving functions is determined and compared by simulating the driving scenarios in a replicable way. In principle, an automated driving function can affect nearly all accidents scenarios. Due to limited resources an investigation of the all situation is not feasible. Therefore, it has been decided to focus on relevant scenarios for the detailed analysis, the so called “Top scenarios”. These scenarios consider driving scenarios, in which the effect of automated driving functions is questionable respectively are of high relevance for the traffic safety, see table 2.

Table 2.

Top Scenarios for the safety impact assessment

Driving Scenario		Proportion in GIDAS accident database
Top 1	Cut-In	16,1%
Top 2	End in lane	1,1%
Top 3	Obstacle in the lane	3,3%
Top 4	Approaching traffic jam	14,4%
Top 5	Motorway entrance	1,8%
Top 6	Rear-end accident	15,8%
Top 7	Single driving accident	20,6%

Environmental impact assessment

The general approach for the environmental impact assessment that is applied to analyse the considered effects (energy demand and travel time) is given in figure 5. It can be expected that different user groups will benefit in different manners. Therefore, the environmental impact

assessment analyses also the benefit of different user groups.

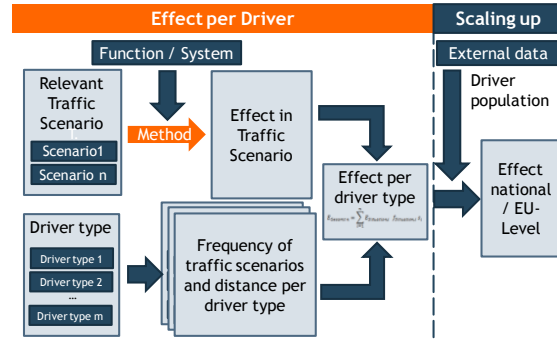


Figure 5. Method for environmental impact assessment

The evaluation is conducted by means of simulation and considers different traffic scenarios. In each traffic scenario the effects are analysed for high numbers of vehicles and a certain road section.

First, the relevant environmental parameters in dependence of the analysed function are identified and aggregated in relevant scenarios. These scenarios are forming the reference and thus the baseline for assessment. Afterwards, the automated driving function which should be assessed is added to the previously defined scenarios in order to estimate the effect in the scenario. The used indicators for the analysis are given in the table 3.

Next to quantification of the effect per traffic scenario, the effects for different driver types are investigated. The different drivers are described based on the travel behaviour (km driven per year and proportion usage of different road types). For each driver type the (spatial) frequency of the different traffic scenarios is obtained. For this purpose different data sources (FOT data, traffic observations, questionnaires, and statistical data) are used.

Table 3.

Evaluation aspects and indicators for environmental impact assessment

Evaluation aspect	Indicators
Energy demand	Positive Kinetic Energy (PKE)
Travel time	Mean velocity
	Mean loss time (urban)
Parking space	Relative change in the number of parking spots

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 Eq. 4. In the last step, the single results for each defined driver type are scaled up on national or European level by means of considering the driver population.

$$E_{\text{DriverType}} = \left(\sum_{i=1}^n E_{\text{scen},i} \times f_{\text{scen},i} \right) \times S_{\text{DriverType}} \quad \text{(Equation 4)}$$

KEY-RESULTS

Technical assessment

In this section the previously presented method for technical assessment is applied to the Adaptive highway demonstrators. First, the changes of frequency of the considered scenarios are analysed between human driving from euroFOT and automated driving. Afterwards, the performance of the automated driving functions is compared to human driving performance in the considered scenarios. For this assessment, euroFOT data [13] from 98 vehicles and in total 8000 h of driving has been clustered in the considered scenarios

Changes of frequency of relevant scenarios

For assessment of automated driving functions first the changes of driving scenarios compared to human driving are analysed, see table 4. The results show that the frequencies for both lane change and cut-in scenarios are increasing.

Table 4.
Change of frequencies of the occurrence of driving scenarios for human and automated driving

Indicator	Human driving	Adaptive highway automation
Lane change per km	0.33 km ⁻¹	0.39 km ⁻¹
Cut-in per km	0.15 km ⁻¹	0.30 km ⁻¹

Changes of performance in relevant scenarios

In this section the effects of automated driving functions within the considered driving scenarios are presented. In the following the driving scenarios “lane change” and “vehicle following” are considered. The effects of automated driving in the scenarios are estimated by calculating the

statistical indicators “effect size”. Regarding the lane change behaviour of automated driving functions, it turned out that there are only slight differences between human driving behaviour. While the maximum lateral acceleration shows a small effect of automated driving (effect size = 0.10), the effect concerning the indicator “manoeuvre time of lane change” is smaller (effect size = 0.18), see figures 6 and 7.

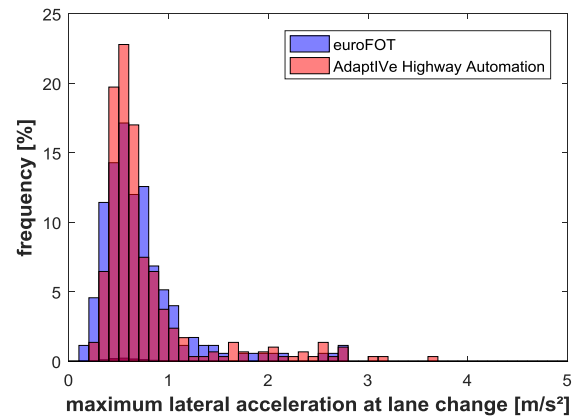


Figure 6. Indicator “maximum lateral acceleration” in scenario lane change

Even more, the share of lane changes with small durations (manoeuvre time < 3 s) can be reduced which leads to a more determined and predictive lane change manoeuvre. This leads to a driving behaviour of automated vehicles which can be more anticipated by other (human) traffic participants which will result in an increase of safety.

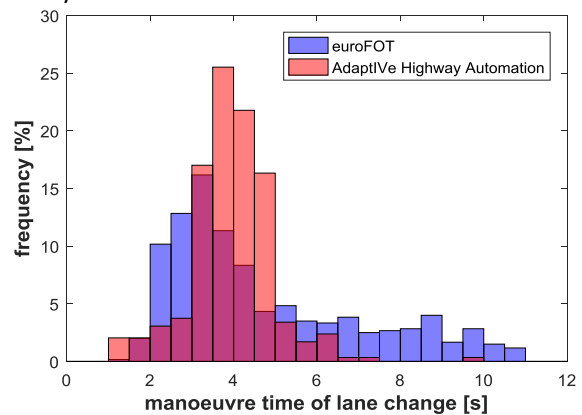


Figure 7. Indicator “manoeuvre time” in scenario lane change

Besides manoeuvre time and maximum lateral acceleration of a lane change, the time headway at initiation of a lane change is assessed as well, see figure 8. Regarding human driving from

euroFOT, the standard deviation is about twice as high compared to automated driving which leads to more lane change manoeuvres initiated at smaller time headways to the front vehicle. This results in an effect size of $d = 0.62$.

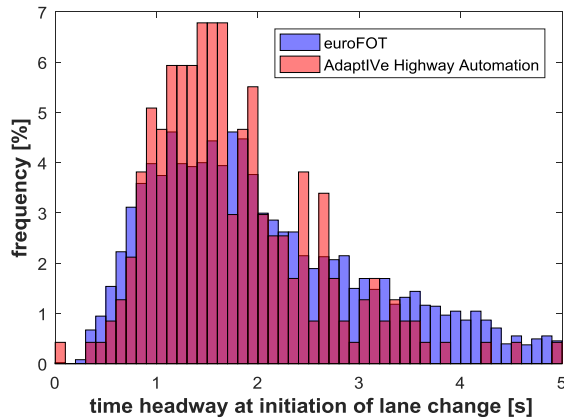


Figure 8. Indicator “minimum time headway” in scenario lane change

For the scenario “vehicle following” the indicator time headway is assessed and compared with human driving behaviour. While the human driver population shows a time headway distribution with a large standard deviation, the automated driving function is showing a smaller standard deviation, see figure 9.

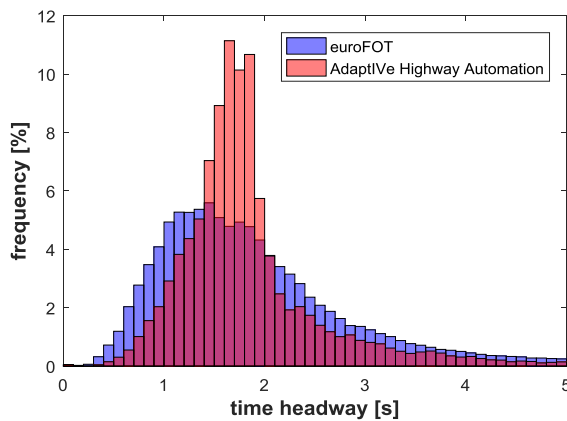


Figure 9. Indicator “time headway” for vehicle following

The small standard deviations of the automated driving function are leading to fewer situations with small headways. Hence, a positive effect on safety induced by automated driving is observed.

User-related assessment

Most participants thought that “the system was a competent performer” and they also had “confidence in the advice given by the system”.

The majority expressed that they “can rely on the system to do its best every time”. Considering whether the driver can depend on the system the majority of the answers were on the “disagree” side and partly neutral, only one respondent agreed strongly that he/she can depend on the system. Considering the statement “I can rely on the system to behave in consistent ways”, most of the responses were in the middle, i.e. close to neutral, however two participants agreed strongly. Considering “trust in the system”, most of the responses were in the middle, i.e. close to neutral, neither agree or disagree, with two participants agreeing strongly. Most participants found the system easy to learn and use, and not unnecessarily complex. They were confident using the system and they would use the system frequently. However, there was not strong support to the statement that the “various functions of the system were well integrated” and there was not much disagreement with the statement that “there was too much inconsistency in this system”.

The total System Usability Scale (SUS) score is 80 which is considered high usability. On the Usefulness/Satisfactoriness scale, the system was perceived as useful (“useful”, “good”, “effective”, “assisting” but not “raising alertness”) and partly satisfactory (“pleasant”, “nice”, but not “desirable” or “likable”). The combined rating of Usefulness – Satisfactoriness is shown in figure 10.

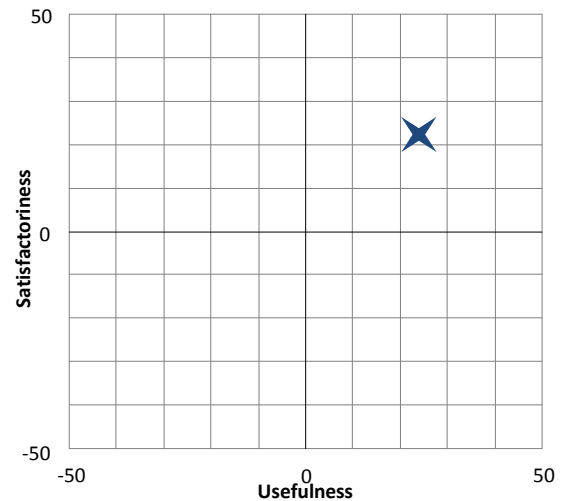


Figure 10. Combined usefulness and satisfactoriness rating of the “Traffic Jam Assist” system

Considering the HMI solution, the participants found that it was easy to activate the function with steering wheel paddles, they found the way to turn on and turn off the system intuitive and they felt safe when enabling the system. The participants felt acceleration and braking while the car drove itself comfortable. Concerning “the comfort of the steering while the car drove itself” and “how good the system was to drive the car on the whole,” there was a wide variance of answers and the “mean” answer cannot be differed from “neither comfortable nor uncomfortable”. The participants found that, the information given in the displays was understandable and the information given in the displays was not distracting.

The participants’ answers indicate that they are not fully aware of the system’s limitations. There are clear expectations in decreased fuel consumption and increased driving comfort among the respondents. The participants estimated the highest usage rate of the system on motorways in their everyday driving. The majority of the participants indicated that they would be willing to pay between 10,000 and 40,000 SEK for purchasing the system.

Answering the question about what they would do while “driving” the autonomous car regularly, a wide range of answers were given, i.e. from full monitoring of driving to completely relaxed presence and doing other things than driving related activities.

Some worries were expressed about relying on the system in real traffic – “does the car constantly handle new and different situations consistently in real traffic with a lot of drivers around who cannot drive a car and do a lot of stupid things”. Also, one respondent felt that driving pleasure disappears with automated driving.

In-traffic assessment

In this section, a simplified application example of the presented in-traffic assessment methodology is presented for a Traffic-Jam-Assist (TJA). Of particular interest is in this case the influence of the Automated Driving Function on its surrounding traffic.

In the example, three cars are driving behind each other. The first vehicle starts with a constant speed and after ten seconds, it decelerates to a

new constant velocity. The third vehicle represents a human driver by means of the Intelligent Driver Model (IDM) [22]. We are interested in the different performance of the third vehicle, depending on whether the second vehicle is operated by a human, modeled by means of the IDM, or the TJA.

Figure 11 shows an example of a velocity profile of the first vehicle. The velocity profile is parameterized using three parameters: the velocity reduction Δv , the total time of braking t_{brake} and the end velocity v_{end} .

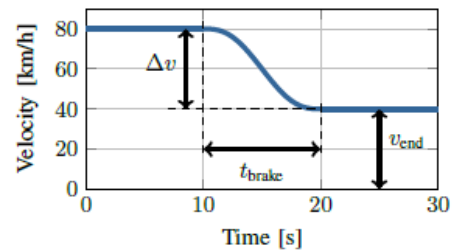


Figure 11. Definition of braking profile of predecessor.

To extract the scenarios from real-life data, 60 hours of naturalistic driving data is used. From this data, approximately 3600 braking scenarios are extracted. KDE is employed to fit the distributions from which infinitely many unique test cases can be generated.

In total 10000 test cases are generated and simulated. Six different Performance Indicators (PIs) are used to measure the performance of the third vehicle for each test case, see table 5.

Table 5. Definition of the Performance Indicators (PI) used for the In Traffic Assessment.

PI	Unit	Description
TTC	s	Minimal Time-To-Collision (TTC), defined as the ratio of the clearance and the relative velocity
d	m	Minimal distance
THW	s	Minimal Time Headway
a_{min}	m/s^2	Minimal acceleration
a_{rms}	m/s^2	Root mean square of acceleration
j_{rms}	m/s^3	Root mean square of jerk

For each performance indication, a Cumulative Probability Distribution (CDF) is constructed using the results of the 10000 simulations. As measure

of the influence of the TJA, the Kolmogorov-Smirnov test is applied for the two different CDFs, i.e. the CDF of the simulations with the second vehicle modelled with TJA and the CDF of the simulations with the second vehicle modelled with IDM. With the Kolmogorov-Smirnov test the maximum difference between two CDFs is estimated.

Figure 12 shows the CDF of the minimal Time-To-Collision (TTC) between the second and third vehicle. As can be seen, the TTC is in general smaller if the second vehicle is equipped with TJA. This maximum difference between the two CDFs is 0.133. The difference can be explained from the fact that the TJA will react later to the braking vehicle, compared to the IDM (mostly due to sensor delay) and therefore, it has to brake harder, which results in a lower TTC between the second and third car. The results for the other PIs are presented in table 6.

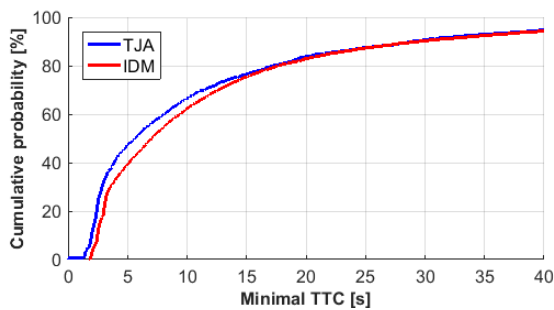


Figure 12. Cumulative probabilities for the minimal Time-To-Collision. The second vehicle is modelled with TJA (blue) or IDM (red).

Table 6. Results of the Kolmogorov-Smirnov tests applied for the different PIs

PI	Results
TTC	0.133
d	0.115
THW	0.015
a_{min}	0.173
a_{rms}	0.052
j_{rms}	0.304

Safety impact assessment

For the impact assessment the described top scenarios have been simulated with the simulation environment OpenPASS, see Figure 13.

The overall simulation results of the simulation show a safety benefit of the automated driven vehicles compared to the by the SCM-model driving vehicles. The results are reported in detail in the Adaptive deliverable D7.3 [23]. In this paper only brief explanation on how the results are derived is given.

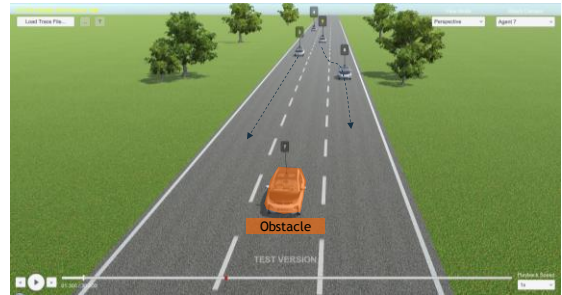


Figure 13. Simulation of the "obstacle in the lane" scenario in OpenPASS.

To analysis the safety effect of automated driving functions for each simulation run it has been analyzed, whether a collision of a relevant vehicle - automated or human driven - and if, at which time point the collision occurs. In the second step it analyses along the simulation flow how many simulation run remain without any collision. An example of the resulting figure in the obstacle in the lane scenario is given in figure 14.

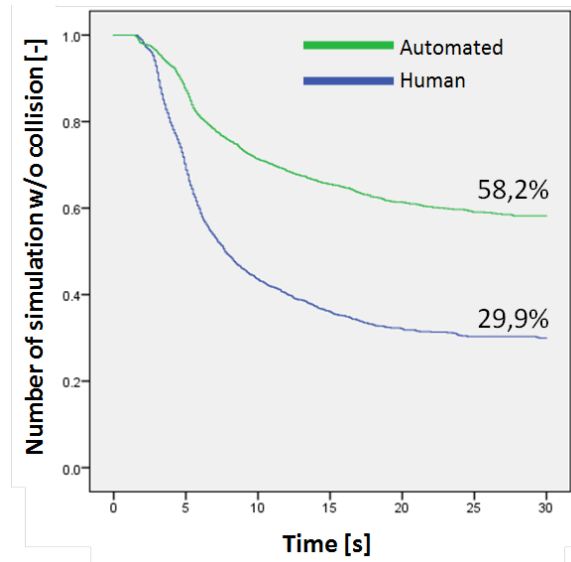


Figure 14. Simulation results for the "obstacle in the lane" scenario (automated vehicle).

By this curves (analogue to a Kaplan-Meier survivorship curve) it cannot only be analyzed, what is the overall benefit, but also at which time

significant safety benefits occur. In the given example major safety effects can be observed in the time span from 2 s to 17 s. This is the time frame, at which the vehicle approaches the obstacle in the lane. After the relevant vehicle has past the obstacle only minor difference between both analyzed simulation configurations (human vs. automated driving) can be observed. The resulting overall benefit in terms of not having an accident in this example is about 28 %.

Environmental impact assessment

From the analysis of the different data sources and the clustering of the people’s driving behaviour driver types can be determined. They are defined by their driving profile, which consists of single traffic scenarios, e.g. intersections, new speed limits or free driving. Figure 15 shows the effect of the automated driving function on the mean velocity of all driver types depending on the daily mileage.

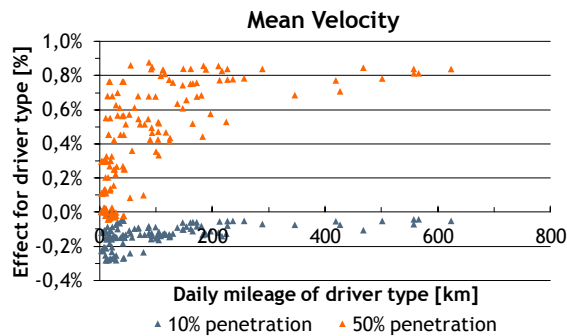


Figure 15. Effects on the mean velocity for all driver types depending on the daily mileage

The chart shows that the mean velocity, in case of a penetration rate of 10%, is slightly reduced for nearly all driver types. For a penetration rate of 50% the mean velocity increases for the most driver types. Particularly for higher daily mileages the effect is relatively high because longer trips have more sections of free driving, which cause a continuously increase of the mean velocity for vehicles with an automated driving function compared to human drivers, whereas scenarios like crossings with priority rules or roundabouts do not raise the mean velocity because they are not addressed by the function.

The following Figure 16 shows the equivalent effects for each driver type concerning the Positive Kinetic Energy (PKE).

The effects of automated driving functions on the PKE are obviously stronger than the effects on the mean velocity. For a penetration rate of 10% the reduction of the PKE is, independently of the daily mileage, between 1% and 2%. It increases up to 16% for driver types who drive high daily mileages when half of the vehicles are equipped with automated driving functions.

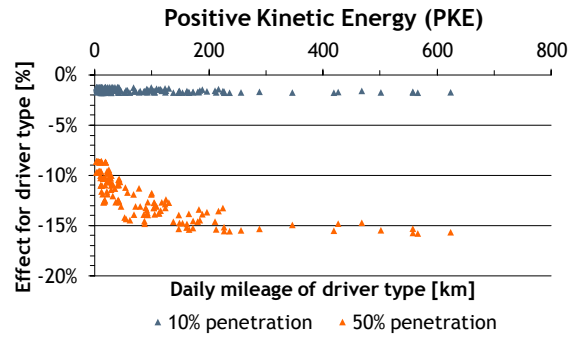


Figure 16. Effects on the PKE for all driver types depending on the daily mileage

To get an overall effect of the automated driving functions for the mean velocity and the PKE, the effects of the different driver types have to be weighted. For this, the mentioned data sources have been used to determine the occurrence of each driver type in the driver population. In Table 7 the effects for the entire driver population are given.

Table 7.

Overall effects of the automated driving function for the whole driver population

	Mean Velocity	Positive Kinetic Energy (PKE)
10% penetration	-0.12%	-1.54%
50% penetration	0.53%	-12.77%

The presented results are based on data sets from Germany because the amount of data there is quite comprehensive. For other countries with a similar data basis the method can be adapted and used as well.

CONCLUSIONS

In this paper the comprehensive framework for evaluation of automated driving which has been developed in the European research

project AdaptIVe is presented. Based on a classification in continuous and event-based operating automated driving functions, the test tools are assigned to the several evaluation aspects. Thus, for technical assessment, the driving behaviour of the automated driving functions is assessed within a small field test and compared with human driving from euroFOT. The results indicate that the driving behaviour of the AdaptIVe automated driving functions is in line with human driving behaviour. Even more, due to their deterministic behaviour, their driving behaviour might be more predictive and thus increasing safety. For user-related assessment tests on a test track were carried out. Here, the results have been collected with a questionnaire. The test persons expressed some worries about relying on the system in real traffic. Overall, good results for satisfactoriness and usefulness of the system have been reached. Concerning the in-traffic assessment, a data-driven approach for assessing the interaction with other vehicles has been developed which is simulation-based. The results have been analysed with the Kolmogorov-Smirnov test. The safety- and environmental impact of automated driving functions developed in AdaptIVe has been analysed by using simulations. While for safety-impact traffic and driving scenarios are considered, for environmental-impact solely traffic scenarios have been analysed. The results indicate that automated driving is leading to a reduction of energy demand.

ACKNOWLEDGEMENT

The research leading to these results has received funding from the European Commission Seventh Framework Programme (FP/2007-2013) under the project AdaptIVe, grant agreement number 610428. Responsibility for the information and views set out in this publication lies entirely with the authors. The authors would like to thank all partners within AdaptIVe for their cooperation and valuable contribution.

REFERENCES

- [1] Rodarius, C., Dufelis, J., Fahrenkrog, F., et al., "Test and Evaluation Plan", AdaptIVe Deliverable D7.1, 2015.
- [2] N.N., Report to Congress, DARPA Prize Authority, DARPA, 2006.
- [3] N.N., DARPA Urban challenge Media, DARPA urban challenge webpage: <http://archive.darpa.mil/grandchallenge/>, 2014.
- [4] Markoff, J., "Google Cars Drive Themselves in Traffic", The New York Times, 2010.
- [5] Ziegler, J, Bender, P, Schreiber, M., et al., "Making Bertha Drive—An Autonomous Journey on a Historic Route", IEEE Intelligent Transportation Systems Magazine, Volume 6 ,Issue 2, 2014.
- [6] Ploeg, J., Shladover, S., Nijmeijer, H., et al., "Introduction to the Special Issue on the 2011 Grand Cooperative Driving Challenge", IEEE Transactions on Intelligent Transportation Systems, Volume 13, Issue 3, 2012.
- [7] Etemad, A., Bartels, A., "A Stepwise Market Introduction of Automated Driving", 21st ITS World Congress, Detroit, 2014.
- [8] Wachenfeld, W., Winner, H., „Die Freigabe des autonomen Fahrens“, Autonomes Fahren. Technische, rechtliche und gesellschaftliche Aspekte. Springer Vieweg, 2015.
- [9] N.N., "Taxonomy and Definitions for Terms Related to On-Road Automated Motor Vehicles", SAE document J3016, 2014.
- [10] Scholliers, J., Hendirks, F., Aust, M.L., et al., "Project final report and recommendation for future assessments", PReVAL project, Deliverable DL IP D12/D16.4, 2008.
- [11] Wilmink, I., Janssen, W., Jonkers, E., et al., "Impact assessment of Intelligent Vehicle Safety Systems", eIMPACT, Deliverable D4, 2008.
- [12] Larsson, P., Esberg, I., van Noort, M., "Test and evaluation plans", interactIVe Deliverable D7.4, 2012.
- [13] Benmimoun, M., Puetz, A., Zlocki, A., and Eckstein, L., "euroFOT: Field Operational Test and Impact Assessment of Advanced Driver Assistance Systems: Final Results", Proceeding of the FISITA 2012 World Automotive Congress, 2013.
- [14] Reichel, M., Botsch, M., Rauschecker, R., Siedersberger, K.-H., et al., "Situation aspect

modeling and classification using a scenario based random forest algorithm for convoy merging situations”, 13th International IEEE Conference on Intelligent Transportation Systems (ITSC 2010), pp. 360-366, 2010.

[15] Roesener, C., Fahrenkrog, F., Uhlig, A., Eckstein, L., “A Scenario-based Assessment Approach for Automated Driving by Using Time Series Classification of Human Driving Behaviour”, IEEE ITSC 2016, Rio de Janeiro, Brasil, 2016.

[16] Coe, R., “It’s the Effect Size, Stupid: What effect size is and why it is important”, Annual conference of the British Educational Research Association, University of Exeter, England, 12-14, 2002.

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

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

[20] Van der Laan, J. D., Heino, A., de Waard, D., A simple procedure for the assessment of acceptance of advanced transport telematics”. Transportation Research C, 5, 1-10., 1997.

[21] Stellet, J. E., Zofka, M. R., Schumacher, J., Schamm, T., Niewels, F., and Zöllner, J. M., “Testing of advanced driver assistance towards automated driving: A survey and taxonomy on existing approaches and open questions,” in 2015 IEEE 18th International Conference on Intelligent Transportation Systems, pp. 1455–1462, IEEE, 2015.

[22] Elrofai, H., Worm, D., and Op den Camp, O., “Scenario Identification for Validation of Automated Driving Functions”, pp. 153–163. Springer International Publishing, 2016.

[23] Rosenblatt, M., “Remarks on Some Nonparametric Estimates of a Density Function,” The Annals of Mathematical Statistics, vol. 27, pp. 832–835, 1956.

[21] Parzen, E., “On Estimation of a Probability Density Function and Mode,” The Annals of Mathematical Statistics, vol. 33, pp. 1065–1076, 1962.

[22] Treiber, M.; Hennecke, A.; Helbing, D., "Congested traffic states in empirical observations and microscopic simulations", Physical Review E, 62 (2): 1805–1824, 2000.

[23] Fahrenkrog, F., Wang, L., Rösener, C. et al., “Impact analysis for supervised automated driving applications”, AdaptIVe Deliverable D7.3, 2017.