ABSTRACT

One of the major challenges for enabling market introduction of automated driving is to identify risks and benefits of these functions. For this purpose, a new framework for assessing the safety impact of automated driving functions (ADFs) has been investigated. This framework is based on accident- and field operational test- (FOT-) data while using simulations for assessment of ADFs with respect to a certain baseline. According to the German Ethics Commission for Automated and Connected Driving, this baseline has to be human manual driver performance. For modelling of this baseline in simulations, so-called driver performance models are introduced in this publication and incorporated in an overall framework for effectiveness assessment.

The main idea of the developed framework is that the types of driving scenarios, respectively physical accident constellations, do not change with automated driving. However, since ADFs are continuously controlling the behavior of the vehicle, it is possible that ADFs will get involved less frequently in accident scenarios playing a major role at human driving, e.g. rear-end accident scenarios. On the other hand, it is likely that other previously irrelevant accident types will rise. Consequently, the frequency of occurrence and the severity of the addressed driving scenarios may change with automated driving although the types of driving scenarios stay the same. To investigate the change of severity in a driving scenario, accident re-simulations are used. The changes in frequency of occurrence of driving scenarios are analyzed by using traffic simulations. In this work, so-called driver performance models are introduced for modelling human baseline in accident re-simulations. Key findings concerning the structure of these driver performance models are presented.

The developed method and models are applied on two generic ADFs, a generic “Motorway-Chauffeur” (SAE level 3) and a generic “Urban Robot-Taxi” (SAE level 4). The results indicate that, e.g. a Motorway-Chauffeur at a market penetration of 50 % has a potential for reducing about 31 % of all accidents on German motorways resulting in personal injury. This equals 2 % of all accidents on German roads.
INTRODUCTION

In the last decade various automotive functions for supporting the driver have been developed. These so-called advanced driver assistance systems (ADAS) are supporting the driver on different levels of the driving task. Driven by recent developments in algorithms for environment perception and decision making, the ultimate goal of vehicle automation seems to be a solvable task as shown by several demonstrations [1].

However, due to an increasing complexity of decision making algorithms of these complex functions, identifying benefits and drawbacks will be challenging. Hence, new safety effectiveness assessment methods have to be designed which are based on detailed accident-, FOT- and simulation data and that are assessing the ADFs with respect to a certain baseline. Since automated driving will not be able to avoid all accidents on roads, e.g. due to the misbehavior of other traffic participants and physical limits, a baseline for assessment has to be defined. According to the German Ethics Commission for Automated and Connected Driving,

“[..] the licensing of automated systems is not justifiable unless it promises to produce at least a diminution in harm compared with human driving, in other words a positive balance of risks […]” [2]

Consequently, the reference for safety impact assessment needs to be human driver performance. In order to assess ADFs with respect to human driver performance, this paper introduces a method for safety effectiveness assessment. The basic idea of this framework is that the types of accident constellations and thus driving scenarios do not change with automated driving. However, the severity and frequency of occurrence of these driving scenarios may change with automated driving.

BACKGROUND

For effectiveness assessment of (advanced) driver assistance systems with environment perception, many different methods have been used in the past. All these methods have in common, that they compare driving situations without the system with driving situations, in which the system is activated. One valid approach for determining the effectiveness of ADAS is the accident re-simulation on basis of in-depth accident data, e.g. as applied in [3]. In this case, reconstructed accident scenarios from detailed accident data, such as the German-in-depth accident database (GIDAS) [4], are simulated with and without the considered function. The difference in performance in the situation, e.g. probability of severe injuries, is considered as the benefit of the function. An alternative to re-simulation of single accident situations is provided by stochastic approaches describing the situational variables of a driving scenario by Monte Carlo sampling of synthetic driving situations from probability distributions as presented in [5]. A disadvantage of accident re-simulations is that new induced driving scenarios by automated driving cannot be considered, because these are not represented in the accident data. Another approach for safety impact assessment based on recorded data is the field operational test (FOT) as presented in [5]. Here, huge amounts of driving data without function (control condition) and with activated function (experimental condition) are collected. The effectiveness of the considered function is analyzed by investigating the change in frequency of occurrence of incidents and near-crashes compared to the baseline. For effectiveness assessment of a function in defined situations, driving simulator studies can be used as well. This approach allows a detailed investigation of human driver performance with and without the considered function as demonstrated in [6], but requires a selection of situation parameters to be presented to the drivers. As described previously, ADFs need to be assessed in the whole entity of possible driving situations in their operational design domain. Hence, simulations of these functions in the whole traffic are a promising approach as presented in [7]. However, validation of these simulations remains challenging because of the variety and complexity of models necessary for safety impact assessment.

Based on the available methods presented previously, a suitable method for assessing the effectiveness of road vehicle automation is defined. Although accident re-simulation based on detailed accident data is a valid approach, it will not be suitable for assessing ADFs since this approach is based on previously recorded detailed accident data from human driving. In order to identify new driving situations induced by ADFs, a FOT would be suitable. However, considering the necessary resources difficult to realize. Thus, a holistic approach including accident re-simulations for investigation of changes in severity and traffic simulations for assessing changes in frequency of driving scenarios is developed for effectiveness assessment of ADFs.

FRAMEWORK FOR EFFECTIVENESS ASSESSMENT

Built on previously recorded accident- and FOT-data and extended by simulation data, the effectiveness of a defined ADF is assessed by considering the changes in severity and frequency of addressed driving scenarios, see Figure 1. Based on a definition of the ADF and the addressed driving scenarios the effectiveness fields – all
addressed accidents and relevant driving situations – are identified in the input data. To this end, the absolute number of accidents per driving scenario is extracted from accident statistics for upscaling. By this, the results derived from detailed data can be projected upon the effectiveness on a national level. The parameters spaces (e.g., probability density function of velocity of involved traffic participants) are extracted from in-depth accident- and FOT-data for determination of the changes in severity of driving scenarios due to the function.

**Figure 1. Framework for Effectiveness Assessment of Road Vehicle Automation.**

Afterwards, the changes in frequencies of occurrence of the defined driving scenarios are assessed by using traffic simulations. Here, so-called driver error models are used to model critical driving situations in traffic simulations. To identify the changes in severity in the defined driving scenarios, these are simulated with and without ADF while the reference performance is modelled by human driver performance models. The effectiveness $E$ of an ADF in terms of safety can be derived based on a consideration of the accident risk $R_{Scenario}$ by the severity $I_{Scenario}$ and the frequency of occurrence $f_{Scenario}$ for each driving scenario.

$$R_{Scenario} = I_{Scenario} \cdot f_{Scenario}$$

The effectiveness $E$ in a scenario $i$ is defined as the difference of risks $\Delta R_i$.

$$E_i = \Delta R_i = R_{ADF,i} - R_{Human,i}$$

Substituting the risk $R_i$ by severity $I_i$ and frequency $f_i$ results in:

$$E_i = I_{ADF,i} \cdot f_{ADF,i} - I_{Human,i} \cdot f_{Human,i}$$

With the change in severity $\Delta I = I_{ADF}/I_{Human}$ and the change in frequency of occurrence $\Delta f = f_{ADF}/f_{Human}$, the effectiveness $E$ is derived for all scenarios $n$ by:

$$E = \sum_{i=1}^{n} I_{Human,i} \cdot f_{Human,i} \cdot (\Delta I_i \cdot \Delta f_i - 1)$$

**Definition of Driving Scenarios based on Accident Type Catalogue**

The developed framework assumes that the defined driving scenarios cover all physical possible accidents constellations. For this purpose, the driving scenarios are derived from the German accident type catalogue [9] that includes a classification scheme for all accidents by a three-digit code built upon decades of experience by the German police. In consequence, almost all accident constellations that are physical possible are included in this catalogue. The considered driving scenarios are derived from this catalogue by assigning each three-digit accident types UTYP3 to a driving scenario. This process is illustrated on the example of a “cut-in” driving scenario in Figure 2.
Figure 2. Derivation of driving scenarios based on three-digit accident classification on the example of a “cut-in” driving scenario.

Description of Automated Driving Function

The previously derived driving scenarios are used to describe the functional scope of the assessed ADF. In this sense, to describe an Urban Robot-Taxi, only driving scenarios on urban roads within the operational design domain of the Urban Robot-Taxi will be linked to the function. In addition, functional limitations, e.g. due to environmental conditions (fog, heavy rain, snow) are included in the description of the ADF and can be used to limit the addressed accidents. An exemplary description of an Urban Robot-Taxi is given in Table 1.

Table 1.
Description of automated driving functions and their operational design domain (ODD) on the example of the Urban Robot-Taxi.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Urban Robot-Taxi</td>
</tr>
<tr>
<td>Level of automation according to [SAE16]</td>
<td>4</td>
</tr>
<tr>
<td>Sensor view range</td>
<td></td>
</tr>
<tr>
<td>Adressed driving scenarios</td>
<td>• Driving without influence from leading vehicle&lt;br&gt;• Approaching static object&lt;br&gt;• Approaching leading vehicle&lt;br&gt;• Approaching lateral moving object&lt;br&gt;• Approaching traffic jam&lt;br&gt;• Cut-in&lt;br&gt;• Lane change&lt;br&gt;• Turning&lt;br&gt;• Crossing&lt;br&gt;• U-Turn</td>
</tr>
<tr>
<td>Road types and speed range</td>
<td>• Inside city-limits: 0 - 50 km/h</td>
</tr>
<tr>
<td>Functional limitations</td>
<td>• None</td>
</tr>
</tbody>
</table>

Driving Scenario-based Identification of Effectiveness Fields

After describing the assessed ADFs including their applicable driving scenarios, the effectiveness fields – the accidents and driving situations where the ADFs have a potential impact - are estimated. For in-depth accident data and national accident statistics, the three-digit accident type can be used to select the driving scenarios. FOT-
data is not labelled with a three-digit accident type as it contains time series data. To cluster FOT-data, the driving scenario classification algorithms based on machine learning introduced in [9] can be used. Here, the relative motion included in the three-digit accident type is used to detect the driving scenarios in time series data.

The effectiveness fields can be limited regarding road types and limitations of the ADFs. The classification of driving scenarios results in a number of accidents per driving scenario (see Figure 4) that enables to investigate the change in frequency of the driving scenarios as well as the parameter spaces necessary to determine the change in severity per driving scenario (see Figure 5). Figure 3 illustrates the whole definition process of a driving scenario-based estimation of the effectiveness fields exemplified for a “cut-in” driving scenario.

For example, from all accidents with personal injuries A(P) occurring within city limits in Germany (70 % of all accidents), an Urban Robot-Taxi is addressing 66 %. The other accidents in the domain cannot be addressed due to the reason that driving scenarios are not covered by the functional scope of the automated driving function (14 %), driver and vehicle related limits such as technical failures or alcohol use (3 %) and no car involvement in the accident (17 %), see Figure 4.

**Figure 3.** Process for driving scenario-based estimation of effectiveness fields due to methodical constraints and description of the ADF. The effectiveness fields include the number of accidents as well as the parameter spaces per driving scenario.

**Figure 4.** Numbers of addressed accidents resulting from effectiveness field of Urban Robot-Taxi in German national accident statistics DESTATIS.

Next to the number of accidents resulting from the effectiveness fields, the parameter spaces describing the driving scenarios for estimation of the changes in severity are extracted from FOT- and in-depth accident data. The
parameter spaces are represented as Kernel Density Estimation (KDE) obtained from the probability density functions of situational variables of in-depth accident data and FOT-data. Using Monte-Carlo sampling techniques according to [11], concrete scenarios that can be simulated are randomly “drawn” from the logical scenarios. Exemplary parameter spaces of situational variables such as “ego velocity” describing the logical scenario “cut-in” are presented in Figure 5.

Figure 5. Parameter spaces for describing the logical scenario “cut-in” for estimation of the changes in severity by simulation.

Both number of accidents per driving scenario and the parameter spaces describing the driving scenario for simulation are used in the following to estimate the effectiveness in terms of a change in accidents per driving scenarios.

Driver Error Models in Traffic Simulations for Changes in Frequency of Driving Scenarios

Traffic simulations are used to identify the changes in frequency of occurrence $\Delta f$ of driving scenarios. For considering the effects within mixed traffic conditions of human driven and automated vehicles, it is distinguished whether a human driven or an automated vehicle has induced or “caused” a certain driving scenario. For example, a human driver cutting-in in front of the automated vehicle can cause a “cut-in” driving situation. In this case, the human driver induced the driving situation while the automated vehicle was involved in it. Based on this principle, a classification scheme for driving situations is introduced, see Table 2.

Table 2. Types of interactions in driving scenarios in mixed-traffic conditions

<table>
<thead>
<tr>
<th>Type of interaction</th>
<th>Type of vehicle driving scenario induced by</th>
<th>Type of vehicle involved:</th>
<th>Illustration</th>
</tr>
</thead>
<tbody>
<tr>
<td>HUM-HUM</td>
<td>Human driver</td>
<td>Human driver</td>
<td></td>
</tr>
<tr>
<td>HUM-ADF</td>
<td>Human driver</td>
<td>Automated driving function</td>
<td></td>
</tr>
<tr>
<td>ADF-HUM</td>
<td>Automated driving function</td>
<td>Human driver</td>
<td></td>
</tr>
<tr>
<td>ADF-ADF</td>
<td>Automated driving function</td>
<td>Automated driving function</td>
<td></td>
</tr>
</tbody>
</table>

The changes of frequencies for all four defined types of interactions are analyzed by using traffic simulation data of human driven and automated vehicles for several market penetration rates of automated vehicles. For traffic simulation, a 26 km long section of the German motorway A2 around Hanover is used, see Figure 7 (left). Modelling the behavior of human traffic participants is one of the most crucial parts in traffic simulations. Although a tremendous variety of driver models is available [12], [13], [14] the main purpose of these existing models are traffic flow investigations and not investigations related to traffic safety. The main limitation of the available models is that they do not require to reflect human behavior in critical and uncommon situations but that they have been designed to represent the trained “normal” driving behavior. Special driver models are therefore needed to realistically represent human driving behavior in incident situations.
Consequently, so-called driver error models are developed that are modelling human errors causing incident driving situations in traffic simulations.

To model how humans induce incident situations in traffic simulations, the principles leading to human errors have to be incorporated in the simulation models. The existing driver models, e.g. [14], that assume an ideal recognition and decision of humans, are extended by probabilistic error models that represent uncertainties in recognition and decision. According to the findings of [15], human drivers are able to perceive Time-to-Collision (TTC) and Time Headway (THW) to other objects in their surroundings. For example, the human eye is capable of perceiving the TTC to an object by detecting changes in its retinal projection [15]. A similar principle is assumed for perceiving the THW [15]. For modelling driver errors, it is assumed that the perception of TTC and THW is afflicted with uncertainties. Therefore, the perceived $TTC_{\text{perceived}}$ might differ from the real $TTC_{\text{real}}$ in a driving situation ending up in a misjudgment of the situation by the driver that can lead to an incident situation, see Figure 6.

![Figure 6. Probabilistic modelling of uncertainties in recognition and decision for the induction of incident driving scenarios.](image)

It is assumed that these uncertainties in recognition of other traffic participants of the driving scenario are gamma distributed. Based on Monte-Carlo sampling [11] of the gamma probability distributions, for each explicit driving situation occurring in simulation uncertainties in THW and TTC can be generated. While most of the sampled uncertainties will be few and not lead to incident situations, potential incident “cut-in” driving situations will occur according to the probability for high uncertainties represented in the gamma probability distributions. The resulting exemplary changes of frequency for an “approaching leading vehicle” driving scenario are shown in Figure 7 (right).

![Figure 7. Traffic scenario for estimation of changes in frequencies of driving scenarios (left) and change of frequency of “approaching leading vehicle” driving scenario (right).](image)

**Driver Performance Models for Changes in Severity of Driving Scenarios**

If an automated vehicle gets involved in an incident driving situation, the changes in severity $\Delta I$ induced by the ADF are assessed. For this purpose, driving situations with explicit parameters are simulated with an ADF and with human driver performance models as a reference. The process is illustrated in Figure 8. The difference in performance between human and ADFs is defined as the change in severity. This is measured by the likelihood for severe injuries (MAIS2+) that is derived by injury risk curves based on the relative collision speed. The
parameter spaces resulting from the driving scenario-based estimation of effectiveness fields (see Figure 5) are used to generate concrete scenarios with explicit parameters that can be simulated.

**Figure 8. Simulation method for estimating the changes in severity in driving scenarios on the example of a “cut-in” driving scenario.**

For human reference performance, quantitative driver models for modelling human driving performance in defined driving scenarios from [16] are used. The structure of the models is split into perception, information processing and action. Human drivers are acting in unexpected driving situations based on their knowledge rather than on the actual situational variables according to [17]. Thus, human action is modelled with an initial feedforward impulse and a feedback control to stabilize the vehicle afterwards. The initial feedforward reaction is described by reaction time and intensity and is sampled from gamma distributions representing a driver population. The structure of the developed driver performance models is validated based on simulator studies with 35 test subjects [16]. Finally, the developed models are verified based on in-depth accident data for ensuring that they can be applied for the respective driving scenario. The structure of the models is presented in Figure 9.

**Figure 9. Framework for human driver performance models consisting of perception, information processing and action.**

These driver performance models are developed for all covered driving scenarios. For example, in the driving scenario “cut-in” the likelihood for severe injuries (MAIS2+) can be reduced by 42.3 % by the Motorway-Chauffeur.

**Effectiveness of Automated Driving Function**

Finally, the effectiveness of the automated driving function in the effectiveness field is derived based on the changes in frequencies of all driving scenarios and the changes in severity in all driving scenarios. This process is illustrated on the example of the “cut-in” driving scenario at 50 % market penetration of the Motorway-Chauffeur.

The results of the analysis of changes in frequency of occurrence based on traffic scenario level showed a decrease in accidents by 28.2 %, as presented in Figure 10. According to the traffic simulation (see Section 7.3), human drivers induced all resulting in 71.8 % of accidents on traffic scenario level. From all accidents on traffic scenario level induced by human drivers, 43.5 % are with involvement of a human driver (“HUM-HUM”) while the remaining incidents are with involvement of an ADF (“HUM-ADF”)

---

Roesener 8
In the next step, the changes in severity on driving scenario level are analyzed. According to the results of the re-simulation (see Section 7.4), all accidents with involvement of ADF can be reduced by 40 % to 23.4 % of accidents. For the case in which only human drivers were involved, the 31.2 % of accidents do not change with ADF. In total, from the initial 100 % of accidents, the ADFs reduces to 53.4 % of accidents which consequently results in an effectiveness of 46.6 % for the driving scenario “cut-in” at a market penetration rate of 50 %.

RESULTS

The simulation-based estimated effectiveness for the different ADFs is scaled-up on national level for the Federal Republic of Germany. Since the effectiveness of the ADF is determined based on detailed GIDAS accident data that is only available for a limited geographical region in Germany the effects have to be corrected and projected by using the national accident statistics. For this purpose, the correction factors per driving scenario are derived based on the frequency of occurrence of the defined driving scenarios in GIDAS detailed accident and national accident statistics by using the three-digit accident type. On basis of the Urban Robot-Taxi the results presented in Figure 8 will be explained. In the operation domain of the Urban Robot-Taxi 205,321 accidents with personal injuries occurred in 2016. Since only ADFs of passenger cars are considered, just those accidents can be addressed where at least one passenger car is among the first two participants of the accidents. These 36,486 accident cannot be addressed (see light gray area). Furthermore, 47,487 accidents per year are outside the functional limits of the Urban Robot-Taxi (see dark gray area) due to not addressed driving scenarios, alcohol and drug use, technical failures and limitations of the Urban Robot-Taxi (rain, fog, ice, construction sites).

Figure 10. Effectiveness of an ADF on traffic- and driving scenario level on the example of a “cut-in” driving scenario and a market penetration of 50 %.

Figure 11. Effectiveness in terms of avoided accidents of Motorway-Chauffeur and Urban Robot-Taxi [18].
The light blue area represents the number of accidents that are potentially addressable, but cannot be avoided according to the simulation results. These are for example accidents that cannot be avoided due to physical constraints. However, the severity of these accidents possibly can be reduced by a reduction of the collision speed. The dark blue area represents the number of avoided accidents. Hence, the Urban Robot-Taxi can avoid 52,517 accidents at a market penetration of 50 %. This equals an effectiveness of 27 % of all accidents in the operation domain.

**DISCUSSION**

In contrast to existing approaches in literature that define driving scenarios based on ontologies created by expert knowledge, in this work the driving scenarios are derived from the three-digit accident type covering all potential physical accident constellations known to accident research for decades. A set of 13 driving scenarios has been identified from the accident type catalogue. The definition of the driving scenarios by the three-digit accident type reveals tremendous gains. Since both national accident statistics (in five German federal states) and GIDAS in depth accident data feature the three digit-accident type, the driving scenarios can be classified in both types of data. Consequently, both, the number of accidents on national level per driving scenario and the parameter spaces for deriving the change in severity induced by ADFs can be determined with the developed concept. The presented concept in this thesis limited the available number of traffic participants by a number of two that covers 90 % of accidents. A possible enlargement of the presented driving scenarios to cover the remaining 10 % of accidents is to extend the number of traffic participants per driving scenario to more than two. Beyond that, a more detailed clustering into more than 13 driving scenarios can be realized. However, it has to be considered that the efforts for assessment are increasing with the number of driving scenarios.

**CONCLUSIONS**

According to the statements in [2], automated driving functions need to show a positive risk-balance compared to human driving in terms of traffic safety. Therefore, a framework for effectiveness assessment of road vehicle automation has been introduced in this work. The basic idea of this framework is that the types of accident constellations and thus driving scenarios do not change with automated driving. Though, the severity and frequency of occurrence of these driving scenarios may change with automated driving. Traffic simulations with automated driving functions are investigating the changes in frequency of occurrence. For determination of the change in severity in relevant driving scenarios, accident re-simulations were used. After determining the effectiveness of the automated driving functions, they are projected and depicted over the whole territory of the Federal Republic of Germany. The results indicate that, e.g. a Motorway-Chauffeur at a market penetration of 50 % has a potential for reducing about 31 % of all accidents on German motorways resulting in personal injury. This equals 2 % of all accidents on German roads. The Urban Robot-Taxi can avoid 27 % of all accidents with personal injury within city-limits at a market penetration of 50 %. This equals 17 % of all accidents on German roads.

**ACKNOWLEDGEMENT**

The research leading to these results has received funding from the German Federal Highway Research Agency (BASt) within the project “Potential Societal Benefits due to increasing Vehicle Automation” FE 82.0626/2015/ERB. 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 this project for their cooperation and valuable contribution.
REFERENCES


