HARMONIZED APPROACHES FOR BASELINE CREATION IN PROSPECTIVE SAFETY PERFORMANCE ASSESSMENT OF DRIVING AUTOMATION SYSTEMS

Peter Wimmer
Virtual Vehicle Research GmbH
Austria

Olaf Op den Camp
TNO
The Netherlands

Hendrik Weber
Institute for Automotive Engineering (ika), RWTH Aachen University
Germany

Henri Chajmowicz
LAB
France

Michael Wagner
Continental Autonomous Mobility Germany GmbH
Germany

Jorge Lorente Mallada
Toyota Motor Europe NV/SA
Belgium

Felix Fahrenkrog
BMW AG
Germany

Florian Denk
THI - CARISSMA Institute of Automated Driving
Germany

Paper Number 23-0032

ABSTRACT

In the last years, virtual simulations have become an indispensable tool for safety performance assessment of driving automation systems (DAS) and pre-crash technologies which are part of advanced driver assistance systems (ADAS). Different approaches and tools are used in this domain, making comparison of results of different studies difficult. Therefore, the P.E.A.R.S. (Prospective Effectiveness Assessment of Road Safety) initiative was founded to harmonize methods for prospective safety performance assessment and by this make results of such studies more trustworthy and comparable. One essential pillar of such a harmonization is the establishment of the baseline, the set of data to which the performance of the technology under study is compared to when performing prospective assessments. Various ways have been presented in literature for setting up a baseline. For harmonization, these ways need to be analyzed and categorized so that recommendations can be given on when and how to use a certain baseline approach. The research objective of this paper is first to develop general approaches to establish a baseline based on existing ways and second to identify areas of application for each baseline approach.

Based on existing ways, we defined general approaches for setting up a simulation baseline. These baseline approaches can structure all existing ways based on their characteristics and requirements and impacts on safety performance assessment results. Relevant information for each baseline approach is discussed, such as the used data type(s), data processing steps, applied variations to the original data, application of simulation models, and statistical methods, etc.

We identified three types of baseline approaches: A) Using concrete real-world scenarios without modifications. B) Using modifications of concrete real-world scenarios. Here, real-world scenarios are the basis, but some of the existing measured properties are altered or even new properties are added. C) Creating synthetic cases where more general data such as distributions of relevant parameters (e.g., from collision, road user behavior, traffic data) and mechanisms possibly leading to collisions are used. The paper will provide examples for each baseline approach.

The three approaches can be clearly distinguished and should be able to cover the generation of a baseline for all studies in the field of prospective safety performance assessment. Each of the approaches has its pros and cons, e.g., with respect to their representativeness, and the effort to obtain the required data. Also, the evaluation objective to be addressed needs to be considered when selecting an appropriate baseline approach as it has a strong influence on this selection. The categorization of the three approaches allows for defining common recommendations on when to use which approach.

By the baseline approaches presented, P.E.A.R.S. contributes to the harmonization and acceptance of virtual safety performance assessment of driving automation systems (DAS) and pre-crash technologies. This will greatly enhance trustworthiness, comparability and, transparency of results of prospective safety performance assessments.

INTRODUCTION

Trends in the number of traffic casualties in the EU [1] show that it will be difficult to meet the target of Vision Zero: reducing road fatalities to almost zero by 2050 [2]. At the same time, the road traffic system is changing rapidly due to, e.g., the introduction of new mobility systems such as connected, cooperative, automated driving and new enabling technologies such as artificial intelligence and wireless V2X-communication [3]. Driving Automation Systems (DAS), including vehicle safety features such as advanced driver assistance systems (ADAS), are introduced with the intention to make road mobility services safe for all road users. Moreover, DAS and ADAS are implemented to make road mobility services available, more comfortable, and safe for drivers and passengers.

Authorities are being asked to allow vehicles equipped with new advanced DAS onto public roads. However, an appropriate methodology for approval to deploy these vehicles onto the road is not yet in place. The EC formulated recommendations [4] to ensure that DAS contribute to road safety improvements. This example shows that there is a clear need for a prospective safety assessment framework that is capable to deal with the great challenges and fast developments in technology. To keep a feasible testing effort, an increasing role of virtual testing is foreseen to handle the seemingly infinite number of situations that DAS may end up into during the lifetime of the vehicle. Although DAS and ADAS are complex and the safety assessment procedure can be complicated, its results should be unambiguous, easily understood by experts in the field, and explainable to policymakers and the general public.

The EU Horizon 2020 project HEADSTART [5] defined testing and validation procedures for the safety of Connected, Cooperative, and Automated driving functions for specific use cases. HEADSTART set important requirements for the use of simulations to test and validate DAS. The recently started EU Horizon Europe project V4SAFETY [6] uses the HEADSTART requirements as starting point to develop comprehensive procedures for conducting computer simulations to determine the long-term performance and impact of road safety solutions, from the identification and collection of the relevant input data to the projection of the results to a region of interest (e.g., the EU).

Harmonization of the assessment framework is essential to achieve explainable and comparable results, independent of the specific simulation tool used. Moreover, it is to the benefit of all stakeholders that the developed safety assessment framework not only conforms with European Union [7] and United Nations regulations [8] but with international standards such as ISO and SAE as well. The lack of harmonization of prospective assessment was already recognized in 2012, which led to the establishment of a harmonization group: Prospective Effectiveness Assessment for Road Safety (P.E.A.R.S. [9]). In the last decade, the P.E.A.R.S. harmonization group, currently consisting of 31 organizations from industry, research organizations, and academia, provided input to the ISO working group “Traffic accident analysis methodology”, resulting in the publication of an ISO Technical Report [10]. P.E.A.R.S. is continuing its work in drafting an ISO Technical Specification “Prospective safety performance assessment of pre-crash technology by virtual simulation — Guidelines for application” [11](under development). This document will provide a general description of the process for prospective safety performance assessment of pre-crash technology by virtual simulation.

All stakeholders in road safety indicate the need [5] for a predictive safety assessment framework that allows fast and extensive evaluation of safety solutions, including DAS, for a large variety of relevant traffic scenarios. This is already envisaged by policy makers and consumer associations [7], [12], considered in state-of-the-art research activities worldwide [5], [13] and foreseen to be adopted by industry partners to manage testing efforts [14].

The framework uses predictive virtual simulation (hereafter simply 'simulation') in addition to the already existing physical tests—not only for type approval but for consumer testing as well. In safety assessment of a vehicle function (e.g., a newly developed DAS), the performance of the vehicle with the function in a set of relevant traffic scenarios is compared to the performance of the vehicle without the function in the same set of traffic scenarios. The simulations of the set of scenarios for the vehicle without the function under test is called the ***baseline***. The selection of relevant, realistic baseline scenarios is of the utmost importance for the quality of the assessment and the results. However, there is little discussion in literature on how to define such a baseline.

The goal of a predictive effectiveness assessment is to make a reliable prediction of the effect a DAS has on traffic safety. The first step for executing such an assessment lies within formulating the evaluation objective, which defines the overall scope of the study [10]. The overall scope of the assessment can be outlined by a safety solution. Within the automotive industry, many of today's safety solutions come from DAS that target crash avoidance but can also consider other in-vehicle systems that potentially increase safety. Moreover, safety solutions can also be represented outside the individual vehicle by either changes in infrastructure or policy decisions (e.g. decreased speed limits). The remainder of this paper will focus on the evaluation of DAS, still the methods can be directly transferred to other safety solutions.

An overview of common relevant terms within in the scope of predictive assessments is given in **Table 1**.

Table 1.
Relevant terms related to prospective safety assessment

Baseline	set of cases to which the performance of the technology under study is compared to when performing prospective assessments
Case	set of specified conditions used as input for the assessment, generally based on concrete scenarios
Simulation model	a computational model which allows the virtual evaluation of the technology, process, or behavior it represents. A simulation model can also contain other simulation models.
Penetration rate	number of vehicles of a certain type equipped with the activated technology under assessment compared to the total number of vehicles of that type in a certain geographic area over a certain period of time.
Research question	question that a research project is designed to answer
Scenario	description of the traffic, infrastructure, and environmental conditions (for example weather and lighting conditions) for the simulation that consists of a sequence of scenes
Scenario category	selection of scenarios that share one or more characteristics

From the initial outline of the evaluation objective, the objective needs to be defined more precisely, by formulating a research question. Formulating this research question requires the consideration of multiple aspects. The assessment should be executed for a defined collection of scenario categories. Herein, each scenario category represents a selection of scenarios that share one or more characteristics. Limitations to the scope of the assessment should be given - e.g., which weather conditions are considered. Moreover, the metric for the evaluation should be stated. A typical metric is the number of crashes avoided by the applied solution. Surrogate measures which may be considered are metrics describing the criticality of a traffic situation.

An example of a precisely formulated research question is: *What is the safety performance of an AEB (warning + autonomous intervention) at a penetration rate of 100% in rear-end car-to-car crashes while approaching an intersection on urban roads in terms of avoided crashes related to the situation in Europe in 2021?*

From the definition of the evaluation objective, the actual assessment can be executed. For this, it is necessary to create a baseline for the assessment, which describes the scenarios to be analyzed without the technology under assessment. This forms the starting point for the simulation with the technology under assessment (treatment). The baseline needs to be created to match the evaluation objective and scope of the assessment. It should represent all relevant elements of the scenarios that potentially have an influence on the performance of the safety solution. The baseline scenarios need to be representative for the safety situation of the baseline condition in the comparison. Therefore, it is necessary to derive the cases from data representing the safety situation under comparison. In the latter, individual concrete scenarios will be referred to as cases.

The choice of data and the process of converting the data to cases as input for the prospective assessment has a large influence on the overall result of the assessment. Differences in the data processing for baseline generation may cause two studies to be incomparable, even though safety solution and models are the same. If baseline approaches can be aligned across different studies, the foundation for a comparison of safety solutions is laid. When creating baseline cases, it needs to be considered, which data is available as input.

Various ways have been presented in literature for setting up a baseline (e.g., in [15], [16], [17], [18], [19], [13], [20], [21], [22], [23], [24], [25]). For harmonization, these ways need to be analyzed and categorized so that recommendations can be given on when and how to use a certain baseline approach. The research objective of this paper is first to develop general approaches to establish a baseline based on existing ways and second to identify areas of application for each baseline approach. All authors of this paper are active in the P.E.A.R.S. consortium [6]. The method presented represents an important common ground of members within P.E.A.R.S., which will lay the foundation for any further harmonization of predictive safety assessment methods, also within the scope of the V4SAFETY Project [9].

In the following, a high-level categorization will be presented, that allows comparison of ways how data may be used for baseline generation. The categorization allows collecting the most suitable baseline approach for the intended assessment. The choice of baseline approach depends heavily on the evaluation objective, the data that is available and the safety solution. For some research questions, such as studies regarding systems which only become active immediately before the crash, cases can be constructed which are close to real-world crashes. Depending on the intended comparison defined by the research question, original cases need to be adapted to a certain extent to enable the intended comparison. After an introduction of the different baseline approaches, a recommendation is given, when which approach is most suitable.

METHOD

In order to come up with different baseline approaches, we propose a method which is based on a high-level categorization taking into account the type of input data source used, how the input data source is used and the processing of the data itself. These categories are structured in different layers, as it can be seen in the schematic view of the method which is shown in **Figure 1**.

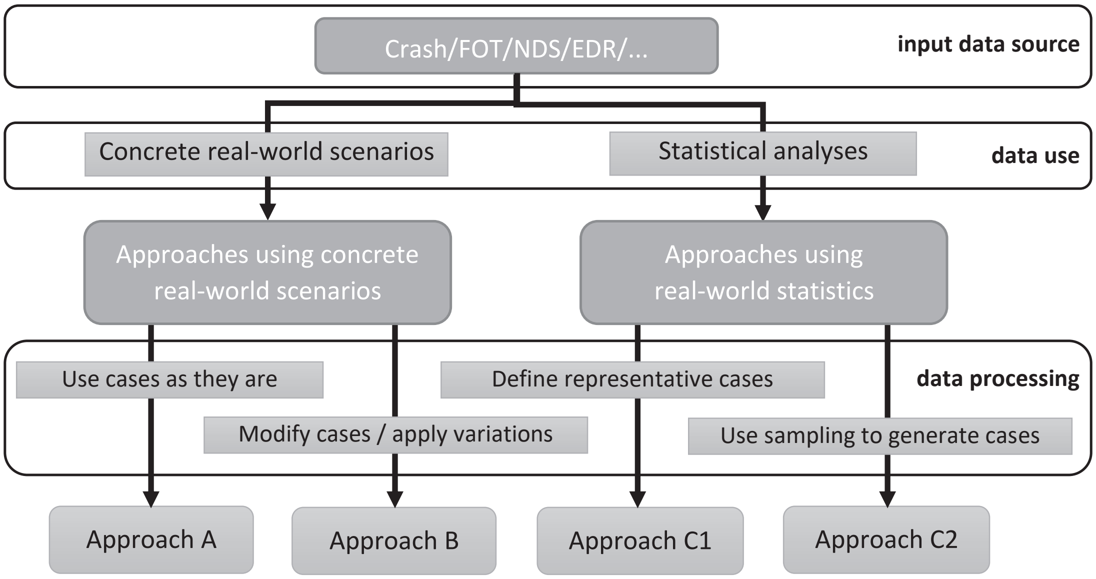


Figure 1. Categorization of the different approaches[1].

---

[1] FOT: field operational test, NDS: naturalistic driving study, EDR: event data recording

Input data source

Generally, input data is associated with reconstructed real-world crashes or with real-world distributions of pre-crash and normal driving conditions [26]. Such input data is needed to generate the baseline scenarios and it can come from one or various data sources, related to crash data, driving data (e.g., FOT, NDS, EDR), experiment data coming from studies in controlled field (e.g., driving simulator, test track) or other sources. It is necessary to understand the type of information that such data contains:

- Crash data contains crash information, considering that a crash consists of any contact with an object, either moving or fixed, at any speed in which kinetic energy is measurably transferred or dissipated. It includes other vehicles, roadside barriers, objects on or off the roadway, pedestrians, cyclists, or animals [27]. Crash data is generally stored in databases (either as aggregated data or case by case), which contain information from real-world crashes gathered by different means, such as detailed investigation teams, police reports, insurance companies, expert reports, or hospital data. Some examples of crash databases are, GIDAS (German In-Depth Accident Study) [28], NASS [29], CARE [30], ITARDA [31], VOIESUR [32], BAAC [33].
- Driving data contains information of real-world traffic situations, covering the range from nominal driving data up to critical data (i.e., near-crash and in-crash). This data may be relevant to derive real-world distributions that can be used to define synthetic scenarios. There are multiple approaches to collect driving data and depending on the source of the recording, it is possible to differentiate:
    - Data from in-vehicle data collection: Either using existing components and sensors from production vehicles or adding additional ones, data is recorded from a vehicle-perspective. Some examples are naturalistic driving studies (NDS) or field operational tests (FOT) [34], [35], [36].
    - Data collected from infrastructure sensors or sensors at a fixed location: Using sensors which are installed under existing or dedicated infrastructure elements, data around a pre-defined area is recorded. Some examples are highway cameras (e.g. focusing on traffic flow measurements), or dedicated cameras at intersections [37]. Recently, the use of drones for driving data collection is also applied [38], providing the flexibility that the sensors do not require to be installed on fixed infrastructure elements, although the data collection approach is similar.
- Experiment data from studies in controlled field encompasses data coming from test track, driving simulator studies or similar experimental data which is gathered in a controlled environment.
- Other: Further data sources may be considered, such as reports, studies, scientific articles which may provide necessary information to build a baseline, as well as complementary sources, such as weather or traffic flow data.

Regardless of the input data type selected, the user performing the prospective safety performance assessment study shall be aware of the details of the data (e.g., quality, data sampling procedure, representativeness), and shall document the used input data sources to ensure transparency of the assessment. This documentation shall contain not only information on the input data used, but also on the selection process of the data, which shall be described as well.

Data use

Once the main input data type has been selected, it is necessary to decide whether to use data directly as it is, or if it is going to be analyzed to derive statistical information from it, such as real-world distributions of relevant parameters. Depending on the selected approach, a distinction can be made between:

- Concrete real-world cases: This approach consists in using the data as it is, so a link exists between the original real-world case and the input data used (e.g., a concrete crash from a crash database is used as input without altering the original recorded case, or a near crash scenario recorded during driving is replayed in the simulation). In some cases, in this step a selection of cases, based on inclusion and/or exclusion criteria may be applied. One example may be to use a group of cases of a crash database [15].
- Results from statistical analysis to determine the baseline: This approach consists in using real-world data as a source for a statistical analysis on the crash/driving data to provide statistics of certain parameters. In a second step, these are used to derive (real-world based) synthetic cases: cases are generated by using distributions of parameter values instead of recorded or reconstructed values and choosing plausible (physically) combinations thereof [24], [13], [21]. In this approach there is only an in-direct link between the synthetic case derived and the original input data.

Data processing

Based on the selected input data and its use, the next step is to consider if any data processing step needs to be applied on the data or not. We distinguish the following options:

For the use of concrete real-world cases:

- Use cases as they are: Collected real-world scenarios are digitized and used as baseline cases. No further modifications or assumptions are applied, besides the ones related to how the data has been collected [15], [17].
- Modify cases / apply variations: The real-world case is the reference, but modifications or variations are applied to it. A distinction is made between:
  - Modification of real-world cases: The complemented information (e.g. speed) can come from an individual value (e.g., posted speed limit) or from distributions of data (e.g., accident data analysis)[18].
  - Variation of the original real-world cases: in this case, a variation of parameter values from the real-world case is applied [20], [19].

  Although information is added or parameters are varied, there is a strong link with the original real-world cases.

For the usage of statistical data to derive synthetic cases, we distinguish two possibilities:

- Define representative cases: In this case, statistical data from various sources is considered to build a limited set of representative cases [21], [22].
- Use sampling to generate cases: The approach consists in generating a (usually) large number of cases following a sampling scheme, in order to cover the whole range of relevant cases [24], [13].

  In the above-mentioned approaches, no direct link exists between the derived synthetic cases and the original data.

As shown in Figure 1, depending on the input data source used, the use of this data and its processing, we identified four different baseline approaches: approach A, approach B, approach C1 and approach C2. A detailed description of the approaches follows in the next section "Results".

RESULTS

As an output of the previously described method, three main types of baseline creation approaches can be distinguished. A graphical overview of the different approaches is given in **Figure 2**, a detailed description follows below.

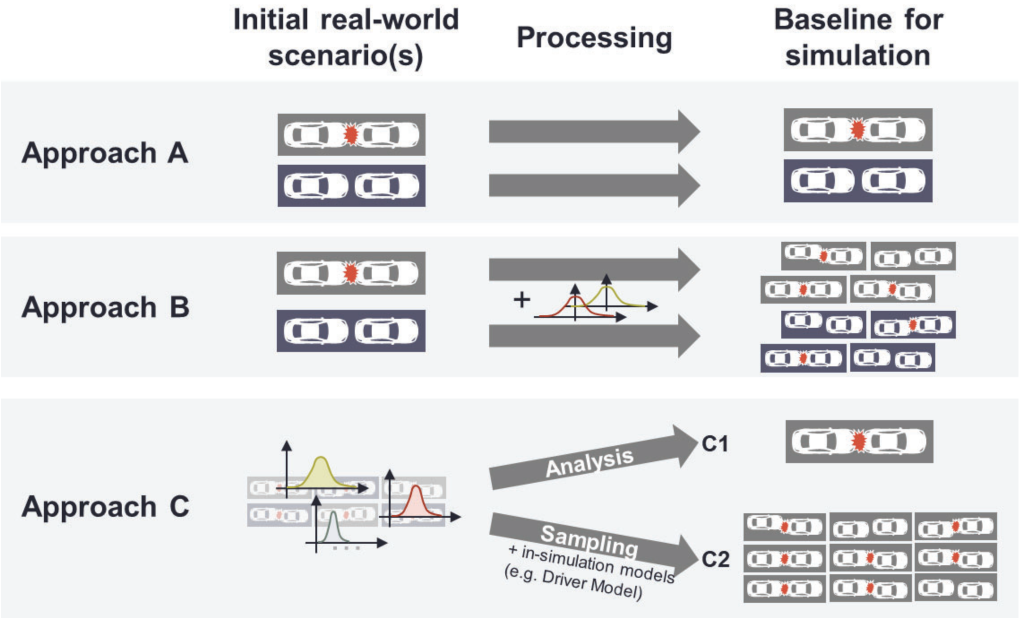


Figure 2. The three baseline approaches, image taken from [11]*(under development).*

Approach A: Digitized real-world scenarios without modification
In this approach, individual real-world scenarios are digitized and used as baseline cases without altering them. The data sources are usually databases consisting of recorded driving data or reconstructed crashes. From these data sources, time series of the dynamic elements (vehicle under test (VuT), surrounding traffic) as well as positions of relevant infrastructural elements and information such as weather, lighting conditions etc. are used to set up each test case, replicating the real-world situation. Although the digitized cases and the possibly resulting crash consequences could in principle be taken directly as baseline results, the authors strongly recommend performing the following steps during safety performance assessment:

- Use the time-series data for the dynamic elements as input for re-simulations in the to be used simulation tool. The results of these re-simulations are the baseline results.
- For the treatment, repeat the baseline re-simulations but with the technology under assessment present. The results are treatment simulation results.
- Derive the differences between both results in a case-by-case analysis as basis for the safety performance assessment.

By this, it is ensured that the differences in the safety performance only result from the influence of technology under assessment and these differences are not artifacts caused by using different methods for obtaining baseline and treatment results.

Approach B: Modification or variation of real-world scenarios
In this approach, real-world scenarios are the basis as well, but modifications to the original data for building the required baseline are made in order to alter existing properties or even to add new ones, e.g., to be able to use an older or less complete dataset and to modify it towards the current-state-of-the-art.

Data sources are usually databases consisting of recorded driving data or reconstructed crashes. In contrast to the previous approach, in this approach the data from the database is enriched by:

- Adding parameters to compensate for unavailable information: The reason for the addition can be missing information from the original real-world case that is needed to define the baseline (e.g., missing speed information). Another aim of this approach is to update the existing data to a specific state of technology. This can be done by using additional in-simulation models (e.g., driver models or technology models such as ABS) and re-simulating the original scenario. In this way, an updated version of an original real-world scenario is generated.
- Modifying existing parameters to compensate for uncertain information: This can be achieved by, for example, adding variations of known parameters to create multiple variants of one single real-world scenario.

In this approach, the number of baseline cases is equal to or higher than the number of real-world scenarios considered in the set-up. The number of treatment simulations is the same as the number of baseline simulations.

Approach C: Baseline consisting of synthetic cases
In this approach statistical information of real-world data is used instead of individual real-world scenarios. This information is used to determine distributions of traffic relevant parameters such as speed, time headway, braking behavior etc. With these distributions synthetic cases can be set up which represent what happens in the real world. These synthetic cases consist of trajectories of all traffic participants of interest for the specific case. Two variations in the application of approach C are distinguished:

- Approach C1: Here the scenario statistics are analyzed to generate a limited set of test cases that are representative for the scenarios that the function under test will encounter in the real world.
- Approach C2: Here the statistical information can be used to set up a usually large number of cases that cover not only the most typical situations but also the rare combinations of parameters (large test space). The generation of cases for the second variant of this approach could be done using sampling techniques or by using a simulation including models that describe human behavior, such as driver models or pedestrian behavior models. The chosen sampling method will determine the number of test cases that result from this approach. Another distinction of this second variant is that here not necessarily the same cases need to be re-simulated in the treatment simulations. Even the number of simulations for baseline and treatment simulation can be different as long as it is ensured that the number of simulations is large enough to ensure a stable result.

Although C1 and C2 differ, they are grouped in approach C as the main difference is only the number of cases, not the methodology. Approaches A, B, and C are distinguished by the use of a different methodology.

Examples for the different baseline approaches

Existing, published ways of setting up a baseline can be attributed to one of these approaches. Following are descriptions of examples for each of the approaches.

Approach A

- In [15], an AEB-Pedestrian was evaluated by virtual simulation of accidents selected from the French accident database, VOIESUR[32]. The database gathers all fatal accidents and a sample of 5% of all injury accidents that occurred in France in 2011, proportionally distributed over the whole French mainland territory. The database is weighted to represent accident severity, involved user type, conflict type and location (by proxy) as described in the national census BAAC[33]. The accident subset used for virtual simulation consists of accidents in which a pedestrian (any age) was hit by the front of a passenger car and pre-crash trajectory / impact speed was available or could be reconstructed from e.g., projection distances contained in reports. Exclusions were: car in loss of control situation, side swipes, pedestrians lying on the road prior to impact and suicides. The final weighted sample consists of 197 fatally injured, 1863 severely injured and 3103 slightly injured pedestrians with injury severity scale of the Police (not AIS).
- SIMPATO (Safety IMPact Assessment Tool)[16]: The SIMPATO tool was used in the EU-funded project "interactIVe" that developed active safety systems for multiple conflict types. The SIMPATO tool focuses on those conflict types that were addressed by most of the interactIVe systems, namely rear-end and run-off conflicts. For the rear-end conflicts, 364 real-world crashes of the GIDAS database have been analyzed. The simulations have been conducted for systems that warned the driver and/or reacted by means of braking or evasive maneuver. The models for the systems' reaction were derived from the interactIVe track tests. For the run-off road conflict, 150 GIDAS accidents were considered. Here, the interactIVe system reaction was always a steering maneuver.

Approach B

- Reference [18] shows an example for adding missing information according to Approach B. The main input data type used is real-world crash data, from a police reports accident database in Germany. The data contains information on the accident conflict situation, collision configuration, geo-coordinates of accident, participants involved as well as injury level of each participant, among other variables. The database does not contain time series information such as trajectories and speed. Trajectory information is added based on the description of the police report. Speed profile information is added based on a statistical analysis of an in-depth accident database (GIDAS), based on participant maneuver, accident location, participant type and injury level sustained. Driving speed, collision speed and deceleration value are extracted to define the speed profile (mean values are assigned for each parameter). Based on the added information, the simulation files can be created. Plausibility checks are done to confirm that the collision is realistic, and that the collision configuration is as reported in the police report.
- An example for adding missing information and creating variations is given in [19]. The main input data type used is real-world crash data, from a police reports accident database in Germany. The data contains information on the accident conflict situation, collision configuration, geo-coordinates of accidents, participants involved as well as injury level of each participant, among other variables. The database does not contain time series information such as trajectories and speed. Trajectory information is added based on the description of the police report. Speed profile information is added based on a statistical analysis of another accident database (GIDAS). The statistical analysis provides 3 variations of 3 parameters (driving speed, collision speed and deceleration value). Considering the 2 participants involved in the accident, a maximum of 729 variations can be generated per each accident. Only the simulation files that confirm a collision exists are considered as plausible data.
- In the L3Pilot safety impact assessment [13] the DAS developed in the project were assessed. The project covered different types of DAS, namely a motorway, an urban and parking DAS. For the motorway DAS, two of the baseline approaches were used. Approach B was used in the counterfactual simulations of rear-end and cut-in conflicts. The real-world cases came from the dataset involving crashes with Volvos (VCTAD [39]), crashes from the Traffic Accident Scenario Community (TASC [18]) database and critical situations from the SHRP2 database[36]. By means of the critical driving scenarios the false positive behavior of the DAS was assessed.

Approach C1

- The CATS project (2014 - 2016) [21] provided a proposal for a test matrix towards Euro NCAP for the testing and safety rating of Autonomous Emergency Braking Systems onboard passenger cars that are capable to avoid or mitigate collisions with cyclists. By studying car-to-cyclist accidents in the EU, obtained from data of France, Germany, Italy, the Netherlands, Sweden, and the United Kingdom, the five most common scenarios for accidents between passenger cars and cyclists were selected. These scenarios describe the trajectories and maneuvers of cyclists and cars for several seconds up to the moment of impact. Next step was to construct test scenarios for the three most dominant accident scenarios. An in-depth study into the accidents was conducted to determine the most relevant parameters and the most relevant ranges of these parameters. Additional input was collected from observation studies that were conducted on specific locations in the Netherlands where many interactions (without collisions) between cyclists and passenger cars were observed. These studies revealed the influence on the cyclist and vehicle speed in an approach of an intersection in the presence of a strong view-blocking obstruction. Based on the accidentology and test parameter studies described above, the set of baseline tests has been proposed. Relevant and realistic test cases were provided based on statistical analysis of thousands of accidents.

Approach C2

- L3Pilot safety impact assessment [13]: Next to the counterfactual simulation described above, which applies baseline approach B, L3Pilot used also approach C2 to assess the safety impact of automated driving. The C2 approach had been applied in typical crash scenario as well as scenarios that pose a challenge for the DAS. For the motorway the following scenarios were considered: lane change conflict, conflict with VRU, minimum risk maneuver, wrong activation, end of lane, obstacle in the lane, lower speed limit and passing a motorway entrance. The number of analyzed cases per scenario varied depending on the considered infrastructure and traffic parameters. Overall, more than 25 000 cases were simulated. The C2 approach was also applied for the urban DAS. Here, all scenarios were generated with a stochastic sampling approach using copulas, which was presented in [40]. Input to the generation of the simulation cases were different sources including accident data, traffic data and data from L3Pilot pilot studies.
- In [23], the effectiveness of a pedestrian protection system implemented in a vehicle was studied. The analyzed scenario was a pedestrian crossing a street unauthorized at an unprotected location. To this end, warning, automated emergency braking and a combination of both were evaluated for varying parametrizations of the algorithm. The approach C2 was employed since both the traffic on the street as well as the pedestrian crossing the street were simulated. The goal of the simulation was to replicate the risk in the described scenario as precisely as possible. Hence, not every simulation resulted in an accident. In order to resolve statistically significant differences in virtual accident numbers, 18 million crossings were simulated in the baseline and 100 million crossings were simulated for the treatment.
- In [25], the effect of a simplified automated driving function with / without external information by an infrastructure-based LiDAR sensor is analyzed. The specific scenario that is studied is a right-turn scenario: cars are turning right and have to yield to straight going cyclists. Thereby, occlusion due to parked cars and a construction site was present. The approach C2 is used in this publication since the authors replicate the crash causation mechanisms in the described scenario and use traffic simulation to create scenarios. In order to resolve an effect size of 10%, 200 million cyclist crossings were simulated in the baseline and for each of the 3 different levels of treatment.

DISCUSSION AND LIMITATIONS

Each of the presented baseline approaches has its advantages and disadvantages, e.g., with respect to the power to generalize to the overall population (“representativeness”) or the demand on the required data. In the following these aspects are discussed.

Approach A has the advantage, that it is based on real-world driving or crash scenarios. Compared to the other approaches, it is relatively easy to derive the baseline from these unaltered scenarios. Basically, the baseline cases can be used directly without any change. The approach might require a conversion of the real-world cases into the required format. However, the scenario should not be changed in terms of trajectories of the involved traffic participants. In the simulation, it needs to be ensured that the traffic participants follow the original real-world trajectory. This is typically done via a so-called trajectory following model. The approach does not need any complex driver behavior models.

The question how likely it is that reconstructed characteristics of an investigated accident will happen again in a comparable manner, can be argued. But the investigated cases represent realistic crash configurations. In this context, we distinguish between cases that are measured (typically NDS or FOT data) and cases that are derived from reconstruction (typically in detailed accident databases). In the latter case the quality and number of available variables of the reconstruction defines how well the case represents the reality and its applicability to investigate safety effects. It is especially difficult to reconstruct the cognitive state of the involved drivers. However, the effect of an ADAS system is often sensitive to these variables.

An often-encountered issue is that there are not enough real-world cases to derive statistically meaningful results. In contrast to the approaches B and C, the number of cases cannot be increased above the number of cases in the used database. This issue is in particular relevant when simulating cases for which crash data is used. In case driving scenarios are derived from NDS or FOT, this issue might be less relevant. However, NDS and FOT data contain typically only a very low number of crashes – if they include crashes at all. This need to be accounted when doing the assessment.

For approach A, it must be considered that the covered time frame of cases – in particular for crash cases – is typically limited (e.g., cases in GIDAS PCM database cover up to 5 s). Therefore, it must be checked whether the effect of the technology under assessment can be evaluated appropriately in this time frame. If the time period of the case is insufficient, one can switch to another approach (e.g., B or C2). Approach B allows to make changes to the original scenarios, which allows also to extend the scenario. However, this extension in approach B will be limited to a couple of seconds since the link to the original case needs still to be given. Approach C2 offers in terms of simulation time more freedom, since typically only the start conditions need to be defined. However, crash mechanisms from the real-world must be still represented correctly in the baseline.

Approach B combines the realism of approach A with the flexibility to adapt the case to specific needs. The needs could have different facets. This could be for instance to consider in the baseline additional technologies to make the cases more representative for today's traffic, to complete missing parameters of the used databases, to enlarge the test space in safety assessment or to consider variations to account for possible shortcomings in the reconstruction of cases. Due to the close relation with approach A, most of the advantages and disadvantages apply also for this approach. For instance, also for approach B the number of suitable real-world cases might be quite low. By variation, the number of simulated cases can be increased. However, the representativity would not be changed, since the number of original baseline cases would stay the same.

The possibility of variation offers quite some degrees of freedom for the assessment. Therefore, approach B can be applied for many different evaluation objectives. In this sense it can be seen as an evolution of approach A. However, it must be considered that the degrees of freedom come with the challenge of ensuring the cases resulting from variations are plausible and representative. For instance, if the driver reaction is changed from the original baseline case, it must be checked how probable this variation is under the traffic conditions in the baseline case to assure adequate weighting in the statistical analysis of the data. The variation of the driver behavior would also mean that the trajectory following model is not sufficient any longer and a more sophisticated model is required.

Baseline approaches C1 and C2 are also quite different from the approaches A and B. Since they rely only on distributions sampled from real-world data and the understanding of crash causation mechanisms and not the concrete real-world data time series, the number of baseline cases can be chosen arbitrarily. Both C approaches (C1 and C2) mainly differ in the number of considered cases for the analysis. While C1 investigates a very limited number of cases, C2 assesses typically a quite high number of cases. Dealing with a high number of cases is typically less a challenge in a virtual assessment than in real-world testing. Thus, it is for C2 much easier to reach a sufficient number of cases for a statistically sound comparison between baseline and treatment. This leads for the virtual assessment often to the choice of C2, since it allows to cover a large scenario space. Nevertheless, there are few evaluation objectives, in which C1 is the choice for the assessment. One example is the round-robin simulation of P.E.A.R.S. [9], in which the difference between several simulation tools in the same simulated scenario is investigated. But it must be considered that this study did not investigate the safety performance of technology. Other evaluation objectives in which C1 would be useful are comparisons of virtual simulation with real-world tests, for which a high number of tests would increase the effort heavily, or in case the simulations of cases are very heavy on computational effort.

The sampling from parameter distributions – if done in a sufficient manner and resulting in much more cases than available in the real-world data source – allows also for a wide coverage of the scenario space ranging from crash via critical scenarios to normal driving. In general, all baseline approaches can cover critical scenarios that both did and did not lead to collision. Collision cases aim at investigating of true positive behavior (cases which required an activation by the technology and in which the technology became active) and false negative behavior (cases which required an activation, but the technology did not become active). By means of non-collision cases false positive (cases which did not require an activation, but in which the technology became active) as well as true negative behavior (cases which did not require an activation and the technology was not activated) can be analyzed. As discussed, exemplary in [23], the rate of false positive activations influences the effectiveness of ADAS systems, since a high number of unnecessary warnings lead to the deactivation of the system by the driver, which clearly undermines the intended positive effects of the system. With respect to the baseline approaches, in the approach C1 and C2 non-collision case can be derived even from databases that contain only collision cases, while approach A and B required the consideration of data source that contains non-collision case. Furthermore, for C2 the non-collision cases are derived implicitly by the number of considered cases and use generation process. For the other baseline approaches A, B and C1, the decision about consideration of non-collision cases in the analysis needs to be made explicitly (e.g. by choosing the real-world case or in the generation).

For C1 and C2 the duration of the simulated cases can also be chosen, which provides an advantage for investigations of a technology that intervenes into the driving dynamics of the vehicles for longer time periods. Surrounding traffic participants can also be considered in baseline approach C1 and C2, although it must be noted that each traffic participant increases the effort and complexity of the simulation. This is in particular of relevance for more complex technology, like e.g. automated driving: the reaction of the technology might depend on the surrounding traffic and vice versa, the flow of the surrounding traffic might depend on the reaction of the technology [41]. But also for ADAS this could be of relevance, e.g. when checking for secondary effects like does the AEB braking lead to more rear-end collisions with the following traffic. For approach B the consideration of surrounding traffic as one variation parameter is also feasible but increases the requirements for the simulation models that represent the traffic in this approach quite heavily. For approach A this is not feasible.

The approaches C1 and in particular C2 offers a very high degree of freedom. However, these opportunities do not come without challenges. For this approach it is vital to understand the mechanisms leading to a crash and the underlying parameter and distributions to ensure that the simulated case represent the real-world cases sufficiently. This requires a deep understanding of the used data as well as of the simulation models, especially whether they reproduce the underlying crash causation mechanisms. A key model for this approach is the driver behavior model, since it decides how critical a case is going to be and what the resulting crashes variables are. A simple trajectory following model is not enough for this approach. Rather a sophisticated driver model is required here. Thus, the big question when using the approach results from the fact that there is no direct link to original real-world cases: does the simulation produce realistic cases? To answer this, an increased effort for the validation and verification of the simulation models and the scenario is required compared to the other approaches. Moreover, this approach requires a high amount of several input data for generating the input data's distributions, which establish the link to the real-world scenarios.

The evaluation objective to be addressed needs to be considered when selecting an appropriate baseline approach as it has a strong influence on this selection. **Table 2** provides exemplary research questions in which the authors would apply a certain approach. The common theme is car-to-car rear-end collisions in an urban environment. The choice of one research question does not mean that no other approach would be suitable. But other approaches were not the preferred option by the authors.

Table 2.
Exemplary possible research questions for the different baseline approaches

Baseline Approach	Exemplary Research Question	Rationale
A	What is the safety performance of an AEB (warning + autonomous intervention) at a penetration rate of 100% in car-to-car crashes on urban roads in terms of MAIS 2+ injuries related to the situation in Germany from 2015 - 2017 as represented in GIDAS PCM?	Approach A is preferred since due to the specific naming of the country, time frame and database to be used. The question implies rather to investigate the performance in particular cases than in the general safety performance of an AEB.
B	What is the safety performance of an AEB (warning + autonomous intervention) at a penetration rate of 100% in car-to-car crashes on urban roads in terms of MAIS 2+ injuries related to the situation in Germany from 2010 - 2017 as represented in GIDAS PCM while considering only ESC-	Approach B is preferred since a direct link to certain crash data is desired (GIDAS-PCM) similar as in the previous example. However, here an altering of the baseline cases is required for those cases in which the vehicle was not equipped with ESC. Now the vehicle needs to be equipped with AEB. The second aspect that hints towards approach B is the consideration of different road frictions. This

	equipped vehicles and considering different road friction?	could also be easier achieved in B by varying the baseline than in approach A, since the different road frictions will not be presented equally, i.e., it is likely that the database contains too few cases with low road frictions.
C1	What is the safety performance of a VRU AEB (warning + autonomous intervention) in car-to-car crashes as defined in the Euro NCAP protocol?	This research question requires only a few simulations, and consequently approach C1 has been chosen. However, the main task is rather to get to the representative crash cases. This step has been done by experts of Euro NCAP when defining the test protocols.
C2	What is the safety performance of an AEB (warning + autonomous intervention) at a penetration rate of 100% in rear-end car to car crashes while approaching an intersection on urban roads in terms of avoided crashes related to the situation in Europe in 2017?	For this research question, Approach C2 is recommended as an insufficient number of real-world crashes is likely to exist in databases. For some countries a reasonable number of crashes in databases might exist. However, for other counties this is not the case. Distributions that are describing the general traffic behavior can be used to generate such conflict cases for different countries.

CONCLUSIONS

This paper presents a methodological analysis of different ways to set up a baseline for prospective safety performance assessment. We found three main elements of the set-up process: input data source selection, data use, and data processing. We distinguish three main approaches for setting up a baseline depending on the choices taken in each of these elements. These approaches should cover any baseline set-up process, some examples from literature are presented in the results section.

The various ways presented in literature for setting up a baseline for prospective safety performance assessment can be attributed to one of these approaches with this methodology. This will help to understand what has been done in past studies and increases comparability and trustworthiness of past and future studies in this field.

Moreover, the paper discusses the advantages and disadvantages of the different approaches as well as the dependency of the approach selection on the evaluation objective of a safety performance assessment study, the data that is available and the safety solution. This will help the readers in the selection of a suitable baseline approach for future studies.

With this work, the authors and P.E.A.R.S. as a whole contribute to the harmonization and acceptance of virtual safety performance assessments of DAS and ADAS. This will greatly enhance trustworthiness, comparability and, transparency of results of such assessments. Furthermore, these baseline set-up approaches will be part of [11] (under development).

REFERENCES

[1] European Commission. 2021a. *European Road Safety Observatory Annual statistical report on road safety in the EU 2021*.

[2] European Commission. 2020b. *Sustainable and Smart Mobility Strategy – putting European transport on track for the future*.

[3] Piao, J., & McDonald, M. 2008c. "Advanced driver assistance systems from autonomous to cooperative approach". *Transport Reviews*, *28*(5), 659–684. https://doi.org/10.1080/01441640801987825

[4] Horizon 2020 Commission Expert Group to advise on specific ethical issues raised by driverless mobility (E03659). 2020d. *Ethics of Connected and Automated Vehicles: recommendations on road safety, privacy, fairness, explainability and responsibility*.

[5] Sakura, SIP-adus, & HEADSTART. 2021e. *Towards the harmonization of safety assessment methods of automated driving*. https://www.headstart-project.eu/wp-content/uploads/2021/12/211209_HEADSTARTSAKURA_WP_Final-draft.pdf

[6] V4SAFETY. 2022f. *Vehicles and VRU Virtual eValuation of Road Safety*. EU Horizon Europe Funded Project. https://www.linkedin.com/company/v4safety/

[7] European Commission. 2022g. "Commission Implementing Regulation (EU) 2022/1426, uniform procedures and technical specifications for the type-approval of the automated driving system (ADS) of fully automated vehicles". *Official Journal of the European Union*, *L221*, 1–64.

[8] United Nations. 2021h. *Regulation No 157, uniform provisions concerning the approval of vehicles with regards to Automated Lane Keeping Systems [2021/389]*.

[9] Wimmer, P., Düring, M., Chajmowicz, H., Granum, F., King, J., Kolk, H., Op Den Camp, O., Scognamiglio, P., & Wagner, M. 2019i. "Toward harmonizing prospective effectiveness assessment for road safety: Comparing tools in standard test case simulations". *Traffic Injury Prevention*, *20*(Supplement 1: Peer-Reviewed Journal for the 26th International Technical Conference on the Enhanced Safety of Vehicles (ESV)), 139–145. https://doi.org/10.1080/15389588.2019.1616086

[10] ISO. 2021j. *ISO/TR 21934-1:2021 Road vehicles — Prospective safety performance assessment of pre-crash technology by virtual simulation — Part 1: State-of-the-art and general method overview*. https://www.iso.org/standard/76497.html

[11] ISO. n.d.-k. *ISO/AWI TS 21934-2: Road vehicles — Prospective safety performance assessment of pre-crash technology by virtual simulation — Part 2: Guidelines for application (under development)*. https://www.iso.org/standard/81790.html

[12] Euro NCAP. 2017l. *Euro NCAP 2030 Roadmap*. https://cdn.euroncap.com/media/74190/euro-ncap-vision-2030.pdf

[13] Bjorvatn, A., Page, Y., Fahrenkrog, F., Weber, H., Aittoniemi, E., Heum, P., Lehtonen, E., Silla, A., & Bärgman, J. 2021m. *L3Pilot D7.4-Impact Evaluation Results*. https://l3pilot.eu/fileadmin/user_upload/Downloads/Deliverables/Update_14102021/L3Pilot-SP7-D7.4-Impact_Evaluation_Results-v1.0-for_website.pdf

[14] Harnarayan, S. S. M.-M. P. T. 2021n. "Simulation Based Virtual Testing for Safety of ADAS Algorithms". In SAE International (Ed.), *SAE WCX Digital Summit*.

[15] Saadé, J., Chajmowicz, H., & Cuny, S. 2019o. "Prospective Evaluation of the Effectiveness of Autonomous Emergency Braking Systems in Increasing Pedestrian Road Safety in France.". *IRCOBI*. http://www.ircobi.org/wordpress/downloads/irc19/pdf-files/37.pdf

[16] Van Noort, M., Bakri, T., Fahrenkrog, F., & Dobberstein, J. 2015p. "SIMPATO-The safety impact assessment tool of interactive". *IEEE Intelligent Transportation Systems Magazine*, *7*(1), 80–90. https://doi.org/10.1109/MITS.2014.2340054

[17] Hasegawa, T., Takahashi, H., & Udaka, S. 2017q. "Clarification of Priority Factors for Reducing Traffic Accident of Fatalities in the U.S. and Benefit Estimation of AEB System for Oncoming Vehicles". *25th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*. https://www-esv.nhtsa.dot.gov/Proceedings/25/25ESV-000171.pdf

[18] Urban, M., Erbsmehl, C., Lorente Mallada, J., Puente Guillen, P., & Tanigushi, S. 2020r. "A Methodology for Building Simulation Files from Police Recorded Accident Data (for ADAS Effectiveness Assessment)". *FISITA World Congress*. https://www.fisita.com/library/fisita-world-congress/2020/f2020-pif-039

[19] Urban, M., Erbsmehl, C., Landgraf, T., Pohle, M., Lorente Mallada, J., Puente Guillen, P., & Tanigushi, S. 2020s. "Parameterization of Standard Test Scenarios of Automated Vehicles Using Accident Simulation Data". *FISITA World Congress*. https://www.fisita.com/library/fisita-world-congress/2020/f2020-pif-038

[20] Leledakis, A., Lindman, M., Östh, J., Wågström, L., Davidsson, J., & Jakobsson, L. 2021t. "A method for predicting crash configurations using counterfactual simulations and real-world data". *Accident Analysis & Prevention*, *150*, 105932. https://doi.org/10.1016/J.AAP.2020.105932

[21] Op den Camp, O., van Montfort, S., Uittenbogaard, J., & Welten, J. 2017u. "Cyclist Target and Test Setup for Evaluation of Cyclist-Autonomous Emergency Braking". *International Journal of Automotive Technology*, *18*(6), 1085–1097. https://doi.org/10.1007/s12239

[22] Stoll, J., Schneider, A., Wisch, M., Seiniger, P., & Schaller, T. 2016v. *PROSPECT D3.1-The addressed VRU scenarios within PROSPECT and associated test catalogue*. https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5a90e757e&appId=PPGMS

[23] Helmer, T. 2014w. *Development of a methodology for the evaluation of active safety using the example of preventive pedestrian protection* [TU Berlin]. https://depositonce.tu-berlin.de/handle/11303/4270

[24] Wang, L., Fahrenkrog, F., Vogt, T., Jung, O., & Kates, R. 2017x. "Prospective safety assessment of highly automated driving functions using stochastic traffic simulation". *25th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*, 1–13. https://www-esv.nhtsa.dot.gov/Proceedings/25/25ESV-000224.pdf

[25] Denk, F., Brunner, P., Huber, W., Margreiter, M., Bogenberger, K., & Kates, R. 2022y. "Assessment of traffic safety interventions using virtual randomized controlled trials: potential of connected and automated driving including V2X for collision reduction at urban intersections". *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, 1183–1190. https://doi.org/10.1109/ITSC55140.2022.9921764

[26] Alvarez, S., Page, Y., Sander, U., Fahrenkrog, F., Helmer, T., Jung, O., Hermitte, T., Duering, M., Doering, S., & Op den Camp, O. 2017z. "Prospective Effectiveness Assessment of ADAS and Active Safety Systems via Virtual Simulation: A Review of the Current Practices". *25th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*. https://www-esv.nhtsa.dot.gov/Proceedings/25/25ESV-000346.pdf

[27] Guo, F., Klauer, S., McGill, M., & Dingus, T. 2010aa. *Evaluating the Relationship Between Near-Crashes and Crashes: Can Near-Crashes Serve as a Surrogate Safety Metric for Crashes?* https://www.nhtsa.gov/sites/nhtsa.gov/files/documents/dot_hs_811_382.pdf

[28] *GIDAS*. n.d.-ab. Https://Www.Gidas.Org/Start-En.Html.

[29] NHTSA. n.d.-ac. *NASS*. Https://Www.Nhtsa.Gov/Crash-Data-Systems/National-Automotive-Sampling-System.

[30] European Commission. n.d.-ad. *CARE*. Https://Road-Safety.Transport.Ec.Europa.Eu/Statistics-and-Analysis/Methodology-and-Research/Care-Database_en.

[31] ITARDA. n.d.-ae. *ITARDA*. Https://Www.Itarda.or.Jp/English/.

[32] Lesire, P., Léopold, F., Chauvel, C., Hervé, V., & Cuny, S. 2015af. "Implication of Children in Road Accidents in France in 2011". *24th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*.

[33] Amoros, E., Lardy, A., Wu, D., Viallon, V., & Martin, J.-L. 2015ag. *Projet VOIESUR - Livrable L3 - Méthodologie : redressement et extrapolation, projet VOIESUR - ANR*. https://hal.archives-ouvertes.fr/hal-01212490

[34] Innamaa, S., Aittoniemi, E., Bjorvatn, A., Fahrenkrog, F., Gwehenberger, J., Lehtonen, E., Louw, T., Malin, F., Penttinen, M., Schindhelm, R., Silla, A., Weber, H., Borrack, M., Lillo, L. Di, Merat, N., Metz, B., Page, Y., Shi, E., & Sintonen, H. 2020ah. *L3Pilot D3.4-Evaluation Plan*. https://l3pilot.eu/fileadmin/user_upload/Downloads/Deliverables/Update_07102021/L3Pilot-SP3-D3.4-Evaluation_plan-v1.0_for_website.pdf

[35] Eenik, R., Barnard, Y., Baumann, M., Augros, X., & Utesch, F. 2014ai. "UDRIVE: the European naturalistic driving study". *Transport Research Arena*.

[36] Hedlund, J. 2015aj. "SHRP 2 naturalistic driving study: The first data analysis studies". *Journal of Safety Research*, *54*, 3–4.

[37] Nakamura, H., Muslim, H., Kato, R., Préfontaine-Watanabe, S., Nakamura, H., Kaneko, H., Imanaga, H., Antona-Makoshi, J., Kitajima, S., Uchida, N., Kitahara, E.Ozawa, K., & Taniguchi, S. 2022ak. "Defining Reasonably Foreseeable Parameter Ranges Using Real-World Traffic Data for Scenario-Based Safety Assessment of Automated Vehicles". *IEEE Access*, *10*, 37743–37760. https://doi.org/10.1109/ACCESS.2022.3162601

[38] Krajewski, R., Bock, J., Kloeker, L., & Eckstein, L. 2018al. "The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems". In IEEE (Ed.), *21st International Conference on Intelligent Transportation Systems (ITSC)* (pp. 2118–2125). https://doi.org/10.1109/ITSC.2018.8569552

[39] Isaksson-Hellman, I., & Norin, H. 2005am. "How thirty years of focused safety development has influenced injury outcome in volvo cars". *Annual Proceedings. Association for the Advancement of Automotive Medicine*, *49*, 63–77. https://pubmed.ncbi.nlm.nih.gov/16179140/

[40] Rösener, C. 2020an. *A Traffic-based Method for Safety Impact Assessment of Road Vehicle Automation* [RWTH Aachen University]. https://doi.org/10.18154/RWTH- 2020-08950

[41] Helmer, T., Kompaß, K., Wang, L., Kühbeck, T., & Kates, R. 2017ao. "Safety Performance Assessment of Assisted and Automated Driving in Traffic: Simulation as Knowledge Synthesis". In *Automated Driving* (pp. 473–494). Springer International Publishing. https://doi.org/10.1007/978-3-319-31895-0_20

VALIDATION AND PLAUSIBILIZATION OF X-IN-THE-LOOP TESTS FOR DRIVING AUTOMATION

Felix Reisgys
Johannes Plaum
Dr. Andreas Schwarzhaupt
Daimler Truck AG
Germany
Prof. Dr.-Ing. Eric Sax
Institut für Technik der Informationsverarbeitung, Karlsruhe Institute of Technology
Germany

Paper Number 23-0064

ABSTRACT

Virtual X-in-the-Loop (XiL) environments are gaining significant importance in the test of Advanced Driver Assistance Systems (ADAS) and Automated Driving Systems (ADS). In order to derive reliable test results, credibility of XiL environments must be evaluated using suitable methods for XiL validation. This typically involves a back-to-back comparison to reference proving ground (PG) tests. Due to uncertainties inherent to PG and XiL, this validation requires the analysis of multiple test executions. Since this may not always be feasible with limited data availability, we define plausibilization as a preliminary step towards validation, comparing two single test executions in PG and XiL. A plausibilization method is presented, combining the evaluation of pass/fail criteria (PFC) and scenario distance measures. Finally, the application of the method in an ADAS series development project by evaluating three example Software-in-the-Loop (SiL) scenarios confirms that this is a reasonable plausibilization approach. Furthermore, it is shown that the method can be adjusted in a flexible way to meet requirements from different automation levels, systems or scenarios.

INTRODUCTION

In recent years, the scope of driving automation has increased significantly to improve vehicle safety, driver comfort and efficiency. Given the growing capabilities of both ADAS and ADS, the required testing depth and width increases as well. Current test strategies will not be able to cover future testing demands for safety validation as they rely heavily on real-world testing [1]. It is therefore necessary to make use of alternative testing methods. A promising approach pursued in research and industry is scenario-based testing in combination with XiL environments [2]. While XiL is a key enabler for efficient scaling of test volume, the credibility of XiL environments has a major impact on the potential to replace or supplement real-world testing [3]. In this paper, we discuss the validation of XiL environments and the role of uncertainties in test environments in this process. Furthermore, we introduce "plausibilization" as an initial step in a validation process. Finally, a plausibilization method is presented and applied to an example dataset from an ADAS series development project.

STATE OF THE ART

Advanced Driver Assistance Systems

Advanced Driver Assistance Systems (ADAS) process environment sensor data to support the driver in longitudinal and/or lateral vehicle control. Even if an ADAS is active, the driver retains full responsibility for vehicle control and can always override the system. [4] In this paper, two specific ADAS are considered:

An **Advanced Emergency Braking System (AEBS)** tracks objects in front of the ego vehicle and triggers driver warnings and brake interventions to avoid or mitigate collisions. AEBS are mandatory for new vehicles in the European Union since 2015. [5] While the minimum legal requirement is a speed reduction of 20 km/h [6], state-of-the-art AEBS are able to avoid collisions up to 80 km/h ego speed on stationary objects[1]. According to the SAE

---

[1] https://www.adac.de/rund-ums-fahrzeug/ausstattung-technik-zubehoer/assistenzsysteme/lkw-notbremsassistenten/

automation levels, an AEBS system is classified as a SAE level 0 system, as it executes a braking action only temporarily in case of a hazard.

Similar to an AEBS, a **Blind Spot Information System (BSIS)** also tracks objects, however in a lateral zone next to the ego vehicle. The main objective of this system is to support the driver in turning scenarios as especially pedestrians and cyclists may be difficult to see for the driver due to the obstructed field-of-view from the truck cabin. Whenever there is an object close to the ego vehicle, a BSIS information is issued. In case of a potential collision during a turning maneuver, the BSIS issues a warning to the driver. [7] The so-called **Active Sideguard Assist (ASGA)** is a system by Daimler Truck, which additionally triggers a braking intervention in parallel to a BSIS warning [8]. BSIS and ASGA can also be classified as SAE level 0.

Metrics and Pass/fail Criteria

Metrics and Pass/fail criteria (PFC) are used to evaluate the behavior of an ADAS or ADS. While we consider metrics as continuous values, a pass/fail criterion yields a binary information [9]. **Performance metrics** cover aspects such as driver comfort or fuel efficiency [10, 11]. For safety-related analysis, a variety of **criticality metrics** exists [12] to evaluate "the combined risk of the involved actors when the traffic situation is continued". [13] An example criticality metric example is Time-to-Collision (TTC) [14].

X-in-the-Loop Testing

In X-in-the-Loop (XiL) testing, different representations of the System under Test (SUT) can be tested in closed loop simulation environments [15]. Depending on SUT representation, different XiL variants such as the following exist:

- Model-in-the-Loop (MiL): A software model is tested in a virtual environment.
- Software-in-the-Loop (SiL): Real software code is tested in a virtual environment.
- Hardware-in-the-Loop (HiL): Real software code is integrated into target hardware and tested in a virtual environment.

Combining XiL and scenario-based testing enables the continuity of test cases across different XiL environments. Testing different representations of the SUT aims to reduce the share of real-world testing in order to cover the growing test width and depth for higher automation levels. [2] Nevertheless, real-world testing environments such as proving ground and field operational testing continue to play an important role for the system release [16] and serve as benchmark for the validation and plausibilization of XiL environments.

Both XiL and PG testing are prone to **uncertainties** and **errors** [3, 17]. While errors represent an explicit deviation of system behavior from a reference behavior, uncertainties relate to the possible and/or expected errors of a system [18]. Consequently, an error can be regarded as a result of an uncertainty. A distinction can be made regarding the type of uncertainties. **Aleatory** uncertainties are inherent to a system and occur stochastically. They cannot be reduced, however it is possible to quantify them with statistical models [18, 19]. In contrast, **epistemic** uncertainties are theoretically reducible as they result from a lack of knowledge or modeling inaccuracy [18, 19]. In general, both types of uncertainties can occur in XiL and PG tests [3].

If XiL test results are used in an overall test statement, **XiL validation** is a crucial contribution to substantiate credibility of the XiL environment. In contrast to validation of an SUT, XiL validation evaluates the behavior of the test environment only. While there exist various approaches to validate simulation models of XiL subsystems such as sensor models [20, 21] or vehicle models [22, 23], the focus of this paper shall be the overall validation of a full XiL environment. This can be achieved by comparing XiL results to a real-world reference, such as PG tests or other real-world driving data. Relevant inputs are trajectory data [3, 24–26] or criticality metrics [3, 24, 26]. It is possible to directly compare time rows as well as features derived from time rows [3, 25]. Another option is to evaluate maneuver similarity [26]. Furthermore, by using statistical means such as standard deviations [27] or tolerance intervals [3], it is possible to cover uncertainties in both XiL and PG.

Scenario-based Testing

Scenario-based testing provides a structured approach to describe real-world traffic through scenarios that form the basis for the test and release process. A **scenario** is defined as the temporal development between several scenes, similar to a storyline [28]. Each **scene** describes a snapshot of static elements and dynamic actors that form the ego vehicle's environment. Three abstraction levels of scenarios were initially introduced by [29] and used in in the PEGASUS project. They have been extended by a fourth level of abstract scenarios in the VVM project [13]. While the abstraction level decreases from functional to concrete scenarios, the number of potential scenarios to be tested increases:

- **Functional scenarios:** A human readable description of a scenario, concentrating on the behavior and relation of included actors. This may include a visualization.
- **Abstract scenarios:** A formalized, machine readable description, including declarative descriptions such as constraints.
- **Logical scenarios:** A parameterized scenario representation, including parameter ranges or distributions as basis for a parameter variation.
- **Concrete scenarios:** A scenario derived from a logical scenario by selecting a specific parameter set, e.g. start velocities and distances.

Bock et al. [30] propose a **6-layer model** for structured traffic and environment description of scenarios that is based on the previous work in the PEGASUS project by [31, 32]. Scholtes et.al. adapt the 6-layer model to allow the separation of spatial (L1-L3) and temporal (L4-L6) scenario elements [33] and to represent the urban operational design domain that is considered in the VVM project:

1. Road Network and Traffic Guidance Objects, e.g. roads, sidewalks, traffic lights.
2. Roadside Structures, e.g. buildings, guardrails, street lamps.
3. Temporary Modifications of L1 and L2, e.g. temporary signs, covered markings.
4. Dynamic Objects, e.g. vehicles, pedestrians, animals.
5. Environmental Conditions, e.g. lighting, wind, road surface condition.
6. Digital Information, e.g. V2X messages, traffic light states.

This paper focuses on layer 4 and SUT behavior in relation to other traffic participants.

In analogy to requirements-based testing, [31] defines a logical **test case** as a logical scenario including PFC and requirements for test execution. A concrete test case is derived from a logical test case by specifying the scenario parameters. PFC for test cases may use performance and criticality metrics including a target or threshold value.

CONSIDERATIONS FOR PLAUSIBILZATION OF XIL TESTS

Validation and Plausibilization

While validation is a term commonly used when it comes to actual product development and testing, the validation of XiL environments relates to the test environment itself. Since a XiL environment is composed of multiple simulation models [34], the definition of validation is based on the definition for simulation models by Schlesinger:

> "[**Validation** is the] substantiation that a [simulation] model [or XiL environment] within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model [or environment]." [35]

A simulation model is an implementation of a conceptual model as illustrated in Figure 1. Prior to validation, the steps qualification and verification are to be applied. We use the same reference to define qualification:

> "[**Qualification** is the] determination of adequacy of the conceptual model to provide an acceptable level of agreement for the domain of intended application." [35]

For verification we use the same reference with adjustments in accordance to Sargent [36]:

> "[**Verification** is the] substantiation that a [simulation] model [is a correct implementation of a conceptional model and] represents [the] conceptual model within specified limits of accuracy." [35]

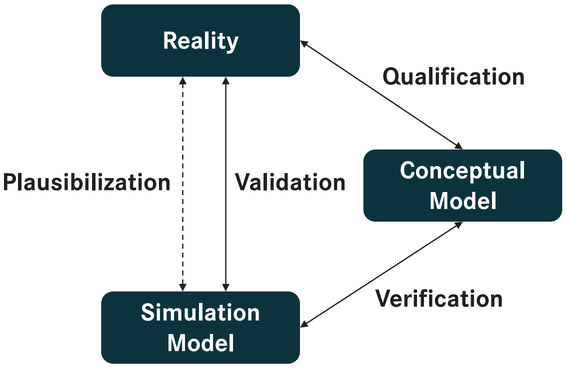


Figure 1. Relation between qualification, validation, verification of simulation models ([35], extended by "Plausibilization").

For the scope of this publication, we only discuss validation and therefore assume that qualification and verification have been performed successfully for all simulation models used in a XiL environment.

Since the intended application of a XiL environment is to reduce real-world tests by generating evidence on real-world behavior of the SUT, existing uncertainties [37] need to be taken into account when it comes to XiL validation. Consequently, XiL validation needs to cover statistical aspects. In the context of scenario-based testing, this requires multiple samples of reference scenarios of both XiL and real-world testing. We therefore extend the definition of XiL validation above for scenario-based testing by the following requirement:

> In the context of scenario-based testing, proving a satisfactory range of accuracy must consider uncertainty of test results and system behavior. This requires data from multiple back-to-back test executions of both real-world and XiL test.

While execution of multiple real-world or XiL tests may be feasible for some applications, there are applications where only a single real-world and XiL scenario execution can be tested and analyzed, e.g. due to economic or practical reasons. Since the term validation as previously specified would not be applicable, we introduce plausibilization as a preliminary step towards validation:

> **Plausibilization** in the context of scenario-based testing is the substantiation that a XiL environment possesses a satisfactory range of accuracy proven for one back-to-back test execution in real world and XiL. It does not consider uncertainty of test results or system behavior.

Validation implicitly requires a successful plausibilization. The rest of this paper will focus on XiL plausibilization in the context of scenario-based testing, without specifying subsequent steps for XiL validation.

Pass/fail Criteria and Test Result

As a prerequisite to the following sections, we further introduce PFC as a binary output of a test evaluation:

$$\mathrm{PFC} = \mathrm{f}_1(\mathrm{x}_1^{\mathrm{out}}, \dots, \mathrm{x}_\mathrm{n}^{\mathrm{out}}) \qquad \text{Equation (1)}$$

With x^{out} representing an output of the test execution, e.g. trajectory data or criticality metrics. PFC are defined based on expert knowledge depending on the SUT and scenario to be tested. In addition, we formally introduce a test result T as the aggregation of all PFC:

$$\mathrm{T} = [\mathrm{PFC}_1, \dots, \mathrm{PFC}_\mathrm{m}] \qquad \text{Equation (2)}$$

A METHOD FOR PLAUSIBILIZATION OF XIL TESTS

Overview

As already discussed in the previous section, plausibilization compares a XiL test execution data to a real-world test execution, represented by PG. Each test execution shall be called a "sample", yielding the terms "XiL sample" and “PG sample”. Plausibilization is successful if the samples (S^{XiL}, S^{PG}) of both test environments are classified as equivalent. Equivalence ($E = 1$) requires the following two criteria:

1. The test results of both test environments are identical.

$$\mathrm{E}_1 = \begin{cases} 1, & \mathrm{T}^{\mathrm{XiL}} = \mathrm{T}^{\mathrm{PG}} \\ 0, & \text{otherwise} \end{cases} \qquad \text{Equation (3)}$$

2. The scenario trajectories which lead to the test result are equivalent. This is evaluated using k scenario distance measures d_k as follows:

$$\mathrm{E}_2 = \begin{cases} 1, & \mathrm{d_k}(\mathrm{S}^{\mathrm{XiL}}, \mathrm{S}^{\mathrm{PG}}) < \mathrm{d_{k,max}} \forall \mathrm{k} \\ 0, & \text{otherwise} \end{cases} \qquad \text{Equation (4)}$$

Equivalence is achieved if both criteria are met:

$$\mathrm{E} = \begin{cases} 1, & \mathrm{E}_1 \wedge \mathrm{E}_2 \\ 0, & \text{otherwise} \end{cases} \qquad \text{Equation (5)}$$

It is assumed that all data required to compute both PFC and scenario distance measures is available, including ego and object trajectories. If there are limitations in data availability in at least one test environment, the method may still be applicable if PFC or scenario distance measures are adapted.

Pass/fail Criteria

PFC are defined and selected based on expert knowledge, given the logical scenario and the SUT. If applicable, PFC are designed in a way that yields “1” for a desired behavior and “0” for undesired behavior. For this publication, we introduce the following PFC:

No collision: Indicates that no collision of the ego vehicle and another dynamic object such as vehicle or pedestrian has occurred in the scenario:

$$\mathrm{noColl} = \begin{cases} 1, & \text{no collision occurred} \\ 0, & \text{collision occurred} \end{cases} \qquad \text{Equation (6)}$$

TTC threshold: Indicates that the minimal TTC has not undercut a specific threshold in the scenario. The threshold needs to be selected based on the logical scenario and the SUT.

$$\mathrm{ttcTh} = \begin{cases} 1, & \mathrm{TTC(t)} \geq \mathrm{TTC_{min}}\ \forall \mathrm{t} \\ 0, & \text{otherwise} \end{cases} \qquad \text{Equation (7)}$$

AEBS warning: Indicates if the AEBS of the ego vehicle has triggered a warning to the driver in the scenario.

$$\mathrm{aebsW} = \begin{cases} 1, & \text{warning triggered} \\ 0, & \text{otherwise} \end{cases} \qquad \text{Equation (8)}$$

AEBS partial braking: Indicates if the AEBS of the ego vehicle has triggered a partial braking in the scenario.

$$\mathrm{aebsPB} = \begin{cases} 1, & \text{partial braking triggered} \\ 0, & \text{otherwise} \end{cases} \qquad \text{Equation (9)}$$

AEBS full braking: Indicates if the AEBS of the ego vehicle has triggered a full braking in the scenario.

$$\mathrm{aebsFB} = \begin{cases} 1, & \text{full braking triggered} \\ 0, & \text{otherwise} \end{cases} \qquad \text{Equation (10)}$$

Blind spot information: Indicates if the BSIS of the ego vehicle has triggered a blind spot information to the driver in the scenario.

$$\mathrm{bsisI} = \begin{cases} 1, & \text{information triggered} \\ 0, & \text{otherwise} \end{cases} \qquad \text{Equation (11)}$$

Blind spot warning: Indicates if the BSIS of the ego vehicle has triggered a blind spot warning to the driver in the scenario.

$$\text{bsisW} = \begin{cases} 1, & \text{warning triggered} \\ 0, & \text{otherwise} \end{cases} \qquad \text{Equation (12)}$$

ASGA braking: Indicates if the ASGA of the ego vehicle has triggered a blind spot braking in the scenario.

$$\text{asgaB} = \begin{cases} 1, & \text{braking triggered} \\ 0, & \text{otherwise} \end{cases} \qquad \text{Equation (13)}$$

Scenario Distance Measures

The general idea of scenario distance measures is to quantify the dissimilarity of two concrete scenarios based on the data generated from their execution [38]. Though being deployed in multiple applications, scenario distance measures have rarely been named explicitly in publications. We focus on those scenario distance measures that can be computed from recorded trajectory data [39, 40] . This requires a definition of three coordinate systems as follows (see Figure 2):

Inertial coordinate system: A coordinate system bound to a static reference point and without any rotation. For the sake of simplicity, the coordinate system is initialized with its x-axis coinciding with the ego vehicles' x-axis at the beginning of a scenario. Variables using this coordinate system are indicated by index "i".

Ego coordinate system: A coordinate system bound to a static reference point and rotating with the ego vehicle. Consequently, both x-axes of coordinate system and ego vehicle coincide at any time. Index "e" indicates this coordinate system.

Object coordinate system: A coordinate system bound to the front of the ego vehicle and rotating with the ego vehicle. Hence, the coordinate system moves with the ego vehicle and has coinciding x-axes with the ego vehicle at any time. This coordinate system is indicated by index "o".

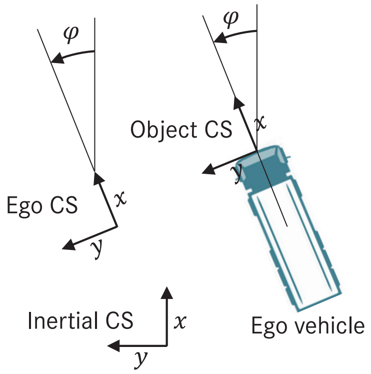


Figure 2. Overview of coordinate systems used.

The following trajectory data are processes for the scenario distance measures:

- Ego trajectory in inertial coordinate system: $\mathrm{r}_{\mathrm{ego}}^{\mathrm{i}} = [\mathrm{x}^{\mathrm{i}}, \mathrm{y}^{\mathrm{i}}]$
- Ego longitudinal velocity in ego coordinate system: $v_{x,ego}^{e}$
- Ego yaw angle in inertial coordinate system: φ_{ego}^{i}
- Object relative trajectory in object coordinate system: $r_q^o = [x_q^o, y_q^o]$

The initial step to compute the scenario distance measures is the application of Dynamic Time Warping (DTW) [41] on the ego trajectory data and the extraction of the respective DTW path. This eliminates time dependency. Given two time series $A = [a_1, \ldots, a_n]$ and $B = [b_1, \ldots, b_m]$, DTW assigns for each element a corresponding element of the other time series: $[a_u, b_v]$. The vectors $u = [u_1, \ldots, u_p]$ and $v = [v_1, \ldots, v_p]$ represent the DTW path.

The DTW path is adjusted in the following way: In order to reduce the number of elements which are assigned multiple times, only the elements of the time series containing fewer elements are assigned to the elements of the time series with more elements. Hence, there is no change in the time series containing more elements. Given $n \leq m$, this yields an alternative assignment $[a_{\hat{u}}, b_{\hat{v},}]$ with $\hat{u} = [\hat{u}_1, \dots, \hat{u}_m]$ and $\hat{v} = [1, \dots, m]$. In case multiple elements $u_{p,1}, \dots, u_{p,k}$ of A are assigned to an element of B, the last element $\hat{u}_p = u_{p,k}$ is selected. Furthermore, if $m < n$, $\hat{u}$ and $\hat{v}$ are determined vice versa.

Using the DTW path, we define three scenario distance measures:

Scenario Distance Measure 1 uses the ego vehicle trajectory and the relative trajectory of an object:

$$d_1(S^{XiL}, S^{PG}) = \max_j 0.5 \cdot (g\left(r^{i,XiL}_{ego,\hat{u}_j}, r^{i,PG}_{ego,\hat{v}_j}\right) + g\left(r^{o,XiL}_{q,\hat{u}_j}, r^{o,PG}_{q,\hat{v}_j}\right)) \qquad \text{Equation (14)}$$

$$g(x, y) = \min(|x - y|, g_{th}) \qquad \text{Equation (15)}$$

The parameter g_{th} defines a threshold for the maximum Euclidean distance of x and y.

Scenario Distance Measure 2 analyzes the ego vehicle longitudinal velocity:

$$d_2(S^{XiL}, S^{PG}) = \frac{1}{\max(n, m)} \sum_{j=1}^{\max(n,m)} g(v^{e,XiL}_{x,ego,\hat{u}_j}, v^{e,PG}_{x,ego,\hat{v}_j}) \qquad \text{Equation (16)}$$

Scenario Distance Measure 3 considers the ego vehicle yaw angle:

$$d_3(S^{XiL}, S^{PG}) = \frac{1}{\max(n, m)} \sum_{j=1}^{\max(n,m)} g(\varphi^{i,XiL}_{ego,\hat{u}_j}, \varphi^{i,PG}_{ego,\hat{v}_j}) \qquad \text{Equation (17)}$$

If there are two or more dynamic objects in the scenario, they have to be assigned to each other. Since in this publication only scenarios with one other object are considered, there is no further assignment algorithm necessary.

Scenario Distance Measure Threshold Determination

In order to apply Equation (4), it is necessary to define suitable scenario distance thresholds which mark equivalence of two samples. As there is no standardized process for this, we propose and apply a new method for this.

As previously mentioned, uncertainties are inherent to both XiL and PG tests. Consequently, test results may differ if an identical concrete scenario is executed repeatedly. Each sample is assigned to a respective test result, yielding groups of samples. For each group, scenario distance measures to samples within the group are computed. It is assumed that there is at least one group of samples where 95 % (Coverage C) of the scenario distance measure population falls below the threshold that indicates equivalence. Since the number of samples in each group is a finite number, a one-sided tolerance interval [42] is used to compute the respective thresholds for each group:

$$P\left(F\left(d_{k,T_i,max}\right) \geq C\right) \geq 1 - \alpha \qquad \text{Equation (18)}$$

The parameter α represents the confidence and is set to 95 %. This method is applied to each test result/sample group occurring, given that there exist at least three samples in the respective group. As a next step, the minimum threshold value of all groups is selected:

$$d_{k,max} \min_{T_i} d_{k,T_i,max} \qquad \text{Equation (19)}$$

APPLICATION OF XIL PLAUSIBILIZATION

Reference SUT and Scenarios

The plausibilization method presented is applied to the software test of a series AUTOSAR-based [43] Electronic Control Unit (ECU) for execution of ADAS algorithms, including those for AEBS, BSIS and ASGA. Environment sensors provide the required information to the SUT, while respective actuator controllers are deployed for vehicle control.

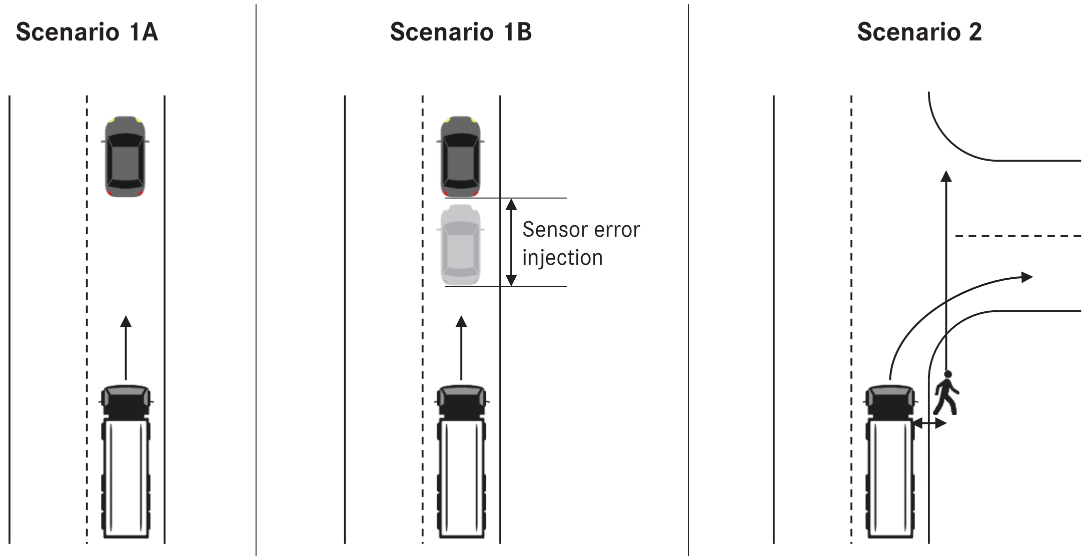


Figure 3. Overview of reference scenarios.

For evaluation, three concrete scenarios are instantiated from the logical scenario illustrated in Figure 3 and described below:

Scenario 1A: The ego vehicle drives with highway speed towards a stationary passenger car.

Scenario 1B: Identical to Scenario 1A, except to an additional sensor error injection for the relative distance of the preceding passenger car. This scenario is only performed in XiL tests.

Scenario 2: The ego vehicle drives with city speed and turns right. A pedestrian walks next to the right side of the ego vehicle, in the same direction as the ego vehicle drives prior to the turn maneuver. Due to the turning maneuver of the ego vehicle, the trajectories of ego vehicle and pedestrian intersect.

While scenarios 1A and 1B are intended to trigger an AEBS reaction, scenario 2 is supposed to lead to a BSIS and ASGA reaction. Based on the SUT, the following PFC are selected for each scenario:

Scenario 1A/1B: $[\mathrm{noColl}, \mathrm{ttcTh}, \mathrm{aebsW}, \mathrm{aebsPB}, \mathrm{aebsFB}]$

Scenario 2: $[\mathrm{noColl}, \mathrm{bsisI}, \mathrm{bsisW}, \mathrm{asgaB}]$

XiL Environment and Proving Ground

A SiL environment is used as XiL representation, including simulation models for vehicle, other ECUs, driver, environment (objects etc.) and sensors. The SUT representation is fully virtual as well and contains the real SUT series application code, while base software is modeled in a simplified way. Scenarios are defined using a combination of Open Drive [44] and proprietary description files. Each concrete scenario is tested once, as there does not exist considerable aleatory uncertainty in this SiL environment.

Proving ground tests are performed for scenario 1A and 2. There are five samples of scenario 1A (indicated by I to V) and three samples of scenario 2 (indicated by VI to VIII), which allows the computation of scenario distance measure thresholds. Scenario 1A tested on PG also serves as a reference for scenario 1B SiL tests.

In both test environments, SUT network communication is recorded consistently. Furthermore, precise trajectory data using Differential-GPS is recorded on PG, while equivalent ground truth information can be extracted from SiL tests. In order to compute scenario distance measures, all tracks are previously cut using predefined start and end conditions.

Plausibilization Results

Scenario distance measure thresholds are determined based on PG samples. Since in this example all test results are identical in the respective PG tests for scenario 1A, there is only one group of samples. Though this does not apply for scenario 2, we still consider all PG samples as part of an equivalent group as there has been a manual driver brake intervention in case of no ASGA brake intervention (applies to sample VI). The analysis as described in Equation (18) and Equation (19) yields individual thresholds $d_{1,max}$, $d_{2,max}$ and $d_{3,max}$ for scenarios 1A/1B and scenario 2:

Table 1.

Scenario distance thresholds

Scenario	$\mathbf{d_{1,max}}$	$\mathbf{d_{2,max}}$	$\mathbf{d_{3,max}}$
1A/1B	2.434	0.752	0.006
2	6.657	0.443	0.028

Plausibilization is applied to all sample combinations of corresponding scenarios, which yields the following results:

Table 2.

Equivalence of Scenarios 1A and 1B (1/0 indicate whether equivalence of PFC or scenario distance measure is achieved)

Sample combination	**noColl**	**ttcTh**	**aebsW**	**aebsPB**	**aebsFB**	d_1	d_2	d_3	E
1A-I	1	1	1	1	1	1	0	0	0
1A-II	1	1	1	1	1	1	1	1	1
1A-III	1	1	1	1	1	1	1	1	1
1A-IV	1	1	1	1	1	1	1	1	1
1A-V	1	1	1	1	1	1	0	1	0
1B-I	1	1	1	1	1	0	0	0	0
1B-II	1	1	1	1	1	0	0	1	0
1B-III	1	1	1	1	1	0	0	1	0
1B-IV	1	1	1	1	1	0	0	1	0
1B-V	1	1	1	1	1	0	0	1	0

Table 3.

Equivalence of Scenario 2 (1/0 indicate whether equivalence of PFC or scenario distance measure is achieved)

Sample combination	$noColl$	$bsisI$	$bsisW$	$asgaB$	d_1	d_2	d_3	E
2-VI	1	1	1	0	1	1	1	0
2-VII	1	1	1	1	1	1	1	1
2-VIII	1	1	1	1	1	1	1	1

In Table 2 and Table 3, 1/0 indicates whether equivalence of PFC or scenario distance measure is achieved. For scenario 1A, three of five sample combinations are classified as plausible, while the same applies to two of three scenario 2 samples. For scenario 1A, the samples with unsuccessful plausibilization were not equivalent in terms of scenario distance measures. In contrast, scenario 2, one sample is not equivalent due to a mismatch of a PFC

($asgaB$). For scenario 1B, plausibilization is not successful for any sample due to failed equivalence of scenario distance measure 1. A plot to illustrate scenario distance measures for both a plausible (1A-II) and not plausible (1B-II) sample combination (both marked in gray in Table 2) is shown in Figure 4 and Figure 5.

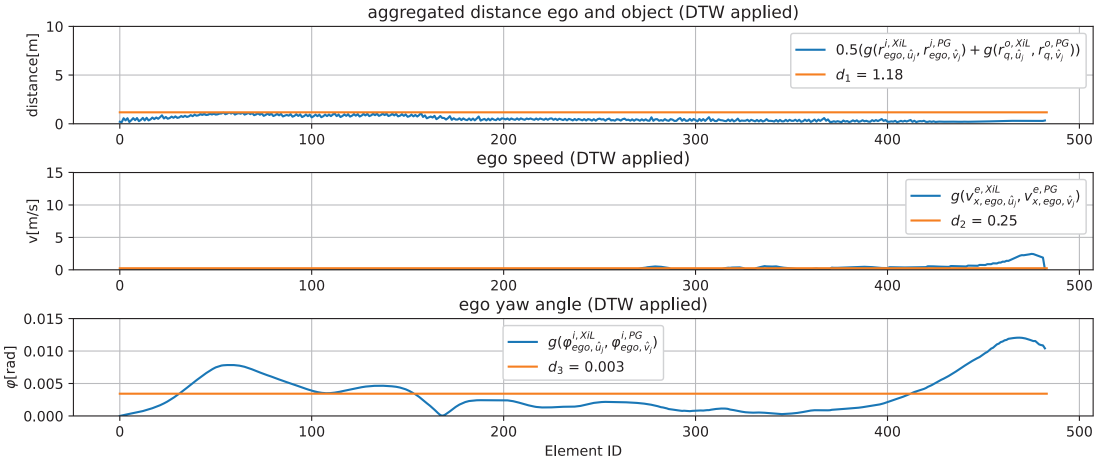


Figure 4. Illustration of scenario distance measures for sample combination 1A-II (plausible).

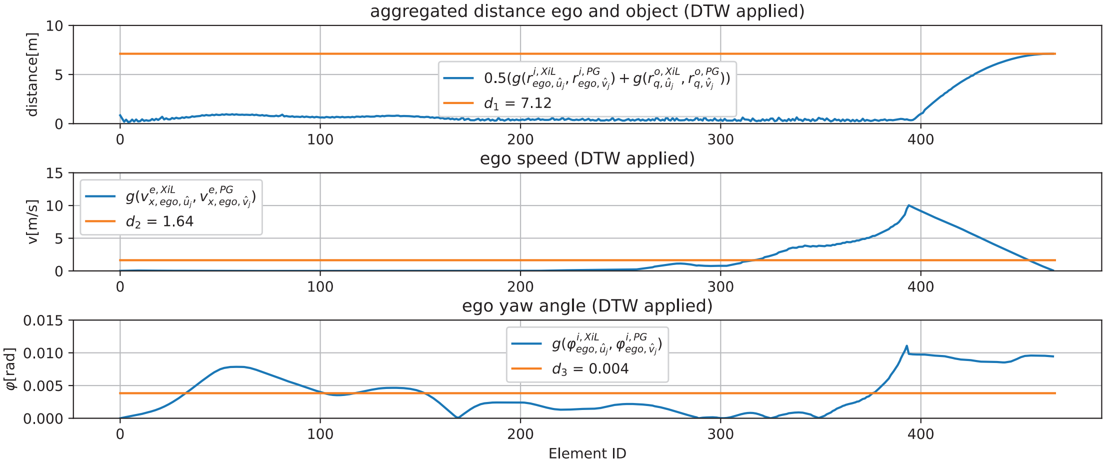


Figure 5. Illustration of scenario distance measures for sample combination 1B-II (not plausible).

The major difference between scenario 1A and 1B is a deviation of d_1, which is caused by a significant increase of the aggregated distance ($0.5 \cdot (\mathrm{g}\left(\mathrm{r}^{\mathrm{i,XiL}}_{\mathrm{ego},\hat{\mathrm{u}}_\mathrm{j}}, \mathrm{r}^{\mathrm{i,PG}}_{\mathrm{ego},\hat{\mathrm{v}}_\mathrm{j}}\right) + \mathrm{g}\left(\mathrm{r}^{\mathrm{o,XiL}}_{\mathrm{q},\hat{\mathrm{u}}_\mathrm{j}}, \mathrm{r}^{\mathrm{o,PG}}_{\mathrm{q},\hat{\mathrm{v}}_\mathrm{j}}\right))$ in Equation (17) in scenario 1B. There is also a higher deviation of the mean ego velocity in scenario 1B. This behavior can be explained by the error injection, which causes the ego vehicle to brake earlier in scenario 1B compared to scenario 1A in both SiL and PG.

Discussion

The presented XiL plausibilization method combines an expert knowledge-based approach (PFC) with a data-driven approach (scenario distance measures). While PFC equivalence ensures that a XiL test yields the same test results as a PG test, scenario distance measures are used to evaluate the trajectories leading to these test results. Both PFC and scenario distance measures can be tailored to application demands, including both testing of ADAS and ADS. Consequently, this dependency may lead to different plausibilization results for the same XiL and PG data. However, this is a desired behavior as requirements for XiL tests may vary by application.

When it comes to the application example presented, the overall plausibilization statement can be considered as reasonable from a visual expert inspection of all scenarios. Both scenario 1A and 2 show a very similar SUT stimulation and behavior in XiL and PG, whereas the plausibilization results of scenario 1B are caused by a different SUT behavior which has been identified in trajectory analysis through scenario distance measures. Even though there is no standardized way to evaluate a plausibilization statement, the presented method is suitable for the use case discussed and can be adapted for further use cases.

While mere plausibilization is not sufficient to statistically compare XiL to PG, it is still a necessary step towards a XiL validation. The fact that plausibilization of scenario 1A does not yield a positive result for all PG samples shows the variance of PG tests and underscores this statement. To argument credibility of a XiL environment it is not expected that all sample combinations are classified as plausible. Statistical evaluation will therefore enhance the plausibilization statement.

CONCLUSIONS AND OUTLOOK

Credibility of XiL tests is a major concern when it comes to their growing deployment to test ADAS and ADS. While direct comparison of a XiL test to a reference PG test is an established procedure, taking into account uncertainty of both XiL and real-world testing is a significant enhancement of existing XiL validation methods. In this paper, we discuss XiL validation and introduce the term plausibilization as a preliminary step towards validation. We further provide a potential method to plausibilize a XiL sample by comparing its PFC and scenario distance measures to a PG sample as a reference. To evaluate usability of this method, it is applied to an ADAS series ECU SiL test. For two concrete scenarios, a successful plausibilization can be executed for a majority of sample combinations. In an additional scenario, which contains an intended error injection, all sample combinations are evaluated as not plausible.

The overall plausibilization method is promising considering the application example discussed. Nevertheless, there is a high dependency on PFC and scenario distance measure selection and definition of scenario distance measure thresholds. As for right now, these steps require a significant amount of expert knowledge and may be biased. For XiL validation, it is furthermore necessary to add steps considering statistical behavior and uncertainty of test and plausibilization results.

REFERENCES

[1] W. Wachenfeld, "How Stochastic can Help to Introduce Automated Driving," PhD Thesis, TU Darmstadt, 2017.

[2] F. Reisgys, J. Plaum, A. Schwarzhaupt, and E. Sax, "Scenario-based X-in-the-Loop Test for Development of Driving Automation," in *14. Workshop Fahrerassistenzsysteme und Automatisiertes Fahren*, Bonlanden, 2022.

[3] F. Reisgys, M. Elgharbawy, A. Schwarzhaupt, and E. Sax, "Argumentation on ADAS Simulation Validity using Aleatory and Epistemic Uncertainty Estimation," in *Proceedings of the Driving Simulation Conference 2021 Europe VR*, Munich, Germany, 2021, pp. 25–32.

[4] T. M. Gasser, A. Seeck, and B. W. Smith, "Framework Conditions for the Development of Driver Assistance Systems," in *Handbook of Driver Assistance Systems*, H. Winner, S. Hakuli, F. Lotz, and C. Singer, Eds., Cham: Springer International Publishing, 2016, pp. 35–68.

[5] *Regulation (EC) No 661/2009 of the European Parliamant and of the Council*, 2009.

[6] *Commission Regulation (EU) No 347/2012*, 2012.

[7] *Uniform provisions concerning the approval of motor vehicles with regard to the Blind Spot Information System for the Detection of Bicycles: UN ECE R 151*, 2020.

[8] Daimler Truck AG, *Mercedes-Benz Trucks presents two worldwide innovations in their trucks for more safety on the road*. Stuttgart, 2020. Accessed: Nov. 30 2022. [Online]. Available: https://media.daimlertruck.com/marsMediaSite/ko/en/47504429

[9] C. King, L. Ries, C. Kober, C. Wohlfahrt, and E. Sax, "Automated Function Assessment in Driving Scenarios," in *IEEE 12th International Conference on Software Testing, Verification and Validation*, Xi'an, China, 2019, pp. 414–419.

[10] C. Esselborn, M. Eckert, M. Holzäpfel, E. Wahl, and E. Sax, "Method for a scenario-based and weighted assessment of map-based advanced driving functions," in *ATZ live, 20. Internationales Stuttgarter*

Symposium, M. Bargende, H.-C. Reuss, and A. Wagner, Eds., Wiesbaden: Springer Vieweg, 2020, pp. 193–207.

[11] T. Plum *et al.,* “A simulation-based case study for powertrain efficiency improvement by automated driving functions,” *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 233, no. 5, pp. 1320–1330, 2019.

[12] L. Westhofen *et al.,* “Criticality Metrics for Automated Driving: A Review and Suitability Analysis of the State of the Art,” *Arch Computat Methods Eng*, 2022.

[13] C. Neurohr, L. Westhofen, M. Butz, M. H. Bollmann, U. Eberle, and R. Galbas, “Criticality Analysis for the Verification and Validation of Automated Vehicles,” *IEEE Access*, vol. 9, pp. 18016–18041, 2021.

[14] J. C. Hayward, “Near miss determination through use of a scale of danger,” 1972.

[15] K. v. Neumann-Cosel, “Virtual Test Drive,” PhD Thesis, TU München, 2014.

[16] C. King, L. Ries, J. Langner, and E. Sax, “A Taxonomy and Survey on Validation Approaches for Automated Driving Systems,” in *IEEE International Symposium on Systems Engineering (ISSE)*, Vienna, 2020, pp. 1–8.

[17] O. Balci, “Principles and techniques of simulation validation, verification, and testing,” in *Winter Simulation Conference Proceedings*, Arlington, VA, USA, 1995, pp. 147–154.

[18] S. Riedmaier, B. Danquah, B. Schick, and F. Diermeyer, “Unified Framework and Survey for Model Verification, Validation and Uncertainty Quantification,” *Arch Computat Methods Eng*, 2020.

[19] W. L. Oberkampf, S. M. DeLand, B. M. Rutherford, K. V. Diegert, and K. F. Alvin, “Error and uncertainty in modeling and simulation,” *Reliability Engineering & System Safety*, vol. 75, no. 3, pp. 333–357, 2002.

[20] S. Bernsteiner, Z. Magosi, D. Lindvai-Soos, and A. Eichberger, “Radarsensormodell für den virtuellen Entwicklungsprozess,” *ATZ Elektron*, vol. 10, no. 2, pp. 72–79, 2015.

[21] A. Schaermann, A. Rauch, N. Hirsenkorn, T. Hanke, R. Rasshofer, and E. Biebl, “Validation of vehicle environment sensor models,” in *IEEE Intelligent Vehicles Symposium (IV)*, Los Angeles, CA, USA, 2017, pp. 405–411.

[22] M. Viehof and H. Winner, “Stand der Technik und der Wissenschaft: Modellvalidierung im Anwendungsbereich der Fahrdynamiksimulation,” TU Darmstadt, Darmstadt, 2017.

[23] *ISO 19365:2016: Passenger cars - Validation of vehicle dynamic simulation - Sine with dwell stability control testing,* ISO, 2016.

[24] K. Groh, S. Wagner, T. Kuehbeck, and A. Knoll, “Simulation and Its Contribution to Evaluate Highly Automated Driving Functions,” *SAE International Journal of Advances and Current Practices in Mobility*, vol. 1, no. 2, pp. 539–549, 2019.

[25] D. Notz *et al.,* “Methods for Improving the Accuracy of the Virtual Assessment of Autonomous Driving,” in *8th IEEE International Conference on Connected Vehicles and Expo (ICCVE)*, Graz, Austria, 2019.

[26] C. Stadler, F. Montanari, W. Baron, C. Sippl, and A. Djanatliev, “A Credibility Assessment Approach for Scenario-Based Virtual Testing of Automated Driving Functions,” *IEEE Open Journal of Intelligent Transportation Systems*, vol. 3, pp. 45–60, 2022.

[27] S. Riedmaier, J. Nesensohn, C. Gutenkunst, T. Düser, B. Schick, and H. Abdellatif, “Validation of X-in-the-Loop Approaches for Virtual Homologation of Automated Driving Functions,” in *11th Graz Symposium Virtual Vehicle*, 2018.

[28] S. Ulbrich, T. Menzel, A. Reschka, F. Schuldt, and M. Maurer, “Defining and Substantiating the Terms Scene, Situation, and Scenario for Automated Driving,” in *18th IEEE International Conference on Intelligent Transportation Systems (ITSC)*, Gran Canaria, Spain, 2015, pp. 982–988.

[29] T. Menzel, G. Bagschik, and M. Maurer, “Scenarios for Development, Test and Validation of Automated Vehicles,” in *IEEE Intelligent Vehicles Symposium (IV)*, Changshu, 2018, pp. 1821–1827.

[30] J. Bock, R. Krajewski, L. Eckstein, J. Klimke, J. Sauerbier, and A. Zlocki, “Data Basis for Scenario-Based Validation of HAD on Highways,” in *27th Aachen Colloquium Automobile and Engine Technology*, Aachen, 2018.

[31] F. Schuldt, “Ein Beitrag für den methodischen Test von automatisierten Fahrfunktionen mit Hilfe von virtuellen Umgebungen,” PhD Thesis, TU Braunschweig, 2017.

[32] G. Bagschik, T. Menzel, and M. Maurer, “Ontology based Scene Creation for the Development of Automated Vehicles,” in *IEEE Intelligent Vehicles Symposium (IV)*, Changshu, 2018, pp. 1813–1820.

[33] M. Scholtes *et al.,* “6-Layer Model for a Structured Description and Categorization of Urban Traffic and Environment,” *IEEE Access*, vol. 9, pp. 59131–59147, 2021.

[34] B. Schütt, M. Steimle, B. Kramer, D. Behnecke, and E. Sax, “A Taxonomy for Quality in Simulation-Based Development and Testing of Automated Driving Systems,” *IEEE Access*, vol. 10, pp. 18631–18644, 2022.

[35] S. Schlesinger *et al.,* “Terminology for model credibility,” *SIMULATION*, vol. 32, no. 3, pp. 103–104, 1979.

[36] R. G. Sargent, “Verification and validation of simulation models,” in *Proceedings of the 2010 Winter Simulation Conference (WSC)*, Baltimore, MD, USA, 2010, pp. 166–183.

[37] J. Langner, K.-L. Bauer, M. Holzapfel, and E. Sax, “A Process Reference Model for the Virtual Application of Predictive Control Features,” in *IEEE Intelligent Vehicles Symposium (IV)*, Las Vegas, NV, USA, 2020, pp. 1759–1764.

[38] K. Backhaus, B. Erichson, S. Gensler, R. Weiber, and T. Weiber, *Multivariate Analysemethoden*. Wiesbaden: Springer Fachmedien Wiesbaden, 2021.

[39] J. Bian, D. Tian, Y. Tang, and D. Tao, “A survey on trajectory clustering analysis,” Feb. 2018. [Online]. Available: http://arxiv.org/pdf/1802.06971v1

[40] L. Ries, P. Rigoll, T. Braun, T. Schulik, J. Daube, and E. Sax, “Trajectory-Based Clustering of Real-World Urban Driving Sequences with Multiple Traffic Objects,” in *IEEE International Intelligent Transportation Systems Conference (ITSC)*, Indianapolis, IN, USA, 2021, pp. 1251–1258.

[41] T. Giorgino, “Computing and Visualizing Dynamic Time Warping Alignments in R : The dtw Package,” *Journal of Statistical Software*, vol. 31, no. 7, 2009.

[42] D. S. Young, “tolerance : An R Package for Estimating Tolerance Intervals,” *J. Stat. Soft.*, vol. 36, no. 5, 2010.

[43] AUTOSAR, “Classic Platform Release Overview,” 2021.

[44] ASAM, “OpenDRIVE Base Standard 1.7,” 2021.

PEER REVIEW PAPER

This paper has been peer-reviewed and published in a special edition of Traffic Injury Prevention 24(S1), by Taylor & Francis Group. The complete paper will be available on the Traffic Injury Prevention website soon. To access ESV Peer-reviewed papers click the link below
https://www.tandfonline.com/toc/gcpi20/24/sup1?nav=tocList

PEER REVIEW PAPER

This paper has been peer-reviewed and published in a special edition of Traffic Injury Prevention 24(S1), by Taylor & Francis Group. The complete paper will be available on the Traffic Injury Prevention website soon. To access ESV Peer-reviewed papers click the link below
https://www.tandfonline.com/toc/gcpi20/24/sup1?nav=tocList

CHARACTERIZATION AND MITIGATION OF INSUFFICIENCIES IN AUTOMATED DRIVING SYSTEMS

Yuting Fu[1], **Jochen Seemann**[1], **Caspar Hanselaar**[2], **Tim Beurskens**[1], **Andrei Terechko**[1], **Emilia Silvas**[2,3], **Maurice Heemels**[2]
[1] – Central Technology Office, NXP Semiconductors, Eindhoven, The Netherlands
[2] – Department of Mechanical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands
[3] – Integrated Vehicle Safety, TNO, Dutch Organization for Applied Scientific Research, Helmond, The Netherlands

ABSTRACT

Automated Driving (AD) systems have the potential to increase safety, comfort and energy efficiency. Recently, major automotive companies have started testing and validating AD systems (ADS) on public roads. Nevertheless, the commercial deployment and wide adoption of ADS have been moderate, partially due to system functional insufficiencies (FI) that undermine passenger safety and lead to hazardous situations on the road.

In contrast to system faults that are analyzed by the automotive functional safety standard ISO 26262, FIs are defined in ISO 21448 Safety Of The Intended Functionality (SOTIF). FIs are insufficiencies in sensors, actuators and algorithm implementations, including neural networks and probabilistic calculations. Examples of FIs in ADS include inaccurate ego-vehicle localization on the road, incorrect prediction of a cyclist maneuver, unreliable detection of a pedestrian in rainy weather using cameras and image processing algorithms, etc.

The main goal of our study is to formulate a generic architectural design pattern, which is compatible with existing methods and ADS, to improve FI mitigation and enable faster commercial deployment of ADS. First, we studied the 2021 autonomous vehicles disengagement reports published by the California Department of Motor Vehicles (DMV). The data clearly show that disengagements are five times more often caused by FIs rather than by system faults. We then made a comprehensive list of insufficiencies and their characteristics by analyzing over 10 hours of publicly available road test videos. In particular, we identified insufficiency types in four major categories: world model, motion plan, traffic rule, and operational design domain. The insufficiency characterization helps making the SOTIF analyses of triggering conditions more systematic and comprehensive.

To handle faults, modern ADS already integrate multiple AD channels, where each channel is composed of sensors and processors running AD software. Our characterization study triggered a hypothesis that these heterogeneous channels can also complement each other's capabilities to mitigate insufficiencies in vehicle operation. To verify the hypothesis, we built an open-loop automated driving simulation environment based on the LG SVL simulator. Three realistic AD channels (Baidu Apollo, Autoware.Auto, and comma.ai openpilot) were tested in the same driving scenario. Our experiments suggest that even advanced AD channels have insufficiencies that can be mitigated by switching control to another (possibly less advanced) AD channel at the right moment.

Based on our FI characterization, simulation experiments and literature survey, we define a novel generic architectural design pattern Daruma to dynamically select the channel that is least likely to have a FI at the moment. The key component of the pattern does cross-channel analysis, in which planned trajectories and world models from different AD channels are mutually evaluated. The output of the cross-channel analysis is combined with more traditional fault detections in a safety fusion component. The safety fusion then feeds an aggregated per-channel safety score to the high-level arbiter, which eventually selects the AD channel to control the vehicle. The formulated architectural pattern can help manufactures of autonomous vehicles in mitigating FIs.

Limitations of our study suggest interesting future work, including algorithmic research on cross-channel analysis and safety fusion, as well as evaluation of the cross-channel analysis in simulations and road tests.

1. INTRODUCTION

Potential advantages of Automated Driving (AD) systems include increased safety, comfort and energy efficiency [1]. Major automotive companies and academic institutions have already started testing and validating AD systems (ADS) on public roads [2, 3, 4, 5]. Nevertheless, the commercial deployment and wide adoption of ADS have been limited, partially due to the challenge of functional insufficiencies (FIs), which is the focus of this paper.

Traditionally, safety-critical systems focused on detecting and mitigating faults, such as a memory cell bit flip or a software deadlock, which are the primary focus of the automotive standard ISO 26262 [6]. Modern ADS deploy redundant AD channels to cope with faults [7]. As illustrated in Figure 1, an AD channel consists of the sensors (such as radar and camera), the ADS display for visualization and human interaction, and the processing modules including perception, localization, maps, prediction, and motion planning module, but excludes actuators for steering and acceleration. The greyed out items in Figure 1 are the human machine interface and the actuators, which are shared among the multiple AD channels of the ADS.

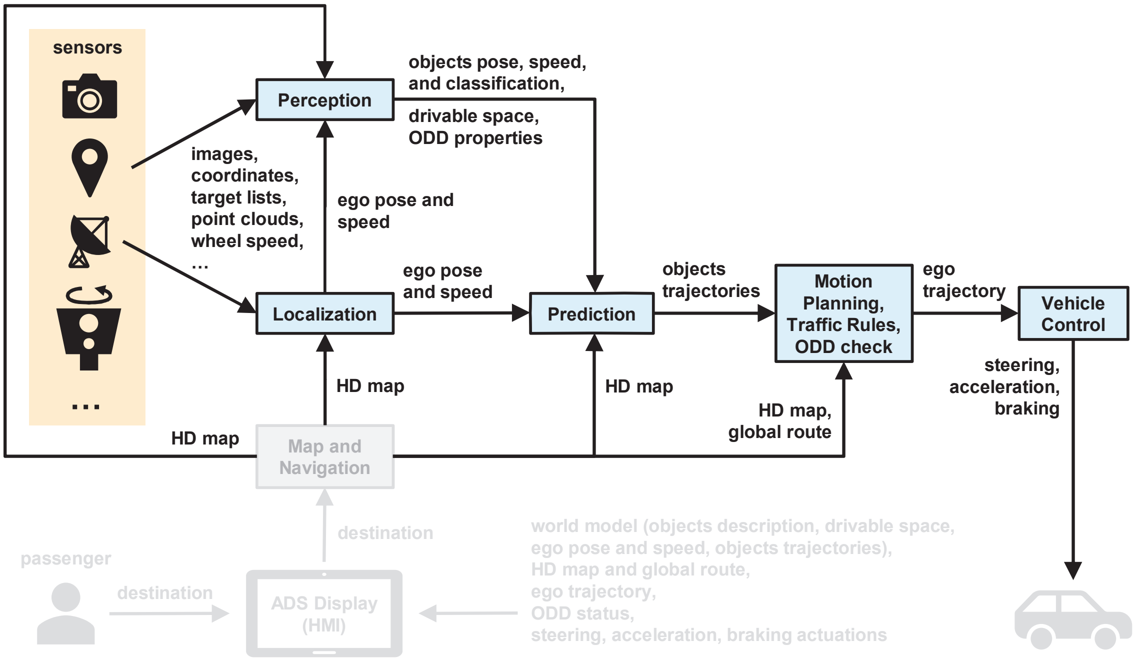


Figure 1. The simplified AD channel architecture.

In our work we do not focus on fault handling, because it has been widely studied and handled by effective safety mechanisms leveraging error-correction codes, watchdog monitors, redundancy, etc. In contrast, advanced environmental sensors, deep neural networks, and probabilistic or non-deterministic algorithms in modern ADS introduce insufficiencies in the implementation of the AD functionalities which have received significantly less attention. Such insufficiencies are called FIs in the ISO 21448 standard Safety Of The Intended Functionality (SOTIF) [8].

An example of a FI is illustrated in Figure 2 below, where the ADS is turning right on the T-shaped intersection and does not properly predict the intention of the pedestrian walking across the street. Consequently, the ADS does not slow down [9]. To avoid collision the human driver overrides the ADS operation and takes over control of the vehicle. This example show-cases the potential hazardous consequences of FIs, as well as the reduced ADS availability due to the disengagement. Note that the ADS and the ego vehicle in this example do not suffer from a fault in the traditional sense of functional safety as defined in the standard ISO 26262 [6].

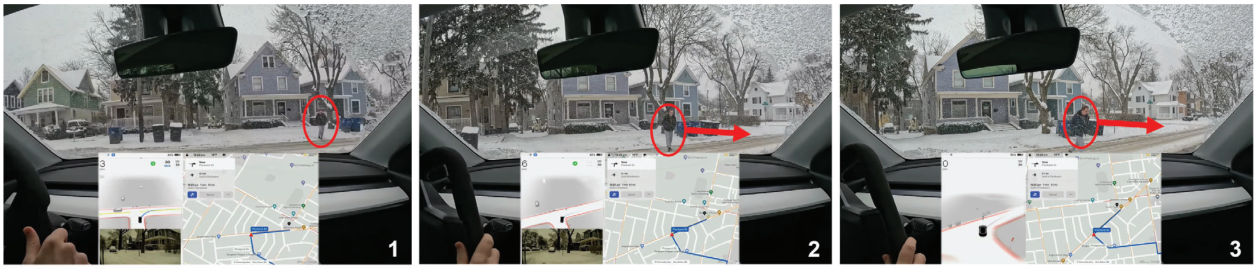


Figure 2. Incorrect prediction of a pedestrian's trajectory is an example of ADS functional insufficiencies. The camera-based perception module of the ADS correctly detects and tracks the pedestrian in step 1 and step 2. However, in step 2 it fails to predict the pedestrian's intention to cross the street, and in step 3 the human driver disengages automated driving.

Some other insufficiency examples are:

1. ADS perception fails to detect a pedestrian;
2. ADS prediction fails to correctly predict the cyclist's maneuver;
3. ADS motion planning fails to yield a trajectory feasible for vehicle actuators;
4. ADS drives the vehicle faster than the speed limit.

Mitigating FIs at runtime with the help of design patterns is the primary focus of this paper. Other approaches, such as product lifecycle management and various other safety techniques recommended by SOTIF at design time are outside of the scope of this paper. A classic example of a FI mitigation technique is sensor fusion [10]. It integrates multi-modal sensors output into a single world model by combining complementary sensor data with varying weights dependent on the situation. For example, the camera sensor will be favored in a clear weather condition, while the radar sensor will be preferred in rainy conditions. Note that these sensors can also have different capabilities, for example, the camera can perceive the traffic light color, while the radar sensor can detect object speeds.

Our work concentrates on two essential problems and contributes to their solutions:

1. There is no comprehensive inventory and characterization of FIs in modern ADS. Providing such an inventory will be beneficial in assessing the prevalence of FIs and their potential impact on ADS safety. Our paper presents a new comprehensive characterization of FIs in ADS, which can be used in the analysis of triggering conditions recommended by the SOTIF.
2. Despite integration of state-of-the-art functional safety mechanisms and redundant AD channels for fail-operational behavior, modern ADS still occasionally suffer from FIs that compromises safety and availability. Consequently, how to reduce the negative impact of FIs in autonomous safety-critical systems [11, 12] is an important problem. We contribute to its solution by introducing Daruma: a generic architectural design pattern extending redundant ADS with cross-channel analysis and arbitration of redundant AD channels to mitigate FIs.

It should be noted that the high-level goal of our study is to support technological evolution to safe AD and reduce road accidents. We do not intend to assess safety of the evaluated ADS, but rather to extract architectural recommendations to reduce the negative impact of FIs.

The remainder of this paper is organized as follows. In the next section we survey literature on FIs and architectural design patterns to mitigate them. Section 3 discusses analysis of our road test studies, presents our characterization table of FIs, and makes a hypothesis about a new architectural design pattern to mitigate FIs. The open-loop simulation method and results in Section 4 are used in the subsequent section to define the design pattern for cross-channel analysis. In Section 5, we conclude our findings and discuss future research directions in Section 6.

2. RELATED WORK

In this section we first present a summary of related work on reports and statistical analyses of FIs, then introduce several existing FI mitigation methods.

ANALYSIS OF FUNCTIONAL AND OUTPUT INSUFFICIENCIES

As stated in [13], hazardous situations caused by FIs are typically not considered by the traditional safety analysis process focusing on ISO 26262 [6]. To identify and evaluate such hazards, SOTIF [8] focuses on identifying triggering conditions of the FIs and the acceptance criteria of the ADS response in the presence of the triggering conditions. Triggering conditions as defined in SOTIF, such as unusual road users, weather conditions, light glare, EMI interference, etc., are external factors that initiate the ADS FIs. Furthermore, internal ADS implementation insufficiencies can also cause FIs [8, 14, 15, 16]. Representation of the external scene and ego vehicle behavior are captured in the state of the ADS components. Erroneous component state, such as missed object or ghost object detection, incorrect predictions of trajectories, etc., is termed output insufficiencies (OI) in SOTIF. Our characterization study focuses on OIs to make a comprehensive characterization of FIs in ADS. The characterization can help employ appropriate mitigation mechanisms to address a wider range of triggering conditions. Note that ISO 26262, SOTIF and UL 4600 [17] standards all focus on technology-neutral design-time processes, while our work concentrates on concrete FIs and a safety mechanism to mitigate FIs at runtime. Figure 3 below presents simplified relations among the terms and concepts defined in ISO 26262 and SOTIF.

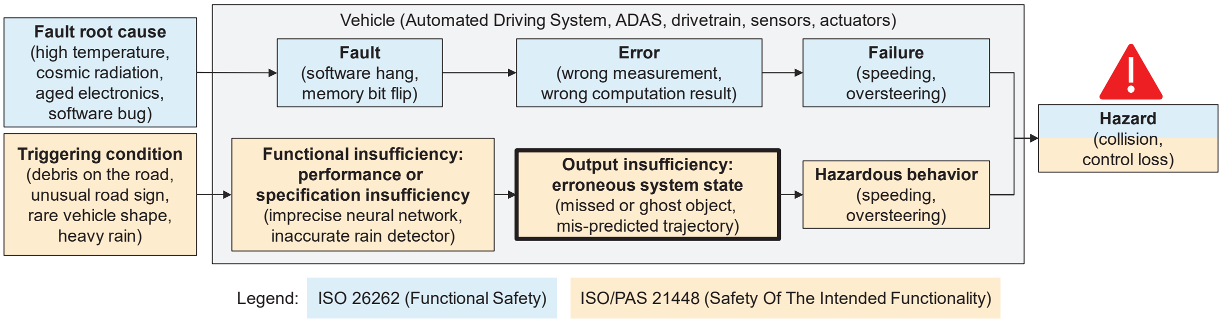


Figure 3. Simplified cause and effect hazard models in ISO 26262 (in blue above) and SOTIF (in yellow below). The grey *Vehicle* box separates what happens inside and outside of the vehicle. Examples are given in brackets. The bold *Output insufficien*cy box identifies the focus of our classification.

FIs in autonomous vehicles have been logged and studied both in real-life road tests and in academic research. For example, a number of real-life consequences of FIs are registered by the California Department of Motor Vehicles (DMV). DMV publishes AV collision reports and disengagement reports from a wide range of AV manufacturers every year [18, 19], in which system failures as well as FIs are reported.

Some statistical analyses have been conducted based on the DMV reports: [20] provides a processed dataset of the DMV reports from 2014 to early 2020; [21] analyzes frequencies rather than causes of AV disengagements and accidents; [22] finds that most disengagements are related to machine-learning-based perception and decision-and-control systems. When categorizing the causes of the disengagements and accidents, however, [22] does not clearly separate faults and FIs. While system faults are much better studied, FIs are less systematically addressed at design time and hence have to be handled at runtime. To focus ADS improvements on biggest issues first, the automotive industry needs classification and prioritization of the FIs contributing to dangerous situations.

There is also research on FIs based on pre-recorded sensor datasets and environment conditions simulated by AD simulators. For example, [23] proposes a methodology on discovering novel triggering conditions for perception FIs with a case study of a LiDAR-based AD perception system; [24] proposes a workflow to identify triggering conditions defined in [8] for LiDAR and camera-based perception module and evaluates the proposal with simulation tools.

To our best knowledge, no detailed classification of FIs of ADS internals have been made in existing studies of FIs and most of the prior works focus on the triggering conditions of FIs. In this paper, we do not study and categorize the triggering conditions, but focus on FIs in the perspective of SOTIF, and distinguish them from systematic failures and errors, which have been well studied in the scope of ISO 26262. Our comprehensive characterization of FIs will be helpful in the design of mitigation mechanisms for ADS FIs as well as in triggering conditions analysis.

MITIGATION OF FUNCTIONAL INSUFFICIENCIES

In this subsection we discuss the existing FI mitigation strategies. One mitigation strategy is to reduce ADS FIs by avoiding the external triggering conditions using restricted Operational Design Domains (ODD) [25, 26]. Alternatively, we can limit the frequency of FIs occurrence by improving the performance of ADS at the component and system levels.

At the component level, FIs can be mitigated by improving the performance of a specific ADS module or using an alternative method. For example, ADS perception can be improved by more accurate sensors; an FI in a lane detection algorithm can be mitigated by exploiting high-definition maps and enhanced localization algorithms [8]; ghost objects can be decreased by using a different detection mechanism; missed objects can be reduced by combing the radar and the vision system. Similar approaches can also be applied to motion planning and actuation algorithms [27]. [10] details the pros and cons of different sensors and sensor fusion algorithms. [16] analyzes FIs of two commonly used sensor fusion methods and provides several suggestions on how to improve the multi-sensor fusion algorithms. A similarity-based incremental learning algorithm is proposed in [28] to improve the learning model for pedestrian prediction over time. [15] presents an extensive survey of existing human motion prediction algorithms and suggests that a combination of multiple prediction algorithms could be the approach to a more robust prediction in unknown situation. Besides applications in the AD prediction module, neural networks are also widely used in the perception module to detect and identify objects. [29] shows that the ensemble performance of multiple bitwise neural networks can surpass the performance of a single high-precision neural network in terms of classification accuracy. The Responsibility-Sensitive Safety (RSS) [30] and the Safety Force Field (SFF) [31] algorithms check if the planned trajectory is compliant with the ADS safety constrains and predefined traffic rules, so as to prevent the AV from being the cause of a road accident from the legal perspective. These algorithms are able to mitigate some hazardous situations caused by FIs in the ADS motion planner without actively detecting or identifying the FIs, because only actuation control commands that are within the safety envelope of the ADS are passed on by the RSS or the SFF modules to the AD actuators. Similarly, [8] mentions applying restrictions on AD functions as a way to prevent or mitigate SOTIF-related risk. However, as [13] points out, such approach may compromise the usability of the ADS.

At the system-level, FI can be addressed by monitoring, redundancy, diversity (heterogeneity), functional restrictions or other measures [8]. The popular monitor/actuator design pattern from [27] enables shutting down the main driving channel in a fail-silent way, if it violates safety criteria according to the monitor. [32] proposes a set of application-specific verification methods to ensure the correct arbitration in a fault-tolerant redundant AD architecture with two heterogeneous AD channels. However, [32] emphasizes that the diversity in AD channels will result in inconsistent output due to the problem of replica indeterminism [12]. As a result, the traditional voting approach such as majority voting or triple modular redundancy (TMR) among the redundant channels is not suitable. To cope with the replica indeterminism, more sophisticated comparison techniques, weighing mechanisms, etc. need to be deployed in the heterogeneous ADS. [33] proposes a Level 3 ADS to mitigate exiting ODD [25, 26], which is a safety-critical FI, by transferring the control to the fallback human driver. [13] proposes an AD architecture where the safety channel and the nominal channel monitor the health status of the other. The safety channel in [13] has its own set of sensors and a simplified world model. Although primarily aiming at mitigating faults and failures in the ADS, [13] mentions that the safety channel might be able to help address certain FIs due to its heterogenous implementation. However, both proposals from [33] and [13] can cause dangerous situations due to possibly insufficient time budget for the human driver to take over [34]. [11] proposed a commander/monitor dual-channel AD architecture to reduce false positives in free space detection. In their proposal the free space used for motion planning is the intersection of free space identified by the commander channel and the monitor channel. In other words, the common free space detected by both AD channels.

Companies in the automotive industry proposed several architectures to address the ADS FIs on the system level as well. Mobileye's True Redundancy [35] ADS uses two independent channels to reduce FIs in AD perception. One channel of the True Redundancy ADS uses only cameras for perception, the other uses only radars and LiDARs. The sensors in different channels do not interact with each other and each channel builds an independent world model. Instead of focusing only on improving the perception module performance, a Safety Shell algorithm is proposed in [36] to calculate the risk of AD channels based on the cross-channel analysis of the world models and trajectories. An arbitration logic is then applied based on the channel last safe intervention time derived from the risk calculation [36]. Instead of a specific algorithm or implementation proposal, our work proposes a generic design pattern to address FIs in ADS based on our FI characterization and experimental results on our open-loop simulation setup with realistic AD frameworks.

From the above overview, we conclude the following. The existing works tend to mix faults and FIs when discussing the improvement proposals. An accurate and extensive classification of FIs is currently missing in the literature. In addition, there are no clear winning design patterns for ADS to handle FIs at the system-level yet. Therefore, we would like to propose a clear characterization and categorization of FIs and a system-level approach to address FIs.

3. CHARACTERIZATION OF OUTPUT INSUFFICIENCIES

This section first motivates for characterization of OIs, then describes the method of video-based analysis of existing ADS in real traffic scenarios to characterize FIs. Finally, the Results subsection presents a table with OI characteristics.

MOTIVATION TO FOCUS CHARACTERIZATION ON OUTPUT INSUFFICINCIES

SOTIF divides FIs (functional insufficiencies) into performance and specification insufficiencies, which can be diverse in nature and origin and, consequently, hard to classify in a comprehensive way. On the other hand, FIs result in OIs (output insufficiencies), as shown in Figure 3. OIs can be easily attributed to the few major internal ADS functions, such as perception or path planning. Moreover, analyzing ADS FI root causes (e.g. triggering conditions or design property) requires detailed information about all internal hardware and software components, which practically complicates the classification of FIs if the ADS is studied as a black box. Therefore, we focused our characterization on OIs, which should cover most of the FIs.

CHARACTERIZATION METHODOLOGY

The methodology of our characterization study is to inspect the DMV reports and public road test videos to enumerate, classify and characterize FIs and to assess their prevalence and severity in automated driving.

First, we performed statistical analysis of over 2500 disengagement registered in the 2021 DMV disengagement report [19] to learn about reasons of disengagements in real-life cases. We classified the disengagement causes and the frequencies of their occurrence. The statistical analysis helped us gain knowledge of the most frequent contributing reasons of disengagement. The details of our DMV report study are presented in the Characterization results subsection below.

While the DMV reports provide a large amount of data, the text reports are often ambiguous because different companies provide different levels of details in the description of facts causing disengagement [19]. Fortunately, recent advances in prototyped ADS also resulted in many public video recordings of autonomous vehicles driving on public roads with real diverse traffic. Therefore we studied such video recordings as well to obtain a more intuitive understanding of the disengagement causes and consequences. Remarkably, the videos often include display screens with environment perception and motion plan of the ego vehicle, see an example in Figure 4. Several recordings even deploy drones [2] to accompany the in-cabin videos with a birds-eye-view perspective on the road and traffic.

By comparing the real-life video of the vehicle surroundings with the ADS state on the display, an attentive human observer can spot differences between the physical environment around the vehicle and the world model the ADS perceives, which are OIs of the ADS. For example, the figure below shows the real-life view with a traffic light pole that is meanwhile absent on the ADS display. To compensate for the ADS insufficiencies the driver in the video recordings sometimes takes over control of the vehicle in order to ensure safe driving. By extrapolating such situations as if no disengagement took place, the human observer of the videos can judge the severity of the ADS output insufficiency in the traffic scenario. Note, that such insufficiencies are not caused by system faults, as faults are often properly detected by the ADS safety monitors resulting in a warning to the driver or other fault handling mechanisms.

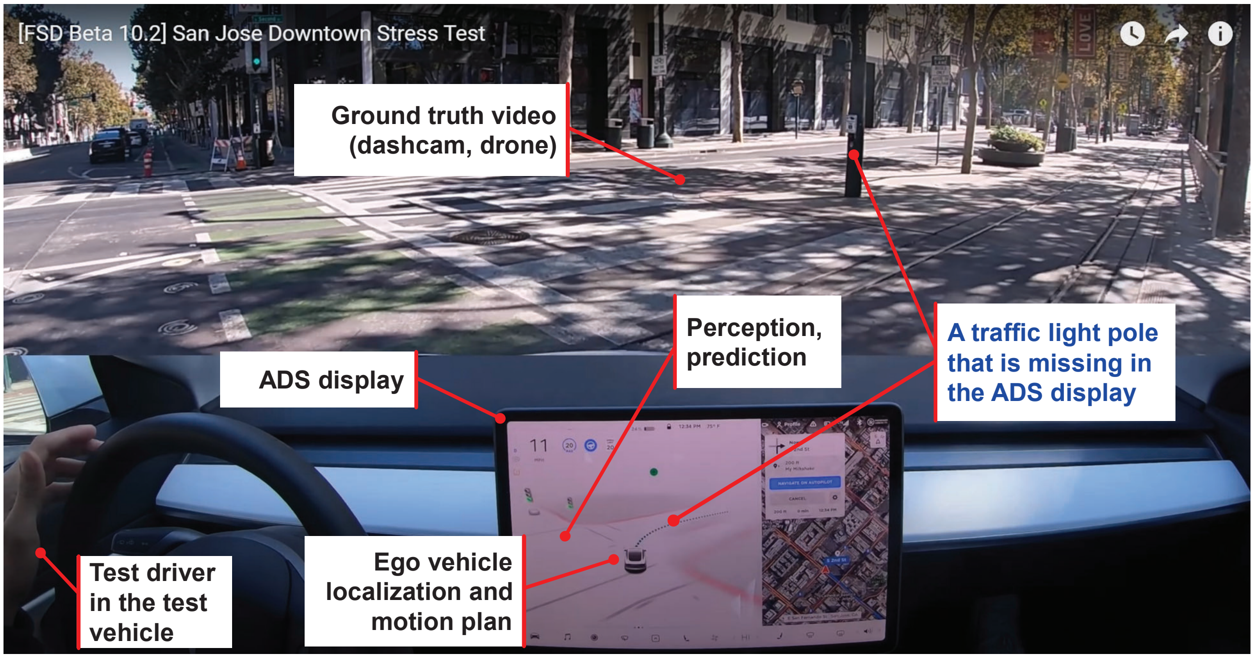


Figure 4. Road test video screenshot with a ground truth view above and an ADS display below identifying an insufficiency.

CHARACTERIZATION RESULTS

The distribution of causes of the disengagement from the 2021 DMV disengagement report [19] is depicted in Figure 5 below. Note that the "out of scope" category means the disengagements were correctly and automatically triggered by the ADS without leading to any hazardous situation; "fault" category includes all the reports that clearly claimed that a software or hardware or other systematic faults had occurred. Figure 5 shows that over 90% of the DMV report data is analyzable for our purpose and obviously disengagements caused by insufficiencies occurred five times more than those caused by system faults.

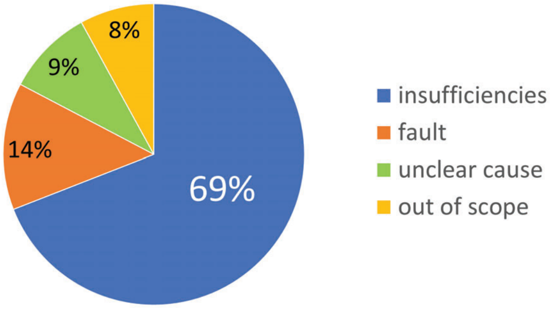


Figure 5. Causes of disengagement in 2021 DMV disengagement report.

Videos of different road test scenarios we studied are publicly available, such as [2], [3], [5], [9]. An overview of the reviewed videos is summarized in Table 1. In those videos, the ADS did not warn the driver about internal systematic faults nor showed any sign of internal systematic faults when the insufficiencies were observed. Notably, as shown in Table 1, 87% of the insufficiencies we registered could have resulted in a collision or a hazardous situation on the road, should the driver had not intervened with AD disengagement, or other road users had behaved differently.

Table 1. Summary of studied videos of road tests.

Number of watched videos	32
Watched hours	10 hours 17 minutes
Total driving scenarios with OIs	71 (100%)
Safety-critical scenarios with OIs	62 (87%)
Number of ADS brands	9

Based on our DMV report analysis and the video observations, we concluded that ADS exhibits OIs against one of the three criterion: ground truth, human intuition and traffic rules. For example, a missed object by the perception system is an ADS deficiency against the ground truth; the inability to reduce the speed on urban roads is a OI against the traffic rule; the too conservative motion plan of an ADS is against the driving style based on human intuition.

We also distinguish 16 OI types and divided them into four categories: world model, traffic rule, motion plan, and ODD as listed in Table 2 with their characteristics. The categories identify the ADS subsystem responsible for the OI. The OI names describe the ADS insufficiency. For example, insufficiency #1 identifies the system's inability to find the location of the ego-vehicle. The reference column specifies what the insufficiency is relative to. The typical timing nature of the insufficiency is captured in the timing column. OIs can exist either for a long time spanning several seconds or sporadic. For example, insufficiency #2 can exist throughout the ADS operation time, while insufficiencies #8 and #11 usually only exist for a short time and refer to the predicted motion in future time. Finally, the sensors column gives examples of sensor modules involved in triggering the insufficiency. Note that several OIs do not involve sensors at all, such as insufficiencies #10 or #13.

Table 2. Output insufficiencies classification and characterization.

ID	category	name	criterion	ADS module	sensors	timing
1	world model	wrong ego-vehicle localization	ground truth	localization	GNSS, IMU, lidar, camera	sporadic, long
2	world model	wrong map	ground truth	map	HD map files	long
3	world model	missed object	ground truth	perception	lidar, radar, camera, ultrasonic sensor, microphone	sporadic, long
4	world model	ghost object	ground truth	perception	lidar, radar, camera, ultrasonic sensor, microphone	sporadic, long
5	world model	wrong object position, orientation, or dimension	ground truth	perception	lidar, radar, camera, ultrasonic sensor, microphone	sporadic, long
6	world model	wrong object classification	ground truth	perception	lidar, radar, camera, ultrasonic sensor, microphone	sporadic, long
7	world model	wrong drivable space identification	ground truth	perception	lidar, radar, camera, ultrasonic sensor	sporadic, long
8	world model	wrong object trajectory	ground truth	prediction	-	future, sporadic
9	traffic rule	wrong traffic sign, light, lane marking or operator recognition	traffic rule	perception	camera, V2X	sporadic, long
10	traffic rule	violation of traffic regulation (e.g. right of way)	traffic rule	motion planning	-	sporadic, long

11	motion plan	counter-intuitive motion plan	human intuition	motion planning	-	future, sporadic
12	motion plan	indeterminate motion plan	human intuition	motion planning	-	sporadic
13	motion plan	unsafe planned trajectory	human intuition	motion planning	-	sporadic, long
14	ODD	wrong weather classification	ground truth	ODD checker	rain and light sensor, visibility range sensor	long
15	ODD	wrong road classification	ground truth	ODD checker	road surface sensors, GNSS	long
16	ODD	wrong traffic classification	ground truth	ODD checker	camera, radar, clock, GNSS, V2X	long

Each OI from Table 2 can lead to hazardous situations. Besides obviously dangerous OIs such as #3 “missed object”, seemingly innocent OIs, such as ODD- or motion planning-related ones, can trigger a sequence of events leading to a hazard. For example, due to a wrong weather classification categorized by insufficiency #14, the camera usage can be allowed in adverse light conditions that can lead to an accident. A counter-intuitive motion plan #11 may confuse other traffic road users and cause a collision due to misunderstanding in motion negotiations.

Figure 6 illustrates the distribution of OIs per OI category as defined in Table 2 from both the 2021 DMV disengagement report and the road tests we studied. The distribution patterns are similar in both DMV report and the real-life video study. The plot also suggests that OIs related to the world model category and the motion plan category are the most prevalent.

In summary, in our studies we observed only a limited number of system faults and, yet, modern ADS often suffered from many OIs that could lead to severe hazards. Therefore, to enable focused ADS improvements, knowledge of the most frequent contributing OIs to dangerous situations are needed.

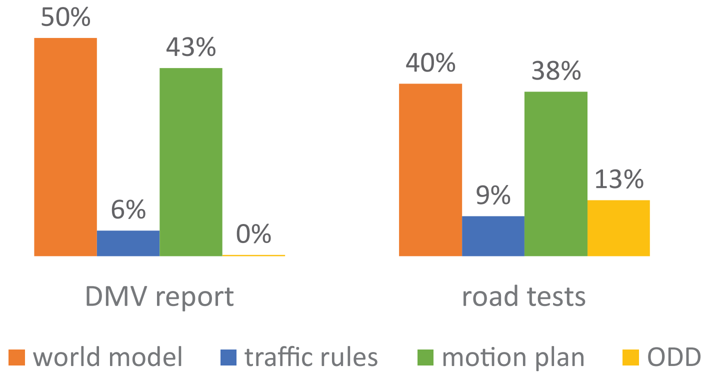


Figure 6. Distribution of functional insufficiencies in the 2021 DMV disengagement report and the road tests we studied.

4. MITIGATION OF FUNCTIONAL INSUFFICIENCIES

This section experimentally defines an architectural design pattern to mitigate FIs.

HYPOTHESIS ABOUT HETEROGENEOUS AD CHANNELS AND FUNCTIONAL INSUFFICIENCIES

Redundancy is commonly implemented by car manufacturers to improve the vehicle safety. For example, BMW includes three AD channels in their ADS [7]. To avoid systematic faults and common cause failures [6], heterogenous channels do not reuse hardware, software, algorithms or, possibly, sensors. AD systems individually developed by different car manufacturers can be seen as heterogenous AD channels. Interestingly, in our OI characterization work described in Section 3 we noticed that the distribution and frequency of OIs reported by different car companies varied from each other. We therefore make the following hypothesis:

There exist times when some AD channels in a heterogenous multi-channel ADS encounter functional insufficiencies, while the other AD channels do not.

A motivation scenario is shown in Table 3 below to illustrate this hypothesis. In this example we assume that all AD channels in the ADS can have FIs at certain times during the driving. For instance, at time steps 3, 5, and 7 one or two channels have FIs but there is still at least one available channel that does not encounter FI. Therefore, if the limited channel is in control of the vehicle at that time, it can switch to a channel not affected by FIs, instead of issuing disengagement or performing an emergency brake. The ADS can then stay in the AD mode and continue driving. In this way the ADS safety and availability are enhanced. If this hypothesis is true, there are opportunities to switch between complementary channels at runtime to mitigate FIs on top of existing implemented methods inside each AD channel, such as sensor fusion. Furthermore, at time step 4, although FIs are triggered in all channels, there still exist possibilities for the ADS to switch to the channel that does not have safety-critical FIs based on the evaluation of the current states of all the AD channels or the ADS can perform an evasive maneuver. Such AD channel evaluation can be done, for example, via the risk and arbitration calculation proposed in [36].

Table 3. A motivation scenario showing functional insufficiencies in different AD channels of the ADS at different time steps.

Time step	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Channel1	OK	OK	OK	FI	FI	OK	OK
Channel2	OK	OK	OK	FI	OK	OK	FI
Channel3	OK	OK	FI	FI	OK	OK	FI

OPEN-LOOP SIMULATION METHODOLOGY TO VERIFY THE HYPOTHESIS

Ideally, we would like to validate our hypothesis by experimenting with multiple heterogenous AD channels operating simultaneously in the same test vehicle. However, we do not have access to such a test vehicle. The available DMV reports and the road test videos do not allow us to evaluating different ADS in the exactly same driving scenario either. First, the multiple AD channels in the test vehicle are not transparent in both the DMV reports and the videos; second, in some videos even when driving multiple times the same vehicle along the same route, the test vehicle was exposed to different environments (such as different lighting and traffic). Therefore, we set up an LG SVL simulation environment for autonomous vehicles [37] with three realistic AD channels: Baidu Apollo 5.0 [38, 39], Autoware.Auto AVP [40], and Comma.AI openpilot [41]. In the simulation environment we can compare the AD channels against the "ground truth" of the simulator, see Figure 7 with screenshots of the simulator and AD channel GUIs. The screenshots of the GUIs are taken at the same simulation moment, when a cyclist is about to cross the ego vehicle's trajectory.

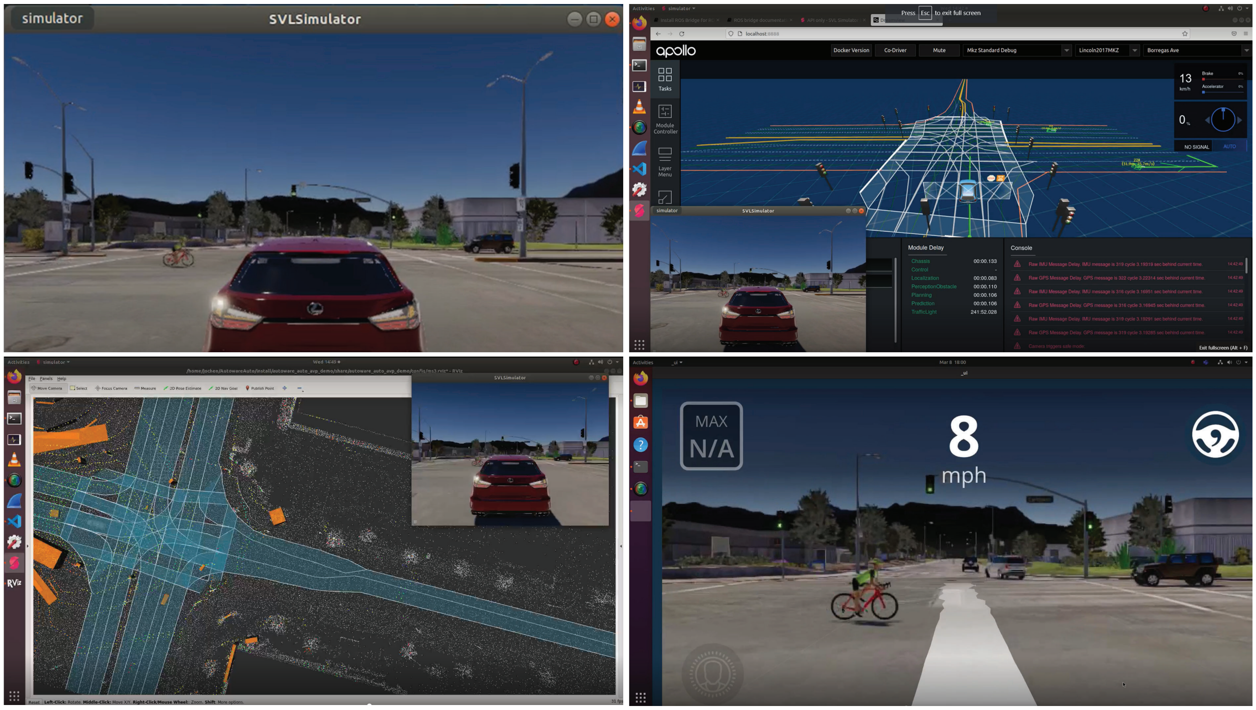


Figure 7. Screenshots of the LG SVL simulator (upper left), Apollo 5.0 Dreamview GUI (upper right), Autoware.auto Rviz GUI (lower left), openpilot GUI (lower right) at the same simulation moment.

The AD channels we chose for our simulation study, despite having different capabilities, have been used to drive real-life vehicles autonomously. Ideally, we would like to have three AD channels running in parallel in the same simulation run. However, due to high engineering complexity and high hardware requirements we resorted to running them sequentially. To ensure that the environment is exactly the same for every undertaken experiment, we employed an open-loop simulation where the AD channel's control over the vehicle was disabled and the ego-vehicle was driven by the same script asserting a fixed throttle (`apply_control()` with `sticky=True` in the LG SVL simulator). Furthermore, the traffic was controlled through a fixed random number generator seed, assuring that each scenario would show exactly the same traffic objects and motions. Finally, we used the same vehicle model in the LG SVL simulator but equipped it with various sensors required for the different AD channels. Figure 8 illustrates our experimental open-loop simulation setup.

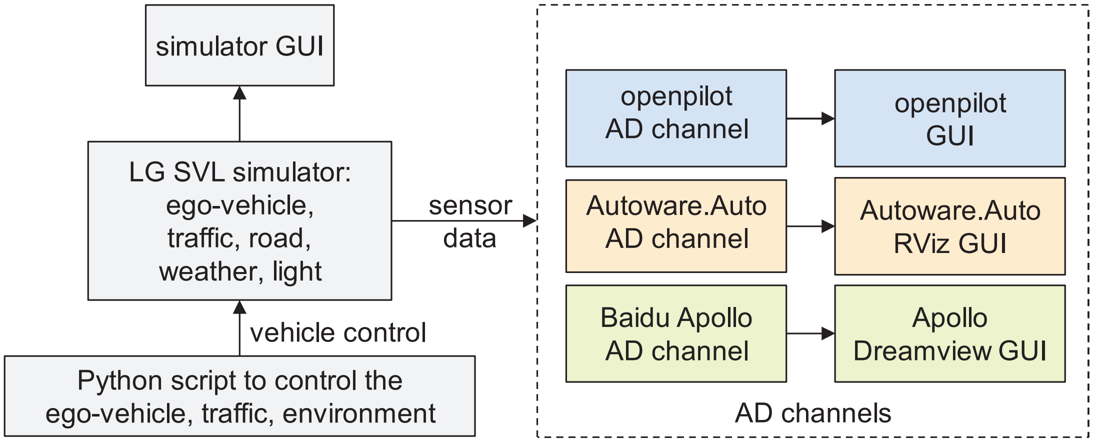


Figure 8. Open-loop simulation setup with sensors feeding AD channels and a Python script controlling the ego-vehicle and the traffic.

Below is an overview of our simulation platform and scenarios with Table 4 contrasting the AD channels we used:

- Alienware Aurora workstation, Intel Core i7-8700 and Nvidia GeForce GTX 1070, 64GiB SDRAM
- Ubuntu 18.04, desktop
- LG SVL simulator version 2021.3, the Borregas Avenue crossroad map
- Driving scenarios:
    - the ego vehicle drives straight with a fixed acceleration from a fixed starting position to a fixed end position using a Python script
    - 10 seeds for random crossing traffic and 1 scenario with one vehicle in front of the ego vehicle performing a left turn
    - all traffic lights are configured to be always green
    - good lighting and clear weather (no rain, fog, wetness, nor cloudiness), 12:00am

Table 4. Overview of the three AD channels used in the experiments.

Channel name	Version/tag	Sensors	Key algorithms & features	ODD
Baidu Apollo	5.0	lidar, camera, radar, GNSS, IMU	Deep neural networks, various computer vision algorithms, Kalman filter, configurable high-level sensor fusion, HD map	Urban
Autoware.Auto	AVP demo	lidar, GNSS, IMU	Euclidian cluster, NDT matching, HD map	Parking
openpilot	0.8.10	camera, (radar not used)	Deep neural network, MPC, Kalman filter	Highway

It is worth mention that the publicly documented LG SVL simulator setup did not use the advanced camera-radar-lidar fusion in Apollo 5.0 and relied purely on the lidar for object detection. In our experiments, we also manually created derivative Apollo channels by creating new configurations to fuse lidar and camera sensor data in Apollo 5.0. But we didn't observe any performance improvements in the Apollo visualization GUI. Also the openpilot was not utilizing the radar sensor in our simulations.

Besides the limited features implemented in the chosen AD channels, the hardware computational and memory resources can also be a triggering condition that causes FIs. In particular, we observed in our experiments that if the GPU power was not sufficient to run the simulator and process the AD channels, a video frame could be dropped resulting in one or more missed objects. Therefore, for demanding use cases we ran the simulator and visualization on a machine separate from the AD channel machine to keep the GPU processor utilization lower than 60% and video memory usage of less than 4GiB. Such hardware considerations must also be included in the analysis of FIs next to external triggering conditions as described in [8]. For example, at a busy intersection with many road users, the computational and memory limits of the hardware can lead to degraded performance of the ADS and, subsequently, to hazards.

EXPERIMENTAL RESULTS

Our open-loop simulations revealed 42 OIs, see Table 5. The result shows that first, some OIs from Table 2 are absent in our simulations due to the limited number of simulated driving scenarios; second, the AD channels have different capabilities which severely influence their OIs distribution per category in Table 5. For example, in our experiments Apollo is the only channel with visualized predictions of road users, and the Autoware.Auto and openpilot channels do not analyze traffic lights, because they are outside of their ODDs. Note that we therefore do not perform the imbalanced comparison between the channels. For example, for openpilot we only counted the objects on the ego vehicle's trajectory and excluded detection of objects outside of the camera field of view. Nevertheless, we can conclude that even the most capable channel, which is Apollo in our experiments, has limitations. Finally, Table 5 clearly suggests that different AD systems have different types of OIs at different frequencies, e.g. Apollo had only 2 missed objects, while Autoware.Auto and openpilot had twice as much.

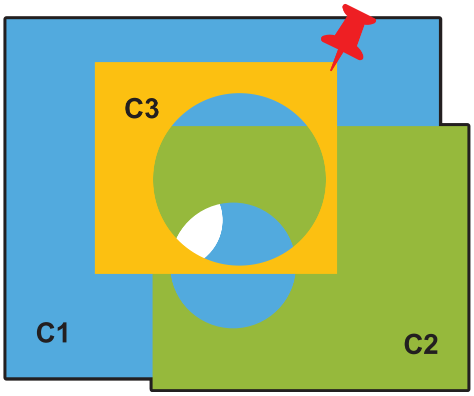
Table 5. Output insufficiencies of the AD channels identified in the open-loop simulations.

Functional insufficiency	Apollo	Autoware.Auto	openpilot	Total
wrong ego-vehicle localization	-	2	-	2
missed object	2	4	17	23
ghost object	-	1	3	4
wrong object position, orientation, or dimension	2	1	-	3
wrong object classification	1	1	-	2
wrong drivable space identification	-	-	4	4
wrong object trajectory	4	-	-	4

The goal of our simulations was to validate the hypothesis that there exist times where not all heterogeneous channels have OIs that lead to a hazardous situation. In general, we observed that the most capable Apollo channel performed very well in all of our simulations and could have been the best channel to drive the vehicle for most of the time. Nevertheless, we spotted several cases when even the best channel encountered insufficiencies, while the others did not:

1. In one random traffic simulation a cyclist was quickly crossing the trajectory of the ego-vehicle, see the illustrations of the AD channels in Figure 7. Apollo detected the bicycle only after it crossed the ego-vehicles route, while Autoware.Auto was able to trace the cyclist earlier, as soon as it reached the crossroad. Consequently, Apollo had a false negative detection in the perception system, which we term "#3 missed object" in our classification from Table 2. Note that false negatives may lead to severe accidents.
2. In a scenario with a single vehicle in front, the Apollo channel overestimates the front vehicle dimensions (insufficiency #5 in Table 2), which can result in sudden braking or evasive maneuvers that are unexpected to the rest of the road users. Such false positives may be less severe than false negatives, but may sacrifice automated driving availability and comfort. In contrast, the Autoware.Auto perception system correctly estimated the size of the front vehicle.
3. In another random traffic simulation Apollo predicted that an approaching truck from the opposite direction would cross the ego vehicle's way (insufficiency #8 in Table 2). Such false positive predictions can cause the ego-vehicle to perform unnecessary emergency braking.

Conclusions from our open-loop simulations are illustrated in Figure 9 in a qualitative way using a Venn diagram based on [36] and the SOTIF classification of driving scenarios into four areas in SOTIF Section 4 [8]. The three channels in Figure 9 have different capabilities, such as traffic light recognition or road user motion prediction, which are identified with colored rectangles. Circles represent FIs of each channel as a hole in its capability. Despite having larger capabilities, FIs still exist in advanced AD channels. And the (sometimes smaller) FIs of the advanced channels could be covered by capabilities of the other less advanced channels. In other words, capabilities of some (less advanced) AD channels can be complementary to the other (more advanced) channels. Therefore, in Figure 9 the shared insufficiencies*,* which are identified by the white hole, represent the same FIs that exist and will be triggered by the same triggering conditions in all the channels. These shared FIs are not covered by the capabilities of any channel in the same ADS. However, the shared FIs are smaller than any of the total FIs of each individual channel (represented by the big circle in each channel). Noteworthy, the common capabilities of the combined multi-channel system can be potentially bigger than that of the most advanced channel alone. The white hole in AD capabilities closely relates to the unknown hazardous scenarios in the SOTIF classification.



Legend:
C1, C2, C3: AD channels
rectangles: AD capabilities
circular holes: channels insufficiencies (lack of capability)
white hole: shared insufficiencies (lack of capability)
black line: boundary of the common ODD
red push pin: capability required at a given moment

Figure 9. Conceptual illustration of overlapping functional insufficiencies and capabilities of multiple AD channels. Capabilities of the channels can be made complementary to enable runtime selection of the channel without insufficiency.

At runtime, the driving situation changes constantly, which can be represented by an imaginary push pin moving on Figure 9 and the pinpoint specifies the currently required capabilities for safe continuous driving. If the pin ends up on a colored block, there is at least one AD channel that has sufficient capability to drive safely under such situation. Note, that the ADS always has to select a channel that has a color dot under the pin. For example, if the pin ends up on the upper right corner of channel C3, as shown in Figure 9, then the ADS should either choose channel C1 or C3, but not channel C2. However, if the pin lands inside the white hole or even in the area outside of the black border line (for example, when the ADS is outside of ODD), no channel has the required capability and the ADS will suffer from the shared insufficiency existing in all the channels. In such cases the ADS has to perform an emergency brake or disengage from the AD mode to hand over control to the human. Just like any AD channel will have certain FIs, the shared FIs are also inevitable. But by using multiple heterogeneous AD channels, we believe it is possible to minimize the shared FIs within the ODD. From the SOTIF perspective, the design time goal is to construct an AD system, which has a minimum white hole of capabilities inside its ODD and which can switch between channels at runtime to choose the most capable channel for the current driving scenario.

In order to mitigate FIs, the multi-channel ADS has to continuously monitor heterogeneous AD channels at runtime and switch to a more capable AD channel when a FI occurs in the current AD channel. Based on this observation in the next section we propose an architectural design pattern used to select the least limited channel at runtime. We named our design pattern Daruma.

DARUMA: ARCHITECTURAL DESIGN PATTERN FOR CROSS-CHANNEL ANALYSIS AND ARBITRATION

Figure 10 below presents our Daruma architectural design pattern for multi-channel ADS. The Daruma design pattern leverages cross-channel analysis to dynamically select a channel that has sufficient capabilities to safely and comfortably continue driving the vehicle in the current situation. The traditional ADS components are shown in grey boxes in Figure 9; the new or modified components are identified as the colored boxes and italics text. The heterogeneous AD channels (channel1 to channelN) include AD functions and (possibly shared) sensors depicted in Figure 1. The output of the Daruma architecture is low-level actuation setpoints that are sent to the vehicle actuators. The monitoring techniques from prior works are part of either the AD channels or the components shown as “fault monitors, ODD checker, …” in Figure 10. Traditionally, the arbitration decision of the multi-channel ADS is based on the output of intra-channel safety mechanisms, platform fault monitors and other metrics such as comfort, efficiency and availability. In general, the high-level arbiter is similar to the Mode Manager from [42] and controls the selection of the driving channel. However, the high-level arbiter in the Daruma architecture also receives output from the cross-channel analysis, which can trigger selection of a different (safer) channel even in the absence of faults.

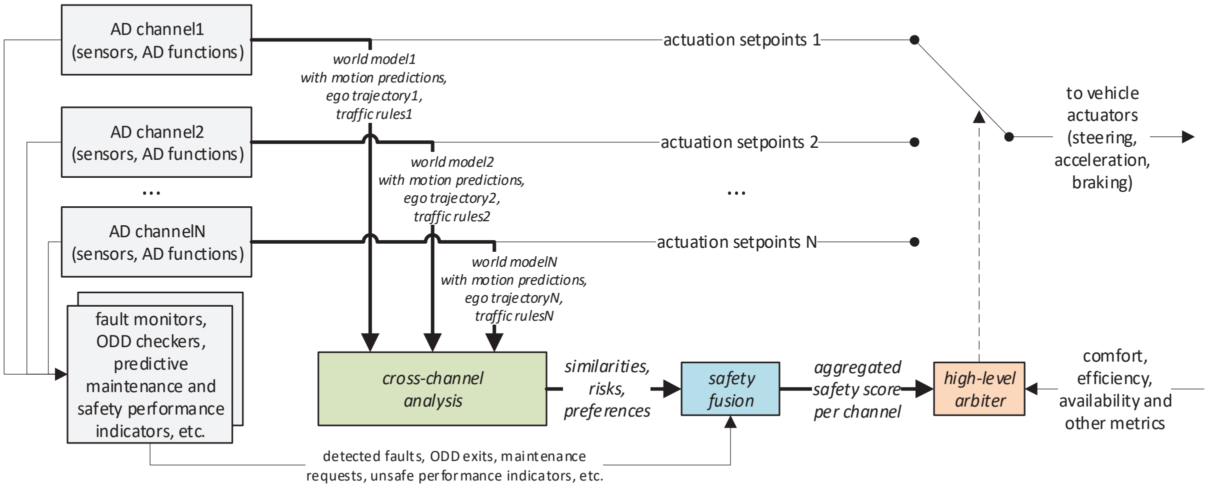


Figure 10. Architectural design pattern with cross-channel analysis and AD channel arbitration to mitigate functional insufficiencies in ADS.

To perform cross-channel analysis, Daruma architecture taps additional information from the AD channels, including the world models, object motion predictions, ego vehicle motion plans and traffic rule assessments. This high-level channel state information is often shown on ADS displays in the autonomous vehicles to inform the passenger about vehicle's internal state and planned actions.

Cross-channel analysis can involve several state and design properties of the multi-channel system:

1. *Similarities* of channels' high-level states. If multiple channels identify an object to be present in the same location with a similar motion plan, there is a higher chance of finding this object in reality, provided that the channels are sufficiently heterogeneous and reliable. Each color in Figure 11 represents output from one channel. Obviously, the three channels output different object locations, shapes and trajectories. Major differences among these channels are:
    a) the green channel doesn't see the vehicle behind the ego,
    b) the child crossing the street is only observed by the red and blue channels,
    c) the red channel plans to stop, the green and the blue channels plan to continue driving straight.
2. Cross-channel *risk* of collision, driving off-road, or losing control over the vehicle. The new cross-channel analysis component performs cross-checking of ego vehicle motion plans against the world model and object predictions generated in each channel. Therefore, for 3 channels we can compute risk as a function of future time for 3 ego motion plans times 3 world models with a total of 9 risk functions, see Figure 11. Even violating the traffic rules can be expressed as a safety risk. For example, one channel outputs an ego stop trajectory, while the other requests an acceleration of the ego vehicle. From the availability perspective, the second channel will be preferred. However, if the first channel complements the ego vehicle stopping trajectory with a red traffic light annotation, the second channel can get a low safety score in the end.
3. Design-time or runtime channel *preferences*. The preference can be expressed in weights to promote the most advanced channel for the current ODD or in terms of advance switch times as described in [36].

Straightforward majority voting using similarities of channels may not yield the safe arbitration. For example, according to difference b) the green and blue channels agree, while only the red channel is correct about the pedestrian in front. Therefore, inter-channel similarity alone as a metric is not sufficient to judge the situation and select the optimal trajectory. On the other hand, similarity can be captured by matching risk profiles from cross-channel analysis or using credit-based temporal matching, where the confidence in having a match increases if the match occurs several times in a row.

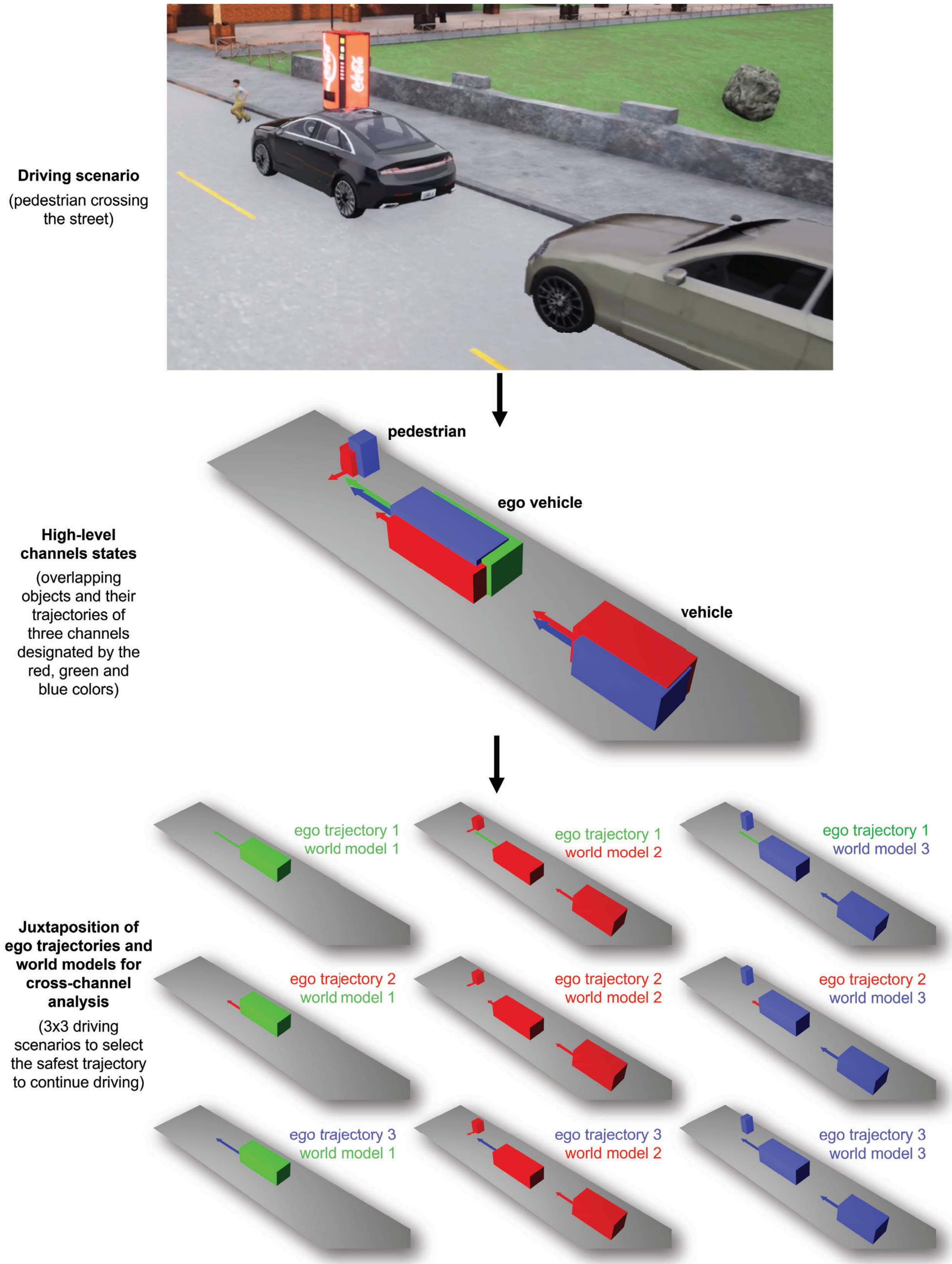


Figure 11. Cross-channel analysis by using juxtaposition of high-level states of three AD channels. The driving scenario on the top is processed by the red, green and blue channels to compute high-level channel states. Then, by applying the Daruma design pattern, the ego vehicle trajectories from each channel are cross-analyzed for hazards in all world models forming a 3x3 matrix of possible driving scenarios.

Algorithmic choices for cross-channel analysis include:

1. *Geometric overlay* of the high-level channel current and predicted states, which is partially used in [32], focuses on detection of collisions and off-road driving based on geometric intersections of 2D or 3D shapes. Remarkably, it avoids AD disengagements when the AD channels differ in a safe way, which improves ADS availability. For example, if the one channel suggests to avoid obstacle on the left and the other – on the right, then the geometric overlay will show that both channels are safe although they disagree.
2. *Risk-based calculation,* where the motion plans from one channel are analyzed against world models and object motion predictions from other channels. This cross-channel analysis computes the probability of collision in future, e.g. using the last safe intervention time [36] to identify a last time moment to switch to another AD channel or to perform an evasive maneuver on time. Note that this approach greatly improves availability of ADS by giving channels the time to correct themselves and filter out intermittent issues.
3. Another interesting method for the cross-channel analysis is *machine learning*. For example, [43] proposes neural network RiskNet to identify collision risks in camera images using optical flow. By analogy, one can define a neural network model operating on the high-level cross-channel state information or its derivatives.

In the Daruma design pattern the inter-channel and cross-channel analysis results (e.g. similarity matches, risk functions, preference weights) are fed into the safety fusion module. The safety fusion module computes an aggregated safety score per AD channel based on the cross-channel analysis results and the traditional metrics. Subsequently, the aggregated safety score allows the high-level arbiter to select a safe channel while taking into account comfort, efficiency, and other metrics. Various techniques can be applied to the safety fusion, such as hierarchical decision tree or a numerical weights calculation. Note that in a simple implementation of the safety fusion, the aggregated safety scores can be binary – either zero or one to clearly indicate preference of the safety subsystem.

The effectiveness of the proposed Daruma design pattern strongly depends on the quality of the AD channels. If a channel is poor in many situations, it can decrease the overall system performance. Therefore, it is important to develop and select diverse channels with complementary capabilities and overall high performance at design time. In other words, the designer has to reduce the white hole in Figure 9 by properly dimensioning and positioning the channel capabilities.

The proposed Daruma architectural design pattern enables advanced cross-channel analysis techniques *without compromising AD availability*. Instead of engaging in a traditional safe stop or disengagement upon detection of disagreement among channels, Daruma architecture enables switching to a safer channel to continue safe driving with minimum discomfort. Another important virtue of the proposed design pattern is scalability. Due to its modular design the number of channels in Daruma can be increased to boost safety without compromising AD availability. Compared to the monitor/actuator approach described in [27], the scalability of Daruma is an important advantage to create additional commercial value. Moreover, the Daruma architectural design pattern is compatible with state-of-the-art safety measures and architectures:

1. classical intra-channel safety mechanisms, such as sensor fusion and RSS [30], can be used as is;
2. if the architecture does not have a multi-channel arbiter or the arbiter's interface is designed differently from ours such as the one proposed in [7], the aggregated safety scores from our cross-channel analysis can still contribute to the Safety Performance Indicators [44] or serve as an independent cross-channel safety monitor output.

5. CONCLUSIONS

Redundant multi-channel AD architectures are used in the automotive industry to boost fault tolerance. Our work analyzes the extension of these redundant architectures to also mitigate FIs in the heterogenous AD channels.

To learn about the causes and consequences of FIs, we studied over 2500 DMV disengagement records from the year 2021 and over 10 hours of road test videos of vehicles from 9 different car companies. We observe that FIs of automated driving systems can be just as hazardous as faults, if not more dangerous. Furthermore, FIs are much more frequently the reason of ADS disengagement than system faults based on our observation and data analysis. Our analysis reveals that FIs are problematic and frequently occur not only in the world model generation involving perception and sensor fusion, but also in other AD algorithms with uncertainty, such as motion planning, localization, and vehicle control. To systematically perform SOTIF analysis of triggering conditions and enable development of FI mitigation techniques, we present a novel classification and characterization of OIs based on the available road test data.

Furthermore, we survey state-of-the-art techniques to mitigate FIs and formulate a hypothesis that in an ADS with multiple heterogenous channels, there exist times when some AD channels encounter FIs while the others do not. Our open-loop simulations with three realistic AD channels (Apollo 5.0, Autoware.Auto, and openpilot) support our hypothesis demonstrating that in the same driving scenario different AD channels can be complementary to each other in terms of capabilities to drive safely. Thus, by evaluating the safety of each channel via cross-channel analysis and switching to a sufficiently safe channel at runtime, the ADS safety and availability can be increased. Finally, we propose and discuss a generic architectural design pattern, called Daruma, integrating cross-channel analysis, safety fusion and arbiter functions. The Daruma architectural design pattern reuses state-of-the-art redundant AD architectures typically deployed for fault tolerance to mitigate FIs, while preserving the benefits of the intra-channel FI mitigation techniques in prior works. The Daruma design pattern also allows scaling up safety without sacrificing AD availability, while providing rich opportunities to mitigate many categories of FIs.

6. LIMITATIONS AND FUTURE WORK

In this paper we laid the groundwork for a novel conceptual framework for mitigation of the impact of FIs. However, we recognize that our analytical and experimental research had several limitations:

1. Some public road test videos were published by independent users, other videos were published by the company developing the ADS. The latter may have selected more favorable scenarios for the ADS in question, which can skew the FI statistics in our work.
2. The AD channels used in our open-loop simulation experiments have limited features, which may have affected the occurrences of FIs.
3. The triggering condition and the origin of the FIs are hard to pinpoint in our study due to limited available information visible in the public road test and on the ADS displays.

Future research directions include study of FIs related to traffic rules, evaluation of the cross-channel analysis architecture pattern in simulations and road tests, and algorithmic research of the cross-channel analysis and safety fusion. Also, we can update our OI characterization by studying newly released road tests reports, such as the initial data on safety performance of advanced vehicle technologies released by the U.S. National Highway Traffic Safety Administration (NHTSA) [45], and experimenting with more advanced AD channels when they become available.

ACKNOWLEDGEMENTS

We thank Gerardo Daalderop and Chenchen Dai from NXP Semiconductors for fruitful discussions about the presented work. This work was partially funded by the Dutch Research Council (NWO) through the NEON project.

REFERENCES

[1] P. Hogeveen, M. Steinbuch, G. Verbong and A. Hoekstra, "The Energy Consumption of Passenger Vehicles in a Transformed Mobility System with Autonomous, Shared and Fit-For-Purpose Electric Vehicles in the Netherlands," *The Open Transportation Journal,* vol. 15, pp. 201-209, 2021.

[2] Mobileye, "Unedited Ride in Mobileye's Self-Driving Car," 2020. [Online]. Available: https://www.youtube.com/watch?app=desktop&v=hCWL0XF_f8Y&ab_channel=MobileyeanIntelCompany. [Accessed October 2021].

[3] Cruise, "105 minutes of Cruise fully autonomous drving in San Francisco," November 2021. [Online]. Available: https://youtu.be/HiG__iqgYHM. [Accessed November 2021].

[4] Waymo, "How the Waymo Driver navigates San Francisco streets," October 2021. [Online]. Available: https://www.youtube.com/watch?v=2CVInKMz9cA. [Accessed March 2022].

[5] "FSD Beta 10.4 Drive to Ponce City Market Atlanta," November 2021. [Online]. Available: https://youtu.be/2T9GUfzQ7HY. [Accessed December 2021].

[6] "ISO 26262:2018 Road vehicles – Functional safety," 2018.

[7] S. Fürst, "Scalable, Safe and Multi-OEM Capable Architecture for Autonomous Driving," 9th Vector Congress, Stuttgart, Germany, 2018.

[8] "ISO 21448:2022 Road vehicles — Safety of the intended functionality," June 2022. [Online].

[9] "Tesla FSD Beta V10.9 and Human Driving in Slippery Snow," January 2022. [Online]. Available: https://www.youtube.com/watch?v=-R1mvUiDuGw&t=832s. [Accessed Jauary 2022].

[10] D. J. Yeong, G. Velasco-Hernandez, J. Barry and J. Walsh, "Sensor and Sensor Fusion Technology in Autonomous Vehicles: A Review," *Sensors,* vol. 21, no. 2140, 2021.

[11] A. Mehmed, W. Steiner and A. Čaušević, "Systematic False Positive Mitigation in Safe Automated Driving Systems," in *International Symposium on Industrial Electronics and Applications (INDEL)*, Banja Luka, Bosnia and Herzegovina, 2020.

[12] S. Poledna, Fault-tolerant real-time systems: The problem of replica determinism, vol. 345, Springer Science & Business Media, 2007.

[13] M. Törngren, X. Zhang, N. Mohan, M. Becker, L. Svensson and X. Tao, "Architecting Safety Supervisors for High Levels of Automated Driving," in *21st International Conference on Intelligent Transportation Systems (ITSC)*, Maui, Hawaii, 2018.

[14] C. Guo, G. Pleiss, Y. Sun and K. Q. Weinberger, "On calibration of modern neural networks," in *the 34th International Conference on Machine Learning (ICML)*, Sydney, Australia, 2017.

[15] A. Rudenko, L. Palmien, M. Herman, K. M. Kitani, D. M. Gavrila and K. O. Arras, "Human motion trajectory prediction: a survey," *The International Journal of Robotics Research,* 2020.

[16] Z. Zhong, Z. Hu, S. Guo, X. Zhang, Z. Zhong and B. Ray, "Detecting Multi-Sensor Fusion Errors in Advanced Driver-Assistance Systems," in *The 31st ACM SIGSOFT International Symposium on Software Testing and Analysis (ISSTA)*, 2022.

[17] UL Standards, "ANSI/UL 4600:2022. Standard for Evaluation of Autonomous Products (2022)," March 2022. [Online]. Available: https://www.shopulstandards.com/ProductDetail.aspx?productId=UL4600_2_S_20220315. [Accessed June 2022].

[18] Carlifornia Department of Motor Vehicles (DMV), "Autonomous Vehicle Collision Reports," [Online]. Available: https://www.dmv.ca.gov/portal/vehicle-industry-services/autonomous-vehicles/autonomous-vehicle-collision-reports/. [Accessed March 2022].

[19] Carlifornia Department of Motor Vehicles (DMV), "2021 Autonoumous Vehicle Disengagement Reports (CSV)," 2022. [Online]. Available: https://www.dmv.ca.gov/portal/file/2021-autonomous-vehicle-disengagement-reports-csv/. [Accessed March 2022].

[20] A. Sinha, S. Chand, V. Vu, H. Chen and V. Dixit, "Crash and disengagement data of autonomous vehicles on public roads in California," scientific data, November 2021. [Online]. Available: https://www.nature.com/articles/s41597-021-01083-7.

[21] V. V. Dixit, S. Chand and D. J. Nair, "Autonomous Vehicles: Disengagements, Accidents and Reaction Times," *PLOS One,* vol. 12, no. 11, 2016.

[22] S. S. Banerjee, S. Jha, J. Cyriac, Z. T. Kalbarczyk and R. K. Iyer, "Hands Off the Wheel in Autonomous Vehicles?: A Systems Perspective on over a Million Miles of Field Data," in *48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, Luxembourg, 2018.

[23] A. Adee, R. Gansch, P. Liggesmeyer, C. Glaeser and F. Drews, "Discovery of Perception Performance Limiting Triggering Conditions in Automated Driving," in *5th International Conference on System Reliability and Safety (ICSRS)*, Palermo, Italy, 2021.

[24] H. Martin, B. Winkler, S. Grubmüller and D. Watzenig, "Identification of performance limitations of sensing technologies for automated driving," in *IEEE International Conference on Connected Vehicles and Expo (ICCVE)*, Graz, Austria, 2019.

[25] "PAS 1883:2020 Operational Design Domain (ODD) taxonomy for an automated driving system (ADS) – Specification," 2020.

[26] "J3016_202104 Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," 2021.

[27] P. Koopman and M. Wagner, "Challenges in Autonomous Vehicle Testing and Validation," *SAE International Journal of Transportation Safety,* vol. 1, no. 4, 2016.

[28] G. Habibi, N. Jaipuria and H. P. Jonathan, "SILA: An Incremental Learning Approach for Pedestrian Trajectory Prediction," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, WA, USA, 2020.

[29] S. Vogel, C. Schorn, A. Guntoro and G. Ascheid, "Efficient Stochastic Inference of Bitwise Deep Neural Networks," in *Advances in Neural Information Processing Systems (NIPS) Workshop on Efficient Methods for Deep Neural Networks (EMDNN)*, Barcelona, Spain, 2016.

[30] S. Shalev-Shwartz, S. Shammah and A. Shashua, "On a Formal Model of Safe and Scalable Self-driving Cars," October 2018. [Online]. Available: https://arxiv.org/abs/1708.06374.

[31] D. Nistér, H.-L. Lee, J. Ng and Y. Wang, "An Introduction to the Safety Force Field," 2019. [Online]. Available: https://www.nvidia.com/content/dam/en-zz/Solutions/self-driving-cars/safety-force-field/an-introduction-to-the-safety-force-field-updated.pdf.

[32] A. Mehmed, W. Steiner, M. Antlanger and S. Punnekkat, "System Architecture and Application-Specific Verification Method for Fault-Tolerant Automated Driving Systems," in *IEEE Intelligent Vehicles Symposium (IV)*, Paris, France, 2019.

[33] T. Fruehling, A. Hailemichael, C. Graves, J. Riehl, E. Nutt and R. Fischer, "Architectural Safety Perspectives & Considerations Regarding the AI-based AV Domain Controller," in *IEEE International Conference on Connected Vehicles and Expo (ICCVE)*, Graz, Austria, 2020.

[34] B. Zhang, J. de Winter, S. Varotto, R. Happee and M. Martens, "Determinants of take-over time from automated driving: A meta-analysis of 129 studies," *Transportation Research Part F: Traffic Psychology and Behaviour,* vol. 64, pp. 285-307, 2019.

[35] Mobileye, "True Redundancy," Mobileye, [Online]. Available: https://www.mobileye.com/true-redundancy/. [Accessed March 2022].

[36] C. A. J. Hanselaar, E. Silvas, A. Terechko and W. P. M. H. Heemels, "Safety Shell: Reducing performance-limits in autonomous vehicles," in *The 25th IEEE International Conference on Intelligent Transportation Systems (ITSC)*, Macau, China, 2022.

[37] "SVL Simulator," LG Electronics America R&D Lab, [Online]. Available: https://wise.svlsimulator.com/simulations. [Accessed March 2022].

[38] Baidu Apollo, "ApolloAuto," 2019. [Online]. Available: https://github.com/ApolloAuto/apollo/tree/v5.0.0. [Accessed February 2022].

[39] SVL Simulator, "Running Apollo 5.0 with SVL Simulator," [Online]. Available: https://www.svlsimulator.com/docs/system-under-test/apollo5-0-instructions/. [Accessed February 2022].

[40] Autoware.Auto, "Autonomous Valet Parking Demonstration," 2020. [Online]. Available: https://autowarefoundation.gitlab.io/autoware.auto/AutowareAuto/avpdemo.html#autotoc_md72.

[41] comma.ai, "openpilot," [Online]. Available: https://github.com/commaai/openpilot. [Accessed December 2021].

[42] Aptiv, Audi, Baidu, BMW Group, Continental, Daimler, FCA, HERE, Infineon, Intel, and VW, "Safety First for Automated Driving," 2019. [Online]. Available: https://www.mercedes-benz.com/content/dam/brandhub/innovation/safety-first-for-automated-driving/safety-first-for-automated-driving-withepaper_en.pdf.

[43] T. J. Schoonbeek, F. J. Piva, H. R. Abdolhay and G. Dubbelman, "Learning to Predict Collision Risk from Simulated Video Data," in *IEEE Intelligent Vehicles Symposium (IV)*, Aachen, Germany, 2022.

[44] P. Koopman, "Safety Performance Indicators (SPIs) for Autonomous Vehicles," 2020. [Online]. Available: https://users.ece.cmu.edu/~koopman/lectures/L124_SPI_vs_KPI.pdf. [Accessed March 2022].

[45] NHTSA, "NHTSA Releases Initial Data on Safety Performance of Advanced Vehicle Technologies," June 2022. [Online]. Available: https://www.nhtsa.gov/press-releases/initial-data-release-advanced-vehicle-technologies. [Accessed June 2022].

EVALUATION OF SAFETY AND MOBILITY AROUND LOW SPEED AUTOMATED VEHICLE THROUGH REAL WORLD DEPLOYMENT IN URBAN ROADWAY SYSTEM

Abhijit Sarkar
Virginia Tech Transportation Institute
Blacksburg, Virginia, USA

Vaibhav Sundharam
Virginia Tech Transportation Institute
Blacksburg, Virginia, USA

Kevin Grove
American Trucking Association
USA

Paper Number 23-0119

ABSTRACT

Low-speed automated vehicles (LSAVs) are a new type of road transportation option that can be deployed in densely populated areas to connect passengers to existing transit systems. These vehicles are designed to operate at low speeds (often in the < 25 mph range) in complex operational design domains and can be retrofitted to accommodate at-risk road users, thereby making transportation even more accessible. Also, as many LSAVs are electric vehicles, they also show great potential for benefiting the climate. Even though there are numerous benefits to deploying LSAVs, several hurdles must be overcome to achieve success. For example, it is unclear how an LSAV deployed in a regular lane with a 25 mph or higher speed limit may affect other traffic. Also, it is unclear how vulnerable road users like pedestrians and bicyclists behave around LSAVs, as human machine interfaces for such interactions are not yet properly developed. These variables pose multiple questions for the safety, mobility, and operation of LSAVs in unrestricted operational domains. For example, an LSAV's low speed may cause other vehicles to operate at a lower speed, causing more vehicles to queue behind it. This may also make some drivers frustrated and lead them to become involved in dangerous driving situations like overtaking and cut-ins. In this work, we studied real world deployment of an LSAV on the US roadway to understand driver behaviors via 360 degree camera views from cameras installed on the LSAV. We examined the problems encountered during the deployment of an EasyMile (EZ-10) LSAV. We specifically investigated events from a real-world deployment during which the EasyMile LSAV needed to stop. The EasyMile deployment studied in this work included cameras that captured the 360 degrees of roadway environment around the vehicle. We developed a scene perception algorithm using computer vision technology to track other roadway agents like cars, pedestrians, and bicyclists around the EasyMile LSAV. We used object detection and tracking algorithms to track the trajectories of each of the roadway agents. Then we used perspective geometry and camera specifications to find the relative distances and speeds of these agents with respect to the EasyMile. This helped us understand the configurations of the traffic around the LSAV and study other drivers' temporal behavior. For example, the collected data shows the approach of any vehicle towards the EasyMile. Finally, we used this information to study other vehicles' maneuvers and show how the information from the cameras can be used to study simple maneuvers of other vehicles such as cut-ins, lane changes, and following behavior.

INTRODUCTION

Recent efforts to design, test, and deploy AVs have resulted in the development of a variety of configurations. These configurations often differ in size, sensor capabilities, sensor integration, capacity, and other features depending on the operational design domain (ODD) being targeted. One particularly interesting category of AV that has emerged is the LSAV, as depicted in Figure 1. What makes this type of vehicle interesting is the relatively complex ODD in which it is designed to operate. LSAVs are often designed to operate around streets with a variety of other road users and with relatively complex traffic controls in place. Many of these vehicles are designed to hold multiple passengers and are deployed with the intent of supplementing or connecting existing transit systems. To accommodate this more complex ODD, LSAVs often operate at relatively low speeds below 25 mph.

LSAVs face several hurdles to deployment but also offer significant benefits if the hurdles can be overcome. The ability of LSAVs to operate on streets means they can connect riders to common destinations. The ability to operate on streets also means LSAVs could be deployed in densely populated locations where there is a larger pool of users. Finally, the ability of LSAVs to supplement transit can improve riders' connectivity, reliability, and access [1]. Importantly, LSAVs can also be designed to accommodate vulnerable road users whose mobility may be limited. This means LSAVs could provide benefits to an underserved segment of transit users.

Figure 1. Photo. A picture of VTTI's EZ-10 LSAV ready for operation.

The LSAV's autonomy can have great implications in certain operating scenarios. For example, LSAVs can be used for transportation of medical patients, older passengers, and other populations unable to drive (or who do not need to drive) [2-4]. Historically, LSAVs have been deployed on predefined short-range routes. In most cases, the vehicle's operating speed is below 20 mph. As a result, when deployed in a mixed-traffic scenario, the LSAV's average speed is often 10 to 15 mph below the average speed of the traffic surrounding it. This brings up questions and concerns about the effect of LSAVs in a mixed-traffic scenario and whether they pose any safety threat. In this study, we looked at multiple examples of other vehicles following the LSAV for an extended period. We specifically looked at vehicles behind the LSAV and vehicle in front of the LSAV. The helps us to evaluate the deployment of LSAV in real world through cameras mounted around the vehicle. The camera captures dynamics of the other roadway objects around the LSAV. We used advanced computer vision algorithm to track the trajectories and kinematics of each of the roadway objects that are present in the scene. we demonstrate how we can use the trajectories solely from the 2D camera data and retrieve 3D information about roadway agents, including their relative distance and speed in world coordinates. We further looked at the aggregated behavior of the traffic behind the LSAV and in front of LSAV to demonstrate the trends of the traffic.

EXPERIMENTATION AND DATA

Fairfax Relay Deployment

Shortly after exploring these issues at VTTI's facilities, there was an opportunity to explore them in a real-world test deployment. A public-private partnership that included Fairfax Country, VA, Dominion Energy, EDENS (Mosaic), The Virginia Department of Rail and Public Transportation, The Virginia Department of Transportation, VTTI, and George Mason University planned a pilot deployment of an LSAV in Northern Virginia for 2020. The goal of the pilot was to learn how the technology could be deployed safely and effectively and provide an opportunity to learn more about real-world interactions. The shuttle was purchased by Dominion Energy and operated by Fairfax Country with TransDev managing the vehicle, its maintenance, and its operators.

The vehicle was branded the Relay and configured to follow a pre-programmed path during operations. The Relay uses localization LIDAR, GPS, and internal sensors to follow the pre-programmed path. The Relay also uses 3D and 2D LIDARs to detect objects around the shuttle. The localization LIDAR has a range of approximately 492 feet, the 3D LIDAR has a range of approximately 262 feet, and the 2D LIDAR has a range of approximately 131 feet. The ranges of these LIDARs give the Relay the ability to detect changes in the environment or potential conflicts within the environment quickly and at a safe distance. VTTI was able to work with the partners to install data recording systems on the Relay. These systems recorded video 360 degrees around the vehicle in high definition. This video allowed VTTI to see the context of conflicts that emerged during operations. VTTI recorded the video on secure encrypted drives that were brought back to Virginia Tech for processing. Figure 2 shows an example of the video recording from the VTTI deployment. The Fairfax deployment had similar arrangements. In this paper we have mainly used camera data from outward looking videos.

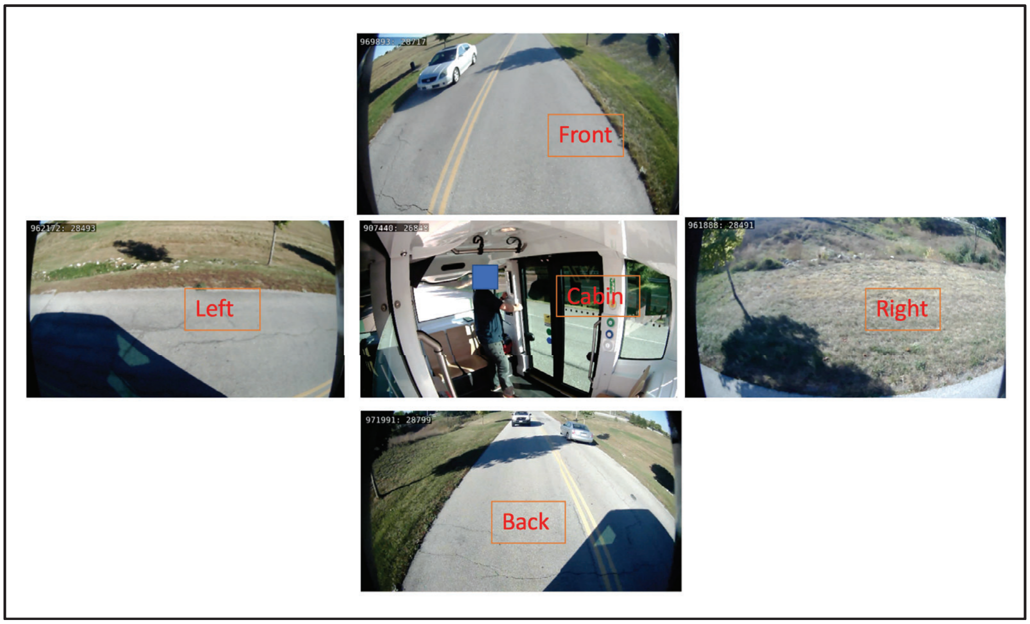


Figure 2. Video images. Example of scene cameras capturing 360-degree information around the LSAV.

COMPUTER VISION BASED CAMERA PROCESSING FOR SCENE PERCEPTION

Scene perception typically involves a detailed understanding of the scene using sensor data. In this case, we used monocular camera images to understand the construction of the roadway scene. For each roadway scene, there may be multiple roadway agents and geometric structures that play a key role. The roadway agents mainly involve the objects present in the scene, such as cars, trucks, pedestrians, and bicycles. Correspondingly, some other object classes also play a part in understanding the road, including road signs, traffic lights, work zone objects, etc. The geometric structural cues are mostly embedded in roadway boundaries, lane lines, crosswalks, etc. Sometimes, depth measurements also help provide an understanding of relative locations of objects and distance from the cameras. Each one of these objects and geometries can be perceived in pixel space by CV algorithms. In this paper, we discuss two specific aspects that are key to understanding the dynamics of other roadway agents around the EasyMile LSAV.

Object Detection and Tracking

Object detection refers to identifying an object in an image and locating that object in the pixel space. Over the last decade, deep learning methods have significantly enhanced the capability to detect objects in an image [5]. A typical object detection algorithm trains on a large set of annotated object classes and builds the capability to specifically localize the sets of pixels in an image to indicate a specific object class, such as a car, person, cat, dog, etc. Often the pixel space of an object is identified by drawing a BBox [6,7], as shown in Figure 3.

Object tracking, on the other hand, refers to the task where each object in a sequence of frames is tracked over time [8-10]. In a more formal definition of the task, each object in an image is assigned a unique object ID, and the object ID is maintained across the frames if the object is present in the FOV. This task mainly belongs to the standard CV task of "multi-object tracking," which also pertains to the subtasks of reidentification (i.e., identifying the same object from one frame to another) and motion prediction. Therefore, an effective algorithm should learn key features of the object from one frame and then should be able to transfer that knowledge into the next frame to reidentify the same object with high efficiency. As shown in Figure 3, each object is assigned a unique object id (e.g., 141, 147 etc.). Thus, through the object detection and tracking we can identify the class of the roadway object and track them over time.

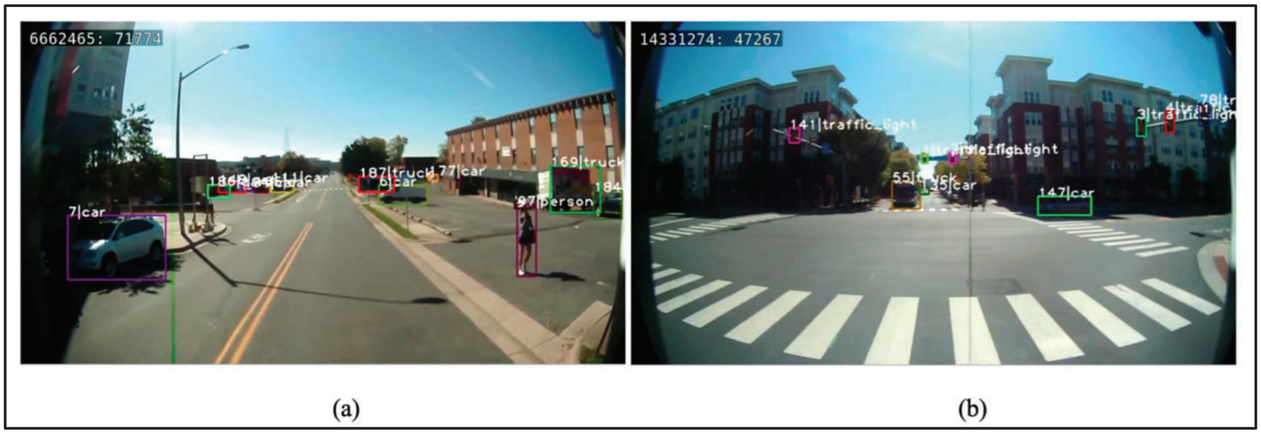


Figure 3. Example of detected objects in the EasyMile dataset.

Lane Detection

Lane lines define the boundary of a specific lane that a vehicle is supposed to follow unless it intends to change lanes. In the context of the project, lane line detection helps us understand the position of the ego vehicle (EasyMile) as well as the position of the other vehicles. We deployed an ultra-fast lane detection algorithm [11] to detect lane lines. It should be noted that the lane detection was performed on the original distorted images, and each lane line from the distorted image was subsequently undistorted and warped to the BEV. An example of detected lane lines on the EasyMile dataset is presented in Figure 4.

Figure 4. Video images. Example of lane detection using ultra-fast lane detection algorithms on the (a) front and (b) rear video from the ego vehicle's camera.

Kinematics of Roadway Objects

Objects like cars, trucks, pedestrians, etc., populate a road scene and interact with each other. Their movements are codependent on other objects present on the road (other cars, pedestrians, etc.). To study the interaction and behavior of these objects around the ego vehicle, it is imperative to measure the distance and speed of such objects. A monocular camera does not preserve the depth. Therefore, we used knowledge of real-world measurements to calculate the distance of any pixel from the ego. To adequately understand the 360-degree behavior we need to study both longitudinal and lateral distance. Assuming a bird eye view of the scene, we need to create a pixel to real world 2D transformation map where we know how distance between consecutive pixels translate to real world coordinate (meter). Hence, we used two references. We used the lane width (12 ft) as the reference for lateral behavior. We used the length of right turn arrow as the reference for longitudinal distance. Once that is done, we use the bounding box from the object detection and tracking to find of distance of any roadway object. Then we used a 4th order Butterworth filter to remove noise. From the distance (both lateral and longitudinal), we used time derivative to compute speed in each direction.

TRAFFIC BEHAVIOR AROUND THE LSAV

In this section, we demonstrate how the information from the previous sections can be useful to understand the traffic behavior around the EasyMile. We specifically look at three scenarios and how each could result in potential safety hazards to the agents surrounding the AV, followed by a visualization of object behavior surrounding the LSAV to determine safety implications:

- A vehicle following the EasyMile AV
- A vehicle cut-in front of the EasyMile AV

Vehicle Following the LSAV

Figure 5 depicts a typical scene where a vehicle is following the AV. The location of the starting point of the detected vehicle along with its relative position with respect to the ego vehicle is plotted. As depicted in Figure 5(a), the vehicle with ID:2 is following the AV at an acceptable distance. However, after a few seconds, there is a rapid change in the velocity of the vehicle, which is evident from Figure 6 (red circle) followed by a 15-second halt. This sudden deceleration is due to the AV abruptly stopping in the middle of the road to give way to an oncoming distracted pedestrian, who is talking on the phone, as evident in Figure 5(e). A human driver would likely have analyzed the situation a priori, and hence such sudden braking would not have been required. Such a situation can lead to safety-critical conditions, possibly causing a crash if the driver following the AV is not maintaining sufficient distance. Also, any delay in the response time of the AV could have caused a serious roadside incident involving pedestrians. To better inform the road users about the deployment of LSAVs, adaptive warning signs can be installed onboard the AV that monitor the traffic right behind the AV and inform drivers to maintain a safe distance from the vehicle.

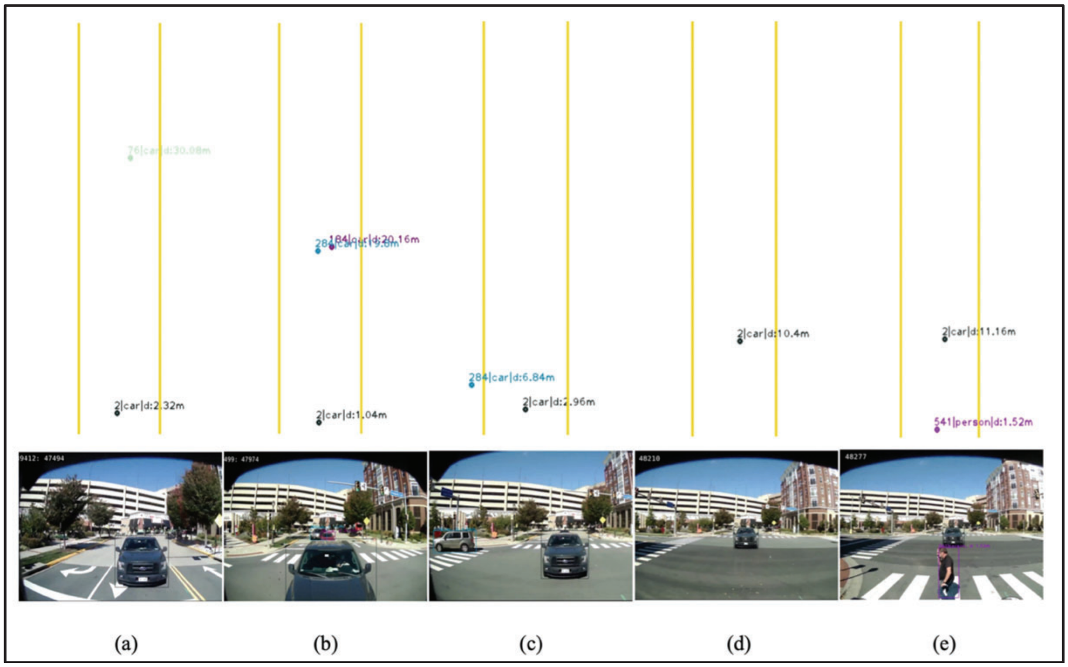


Figure 5. Diagrams with video images. Frames from a sample video taken at different time intervals depicting the trajectory of a vehicle following the AV.

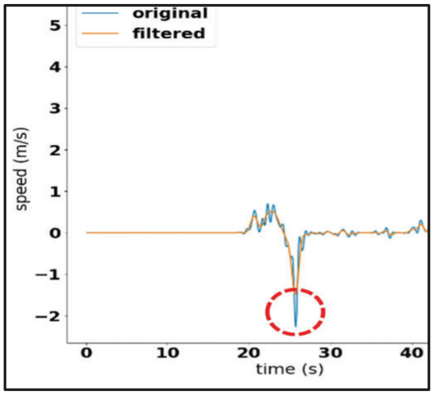


Figure 6. Rapid change in velocity of the vehicle (red circle) suggesting sudden deceleration.

Vehicle Cut-In Front of the LSAV

Being in the research and development phase, the EasyMile AV moves at speeds slower than recommended. Due to this, the drivers following the AV tended to get frustrated and try to overtake the AV, possibly under dangerous conditions. During our analysis of rear videos, we found multiple scenarios where the driver following the AV cut right in front of it to overtake it. Cut in refer to the behavior of another vehicle when then come in front of the another vehicle or between two vehicles [12]. In some situations, we saw drivers cutting into oncoming traffic to overtake the AV. One such case is presented in Figure 7, where the vehicle with ID:631 cuts in the front of the AV by crossing over to the opposite lane. A situation like this where individuals are involved in aggressive driving puts their lives and others' lives in jeopardy.

To mitigate such problems, the LSAV could be operated during non-peak hours when traffic is low. Also, during the initial stages of development, the LSAV can be operated on roads that have multiple lanes. This will give drivers additional room to maneuver their vehicles to overtake the AV, preventing safety-critical conditions. The data collected during this stage can be used to fine-tune the control systems onboard AVs to make them more robust for deployment in complex environments.

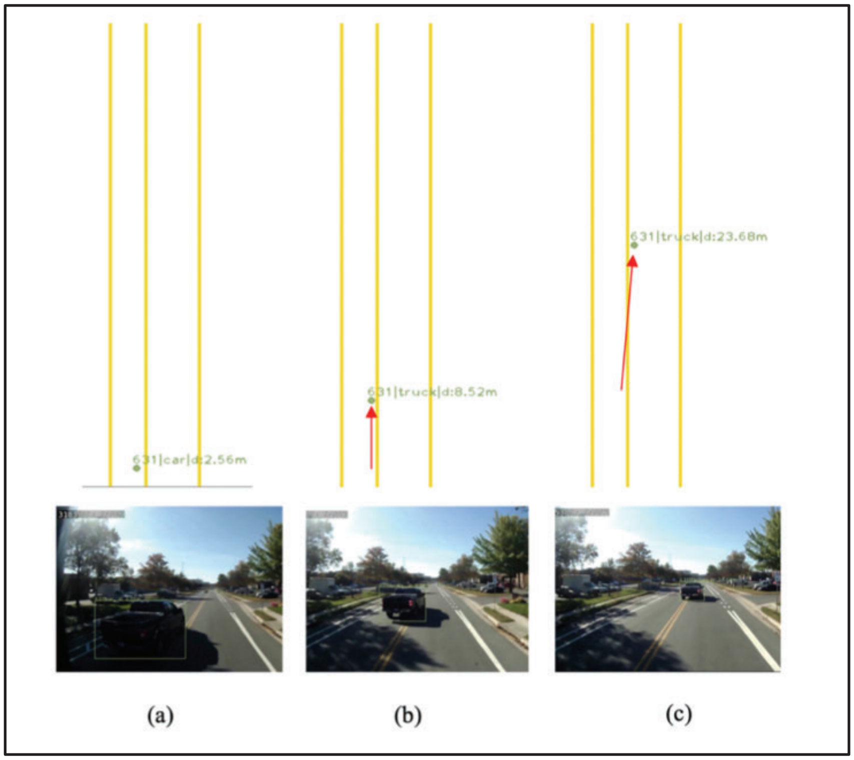
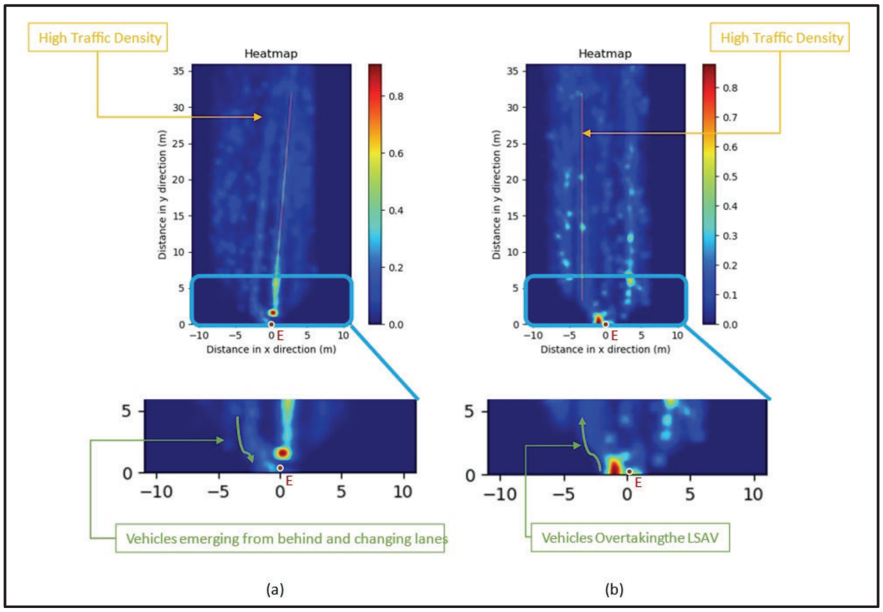


Figure 7. Diagrams with video images. Example of a vehicle cut-in front of the AV.

ANALYSIS OF TRAFFIC TRENDS

On the road, the LSAV is surrounded by numerous objects, and the way that these objects interact with the LSAV can help determine deployment safety implications. To analyze the interaction of LSAVs with various road agents, we plotted the traffic flow around the LSAV as a heatmap. To generate the holistic view of traffic density in the front and the rear of the LASV, we used 115 front camera and 131 rear camera videos. Furthermore, we aggregated 115 density arrays generated for front and 131 density arrays generated for the rear videos to get two 2D arrays (DF, DR) of shape (height, width, 1). We then normalized DF and DR independently, followed by Gaussian blurring with a 15x15 Gaussian kernel to get the final "flow" heatmap as depicted in Figure 8(a-b). Here, each pixel incorporates the density of traffic at that location for all front/rear videos. The distance resolution for both front and rear cameras is 36 meters. In each image, the location of the LSAV is shown as a brown dot marked "E" at location (0, 0). Followings are some of the observations from analyzing the front/rear traffic flow heatmap:

- From Figure 8(a), it is evident that there is a large concentration of traffic behind the LSAV. The vehicles stuck behind the LSAV would cut the lane to get ahead ("green arrow"). Additionally, the traffic building up behind the LSAV hindered the normal operation of the traffic by creating bottlenecks on the road.
- The previous observation is reinforced from the traffic distribution we achieved from the front camera. From Figure 8(b) we see that there is a large concentration of traffic ("red") right beside the LSAV as compared to any other area. This suggests a vehicle changing lanes from behind the LSAV ("green arrow") to overtake it due to the LSAV's slow operating speed.
- Additionally, from Figure 8(b), we can see that the traffic density is low in the lane where the LSAV is operating (i.e., vehicles are avoiding being in the same lane as the LSAV), increasing the traffic in other lanes and creating possible bottlenecks.



(a) (b)

Figure 8. Traffic flow heatmap. (a) Rear camera and (b) front camera. "E" denotes the location of the LSAV.

CONCLUSION

LSAVs are expected to be deployed across the market very soon. In this work, we studied a real-world deployment of EasyMile LSAVs in Fairfax County. We developed and tested advanced CV algorithms to understand the interactions of LSAVs with various road agents. We specifically looked at vehicles in front and behind the LSAV. We investigated the behavior and movement of road agents such as cars and pedestrians around the LSAV. In the process, we developed a scene perception algorithm based on object detection and object tracking to measure kinematic behavior of these roadway agents with respect to the LSAV. We further demonstrated a process to summarize behavior of each of the agents over time and developed a composite understanding of the interaction between the two agents on the road.

Finally, we used these methods to demonstrate some key events such as following, cut-in front of vehicle, yield to pedestrian, etc.

We specifically looked at following behavior and Through experimentation, we quantitatively evaluated how (a) a long following vehicle queue may be generated behind the LSAV and (b) how vehicles try to change lanes and overtake the LSAV in front of the ego. In our experiment, we saw examples of both cases. The long queue suggests that LSAVs may reduce the overall traffic speed of the lane, hence creating traffic jams and reducing volumetric throughput. Also we have seen close following, occasional lane change, and overtaking by vehicles coming from behind; thus LSAV may create safety-critical situations [13] by encouraging such unnecessary driving maneuvers. This adds to LSAVs' operating principles and underlying algorithms, which define their perception and control methodologies. Our study shows that the LSAVs are often too overprotective and apply brakes during non-critical cut-in events. These conditions also raise the question of adaptability and acceptance of this new technology by the public and how other drivers should behave around an LSAV. Therefore, the entities deploying LSAVs need to consider these pros and cons before finalizing LSAVs' path and deployment plan. Similarly, LSAV developers should invest in developing technologies that would help these vehicles operate in mixed-traffic scenarios.

ACKNOWLEDGEMENT

The authors like to thank National Surface Transportation Safety Center for Excellence (NSTSCE) for funding this effort [14]. The team also like to thank Dr. Aditi Manke, Dr. Jon Antin, and Dr. Jon Hankey for their valuable feedback; Dr. Mike Mollenhauer, Jean Paul Talledo Vilela for their help in accessing the data.

REFERENCES

[1] Litman, T. (2021). *Autonomous vehicle implementation predictions implications for transport planning.* Victoria Transport Policy Institute.

[2] National Highway Traffic Safety Administration. (2020). *Automated vehicles for safety.* https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety

[3] Claypool, H., Bin-Nun, A., & Gerlach, J. (2017). *Self-driving cars: The impact on people with disabilities.* Ruderman Family Foundation.

[4] Coyner, K., Blackmer, S., Good, J., Lewis, P., & Grossman, A. (2021). *Low-speed automated vehicles (LSAVs) in public transportation* (TRCP Research Report 220). Transit Cooperative Research Program.

[5] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436-444.

[6] Cai, Z., & Vasconcelos, N. (2019). Cascade R-CNN: High quality object detection and instance segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 43*(5), 1483–1498.

[7] Chen, K., Wang, J. Pang, J. Cao, Y. Xiong, Y. Li, X. Sun, S. Feng, W. Liu, Z. Xu, J., Zhang, Z., Cheng, D., Zhu, C., Cheng, T., Zhao, Q., Li, B., Lu, X., Zhu, R., Wu, Y., … Lin, D. (2019). *MMDetection: Open MMLab detection toolbox and benchmark*. arXiv preprint arXiv:1906.07155.

[8] Papakis, I., Sarkar, A., & Karpatne, A. (2020.) GCNNMatch: Graph convolutional neural networks for multi-object tracking via sinkhorn normalization. arXiv preprint arXiv:2010.00067

[9] Bergmann, P., Meinhardt, T., & Leal-Taixe, L. (2019). Tracking without bells and whistles. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 941-951).

[10] Papakis, I., Sarkar, A., & Karpatne, A. (2021, September). A Graph Convolutional Neural Network Based Approach for Traffic Monitoring Using Augmented Detections with Optical Flow. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)* (pp. 2980-2986). IEEE.

[11] Qin, Z., Wang, H., & Li, X. (2020). Ultra-fast structure-aware deep lane detection. arXiv:2004.11757.

[12] Sarkar, A., Engstrom, J., & Hanowski, R. (2022). Analysis of Car Cut-ins Between Trucks Based on Existing Naturalistic Driving Data. National Surface Transportation Safety Center for Excellence (NSTSCE)

[13] Sarkar, A., Krum, A., Hanowski, R., & Hickman, J. (2021, July). Responsibility Sensitive Safety Analysis of Truck Following in US Highway. In International Conference on Applied Human Factors and Ergonomics (pp. 119-126). Springer, Cham.

[14] Sarkar, A., Sundharam, V., Manke, A., & Grove, K. (2022). Camera-based Feature Identification for EasyMile Operation. National Surface Transportation Safety Center for Excellence (NSTSCE)

INVESTIGATION ON THE CONDITIONS FOR DISTURBANCES IN THE SAFETY PERFORMANCE ASSESSMENT ON THE PERCEPTION FUNCTION OF AUTOMATED DRIVING VEHICLES

Masao Nakagawa
Tetsuya Niikuni
Hiroyuki Yamamoto
National Traffic Safety and Environment Laboratory
JAPAN

Paper Number 23-0128

ABSTRACT

This study focuses on perception as a fundamental part of the function chain of automated driving systems. The dynamic control of automated driving vehicles will be operated based on the perception function resulting from processing information gathered by sensors. Factors influencing perception should be identified and determined for the safety performance assessment because such factors consequence the behavior of automated driving vehicle. Especially, the characteristics of radar on perception function in ADS was investigated, and conditions for disturbances to developing safety performance assessments was discussed.

1. INTRODUCTION

The vehicle regulation for Automated Lane Keeping Systems (ALKS) under the framework of the 1958 agreement in UNECE came into force in January 2021[1]. This regulation will promote the mass production of automated driving vehicles (ADVs) in markets. At the same time, evaluation methods for ADVs respecting technology neutrality will be desired to ensure vehicles' (including ADVs) safety on roads.

The regulation defines technical requirements for the vehicles with ALKS. However, test methods, including such testing conditions and criteria, are not clarified to produce sufficient reproducibility between testers yet.

This study focuses on perception as a fundamental part of the function chain of automated driving systems (ADSs). The dynamic control of ADVs will be operated based on the perception function resulting from processing information gathered by sensors. Factors influencing perception should be identified and determined for the safety performance assessment because such factors consequence the behavior of ADV.

Various types of sensors such as cameras, LiDAR, radar, and GNSS are involved in the perception function in ADS. As the first step, the characteristics of radar on perception function in ADS was investigated. A radar unit was installed in the test vehicle, and the distance to the target truck forward was detected in the proving ground. The distance was tuned as the test condition beforehand. The error between the actual and radar detected distances was evaluated. Furthermore, in the rainfall facility, the influence of rain as a natural disturbance is investigated.

2. FIELD OF VIEW TEST FOR EVALUATION ON SENSOR PERFORMANCE

The ALKS United Nations Regulation (UNR 157) describes requirements for the recognition performance of automated vehicles and their test methods.

Chapter 7 of the Regulation contains provisions on object and event detection and response, which defines the detection range for forward and side detection. As shown in Fig.1, this regulation specifies a frontal detection range of 46 m or more and a side detection range up to the outer edge of both adjacent lanes. In addition, the test requires detection of other road users within a range equal to or greater than this range.

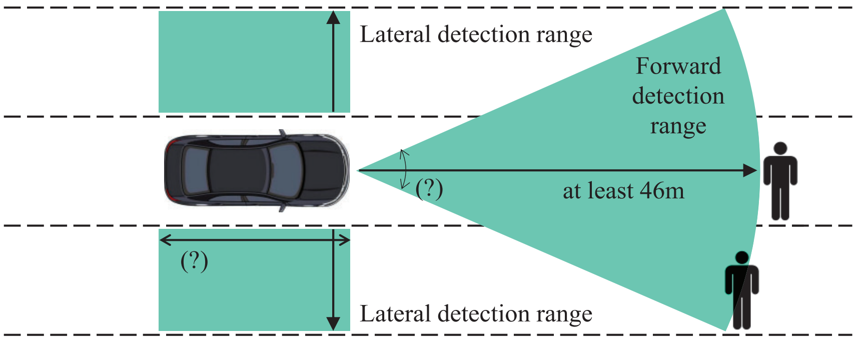


Fig.1 Range of recognition as defined in the ALKS UN Regulations

Chapter 5 of the appendix to the regulations describes the test method, which includes a field of view test, which describes the test requirements for recognition performance. Although these requirements include descriptions of recognition performance for stationary pedestrians at the outside edge of adjacent lanes and stationary motorcycles in their ego lane, they do not specify specific measurement conditions or evaluation criteria.

In a typical driving environment, such as road types, static or dynamic conditions, kinds of objectives and weather, etc. are considered to affect recognition performance, and depending on the combination of these items, there is a possibility that the object may not be recognized.

3. EXPERIMENTAL SETUP

For the experiment, a 79 GHz band stand-alone radar (RAD98, manufactured by Keycom) were used for verification. Figure 2 shows the appearance of the radar for verification, and Table 1 shows the specifications of the radar for verification. Although these parameters are design values, we also confirmed that the parameters were met by the measurement results.

In the experiment, the distance to the recognition target (hereinafter referred to as "preceding vehicle") was obtained by setting the origin at the radar for verification. The verification radar was attached to the leading edge of the vehicle, and the radar control software was used to acquire the radar recognition results. Medium-duty truck with manual transmission, as shown in Fig.3 were used as the preceding vehicles.

Experiments were measured under the evaluation conditions shown in Table 2 in a straight road geometry, referring to the conditions of the ALKS UN rules among the typical driving environment noted above.

Fig.2 The verification radar

Fig.3 Medium-duty truck as preceding vehicle

Table 1 Specifications of the verification radar

Frequency range	Short: 76~77GHz Middle and Long: 77~81GHz
Distance range/ accuracy	Short: 3~20m / ±0.24m Middle: 20~83m / ±0.40m Long: 83~250m / +1.00~-3.00m
Distance resolution	Short: 0.345m Middle: 1.055m Long: 5.751m
Velocity range/ accuracy	Short: ±10.8km/h / ±1.0km/h Middle: ±55.0km/h / ±2.0km/h Long: ±153.4km/h / ±2.7km/h
Horizonal angle range/ accuracy	Short: ±50° / ±2.790°~±3.789° Middle: ±50° / ±4.581°~±6.571° Long: ±40° / ±4.581°~±5.675°
Horizontal angle resolution	35°
Sampling frequency	10Hz

Table 2 Experimental conditions

	Conditions	
	Static	Dynamic
Status	Stationary in front	Moves at given speed
Lanes	Ego or adjacent lane	Ego or adjacent lane
Distance from ego vehicle	・3~20m (every 1m) ・20~80m (every 5m) ・80~200m (every 10m) (Total 42 conditions)	300m towards origin
Speed		Forward: Constant velocity at idling 1st gear: 7km/h 2nd gear: 12km/h 3rd gear: 18km/h Backward: Constant velocity at idling Back gear: 6km/h

4. RESULTS AND DISCUSSIONS

After describing recognition results under basic radar evaluation conditions, the effects of disturbances such as rainfall and structures on recognition performance are discussed as test condition.

4.1 Measurement results under static conditions

The measurement results for the static condition, in which the preceding vehicle is stationary in front of the ego vehicle, are shown. In the static condition, accurate distance (true value of distance) was obtained by placing the preceding vehicle at each landmark position actually measured from the origin. Figs.4 and 5 show the measured distance and measurement error against the true value of the distance measured by the verification radar.

Figure 4 shows the distance recognized by the radar relative to the true value (average value measured for 10 seconds at a frame rate of 10fps), and ideally, the measurement points should line up on a straight line where Y=X. While the radar was able to measure up to 100m in the condition where the vehicle ahead was in its ego lane, it was unable to measure in the range of 50m to 60m and in the range of 90m or more in the condition where the vehicle ahead was in the adjacent lane.

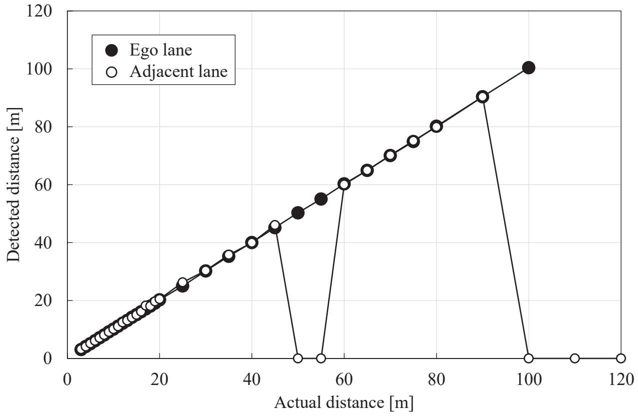
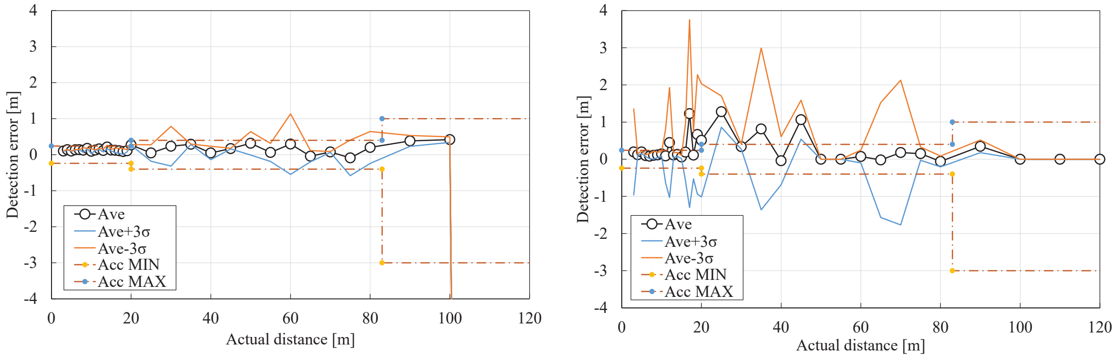


Fig.4 Detected distance at static conditions

Figure 5 shows the measurement error (mean and ±3σ) relative to the actual distance between vehicles. The error is acceptable if it is within the accuracy of the radar parameters listed in Table 2 (single dotted line in Fig.5). Under the condition that the vehicle ahead is in its ego lane, the error is within the accuracy. On the other hand, when the preceding vehicle is in the adjacent lane, the error is out of the range of accuracy for distances up to 50m.

The above results indicate that, under static conditions, the system can recognize a vehicle with a small error up to 100 m when the preceding vehicle is in its ego lane, but when the preceding vehicle is in a adjacent lane, the error becomes large at distances up to 50 m, resulting in poor recognition performance. When the preceding vehicle is in the adjacent lane, the measurement error may become larger due to the influence of radar reflected from the side of the preceding vehicle.



(a) Ego lane (b) Adjacent lane

Fig.5 Detection errors on static conditions

4.2 Measurement results under dynamic conditions

The measurement results are shown for dynamic conditions in which the vehicle ahead moves at a given relative speed. The RCS (radar cross section [dBsm][2]) is a value that expresses the size of the reflective surface and the ability of the radar to reflect the radio waves emitted by the radar and return the reflected waves to the radar detector. A large RCS value indicates that the radar can easily detect the target.

Figure 6 shows how the figure is viewed. The colors of the plots in the figure indicate the order of RCS, with blue, green, and red indicating higher RCS. Here, the color of the plot is only a rank order and does not correspond to the RCS value. The origin is the front end of the radar, and the blue line is the travel path of preceding vehicle (if it is in its ego lane, it is on the Y-axis origin; if it is outside its ego lane (adjacent lane), it is offset in the X-axis direction by that amount). The points around the trajectory are the preceding vehicle, and other objects in the vicinity, such as buildings and guardrails, are detected.

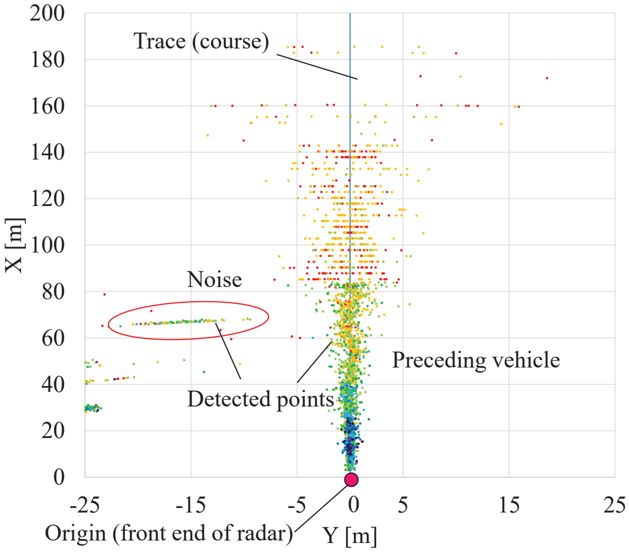
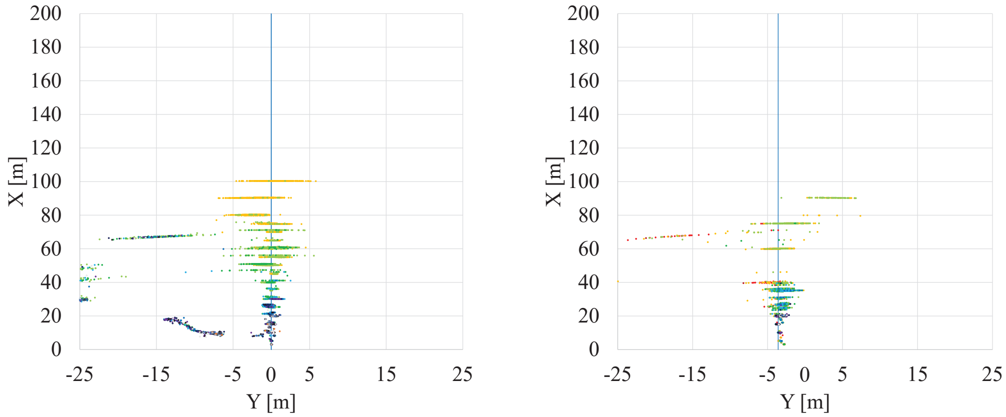


Fig.6 How to read the figure of detected points

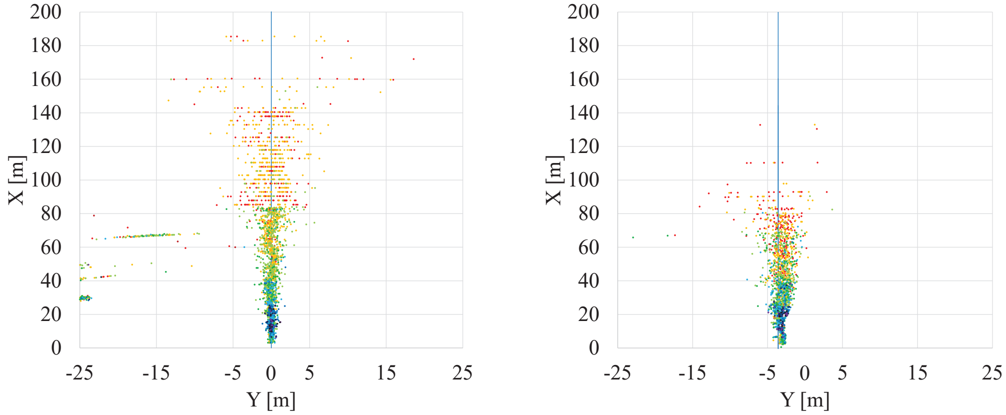
Figure 7 shows a comparison of static and dynamic conditions. In the static condition, the measurable range is about 100m when preceding vehicle is in the ego lane (as also shown in Fig.4). However, when there is a preceding vehicle in the adjacent lane, the distance that can be measured with good sensitivity becomes shorter, and the measurement stability in the middle part is significantly impaired. In addition, under the condition where the preceding vehicle was in the adjacent lane, the effective measurement points varied in the Y-direction due to the influence of lane position. Next, comparing the results with those of the dynamic condition, clear differences were observed between the dynamic condition and the static condition. The maximum measurable distance was about 100m in the static condition, whereas it was about 200m in the dynamic condition. It is also clear that the static condition has a larger variation in the Y-direction and noise in areas where no recognition target exists compared to the dynamic condition. On the other hand, there is less difference among the velocities of the preceding vehicle. In general, radar detects the phase difference caused by the RCS and the relative speeds of the ego vehicle and the preceding vehicle, and the combination of RCS and phase difference is used to identify the target. Under static conditions, the phase difference information is not available and noise sources such as guardrails are stationary, making it impossible to discriminate between the preceding vehicle and the noise based on the phase difference. Hence, under static conditions, the RCS of the preceding vehicle is easily buried in the noise and is considered to be difficult to recognize. On the other hand, under dynamic conditions, the preceding vehicle is considered easier to identify because it has a different phase difference from the noise component. The horizontal variation decreased as the relative speed increased, but there was no significant difference among the speeds.

For these reasons, it is considered that vehicle-mounted radar can recognize a moving preceding vehicle more favorably than a stationary preceding vehicle.



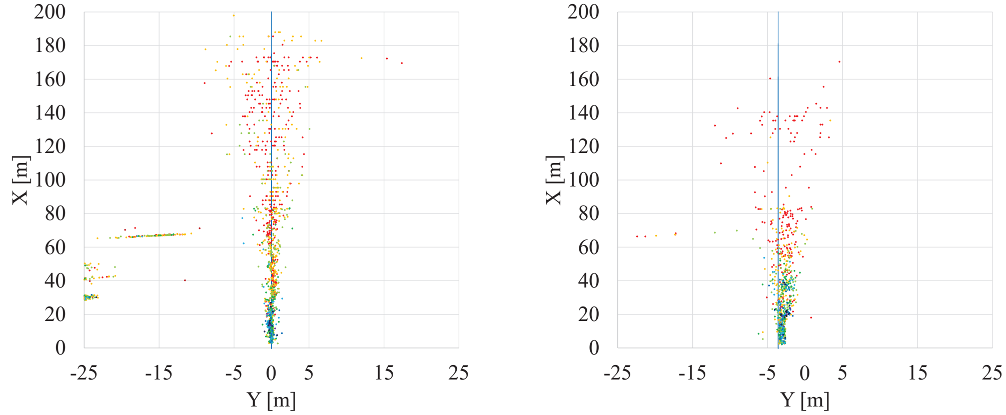
(a) Static condition, ego lane

(b) Static condition, adjacent lane



(c) Dynamic condition, 1st gear, ego lane

(d) Dynamic condition, 1st gear, adjacent lane



(e) Dynamic condition, 3rd gear, ego lane

(f) Dynamic condition, 3rd gear, adjacent lane

Fig.7 Comparison between static condition and dynamic condition

4.3 Effects of Disturbances from Buildings and Other Artificial Objects on Recognition Performance

The effect of disturbance from buildings and structures on recognition performance is evaluated. Comparison of the maximum measurable distance, etc. in the area lined with buildings (No.3 post location), the area with few buildings (No.6 post location) at the Kumagaya 1st proving ground of the National Traffic Safety and Environment Laboratory (NTSEL), and the specific environment test facility of the Japan Automobile Research Institute (JARI) is performed. Each evaluation environment is shown in Fig.8. Measurements were taken at the center of the second lane from the left edge of each environment.

(a) No.3 post location (b) No.6 post location

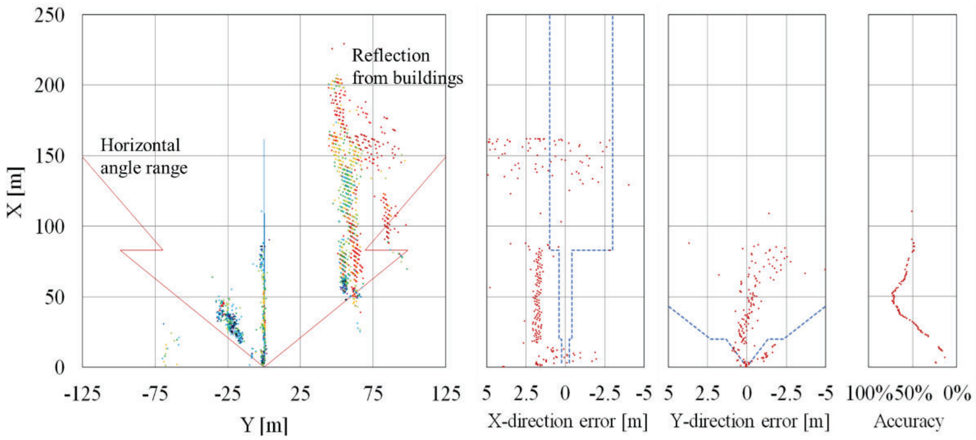
(c) Specific environment test facility

Fig.8 Test environments

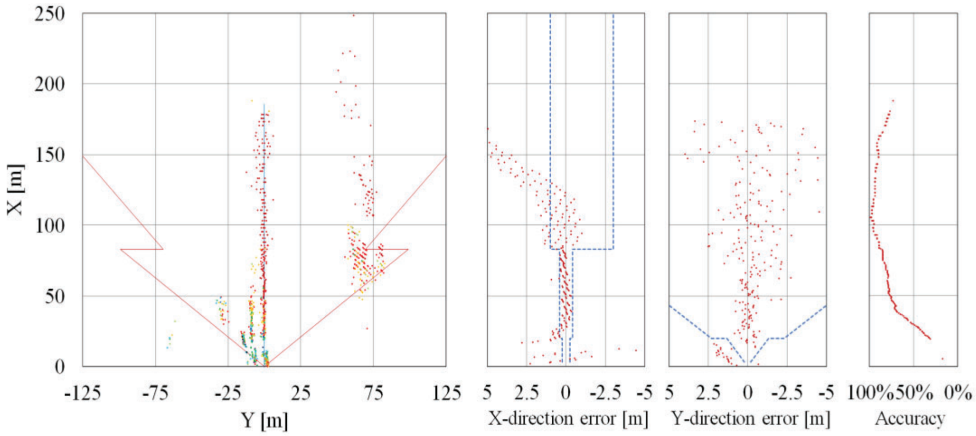
At the No.3 post location, buildings and steel walls are considered to be disturbances that affect recognition performance, while at the No. 6 post location, there are relatively few buildings, and recognition performance is considered to be less disturbed. In the specific environment test facility, the steel frame of the building is considered to affect recognition performance. Here, based on the results of Section 4.2, we verified the dynamic condition in which preceding vehicle travels ahead of its ego lane. The measurement results for each evaluation environment are shown in Fig.9. Figure 9 shows not only the error between the actual position and the position detected by the radar, but also the accuracy of recognition.

From Fig.9, comparing the No.3 and No.6 post locations, it can be seen that the group of points around Y=75m at the No. 3 post position is darker than that at the No. 6 post position. This is due to the influence of reflected waves from buildings, and is considered to be a clear difference in the evaluation environment. The maximum measurable distance at the No.3 post was about 160m, which was shorter than the distance at the No.6 post, about 180m. Furthermore, the maximum measurable distance at the No.3 post was about 60m, which was shorter than the distance at the No.6 post. The error in the X- and Y-directions at the No.3 post position was larger than that at the No.6 post position, suggesting that disturbance from buildings and other objects may cause a larger error.

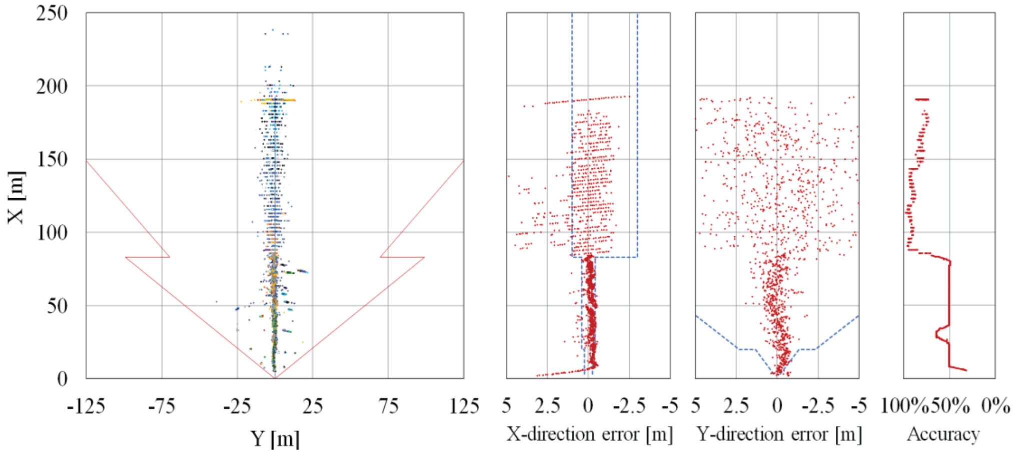
Furthermore, although the maximum measurable distance of the specific environment test facility was approximately 200m, the errors in the X and Y directions were large. The specific environment test facility is surrounded by a steel frame and is narrower in the horizontal direction than the Kumagaya No.1 proving ground, which is thought to make it easier for reflected waves from recognition targets to be buried by reflected waves from the walls of the facility. Therefore, it was found that artificial objects such as buildings affect the recognition performance such as the maximum measurable distance of the radar.



(a) No.3 post location



(b) No.6 post location



(c) Specific environment test facility

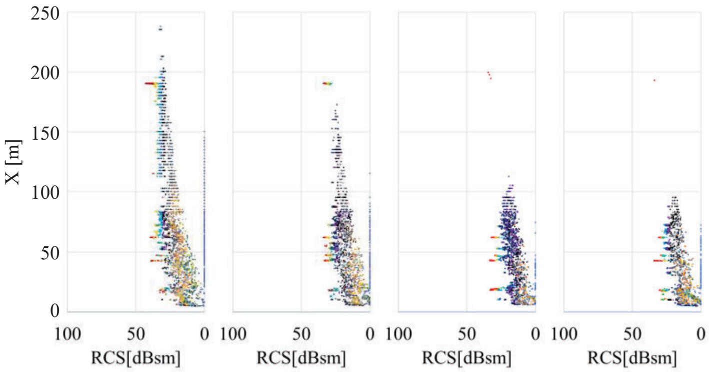
Fig.9 Detection results in each test environments

4.4 Effects of natural disturbances such as rainfall on recognition performance

To evaluate the effect of rainfall on radar recognition performance, measurements were conducted under artificially rainy conditions at the specific environment test facility. Figure 10 shows the rainfall in the facility.

Fig.10 Rainfall in specific environment test facility

This facility can reproduce precipitation amounts of 30mm/h, 50mm/h, and 80mm/h. At each precipitation level, measurements were taken under dynamic conditions. Figure 11 shows the relationship between precipitation and RCS. With increasing precipitation, the RCS, which represents the energy of the reflected wave, attenuated and the maximum measurable distance became shorter. Therefore, it was found that rainfall degrades radar recognition performance, and the effect increases with increasing precipitation.



No rainfall 30mm/h 50mm/h 80mm/h

Fig.11 Precipitations and radar cross section

5. CONCLUSIONS

The perception performance of the radar differs among static and dynamic conditions. The error depended on test environments and conditions, especially safety walls and rain were significantly influenced. Under the test condition, without any roadworthy objects such as safety fences, walls, or poles of traffic lights, the radar detected the preceding vehicle and derived the distance with low errors. On the other hand, when the preceding vehicle was located close to the safety wall made of steel, the radar could not detect the preceding vehicle and derived the distance with many errors. In addition, rainfall was found to degrade radar detection performance.

From the perspective of ensuring the safety of ADVs, it is important to determine the static, dynamic, and disturbance conditions to ensure technological neutrality in evaluating the suitability of ADVs for the required performance. In particular, in the recent dynamometer system [3] for validating perception performance, the conditions above under which the vehicle is in driving condition but the surroundings are static will be important.

REFERENCES

[1] UNECE/TRANS/505/Rev.3/Add.156 (2021)

[2] V. G. Borkar, A. Ghosh, R. K. Singh, and N. Chourasia, Radar cross-section measurement techniques, Defence science journal, Vol.60, No.2 (2010), pp.204-212.

[3] S. Tang, Z. Zhang, Y. Zhang, J. Zhou, Y. Guo, S. Liu, S. Guo, Y. F. Li, L. Ma, Y. Xue, Y. Liu, A survey on automated driving system testing: landscapes and trends, ACM forthcoming, Vol.1, No.1 (2022), arXiv:2206.05961v1.

EVALUATING AND RATING THE SAFETY BENEFITS OF ADVANCED VEHICLE TECHNOLOGIES: DEVELOPING A TRANSPARENT APPROACH AND CONSUMER MESSAGING TO MAXIMIZE BENEFIT

Bruce Mehler
Pnina Gershon
Bryan Reimer
Massachusetts Institute of Technology, Center for Transportation & Logistics, AgeLab
United States of America

Paper Number 23-0150

ABSTRACT

In 2012, a major traffic safety organization tasked the MIT AgeLab with developing a data-driven system for rating the effectiveness of new technologies intended to improve safety. Such a system was envisioned as having the potential to educate and guide consumers towards more confident and strategic purchasing decisions, ideally encouraging adoption of technologies with demonstrated safety benefit. In addition, an evaluation of the status and extent of existing data was seen as a way of identifying research gaps in the state of knowledge about safety systems. The focus was on technologies as a class, not on a rating review of individual vehicle model implementations. As conceptualized, the system aimed to complement traditional NCAP style ratings as well as to provide consumers with transparent information on early stage and often improving safety technologies. Development of the rating system and identification of data was undertaken in consultation with a range of academic, industrial, consumer, NGO, and governmental experts as well as with representatives of many of the major automotive manufacturers and suppliers.

A key observation that emerged was that data on objectively demonstrable real-world benefits were generally sparse and often lower than expectations based on theoretical considerations, simulation studies, or pre-production evaluations. A number of experts and industry representatives expressed some surprise at both the divergence between theoretical and observed benefits and the relative scarcity of data upon which to make objective assessments, while others were quite aware of these issues and the need for the development of objective data under real-world operating conditions. A number of factors that might be relevant to understanding why such differences between expected and observed benefits exist were identified. One outcome of this effort was the founding of the Advanced Vehicle Technology (AVT) consortium to collect and examine objective data under naturalistic driving conditions of how drivers interact with, engage or don't engage, various production safety and driver assistance systems. This ongoing effort is contributing to insights concerning actual benefits and reasons for benefit gaps.

Drawing from our initial work, as well as newer sources of data, we argue that the evaluation and rating of safety and driver assistance technologies for informing the consumer and the public at large should consider both theoretical potential and existing demonstrated benefit of specific technologies. This position is increasingly relevant as the effectiveness of many newer technologies have the potential to actually improve over the lifecycle of a vehicle through software updates. The emphasis on ratings based on observed benefit for actual drivers under real-world conditions is proposed to be complementary, rather than competing with, ratings focused largely on controlled test track evaluations of engineered capability. In addition, a case is made for providing ratings that assesses benefit relative to overall crash, injury, and fatality rates – and in relation to the specific scenario / crash event type that a given technology is intended to address. This approach should aid consumers in considering the extent to which a specific technology is or is not relevant to their particular driving needs.

INTRODUCTION

At first glance it might be assumed that the introduction of new automotive technologies designed to reduce the incidence of crashes, the severity of injuries, and in the number of deaths on our roadways should led to ready adoption and utilization. However, that is often not the case. Most new technologies come with a monetary cost and are generally introduced as upgrade options at the time of new vehicle purchase. What information concerning potential benefits is likely to encourage a consumer to decide to spend the money to acquire a given feature?

Moreover, if a technology is included in a package that a consumer has purchased for other reasons and is a feature that the consumer has to actively use to realize any benefit, what information is necessary to encourage the consumer to use the technology? In the latter case, think of the long-term research and messaging investment required to get the majority of the driving and riding population to regularly use seatbelts even after they became standard required equipment in most countries [1-3].

Governmental entities at the local, state, national, and international levels as well as numerous non-governmental entities have been concerned with these issues for some time. Approximately a decade ago, the AAA Foundation for Traffic Safety (AAA-FTS) issued a request for proposals to support research to identify and develop objective measures that could be used to construct a rating system designed to compare and contrast the effectiveness of relatively new in-vehicle technologies relevant to enhancing driver safety. Building on previous work supported by the U.S. Department of Transportation's funded New England University Transportation Center (NE-UTLC), the Santos Family Foundation, and other sponsors, the Massachusetts Institute of Technology (MIT) AgeLab believed that this undertaking had the potential to provide a useful tool to aid consumers in making informed decisions about the purchase and use of safety systems. We also saw this as an opportunity to stimulate discussion and possible action within the safety research and manufacturer communities relative to developing a more comprehensive approach to thinking about safety and methods for objectively evaluating new technologies. We proposed, and AAA-FTS approved, an approach to the project that, in addition to undertaking a review of the existing safety benefit literature, included engaging and actively consulting with a wide-range of academic, industrial, consumer, NGO, and governmental experts as well as with representatives of many of the major automotive manufacturers and suppliers. Details concerning the methods employed, feedback from individuals and entities consulted, the literature review, findings and recommendations can be found in our project report issued in 2014 [4]; a somewhat modified and abbreviated form of the report was subsequently issued by AAA-FTS [5].

It is important to note that the focus of the project was on technologies as a class, not a rating review of individual vehicle model implementations, and an emphasis on considering observed benefits under naturalistic conditions. Thus, the approach was seen as complementing rather than competing with ratings developed by the Insurance Institute for Highway Safety (IIHS) and the United States' National Highway Transportation Safety Administration (NHTSA) New Car Assessment Program (NCAP) which focus largely on controlled test evaluations of engineered capability of specific vehicle models.

The ratings approach proposed in the report [4,5] suggests that there is value in considering both the potential safety benefit of a technology (based on modeling, simulation, and/or test track evaluations] and observed benefit under actual real-world driving conditions of production systems. Two key findings from the literature review were that a) the latter values are often found to be lower than theoretical expectations and pre-production testing data, and b) that publicly available data on actual system performance in the hands of consumers under naturalistic driving conditions was extremely sparse or non-existent for many technologies. The case for why it is important for consumers to have ready access to information on both the theoretical potential benefit of a technology and information on what is known to date on performance benefits of actual production systems driven by consumers is developed further in the current paper.

An important outcome of the project was the dialog initiated between the MIT AgeLab research team and many of the individuals and groups engaged during the project. The identification of the major research gap between what was known about safety technologies from theoretical and experimental testing efforts and the need to develop richer objective data on how systems were actually being used (or not used) and observed objective benefit in consumers driving production vehicles under naturalistic conditions led to discussions concerning how the development of such research might be supported. This led to the launch of the Advanced Vehicle Technology (AVT) consortium at MIT in 2015 as an academic-industry partnership with the goal of developing a data-driven understanding of how, and to what extent, consumers engage with and leverage advanced driver assistance systems, automation features, and a range of in-vehicle and portable technologies for connectivity and infotainment appearing in modern vehicles. Now in its 8th year, the consortium brings together automakers, insurance companies, tier-1 suppliers, and research and consumer-oriented organizations in a unique collaborative effort. At the time of this writing, members included: Agero, Allstate, Aptiv, Arriver, Audi, Autoliv, BMW, Bosch, Cariad, Consumer Reports, Google, Honda, Insurance Institute for Highway Safety, Jaguar Land Rover, J.D. Power, Nissan, Progressive, Seeing Machines, Smart Eye, Subaru, The LAB (GIE Stellantis & Groupe Renault), Toyota, Travelers, Veoneer, Volvo Car, and Zenseact. (See https://agelab.mit.edu/avt for additional details and an up-to-date listing of sponsoring organizations.) Findings from AVT's efforts are both confirming and extending of the observations and concerns raised in the original ratings project work. Some of these are highlighted below.

EVALUATING AND COMMUNICATING SAFETY BENEFIT INFORMATION

Evaluating Safety Benefits: Methods & Stages

In many ways, the historic "gold standard" in safety benefit evaluation has been the collection of actuarial / epidemiological data on the instances of crashes, injuries, and/or deaths associated with otherwise comparable vehicles equipped with and without a given technology. In the U.S., data might be obtained from governmental sources (e.g., Fatality Analysis Reporting System, National Automotive Sampling System, National Motor Vehicle Crash Causation Survey, General Estimates System) or from insurance industry sources (e.g., Highway Loss Data Institute). As a prime example, electronic stability control (ESC) is a technology for which a significant body of such data and analysis has been developed over time. ESC was first introduced in the U.S. in 1995 and a 2004 analysis of the Fatality Analysis Reporting System (FARS) for all fatal crashes in the United States over 3 years (2001–2003) [6] found that it reduced single-vehicle fatal crash involvement risk by 56%. A follow-up analysis considering a ten-year period [7] reported somewhat smaller, but still substantial, effectiveness values for reductions in fatal cash involvement of 49% for single vehicle crashes and 20% for multiple vehicle crashes. Analyses drawing on the NASS General Estimates System and other sources that considered specific vehicle types found particularly marked reductions in crash events for sports utility vehicles (SUVs) as high as the 70% range [8,9]. Other work based on actuarial / epidemiological estimates further strengthened the safety case for ESC technology [10-13]. Such data supported efforts that led to legislation in the U.S. requiring ESC to be mandatory standard equipment for all passenger cars and most other vehicles by model year 2012. Follow-on actuarial work has continued to make a strong case for the safety benefits of vehicles equipped with ESC [14-16]

Simulation & pre-production testing - While the accumulation of objective actuarial data, as in the case for ESC, presents perhaps the ideal case for establishing safety benefit information that both the general public and regulators can act upon, there are important considerations to keep in mind in terms of evaluating technologies in their early phases of introduction. The most obvious is lack of such data when a technology is initially introduced as a production feature. When ESC was first released to the market in 1995, there were no actuarial data on ESC equipped vs. non-equipped vehicles to draw upon; it took about nine years for substantive analysis such as [6,10] to appear. Consequently, in the early stages of a technology's introduction, evaluations of its potential benefits may draw upon theoretical calculations, simulations, test track data, or, less frequently conducted, structured study of preproduction prototypes driven by research subjects (as opposed to experienced test drivers) on actual roadways. To the extent that such data is made available for outside evaluation, if appropriately framed, can provide useful forecasting information that can be shared with the public.

Post-production testing & naturalistic observation – It is important to recognize that many new technologies tend not to show the same level of benefit that might be anticipated from simulation or test track evaluations when studied as production systems in the field. Many factors can be at play when this occurs such as differences between the implementation of prototypes vs. production versions, a wider range of conditions (e.g., roadway characteristics, weather, traffic, etc.) in the real-world vs. idealized scenarios and experimental tests, sensor and actuator limitations, and driver behavior variables including misunderstanding, misuse, behavioral adaptation, turning a feature off or rarely or never turning a feature on – examples are presented in subsequent sections. Consequently, functional testing of production systems' performance when vehicles are driven by consumers and naturalistic observation of how consumers interact with them is a critical next stage in evaluating actual safety benefit. Besides providing a refined picture of potential safety benefit, learning from such observations can provide invaluable input to the industry to guide more targeted simulation and pre-production testing as well as overall system refinement and functional improvements that lead to increased safety benefits.

It is in this area of developing a data-driven understanding of how, and to what extent, consumers engage with and leverage advanced driver assistance systems, automation features, and a range of in-vehicle and portable technologies that the AVT consortium has been actively engaged in collecting naturalistic data in instrumented production vehicles. While the intent here is not to delve deeply into details of what has been done to date, some representative examples of work coming directly out of the consortium and through collaborations with consortium members may be useful to mention. One analysis coming out of the data collection effort considered FCW systems [17], a technology which has demonstrated safety benefits. However, there are questions around the extent to which the sensitivity characteristics and settings of an FCW system may trigger a high rate of alarms in situations where there appears to be no or a low probability of a collision, which may negatively impact driver responsiveness and satisfaction. Road type, speed, traffic density and deceleration profiles distinguished been categories of event severity. Modeling outcomes suggested patterns than may prove useful in adjusting algorithms to reduce nuisance alerts. Other analyses have looked at drivers' use of ACC as a function of roadway class and at how their frequency of use of the feature changed over their first four weeks of exposure [18,19]. A series of papers represent a portion of the work that has been carried out to date examining driver glance and hand control before, during, and after disengagement of the Tesla Autopilot ADAS feature [20-22]. Complementary analyses looking at driver hand and glance behavior associated with transfers of control in the Cadillac CT6's Super-Cruise implementation of an L2 ADAS feature have been released [23,24]. Speeding behavior during manual driving and with ACC engaged is being investigated [25-27]. Work characterizing the use of Tesla's auto lane change feature is underway [28].

Actuarial data: ideals & limitations – As already discussed, analysis of actuarial / epidemiological data that compares the relative rates of crashes, injuries, and/or deaths associated with vehicles equipped with a technology vs. those without represents the ideal safety benefit standard. While reaching this stage of evaluation takes time, when such data can be obtained, as has been the case with ESC, it can provide a solid basis for establishing objective guidance to the buying public and regulators. That being said, besides the issue of taking years to accumulate sufficient actuarial data, several qualifiers or limitations also need to be kept in mind.

Actuarial data may, in some instances, underestimate the potential benefit of a technology. ESC is a technology that is on by default in most, if not all, passenger vehicles. It has to be actively turned off and it is unadvisable to do so except in very specific cases such as when stuck in mud or snow. Even in these instances, many drivers may be unaware of the appropriateness of doing so under such conditions. Consequently, actuarial data for vehicles with and without the technology installed provide a fairly reasonable comparison of the degree of safety benefit inherent in the technology since it is functionally available for the vast majority of time a vehicle is driven. In contrast, other safety technologies, such as lane departure warning systems, can actively be turned-off and a significant number of drivers may do so if they find the warnings too frequent and/or annoying. An IIHS study [29] that looked at the settings of vehicles from nine manufactures that were taken to dealers for service found that 49% had their lane maintenance features turned off. For lane departure warning systems specifically, the percentage that were turned off ranged from 40% to 67% depending on the vehicle brand. Thus, simply because a vehicle is equipped with a technology does not mean that is actively functioning. As a result, any reduction in crashes, etc. in equipped vs. non-equipped vehicles may be less than what would be the case if the technology of interest was always active. An unused technology can confer no benefit. This is likely an even bigger issue when considering technologies that are off by default and must be activated by a driver. Some drivers may be unaware that a given technology is present in their vehicle, they may be unaware of circumstances where it is intended to be used or of how to turn it on, etc. Again, cases of limited or non-use of the technology will result in an underestimation of potential benefits in an actuarial comparison of vehicles with and without a technology installed. For some types of technologies, alternate approaches conceptually similar to efforts such as [30-33] but drawing upon field operational tests and related naturalistic data collection on post-production systems where a technology is known to be active vs. not active may provide useful alternatives or complementary perspectives.

A further factor to be considered is the potential for improvements in a technology. It is possible that changes in various characteristics of a technology ranging from improvements in sensor technology, algorithm adjustments, to enhancements in a user interface may result in initial actuarial data providing an underestimate of the benefits that may be present in newer releases. If simulation or experimental testing show evidence of a substantive improvement in a technology, naturalistic observation suggests enhanced understanding and/or use of a refined implementation, etc., there may be sufficient grounds for investing in an updated safety benefit assessment based on more recent actuarial data.

In summary, there are important roles for each stage and method of assessing safety benefit - for theoretical and experimental estimates of safety potential, for post-production testing of systems and naturalistic observation of actual consumers, and actuarial assessment of safety benefits throughout the lifecycle of a technology. How this might be translated into information and consumer messaging for the public is taken-up in sections that follow.

Rating a Technology Class vs. Specific Vehicle Implementations: Consumer Value in Both Approaches

In the development of the original AAA-FTS project, there were extensive discussions around whether the focus of the ratings should be on what would in essence be comparative safety benefit ratings of different technology types (e.g., land departure warnings vs. adaptive headlights) or on specific implementations of a technology (e.g., lane departure warnings on model A vs. model B. While we considered the question of to what extent and how specific implementations of a given technology type vary in effectiveness to be of significant interest, a number of considerations led to the decision to focus on a broad or top-level rating evaluation comparing different technology types.

From one perspective, the most appropriate assessment of actual safety benefit should be based on actuarial / epidemiological data examining the extent to which a given technology impacts crashes, injuries, and/or fatalities in the actual driving population. While the total number of adverse events per year is unacceptably high, the number of events per individual vehicle model, identifiable as being equipped with or without a given technology, is relatively low for purposes of calculating effect statistics. This is one consideration in seeing value in developing ratings focused broadly on a class of technology. Additionally, through NCAP, NHTSA provides the consumer with a resource listing for obtaining information on whether various vehicle models do or do not offer selected safety technologies that meet a minimum level of performance. EuroNCAP provides similar function in Europe. Also, IIHS had recently expanded their testing programs to begin to consider model specific performance of safety technologies, beginning with scenario specific test track evaluations of forward collision braking / mitigation systems. Early in the AAA-FTS project it was concluded that there is a place for developing technology level ratings that complement rather than attempt to duplicate these latter efforts.

A reasonable question that might be asked is “How is a technology type rating system going to help a consumer choose one car over another?”. In brief, the answer is that a top-level rating system is not intended as a car model buying guide, but rather is intended to serve as a technology feature buying guide. It is intended to assist the consumer in identifying safety technologies that they may wish to look for in their next vehicle or consider as options within specific vehicle models of interest to them. After a consumer identifies a safety technology of interest, resources such as NCAP listings and IIHS evaluations provide a means to obtain vehicle model specific information generally tied to defined testing scenarios. Another source of model / implementation level information that are consumer oriented are “expert” based evaluations of organizations such as Consumer Reports.

The original rating project report [4,5] included concise examples of content at the technology type level that it was hoped would help support the development of consumer oriented educational support materials. Each was introduced with a section called “What Is It?”, a short one to two paragraph description of what a technology is and the conditions under which it might be relevant. This was intended as a very brief, high level orientation to the technology. Subsequent sections included “Why Would I Use This Technology?”, “How Well Does It Work?”, “Who Benefits Most?”, “In What Situations Doesn’t It Work?”, “Not All Systems Are Alike”, and “Different Names, Same Idea”. Many of these elements that can help educate and support consumers at the technology type level can be found today in the “MyCarDoesWhat?” website (mycardoeswhat.org), initially developed with funding from the Toyota Safety Research and Education Program Settlement and currently operated by the National Safety Council.

The authors of the present paper continue to see important roles for providing the consumer (and the automotive industry) with both technology type and implementation level specific information on emerging as well as established technologies. Current efforts within the MIT AgeLab AVT consortium are invested in developing objective data on driver engagement and system performance from consumers collected under naturalistic conditions that can contribute to both types of technology assessment.

Projected Benefit vs. Observed Benefit

As already discussed, it takes many years to develop objective actuarial data on the extent to which benefits are (or are not) observed with a given technology in the real-world. In the interim, it is important to provide consumers with some guidance on new technologies for which epidemiological data are not yet available. We believe it is important to promote "projected safety benefit" while being "truthful" to the consumer regarding actual demonstrated value. Being truthful during the initial introduction of a technology may largely consist of making it clear that current projected benefits are based on simulations and/or controlled tests that are likely to represent best case scenarios and that actual benefits may be more modest during a technology's initial deployment or when operational conditions are limited, etc. As post-production studies and naturalistic observation of consumer experience and interaction with a technology provide additional data and insight, being truthful will likely involve providing the background information on ratings and newly acquired understanding of the conditions or other factors that impact the extent to which actual benefits are realized

Besides being truthful to the consumer, characterizing the extent to which there is a divergence between projected safety benefits and actual observed benefits can play a critically important role in identifying areas where, and ways in which, improvements in a technology may be possible. Are potential benefits being lost because drivers are turning-off a system and, if so, why? Are drivers unaware that their vehicle is equipped with a selectable technology or are they unsure of how it works, or what its benefits might be such that they never turn it on – or are confused by its functioning – and thus turn it off never to try it again? Are they unsure about what conditions under which the technology should provide benefits vs. conditions where it should not be used or relied upon? Are status icons or messages confusing? Are alerts or status displays annoying, distracting, or otherwise creating unintended negative consequences? Are drivers adapting to the presence of a technology such that they over-rely on or otherwise change their behavior in ways that reduce the net benefit? Such questions are important to ask as humans may frequently interact with systems and interface designs in ways that an engineering team did not anticipate. Another point worth noting is that some of the aforementioned reasons for possible gaps between potential and observed safety benefits is that in certain cases it may be possible to address them through training and educational efforts and consumers do not necessarily have to wait for changes in hardware to realize increased benefits from existing implementations.

Gaps between projected benefit and observed benefit are not, of course, always traceable to driver / system interactions. The presence of a difference between expected benefit and observed benefit may provide clues to the need of adjustment in algorithms, improvements in sensors or actuators, etc. It may sometimes be the case that careful analysis of observed benefits should led to a realistic reevaluation of the original projective benefit estimates. Sometimes a highly anticipated technology that was expected to provide one level of benefit actually proves to provide a more modest benefit. In such cases, ratings of expected benefit should be adjusted accordingly.

In the original project report [4,5], divergences between estimates of potential benefits and observed benefits of several technologies stood out. In the case of back-up cameras, observed scenario specific benefits were assigned a proposed rating one level less than the rating of anticipated benefits based on data available at that time. Averaged across three controlled experimental studies with research participants, Mazzae [34-36] reported that the use of back-up cameras reduced crashes during unexpected collision trials by approximately 30%. In proposed rules in the Federal Register, NHTSA [37] estimated that annual fatalities occurring from backing crashes could be reduced by 46% and a similar percentage was estimated for reductions in annual injuries if all vehicles were equipped with rearview camera technology. In contrast, actuarial assessments of observed performance during this time frame were mixed. Data from the Highway Loss Data Institute [38] considering an initial assessment of one vehicle brand with and without backup cameras showed no significant effect on any insurance coverage; however, this was considered a relatively weak analysis for injury effects involving pedestrians. An initial analysis considering another brand of vehicles [39] found that, contrary to expectations, there was an increase in collision frequency claims (3.1%), severity, and overall losses ($18), but a non-significant reduction in property damage / liability claims. Most relevant from a safety perspective, there was a reduction in the frequency of high severity bodily injury claims of 22.2%. To the extent this 22.2% reduction in high severity injury claims represented a then best-case assessment based on real-world observed safety benefit, a case could be made for consumers to consider looking for such technology in their next vehicle. It also made a case for researchers and the industry to dig into whether the estimates of expected benefits needed to be reassessed and/or whether enhancements in production systems or other factors could be identified to improve real-world benefits. Delving into what has emerged since in our understanding of this particular technology is beyond the scope of the current paper, but this brief summary of part of what was covered in the original project report should provide a useful example of some aspects of why we are interested in both types of benefit assessment.

Once again, the optimistic reason for studying any divergence between estimated potential and observed benefit is that such analyses may lead to insights that ultimately result in improvements in the technology. As discussed in more detail in a later section, this is particularly relevant to today's consumers as updateable software increasingly plays a role in some safety technologies, a consumer may not necessarily have to wait to purchase a future year's model to see improvements in a technology they purchase today. While this is not a certain proposition, it is increasingly something some consumers may wish to take into consideration in evaluating current ratings of potential vs. currently observed benefits – and – hence, another reason for providing the consumer with both types of information.

Overall Safety Benefit vs. Situation Specific Benefit

As highlighted in the 2014 report [4,5], a challenge that we spent significant time considering was how best to represent the safety benefit of one technology relative to another. From one perspective, a technology that offers the potential to save the largest absolute number of lives should logically receive a higher rating than a technology that may save a much smaller number of lives. However, what if the one technology is relevant to a large percentage of all possible crash events, but only actually works successfully in a modest percentage of those cases – while another technology is designed to function in a much more limited number of situations, but is highly successful in preventing loss of life under those conditions. It thus seems "unfair" in a sense to down-rate the second technology relative to the first. This may particularly be the case if the scenario that the second technology is designed to mitigate or eliminate is of particular interest or relevance to a subset of consumers.

These considerations led to the proposal to rate technologies both in terms of **Overall Safety Benefit** (considering the maximum number of lives, injuries, or crash events) and in terms of benefit within **Specific Scenarios** the technology is designed to address. We continue to feel that consumers are best informed to make personally relevant purchasing choices if safety benefit information is considered from both perspectives.

As already discussed, data collected on ESC made a clear case for a substantial overall safety benefit of purchasing a vehicle with the technology before it became a standard feature. Analyses such as [9] which reported a 56-67% reduction in single vehicle crash risk for individuals interested in purchasing an SUV made a particularly strong case of ESC, single vehicle crash risk reduction in the 33-35% range for standard passenger vehicles made a strong case for the technology regardless of the class of vehicle of interest. In contrast, a technology such as adaptive headlights offer relatively modest overall safety benefit when considered against all possible event types. Jamakian [40] estimated a potential reduction in all crash events of 2% and fatal crash events of 8%. However, when considering scenarios related to improving visibility when negotiating cures in darkness or twilight, adaptive headlights were estimated to have theoretical relevance to 90% of crashes that occur on curves at night - 91% for nonfatal injury crashes and 88% for fatal crashes. Thus, for individuals who drive primarily on highways and during daylight hours, their relative weighting of the personal value of such a technology might be quite different than that of an individual who frequently drives at night in rural or suburban locations with frequent curves, turns, and hills.

Another example of the argument for developing information for consumers that consider both overall and scenario specific benefits of a particular technology is that of back-up / rear-view cameras. In 2007, there were a reported 41,059 people killed in motor vehicle traffic crashes in the U.S. [41]. For the same period, there were approximately 71 deaths due to back-over events on public roadways and 221 non-traffic related back-up fatalities. This combined 292 deaths represents a relatively minor percentage of total vehicle related fatalities (0.7%). Consequently, a rating of the overall safety benefit of a technology such as back-up cameras that are intended, in part, to reduce the likelihood of such events would be fairly modest compared to the overall benefit rating of ESC. At the same time, there is a particularly high emotional cost associated with this type of event. NHTSA [42] estimated that 31% of all backup event fatalities involved children under 5 years of age and another 26% were adults 70 years and older; these events often involve family members or other close personal associations. Another NHTSA report [37] estimated that if all vehicles were equipped with rearview video technology that annual fatalities from back-over crashes could be reduced by 46%. A similar percentage reduction in annual injuries was also estimated along with substantive reductions in property damage and vehicle repair costs. With these scenario specific considerations in mind, such information might have been particularly relevant to parents with young children who park their vehicles in garages and/or have long driveways or older adults or others with restricted capacity to easily turnaround to look out the rear window when backing-up prior to the issuing of the requirement that all new cars were to be equipped with back-up cameras by 2018.

Again, see [5] for additional detail on the assessment and reporting on overall vs. scenario specific benefits in the cases of ESC, adaptive headlights, lane departure warning (LDW), forward collision warning (FCW), forward collision mitigation / automatic emergency braking (AEB), and adaptive cruise control (ACC).

Software Updates: Opportunities & Challenges

Historically, physically replacing hardware components in a customer's vehicle to improve a system's functionally has generally been limited, with some exceptions, to recall replacements to address defects. As advanced driver assistance systems (ADAS) and safety technologies have increasingly been made-up in part of software algorithms and software driven displays, the capability to potentially improve the functionality of a system through over-the-air (OTA) or dealer installed software updates is increasingly a part of the technology and user experience landscape. The positive side of this capability is the opportunity to provide a customer with an upgrade, often at no cost, that offers enhanced functional benefit in their existing vehicle. Over the course of AVT's FOT activity, we have observed software updates in our vehicles that have noticeably enhanced aspects of various ADAS systems. This capability can represent a significant win for both the customer and the automotive industry.

While software updates provide the opportunity for enhancing important features in a customer's vehicle, there are challenges that need to be appropriately addressed. At least one manufacturer is known for issuing very frequent OTAs, sometimes as often as multiples per month. Regardless of the frequency with which updates are provided, reasonable and appropriate testing procedures need to be followed to minimize the likelihood of errors or unintended consequences being introduced. Transparency should be provided to the user regarding important details of the changes associated with the update and the impact on safety benefits. What is the purpose of the update? Have the conditions under which a feature is active or inactive changed? Has the appearance or location of icons changed? Has a feature been temporarily or permanently disabled? Our experience in the AVT consortium with a fleet of vehicles across a number of manufacturers and over a number of years has been that, in many instances, clear information on the extent and implications of both OTA and dealer installed software updates as provided to us are generally very limited and, in some cases, lacking all together. Ultimately, both the consumer and the manufacturer will benefit from providing concise and clear information on what features have changed or been added with an update.

An additional, but hopefully positive challenge of software updates, are the implications for ratings of the safety benefit offered by a technology. As software updates ideally improve functionality, ongoing assessments will need to keep-up to date in making data driven information available to the public. As conceptualized, the original rating system allows room for characterizing changes in apparent benefits due to implementation advancements / refinements. The expectation would normally be that the long term "goal posts" for a technology (e.g., rated potential benefits) remain relatively fixed while observed benefit ratings are transparent concerning actual benefits that are documented over time.

Other Aspects of Consumer Messaging

Standardized naming of technologies – It does little good for governmental and non-governmental organizations to invest in developing quality information on various technologies if the consumer becomes confused about what to look for when they move to investigating specific vehicles online or at a dealership. Even a technology that is broadly available and fairly readily identifiable to the initiated, such as adaptive cruise control, may be called by a wide range of brand names that a consumer may or may not be able to connect with a feature that they would like to have in their next vehicle. This issue has been raised in both academic research [e.g., 43] and by consumer advocacy groups [e.g., 44;45] including an ongoing collaborative effort by AAA, Consumer Reports, J.D. Power, the National Safety Council, and SAE International called "Clearing the Confusion" [46,47]. We have had the experience within the AVT consortium of participants who have driven one of our FOT vehicles for a month and particularly appreciated an ADAS feature, later contact us for help, expressing frustration at trying to locate the same feature when looking to purchase a new vehicle. When looking for the feature in her preferred brand on-line, one former participant was confused as to whether or not it was available. Further, when contacting a local dealer, she was told the feature was not available when, in fact, the it was but was only identified under a brand specific name. It is unfortunate when consumers are actively looking for a technology that they have learned something useful about that may offer them a safety benefit that they are end-up purchasing a model without the option because of such confusion. While, as clearing the confusion suggests, manufactures do not necessarily need to entirely give-up the use of brand names for safety and ADAS technology implementations, we would argue that, if used, they should ideally be linked to a standardized technology naming convention.

Branding & capability confusion – Another issue with technology branding, and aspects of the advertising that may be associated with it, is the potential for some consumers to assume that a technology is more capable than it actually is. Perhaps the best-known examples are the terms “Autopilot” and “Full Self-Driving Capability” as used by Tesla to market its ACC (“Traffic-Aware Cruise Control”) and lane centering (“Autosteer”) features. While current Tesla webpages [e.g., 47] include statements such as “Autopilot, Enhanced Autopilot and Full Self-Driving Capability are intended for use with a fully attentive driver, who has their hands on the wheel and is prepared to take over at any moment”, many consumers appear to read more into the names than is spelled-out in the qualifying text quoted. Such branding may also contribute to many in the general public being under the impression that truly self-driving or autonomous vehicles are available for purchase today [e.g., 43,49,50] and concerns have been raised that such misconceptions around terminology could not only impact proper use of current systems but also impact future uptake of increasingly advanced technologies as they are introduced [51-53]. The issue of potentially misleading naming is attracting some regulatory and other legal attention. The State of California recently passed Senate Bill No 1398 that dictates that a manufacture or dealer is not to name or describe a partial automation feature in a way that would “lead a reasonable person to believe” it can function autonomously “or otherwise has functionality not actually included in the feature” [54].

Obtaining accurate information & training at dealerships – Beyond the issue of standardization of naming for identifying a core technology of interest, consumers may go to a dealership to be told that a safety technology or ADAS system of interest to them is not available (when in fact it is), given inaccurate or no information about the operation of feature, or even actively discouraged from considering the option by a salesperson [e.g., 55,56]. Such interactions can lead to consumers not buying a safety technology that they otherwise would have, or being confused or otherwise misunderstanding the proper use of the technology such that they underuse or misuse the technology that undercuts its safety potential. This need not be a universal problem. The Abraham et al. [55] study found that sales staff from dealerships for some vehicle brands provided much more consistent and accurate information on such features than staff for other brands. Investment by manufactures in providing messaging consistent supporting materials (displays, handouts, user manuals, informational & training videos) and training programs for sales and support staff is part of the equation. Equally important is an investment at the dealership level to make use of such resources and to incentivize sales people and other staff to become knowledgeable and pass along clear and useful information to consumers. Such issues will only become more complex as software enabled features and services expand. As such, efforts to consistently align dealerships with manufacturer recommendations need to be accelerated.

Other sources of information & training – As much as dealerships can play a significant role in informing or misinforming consumers, it is not realistic to rely solely on improvements in dealer-based training to fully meet the needs of consumers in learning about the proliferation of safety and ADAS technologies since a consumer purchased their last vehicle. Realistically, there is generally only a limited amount of time and consumer energy and attention available to learn about all the features in a new vehicle at the time of delivery, particularly after what may be a stressful purchasing negotiation. Moreover, consumers vary in their preferred method of leaning [57] and proving alternate and multiple methods are likely to be to everyone’s benefit. Abraham et al. [57] found that drivers who learned through their preferred methods of learning reported higher understanding and use of in-vehicle systems. Beyond direct dealer training during the sales process and/or at vehicle delivery, other potentially useful sources of information include on-line or in-car videos (for viewing when the car is not in motion), in-car training dialogs, the user’s manual, and other manufacturer supplied materials that can be accessed at a consumer’s preferred time, place, and pace.

The concept that one's car can potentially play a useful role in informing and educating a driver in how to use a technology is worthy of further attention. In [57], 25% of participants indicated an interest in the concept of "the car teaches me" while just under 5% indicated that they learned in that manner, likely due to the lack of such a feature. A perhaps less ideal finding was that over 50% of participants reported learning by trial and error.

A challenging but important service that manufacturers' need to invest in is keeping information on a technology consistent across information sources and up-to-date. It may be advantageous for printed materials, particularly printed user's guides supplied at the time of purchase, to include notes encouraging users to check on-line or dynamically updateable in-vehicle sources for the most up-to-date information, since the information in some printed materials may tend to lag behind an actual installed implementation. Being upfront about this reality is another important form of transparency. As already discussed, the increased availability of OTA and dealer installed software updates increases that challenge to keep user information up-to-date while also providing the valuable opportunity of on-going enhancement of systems across a vehicle's lifespan.

Leveraging consumer input – While the focus of this paper is on the use of objective data to evaluate and rate the safety benefits of advanced vehicle technologies that can be used to inform consumers, we also recognize the important role that subjective input from consumers can play in identifying possible explanations for gaps between projected and observed benefits. The AVT consortium uses questionnaires before and immediately after FOT participants' training drives in our instrumented production vehicles to gather data on their understanding of the function and limitations of various safety and ADAS technologies prior to and after what is designed to be reasonably equivalent to a comprehensive dealer introduction to key features of the vehicle. Following a month-long experience driving a vehicle, participants complete a post-experiential questionnaire and take part in a follow-up interview. Data sources such as these can help identify the extent to which there are pre and-post experience understanding / misunderstanding about safety relevant advanced technologies in various production vehicles, along with gathering subjective input on whether a participant feels a technology and a specific implementation enhances safety and whether or not they would be interested in having when purchasing their next vehicle. This input can also be considered in the context of the objective data collected on the extent to which these consumers used selectable features, experienced alerts, etc. Lack of use of a selectable feature or identifying confusion around where to use or how engage it can contribute to understanding the extent to which there are gaps between expected and actual safety benefits. Ideally, such data can also support the development insights that help address such gaps through refinements in the underlying technology, algorithm adjustments, user interface design enhancements, and/or updating marketing and training materials.

The self-report and questionnaire data collected as part of an FOT effort can be very rich, but is inherently limited in terms of sample size. Consequently, we see significant complementary value in survey-based means of gathering consumer input or, as J.D. Power puts it, "listening to the voice of the consumer". The MIT AgeLab and AVT have collected large scale survey data on consumer understanding and interest in driver support and highly automated technologies for a number of years [55,58,50] and, as previously mentioned, have recently joined in a collaborative effort in this area with J.D. Power and PAVE [52-53]. AAA and Consumer Reports are well known for collecting survey data on a regular basis. A recent report by Consumer Reports considers ADAS usability through the lens of staff testing in combination with a large-scale owner survey [59]. Integrating such "messaging" from consumers with the insights gained from hard data collection provides an opportunity for manufacturers, dealers, and governmental and non-governmental consumer facing organizations, as well as researchers, to enhance to messaging that we share with consumers in support of the common goal of improving safety.

CONCLUSIONS

Numerous technologies have been developed specifically to increase safety in modern vehicles or as ADAS marketed primarily as convenience features, but with the expectation that they may provide safety benefits as well. Such technologies are generally introduced as optional features that consumers need to evaluate as to whether they should spend additional money to acquire in their next vehicle. Consequently, there is a fundamental need for consumers to have access to realistic and understandable information on the relative safety benefits any given technology offers, so that they can make personally informed decisions. The need for such information has grown in recent years as number of different features available and the functional details that a user needs to understand to maximally benefit from some ADAS features has increased.

The growing number of safety relevant technologies and the evolving nature of their implementations presents a challenge for rating information and systems to stay current. On the positive side, a number of key organizations are openly invested in evolving and updating their approaches to providing information on and the rating of these technologies. NHTSA has actively solicited input on upgrades to NCAP [60], EuroNCAP updates target standards on a regular basis and has shared a vision for encouraging enhancements in safety [61], IIHS has recently announced the development of a new rating program that evaluates safeguards that help drivers stay focused on the road [62] and engaged in briefing stakeholders ranging from manufactures to academic researchers to other consumer-oriented organizations on plans to make other enhancements to their ratings. Consumer Reports also has a history of sharing information on ongoing considerations for future ratings [e.g., 44].

From a safety advocacy perspective, there is a case to be made that consumers are likely to maximally benefit if they have access to quality information to support two steps in the decision-making process. The first involves identifying technologies that offer a level of safety benefit that is meaningful to their driving situation and personal values. Next, with this information in mind, a consumer can then look into specific vehicle models to see if they offer the technologies of interest to them as well as consider any available information on the relative effectiveness of the implementations in those models.

In the body of this paper we argue that the evaluation and rating of safety and driver assistance technologies for informing the consumer and the public at large should consider both the estimated potential and the existing demonstrated benefit of such technologies. Data on objectively demonstratable real-world benefits are generally spare and often lower than expectations based on theoretical considerations, simulation studies, or pre-productions studies. Nonetheless, estimates of the safety potential of a technology provide important initial guidance during the introduction of a new technology when objective data from naturalistic studies is limited or unavailable, and there has not yet been sufficient time for actuarial data to have been accumulated. Moreover, as data from studies of production systems used in the field are developed and actuarial data is obtained, the identification of differences between expected potential benefit and objectively observed benefit are important in several ways. One fundamental is transparency; the consumer and the public at large deserve an honest presentation of what our current, best understanding of what a given technology is providing in terms of benefit. Without such transparency, trust in information sources may be impacted and effective technology uptake may decline. At the same time, the identification of gaps between expected and observed benefit can be critical in the process of improving the safety benefit of a technology. Among other considerations, identification of gaps can motivate a deeper dive into investigating why the gaps exist and how they might be addressed.

As we have discussed and provided examples from various sources including our AVT consortium work, post-production testing of systems and naturalistic observation of actual consumers' interaction with a technology can bring to light a range of areas where actions can be taken to improve net effectiveness. These may consist of improvements in or adjustments to sensor systems, actuators, or algorithms, as well as identifying driver related issues that may be involved. While consumers typically need to wait for the a new vehicle model to benefit from physical improvements in sensors or actuators, the identification of gaps due to driver related factors such as misunderstanding of how to optimally interact with a technology, confusion around conditions and situations where a technology is or is not useful, lack of knowledge of the availability of a system in their vehicle or even how to engage a feature, may be addressed through improvements in informational sources and other forms of training / education. As we have also highlighted, the increasing capability to address gaps through software updates that refine algorithms or even make improvements in software-based user displays have the potential to actually improve a technology over the lifecycle of an existing vehicle. These opportunities will be lost, however, if investments in studying, and comparing across implementations, actual performance and consumer experience and behavior in production systems are not made.

We have also argued that consumers will be able to make more informed decisions about what technologies to seek out and utilize if useful information is available on both the overall safety benefit of a technology and situation or scenario specific benefits. Such information may lead many consumers to consider a given technology because of a relatively broad and overall high safety benefit. At the same time, there are likely cases where some consumers may decide to also invest in a safety technology that they may not have considered otherwise because the scenario specific benefit is particularly relevant to their personal situation and driving needs. Since many consumers have limited resources, making available information that can aid individuals in making personally strategic buying decisions can increase the equity of the process.

Overall, investing in understanding where and why a given safety technology is or is not presently meeting its presumed safety potential is to the benefit of the consumer, the industry, and society as a whole. Investigating apparent gaps can lead to improvements in the underlying technology / implementation and in consumer understanding and appropriate utilization. Well informed consumers are more likely to consider investing in technologies that may increase their personal safety and the safety of those around them, both inside and outside their vehicle. Transparency regarding what given technologies may or may not provide is important in building trust and in proper use. Customers who are well positioned to actually realize the safety benefit of a given technology are more likely to invest in that acquiring and/or using the technology in future and in encouraging their family and friends to do likewise.

ACKNOWLEDGMENTS

The original MIT AgeLab work on evaluating and rating safety benefits was funded by the AAA Foundation for Traffic Safety as well as through additional support from the U.S. Department of Transportation's New England University Transportation Center (NEUTC). Bridge funding was provided by the NEUTC and the Santos Family Foundation that allowed further work on this topic and exploration of ways in which a collaborative effort might be developed to address the identified research gap due to the relative sparseness of data on post-production safety technology and ADAS use by consumers. These efforts contributed to the establishment of the AVT Consortium at MIT. The development of this paper was undertaken while the authors were funded in part by the AVT Consortium, while the manuscript was primarily supported through the Santos Family Foundation; however, the perspectives and opinions presented are those of the authors who are solely responsible for the content.

REFERENCES

[1] Carter, P. R., & Maker, V. K. 2010. "Changing paradigms of seat belt and air bag injuries: what we have learned in the past 3 decades". Journal of the American College of Surgeons, 210(2), 240-252.

[2] O'Neill, B. 2009. "Preventing passenger vehicle occupant injuries by vehicle design—a historical perspective from IIHS". Traffic Injury Prevention, 10(2), 113-126.

[3] Vasudevan, V., Nambisan, S. S., Singh, A. K., & Pearl, T. 2009. "Effectiveness of media and enforcement campaigns in increasing seat belt usage rates in a state with a secondary seat belt law". Traffic Injury Prevention, 10(4), 330-339.

[4] Mehler, B., Reimer, B., Lavallière, M., Dobres, J. and Coughlin, J.F. 2014a. "Proposed System for the Objective Evaluation of the Safety Benefits of Technologies and Initial Rating Values". Technical Report No. 2013-27c, Massachusetts Institute of Technology AgeLab, Cambridge, MA. DOI:10.13140/2.1.2323.6164.

[5] Mehler, B., Reimer, B., Lavallière, M., Dobres, J. and Coughlin, J.F. 2014b. "Evaluating Technologies Relevant to the Enhancement of Driver Safety". AAA Foundation for Traffic Safety, Washington, DC.

[6] Farmer, C. 2004. "Effect of electronic stability control on automobile crash risk". Traffic Injury Prevention, 5(4), 317-325.

[7] Farmer, C. M. 2010. "Effects of Electronic Stability Control on Fatal Crash Risk". Insurance Institute for Highway Safety. Arlington, VA.

[8] Green, P. E., & Woodrooffe, J. 2006. "The estimated reduction in the odds of loss-of-control type crashes for sport utility vehicles equipped with electronic stability control". Journal of Safety Research, 37(5), 493-499.

[9] Ferguson, S. A. 2007. "The effectiveness of electronic stability control in reducing real-world crashes: a literature review". Traffic Injury Prevention, 8(4), 329-338.

[10] Dang, J. N. 2004. "Preliminary results analyzing the effectiveness of electronic stability control (ESC) systems Evaluation Note". National Highway Traffic Safety Administration Washington, D.C. DOT HS 809 790.

[11] Erke, A. 2008. “Effects of electronic stability control (ESC) on accidents: a review of empirical evidence”. Accident Analysis and Prevention, 40(1), 167-173.

[12] Chouinard, A., & Lecuyer, J. F. 2011. “A study of the effectiveness of Electronic Stability Control in Canada”. Accident Analysis and Prevention, 43(1), 451-460.

[13] Sivinski, R. 2011. “Crash Prevention Effectiveness of Light-Vehicle Electronic Stability Control: An Update of the 2007 NHTSA Evaluation”. NHTSA. Washington, DC. 29.

[14] Webb, C. 2017. “Estimating lives saved by electronic stability control, 2011-2015”. DOT HS 812 391. Washington, DC: National Highway Traffic Safety Administration.crashstats.nhtsa.dot.gov/Api/Public/Publication/812391.

[15] Koisaari, T., Kari, T., Vahlberg, T., Sihvola, N., & Tervo, T. 2019. “Crash risk of ESC-fitted passenger cars”. Traffic Injury Prevention, 20(3), 325-331.

[16] Fitzharris, M. 2020. “Electronic stability control and side impact crashes: 100% cure or a case of realigning safety priorities”. Safety, 2011, 23-25.

[17] Seaman, S., Gershon, P., Angell, L., Mehler, B., & Reimer, B. 2022. “Evaluating the associations between forward collision warning severity and context”. Safety, 8, 5. https://doi.org/10.3390/safety8010005

[18] Reagan, I.J., Hu, W., Cicchino, J., Seppelt, B., Fridman, L. & Glazer, M. 2019. “Measuring Adult Drivers’ Use of Level 1 and 2 Driving Automation by Roadway Function Class”. Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

[19] Reagan, I.J., Cicchini, J.B., Teoh, E.R., Reimer, B., Mehler, B., & Gershon, P. 2022. “Behavior change over time when driving with adaptive cruise control (LECT299s1)”. Proceedings of the 66th Annual Meeting of the Human Factors & Ergonomics Society, October 10-14, 2022, Atlanta, Georgia, USA.

[20] Morando, A., Gershon, P., Mehler, B. & Reimer, B. 2020. “Driver-initiated Tesla Autopilot Disengagements in Naturalistic Driving”. Proceedings of the 12th International Conference on Automotive User Interfaces and Interactive Vehicle Applications (AutomotiveUI'20), Virtual Event. pp. 57-65.

[21] Morando, A., Gershon, P., Mehler, B., & Reimer, B. [2021]. “A model for naturalistic glance behavior around Tesla Autopilot disengagements”. Accident Analysis and Prevention. 161, 106348.

[22] Morando, A., Gershon, P., Mehler, B., & Reimer, B. 2021. “Visual attention and steering wheel control: From engagement to disengagement of Tesla Autopilot”. Proceedings of the 65th Annual Meeting of the Human Factors and Ergonomics Society.

[23] Gershon, P., Seaman, S., Mehler, B., Reimer, B. & Coughlin, J. 2021. “Driver behavior and the use of automation in real-world driving”. Accident Analysis and Prevention. 158, 106319.

[24] Gershon, P., Mehler, B., & Reimer, B. 2023. “Drivers’ response to automation initiated disengagement in real-world hands-free driving”. Proceedings of the 27th International Technical Conference on the Enhanced Safety of Vehicles (ESV), Yokohama, Japan. Paper 23-0137.

[25] Monfort, S.S., Reagan, I.J., Cicchino, J.B., Hu, W., Gershon, P., Mehler, B. & Reimer, B. 2022. “Speeding behavior while using adaptive cruise control and lane centering in free flow traffic”. Traffic Injury Prevention, 23(2), 85-90.

[26] Haus, S.H., Gershon, P., Mehler, B., & Reimer, B. 2022. “Characterizing driver speeding behavior when using partial-automation in real-world driving”. Traffic Injury Prevention. https://doi.org/10.1080/15389588.2022.2089664

[27] Haus, S.H., Gershon, P., Mehler, B., & Reimer, B. 2022. “Speeding behavior when using automation: a descriptive analysis of naturalistic driving data (LECT346s1)”. Proceedings of the 66th Annual Meeting of the Human Factors & Ergonomics Society, October 10-14, 2022, Atlanta, Georgia, USA.

[28] Noonan, T.Z., Gershon, P., Mehler, B., & Reimer, B. 2022. “Characterizing the use of Tesla’s auto lane change feature in driver-initiated maneuvers. (LECT394s1)”. Proceedings of the 66th Annual Meeting of the Human Factors & Ergonomics Society, October 10-14, 2022, Atlanta, Georgia, USA.

[29] Reagan, I. J., Cicchino, J. B., Kerfoot, L. B., & Weast, R. A. 2018. “Crash avoidance and driver assistance technologies — Are they used?”. Transportation Research Part F, 52, 176-190.

[30] Najm, W. G., daSilva, M. P., & Wiacek, C. J. 2000. “Estimation of crash injury severity reduction for intelligent vehicle safety systems”. SAE Transactions, 1859-1865.

[31] Najm, W., Stearns, M., Howarth, H., Koopmann, J., and Hitz, J. 2006. “Evaluation of an Automotive Rear-End Collision Avoidance System”, DOT VNTSC-NHTSA-06-01, DOT HS 810 569, Performed by John A.Volpe National Transportation System Center, Cambridge, MA, Sponsored by National Highway Traffic Safety Administration, Washington D.C, March 2006.

[32] Carter, A. A., Burgett, A., Srinivasan, G., & Ranganathan, R. 2009. Safety impact methodology (SIM): evaluation of pre-production systems. In Proceedings of the 21st International Technical Conference on the Enhanced Safety of Vehicles, Paper (No. 09-0259).

[33] Srinivasan, G., Ranganathan, R., & Burgett, A. 2011. “Safety impact methodology (SIM): application and results of the Advanced Crash Avoidance Technologies (ACAT) Program”. In Proceedings of the 22nd International Technical Conference on the Enhanced Safety of Vehicles (ESV).

[34] Mazzae, E. N. 2008. “On-Road Study of Drivers' Use of Rearview Video Systems (ORSDURVS)”. National Highway Traffic Safety Administration. Washington, D.C. No. DOT HS 811 024, 140.

[35] Mazzae, E. N. 2010. “Drivers' Use of Rearview Video and Sensor-Based Backing Aid Systems in a Non-Laboratory Setting”. National Highway Traffic Safety Administration. Washington, D.C. No. NHTSA-2010-0162, 16.

[36] Mazzae, E. N. 2013. “Rearview Video System Use by Drivers of a Sedan in an Unexpected Obstacle Event. National Highway Traffic Safety Administration. Washington, D.C. No. NHTSA-2010-0162, 39.

[37] NHTSA. 2010. “Federal Motor Vehicle Safety Standard, Rearview Mirrors; Federal Motor Vehicle Safety Standard, Low-Speed Vehicles Phase-In Reporting Requirements; Proposed Rules”. U.S. Department of Transportation. Federal Register, 75(234), 76184-76250.

[38] HLDI. 2012. “Mercedes-Benz collision avoidance features: initial results”. Highway Loss Data Institute. 29(7).

[39] HLDI. 2011. “Mazda collision avoidance features: initial results”. Highway Loss Data Institute. 28(13).

[40] Jermakian, J. S. 2011. “Crash avoidance potential of four passenger vehicle technologies”. Accident Analysis and Prevention, 43(3), 732-740.

[41] NHTSA. 2008. “Fatalities and injuries in motor vehicle backing crashes”. U.S. Department of Transportation. Washington, DC. DOT HS 811 144.

[42] NHTSA. 2014. “Federal Motor Vehicle Safety Standards; Rear Visibility; Final Rule”. U.S. Department of Transportation. Federal Register, 79(66), 19178-19250.

[43] Abraham, H., Seppelt, B., Reimer, B. & Mehler, B. 2017. “What’s in a Name: Vehicle Technology Branding & Consumer Expectations for Automation”. Proceedings of the ACM International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI‘17). Oldenburg, Germany.

[44] Consumer Reports. 2017. “Auto Safety Report: Consumer Clarity and Safety for Today’s Advanced Driver Systems”. https://advocacy.consumerreports.org/research/clearing-the-confusion-recommended-common-naming-for-advanced-driver-assistance-technologies/

[45] AAA 2019. “Advanced Driver Assistance Technology Names: AAA’s recommendation for common naming of advanced safety systems”. https://newsroom.aaa.com/wp-content/uploads/2019/06/ADAS-Technology-Names-Research-Report.pdf

[46] Consumer Reports. 2019. “Clearing the Confusion: Recommended Common Naming for Advanced Driver Assistance Technologies”. https://article.images.consumerreports.org/prod/content/dam/cro/Consumer%20Clarity%20and%20Safety%20for%20Todays%20Advanced%20Driving%20Systems

[47] Consumer Reports. 2022. “Clearing the Confusion: Updated Common Naming for Advanced Driver Assistance Technologies”. https://advocacy.consumerreports.org/research/clearing-the-confusion-updated-common-naming-for-advanced-driver-assistance-technologies/

[48] Tesla. 2022. “Autopilot and Full Self-Driving Capability”. Accessed 12/26/2022: https://www.tesla.com/en_eu/support/autopilot

[49] Seppelt, B., Reimer, B., Fisher, J. & Friedman, D. 2019. “Consumer Confusion with Levels of Vehicle Automation”. Proceedings of the 10th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design. University of Iowa.

[50] Lee, C., Gershon, P., Reimer, B., Mehler, B., & Coughlin, J.F. 2021. “Consumer knowledge and acceptance of driving automation: changes over time and across age groups”. Proceedings of the 65th Annual Meeting of the Human Factors and Ergonomics Society, 65(1), 1395-1399. SAGE Publications.

[51] Kassens-Noor, E., Wilson, M., Cai, M., Durst, N., & Decaminada, T. 2021. "Autonomous vs. self-driving vehicles: the power of language to shape public perceptions." Journal of Urban Technology, 28(3-4), 5-24.

[52] Boor, L., Kolodge, K., Gershon, P., Lee, C., Mehler, B., Reimer, B., Adringa, T., Colter, T., & Niedermeyer, E. 2021. “2021 Mobility Confidence Index (MCI) Study”. Joint white paper: J.D. Power, MIT AgeLab, Partners for Advanced Vehicle Education (PAVE): https://discover.jdpa.com/hubfs/Files/Industry%20Campaigns/Automotive/2021-JDP-MCI-White-Paper_vFinal-Pages-120721.pdf

[53] Boor, L., Rizk, K., Andringa, T., Howell, A., Gershon, P., Lee, C., Mehler, B., Reimer, B. 2022. “J.D. Power 2022 Mobility Confidence Index (MCI) Study”. Joint white paper: J.D. Power, MIT AgeLab, Partners for Advanced Vehicle Education (PAVE): https://ctl.mit.edu/sites/ctl.mit.edu/files/JD-MIT-AVT-2022-US-MCI-Whitepaper-100422.pdf

[54] State of California. 2022. “Senate Bill No. 1398. Chapter 308. An act to add Section 24011.5 to the Vehicle Code, relating to vehicles.” https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202120220SB1398

[55] Abraham, H., McAnulty, H., Mehler, B. & Reimer, B. 2017. “A Case Study of Today’s Automotive Dealerships: The Introduction and Delivery of Advanced Driver Assistance Systems”. Transportation Research Record. 2660. DOI: 10.3141/2660-02.

[56] Boelhouwer, A., Van den Beukel, A. P., Van der Voort, M. C., Hottentot, C., De Wit, R. Q., & Martens, M. H. 2020. “How are car buyers and car sellers currently informed about ADAS? An investigation among drivers and car sellers in the Netherlands”. Transportation Research Interdisciplinary Perspectives, 4, 100103.

[57] Abraham, H., Mehler, B. & Reimer, B. 2018. “Learning to Use In-Vehicle Technologies: Consumer Preferences and Effects on Understanding”. Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society, 62(1), 1589-1593. SAGE Publications.

[58] Abraham, H., Mehler, B. & Reimer, B. 2018. “Learning to Use In-Vehicle Technologies: Consumer Preferences and Effects on Understanding”. Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society, 62(1), 1589-1593. SAGE Publications.

[59] Consumer Reports. 2022. “Consumer Reports’ Guide to ADAS Usability: Consumer insights on understanding, use, and satisfaction of ADAS”. https://data.consumerreports.org/reports/crs-guide-to-adas-usability/

[60] NHTSA. 2022. “Request for comments (RFC); New Car Assessment Program (NCAP), NHTSA Docket No. 2021-0002, 87 FR 13452 (March 9, 2022)”. U.S. Department of Transportation. Federal Register, 87(46), 13452-13521.

[61] EuroNCAP, 2022. “EuroNCAP Vision 2030: A Safer Future for Mobility”. https://cdn.euroncap.com/media/74468/euro-ncap-roadmap-vision-2030.pdf

[62] IIHS. 2022. “IIHS creates safeguard ratings for partial automation”. https://www.iihs.org/news/detail/iihs-creates-safeguard-ratings-for-partial-automation

A METHOD FOR THE CHARACTERIZATION OF PERCEPTION SENSORS

Loren Stowe, Greg Beale, Matt Palmer
Virginia Tech Transportation Institute (VTTI)
United States

Steven Huggins
Global Center for Automotive Performance Simulation (GCAPS)
United States

Sebastian Silvani
National Highway Traffic Safety Administration (NHTSA)
United States

Paper Number 23-0160

ABSTRACT

Vehicle perception systems for both advanced driver assistance systems (ADAS) and automated driving systems (ADS) rely on a plurality of sensors and sensor modalities to “see” the surrounding environment. Each sensor type has its own inherent strengths and weaknesses (e.g., cameras perceive color but need ambient light, radar is insensitive to light but does not perceive color). The goal of this study was to develop systematic and adaptable tests and analysis methods that would allow the performance characterization for a variety of sensors and sensor types. Three common sensor modalities (radar, LiDAR, and camera) were selected for demonstrating the application of the methodology.

Three test maneuver templates were developed to exercise the relative motion of target objects within the sensor’s field of view (FOV). These allowed a broad set of conditions to be configured that corresponded to ones that might occur during real-world driving. These were combined with external conditions (e.g., simulated rain, variable ambient light, sensor degradations) to identify compounding factors that may affect sensor performance. Sensor characteristics and test factor sensitivities were then calculated across different metrics, including distance accuracy and maximum detection range, to demonstrate the process and efficacy of the method in characterizing perception sensor performance.

INTRODUCTION

Safe operation of a vehicle is dependent on an accurate awareness of the driving environment. ADS, as well as lower levels of automation found in ADAS, rely on varied sensor modalities and technologies to gain this awareness. Currently, a review of the Voluntary Safety Self-Assessment (VSSA)[1] reports published by ADS developers indicate that radar, camera and LiDAR sensors are common sensors being used in ADS perception systems. Though each original equipment manufacturer (OEM) may develop unique algorithms for object perception and classification, perception starts with and acts on the output from sensors.

While the sensor manufacturer provides performance specifications for their products, these are typically based on standardized test specifications that are performed in a highly controlled environment. These results are not necessarily indicative of the performance realized in the final operating environment. Other studies [1] have demonstrated that the lab-based methods for defining spec sheet performance can be inadequate at capturing the applied performance of the sensor (Figure 1).

---

[1] These can be accessed at https://www.nhtsa.gov/automated-driving-systems/voluntary-safety-self-assessment.

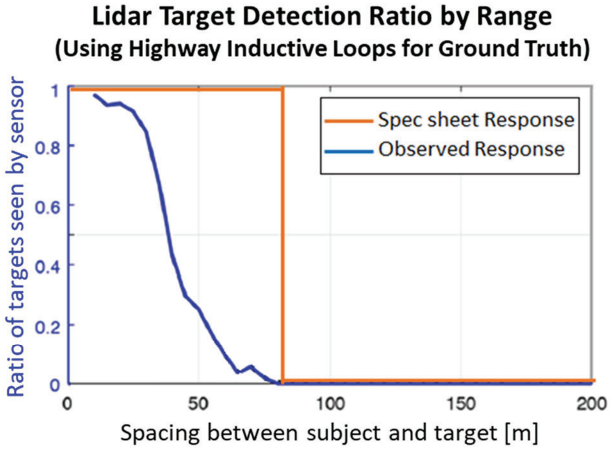


Figure 1. Actual LiDAR target detection ratio versus expected from specification sheet [1].

RESEARCH QUESTION / OBJECTIVES

The objective of this study was to develop test and analysis methods that can be used to characterize the performance of a variety of perception sensors in a variety of operational conditions. This was accomplished by

- Surveying test procedures and measures used by industry
- Identifying potential gaps in current practices for characterizing performance in operational conditions
- Developing a method for efficiently testing and characterizing a variety of sensors under a range of conditions

METHOD

The study sought to develop flexible and efficient characterization methods that could apply to a range of sensors with a variety of inherent sensitivities (e.g., camera to light, LiDAR to precipitation) and data outputs (e.g., object information, point cloud) that were designed for different functions (e.g., forward safety applications, parking assist). To facilitate this, the study explored functional testing to ensure a more direct link between test results and sensor performance in the applicable operational design domain (ODD). The development of the methodology consisted of three primary objectives.

1. Develop repeatable and adaptable tests based on the expected operational conditions such as relative motion with objects of interest and environmental factors.
2. Characterize sensor performance with a common set of metrics defined as a function of the test inputs to facilitate comparisons of sensors and sensor modalities.
3. Capture key sensor performance traits including the following:
    a. Nonlinear responses
    b. Temporal effects
    c. Interactions
    d. Degradation effects

Test Templates

To support these objectives, three test templates were defined to allow a wide range of configurations that could be used to capture primarily longitudinal, lateral, and field of view assessment. For the dynamic tests, ground truth measurements were obtained using differential global position satellite (DGPS) receivers on the subject vehicle (SV) and the target vehicle (TV). This functionality is indicated as the vehicle reference system (VRS) in the images below. For the static objects (e.g., signs, lane markings), the positions were previously mapped with DGPS.

The range sweep test (Figure 2) varied the distance between the SV and TV by adjusting the relative speed between the two vehicles. The SV and TV started at the minimum following distance at the baseline test speed. The SV decreased its speed by a defined amount below the nominal speed to increase the headway for the first half of the trial. When the sensors' maximum detection range was reached, the SV increased its speed above the baseline speed to reduce the headway distance, down to a predefined minimum safe following distance. The baseline speed, relative speed, lane offset, and road configuration can be adjusted to evaluate the effect of these variables on sensor performance.

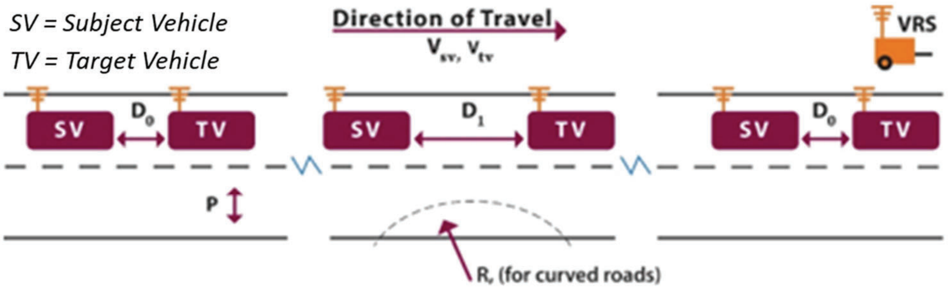


Figure 2. Schematic of Range Sweep Test Template

The lane slalom trials were designed to sweep the target laterally in a sensor's horizontal FOV. In a lane slalom trial, the SV repeated s-turns while approaching the stationary TV, which was parked one or more lanes over from the SV's path (Figure 3*)*. The frequency (cycles per second) and amplitude of the s-turns was varied to create increase the rate the TV moved in the FOV and the distance it moved within the FOV. Expressed as a percentage of lane width, the amplitude of these s-turns ranged from 50 percent of lane width to 200 percent or two-lane widths. The TV offset and road geometry can also be varied.

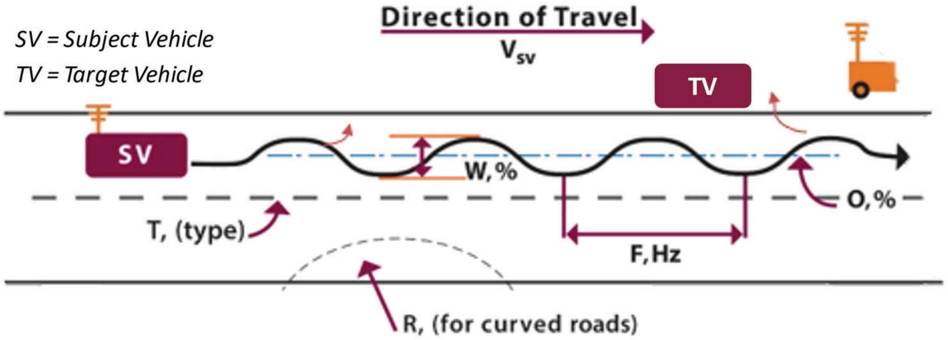


Figure 3. Schematic of Lane Slalom Test Template

The turning circle test was designed to test a sensor's ability to follow a turning vehicle at several viewing angles and range locations (Figure 4). The SV was stationary while it observed the TV traveling in a circle at a constant radius (R_{TV}). The SV's distance from the TV's circle and speed of the TV can be altered. In addition, the orientation of the SV relative to the circle can be changed to test the performance in different regions of the sensor FOV including the edge.

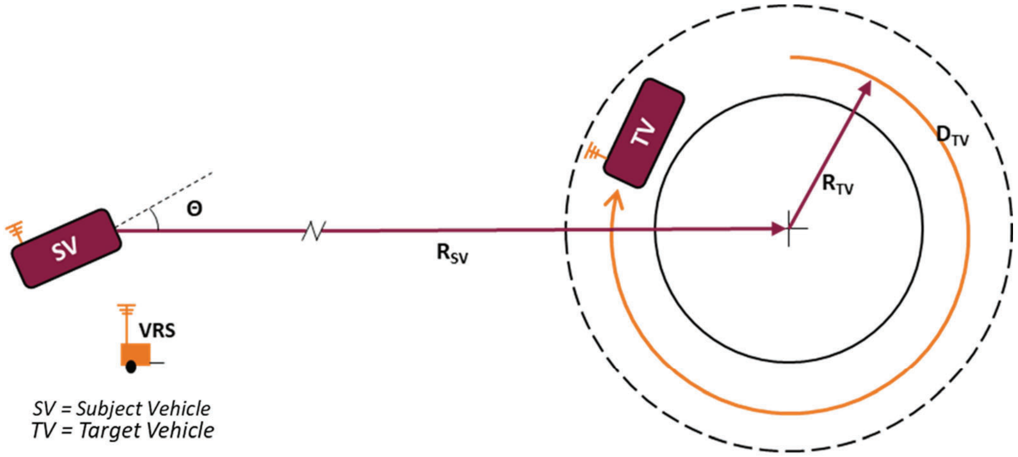


Figure 4. Schematic of Turning Circle Test Template

Table 1 provides a list of the parameters associated with the three templates and an example set of values used during the study. Bold values indicate values used for baseline conditions.

Table 1. Example Parameter Definitions for Test Templates

Variable	Range Sweep Test	Slalom Test	Turning Circle
Distance (D)	*40 -> 2 -> 40 (m)*		
SV Velocity (Vsv)	*Vtv +/- dD*	10, **15**, 20 (mph)	
TV Velocity (Vtv)	10, **15**, 20 (mph)		5, 10, **15** mph
Rate of Change of D (dD)	3, **5**, 10 (mph)		
Position (Lane) of TV (P)	-100%, **0**, +100%		
Radius of Curvature (R)	**0**, 15 (m)	**0**, 15 (m)	
Type (T)		*Solid, Dashed, Straight, Curved, White, Yellow, Ped X-ing, Stop Bar*	
Amplitude/Width (W)		**50**, 100, 150 (%)	
Frequency (F)		**0.4**, 0.75, 1 (Hz)	
Offset (O)		**0**, 25, 50 (%)	
Turning Radius (Rtv)			30, **40**, 50 m
Viewing Range (Rsv)			15, 25, **35**, 45 m
Viewing Angle (Θ)			-30, **0**, 30 deg

Workflow for Test Methodology

The test maneuvers provided a set of parameters that could be varied. Additional elements were included (e.g., rain, sensor degradations) to provide a characterization that could be more representative of the real-world performance. This, along with the different sensor modalities, creates the potential for a large test space. To help make the test space more manageable and relevant for a given sensor and/or application, the following process (Figure 5) was developed to help guide the development of a test plan, resulting in a final test matrix.

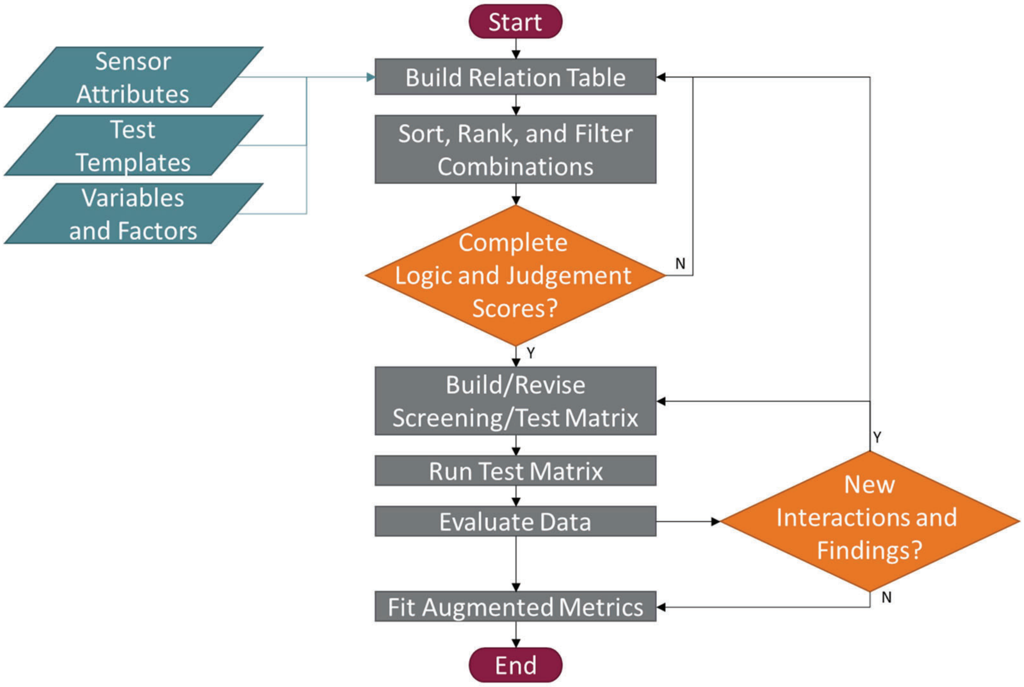


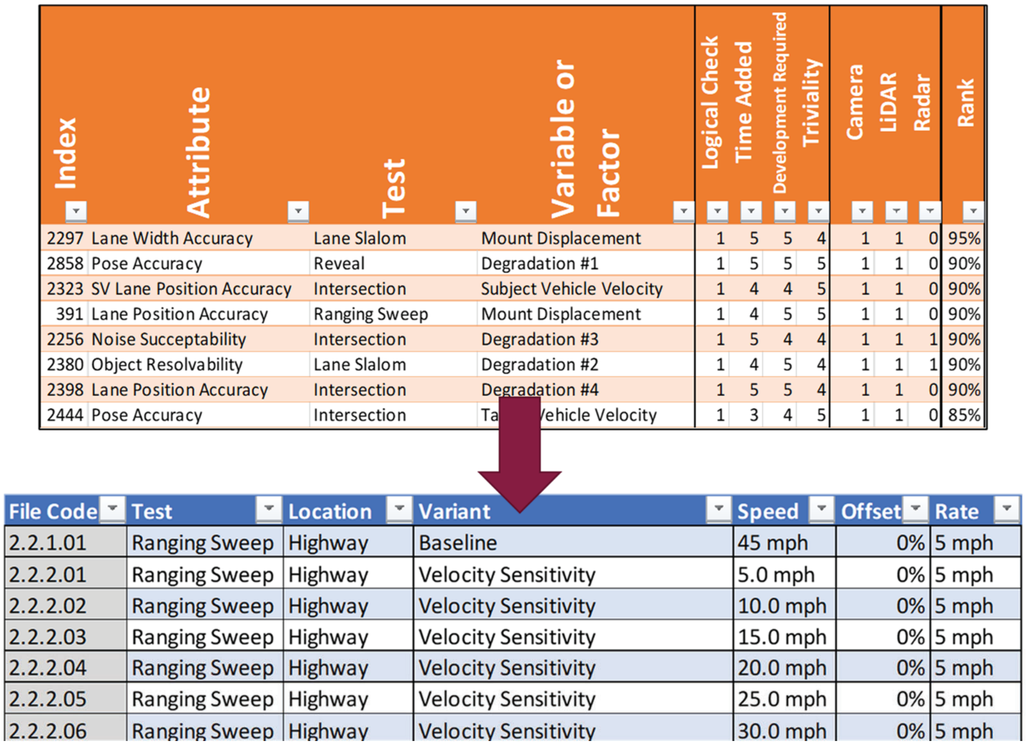
Figure 5. Process flow for the development and execution of the sensor characterization test plan.

Sensor attributes help define the measures of interest and the test conditions. The intended application for the sensor and the environment in which it may operate are important when defining the test template parameters and test conditions. For example, a radar and a camera intended for forward safety applications will likely have different test conditions (e.g., lighting) that are useful for characterizing each sensor's performance efficiently and accurately. Similarly, the same sensor modalities intended for parking assist applications will have different test conditions and potentially different measures of interest i.e., object position and pose accuracy may be more relevant than max range and object velocity.

A key part of the framework is the test templates introduced above that can be configured to specify a variety of maneuvers, relative motions, and road definitions The framework considers general sensor attributes as well as test variables and factors to define the test template parameters and the conditions which these maneuvers will be executed.

In addition to the sensor considerations, the intended operational conditions help identify test variables and factors of interest. While the influence of variables and factors are treated the same in the analysis, they were kept separate to distinguish between test maneuvers and the conditions in which the maneuvers were carried out. Test variables were used in the test templates to define the test maneuvers (e.g., vehicle velocity, following distance, radius of curvature, lane position). Test factors define the conditions the tests are performed (e.g., ambient light, rain rate, sensor surface degradation). Taken together, variables and factors are referred to as parameters.

Once the sensor attributes, test maneuvers, and parameters of interest are defined, they are organized into a relation table to help prioritization the test conditions. Since the number of test conditions has the potential to be very large, this step facilitates a structured down selection process to improve testing efficiency. It also provides an intermediate step in defining the test matrix. Figure 6 shows an example of an excerpt from the relation table and subsequent test matrix.



Index	Attribute	Test	Variable or Factor	Logical Check	Time Added	Development Required	Triviality	Camera	LiDAR	Radar	Rank
2297	Lane Width Accuracy	Lane Slalom	Mount Displacement	1	5	5	4	1	1	0	95%
2858	Pose Accuracy	Reveal	Degradation #1	1	5	5	5	1	1	0	90%
2323	SV Lane Position Accuracy	Intersection	Subject Vehicle Velocity	1	4	4	5	1	1	0	90%
391	Lane Position Accuracy	Ranging Sweep	Mount Displacement	1	4	5	5	1	1	0	90%
2256	Noise Succeptability	Intersection	Degradation #3	1	5	4	4	1	1	1	90%
2380	Object Resolvability	Lane Slalom	Degradation #2	1	4	5	4	1	1	1	90%
2398	Lane Position Accuracy	Intersection	Degradation #4	1	5	5	4	1	1	0	90%
2444	Pose Accuracy	Intersection	Ta ehicle Velocity	1	3	4	5	1	1	0	85%

File Code	Test	Location	Variant	Speed	Offset	Rate
2.2.1.01	Ranging Sweep	Highway	Baseline	45 mph	0%	5 mph
2.2.2.01	Ranging Sweep	Highway	Velocity Sensitivity	5.0 mph	0%	5 mph
2.2.2.02	Ranging Sweep	Highway	Velocity Sensitivity	10.0 mph	0%	5 mph
2.2.2.03	Ranging Sweep	Highway	Velocity Sensitivity	15.0 mph	0%	5 mph
2.2.2.04	Ranging Sweep	Highway	Velocity Sensitivity	20.0 mph	0%	5 mph
2.2.2.05	Ranging Sweep	Highway	Velocity Sensitivity	25.0 mph	0%	5 mph
2.2.2.06	Ranging Sweep	Highway	Velocity Sensitivity	30.0 mph	0%	5 mph

Figure 6. Excerpt from a relation table (top) and test matrix (bottom).

Construction of the final test matrix was an iterative process. The test configurations identified in the test plan framework were used to populate a relation table that provided a means to qualitatively assess different test parameter combinations and prioritize which configurations to include in the initial testing. The first round of data collection employed a screening matrix consisting of two to four levels for a given set of parameters to help identify interactions, sensor sensitivities, and areas that may need additional levels. Figure 7 provides and examples of how the hypothetical or expected response can be used to identify initial parameter values. The left graph shows that the expected position accuracy will have a non-linear relationship to range and rain intensity and there will be an interaction between these two variables. The right graph provides a sampling strategy for the initial values for the screening matrix (levels shown in orange). Results from the initial test are then used to identify areas that require additional levels of rain intensity and range (shown in maroon) to be able to characterize the performance of the sensor more accurately.

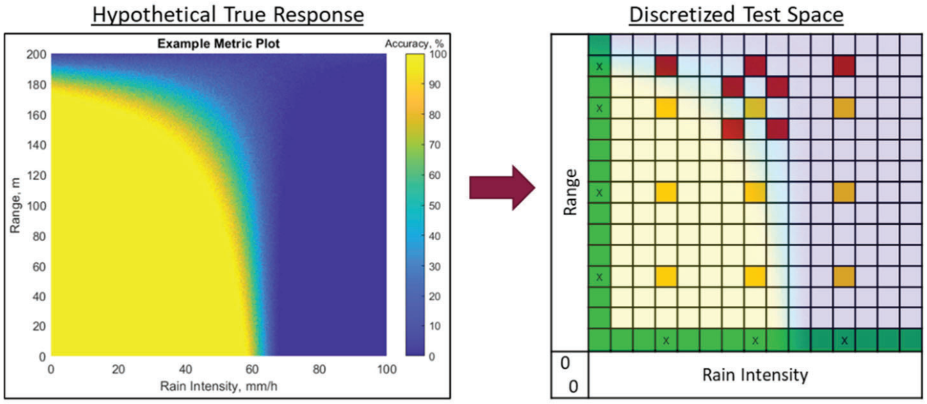

Hypothetical True Response
Discretized Test Space

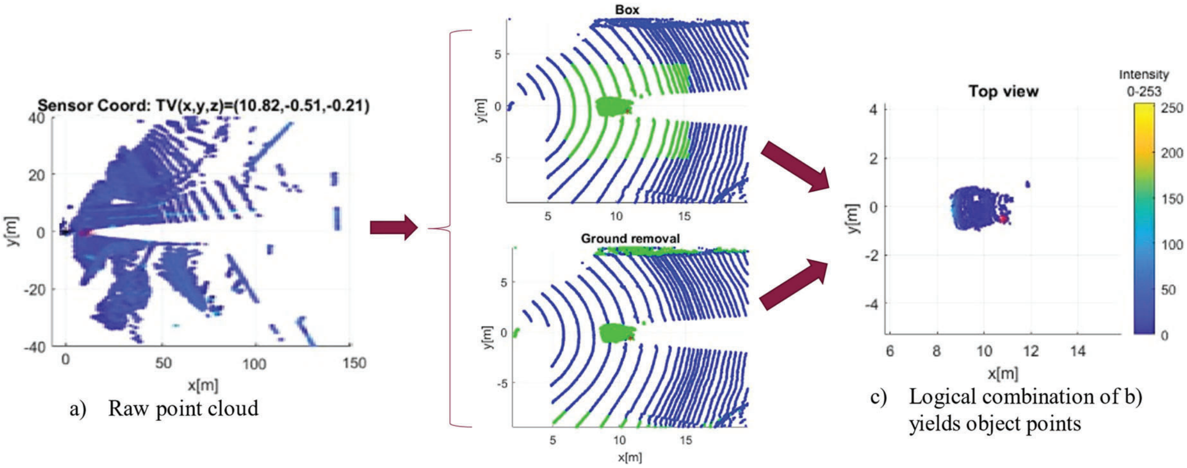
Figure 7. Example of how a screening matrix can be used to select relevant parameter levels.

Data Processing

Figure 7 also provides a visual representation of what data points might be needed to create a relationship for the multi-way dependencies. By fitting an equation to the surface response, sensor accuracy as a function of both range and rain intensity can be found. This can be extended to higher dimensions as well.

As noted above, one of the objectives for the test plan was to facilitate comparison of different sensors and sensor types. However, not all sensors provide the same native output. Many sensors have on-board processing that allows them to provide a list of objects with relevant measures. Consequently, data from sensors that provided less processed data e.g., point cloud from LiDAR or pixel information from a camera, had to be processed to extract object information. The goal of the study was not an evaluation of different object identification algorithms. Therefore, the team used a priori knowledge of the location of the object of interest from the DGPS as a basis for distinguishing which points in the sensors' output were associated with the TV.

Error! Reference source not found. shows an example of the steps used to identify the TV. The top figures provide a bird's eye view of the point cloud. Subplot (a) shows the full FOV for this LiDAR. Subplot (b) is zoomed into the region around the SV and TV. In both, the SV is located at (0, 0) and the TV, represented by the red dot, is located at (10, 0). The first frame (a) shows the density of information provided by the LiDAR. Subplot *b* (right) is zoomed closer to the TV position (note the different axis scale) and highlights the steps used to extract the points associated with the TV. First a bounding box was created based on the DGPS position (shown in green). An algorithm was used to identify the points in this region that were associated with the ground plane. Any points in the point cloud not in this plane were then separated. The ground plane points are shown in blue in the lower image of (b) and all other points are shown in green. The final step extracts the points around the TV position from the DGPS that are not part of the ground plane. The resulting points are shown in subplot (c) where the color indicates the intensity of the returned signal with the rear of the vehicle having a stronger return than the other surfaces. Using this approach allowed the team to have object data from the LiDAR sensors structured in a manner consistent with radars and other object level sensors.



a) Raw point cloud

b) DGPS centered bounding box and rejected ground plane

c) Logical combination of b) yields object points

Figure 8. Identification of the target vehicle using the DGPS-seeded bounding box.

DATA SOURCES

The sensor types for this research were limited to perception sensors (rather than localization sensors), including various types of radar, LiDAR, and cameras intended for use in ADS-equipped vehicles. The method developed and applied in this research effort would also apply to other perception sensing technology. Though localization sensors are an important part of ADS operation, GPS and inertial sensors were not included in the testing portion of this research.

The sensors used in this research included four radar, four LiDAR, and two camera sensors. Table 2 lists the sensors, including the type, reference name, a brief description of the sensor, and the native output from the sensor. One of the 77-GHz radars produced two datasets: firstly, a short range (SR) set with a wide field of view (FOV) without any vertical information, and secondly a long range (LR) set with a narrow FOV and vertical location information. As noted previously, DGPS was included in both the SV and TV to provide ground truth measurements.

Table 2. Sensors Included in Research

Sensor Type	Anonymized Name	Brief Description	Sensor Output
Camera + processor	Object Camera	Mono camera with on-board processing for object and lane tracking	Identified Objects
Camera	Machine Vision	Mono camera, Color	Pixels (photos)
LiDAR	32	32 beam,360 degrees FOV	Point Cloud
LiDAR	128	128 beam,360 degrees FOV	Point Cloud
LiDAR	Solid State	Solid state LiDAR, forward FOV, multiple beams	Point Cloud
LiDAR	Budget	Mechanical LiDAR, forward FOV, single beam	Point Cloud
Radar	SR 24 GHz	Short range radar operating at 24 GHz.	Bulk Objects
Radar	SR 77 GHz	Short range radar operating at 77 GHz.	Point Cloud
Radar	LR 77 GHz	Long-range radar operating at 77 GHz with two scan formats: 1. Wide near field and 2. Narrow far field	Point Cloud

A t-slot frame was used to mount the sensors to the SV. The frame provided a means to mount and align a variety of sensors and simultaneous collection data from all sensors. This ensured the test conditions were the same for each sensor, improving collection efficiency.

The data were collected asynchronously with their own time stamp to utilize each sensor's maximum sample rate. This flexibility required additional steps during data processing but allowed the team to collect data from a variety of sensors simultaneously during each trial.

RESULTS

The research resulted in several key findings. First, it confirmed that the sensor performance observed in representative driving conditions differed from the performance listed in sensor specification sheets (Table 3). Similarly, degraded conditions tended to negatively impact sensor output as can be seen in the last two columns.

Table 3. Comparison of Measured Maximum Range to Specification Sheet

	Sensor	Specification Sheet (m)	Baseline (m)	Percent Change	In Rain (m)	Percent Change
	Object Camera	250	133.3	**-47%**	116.3	**-53%**
LiDAR	32-Beam	100	37.5	**-63%**	29.2	**-71%**
	Solid State	600	145.0	**-76%**	113.7	**-81%**
Radar	SR 77 GHz	64	66.4	**4%**	66.2	**3%**
	SR 24 GHz	95	98.0	**3%**	24.8	**-74%**
	LR 77 GHz (wide)	100	102.2	**2%**	100.5	**1%**
	LR 77 GHz (far)	250	203.2	**-19%**	203.2	**-19%**

The effect of degradation on performance were often non-linear and dependent on the type of degradation and the sensor. Rain resulted in a decrease in the detection range by at least ten percent for the nine sensors tested. While lighting primarily affects camera performance, on-coming headlights at dusk did result in a measurable difference in distance accuracy for one LiDAR. Surface degradations, which attenuate or distort the signal, tended to reduce the performance of the sensor. However, the effect was dependent on the sensor and the degradation (both type and level).

The following plots further explore the effect of rain. The baseline ranging sweep test was performed along with four lane offset conditions from one half lane (50%) to two full lanes (200%). Rain towers were positioned along the roadway and the vehicles executed the test maneuver with and without rain.

Camera performance was particularly susceptible to rain conditions (Figure 9). In addition to attenuating the signal, water droplets distort the light. In addition, while the lateral offset did not have a strong influence on distance error without rain, there was an influence in the presence of rain.

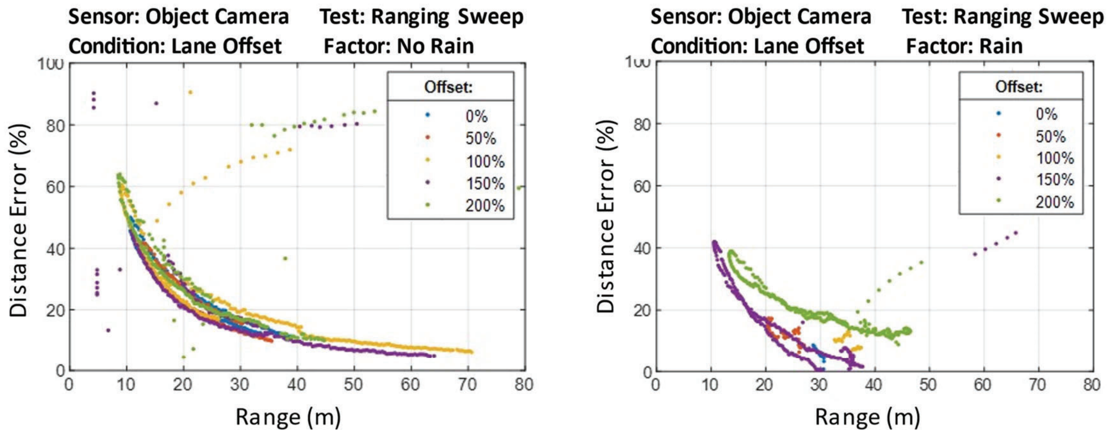


Figure 9. Object camera detection accuracy plots without rain (left) and with rain (right).

LiDAR performance also deteriorated with increased rain (Figure 10). Water has similar detrimental effects for LiDAR and camera. Rain attenuates and blocks the laser pulses to and from the target. This was observed with both the solid-state and 32-beam LiDAR. The LiDAR does not display the sensitivity to lane offset in the presence of rain as was observed for the camera.

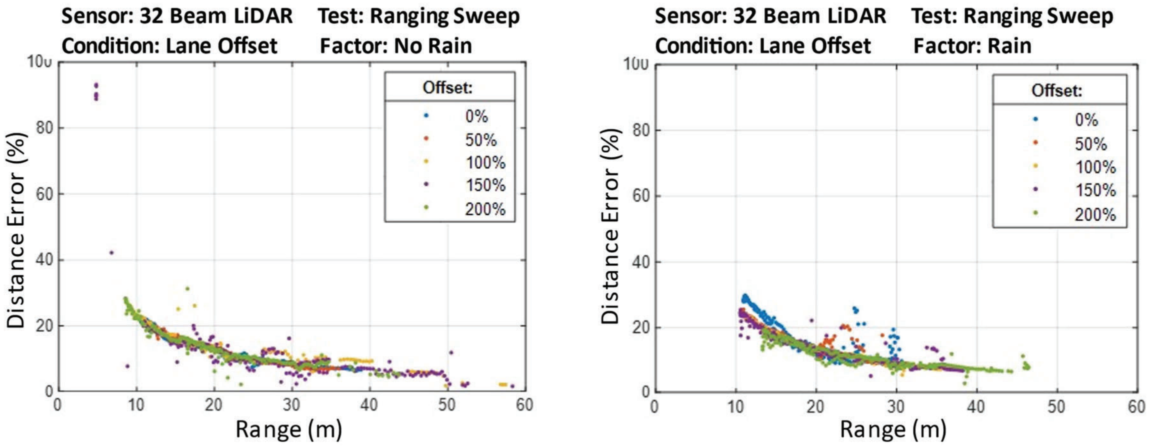


Figure 10. Thirty-two beam LiDAR detection accuracy plots without rain (left) and with rain (right).

As noted earlier, the 24 GHz sensor was the only radar that was significantly affected during the rain trials (Figure 11). Because the 24 GHz radar reports objects, not point clouds, the object tracking software within the object radars could be more susceptible to reduced detection range than those radars that just report point clouds.

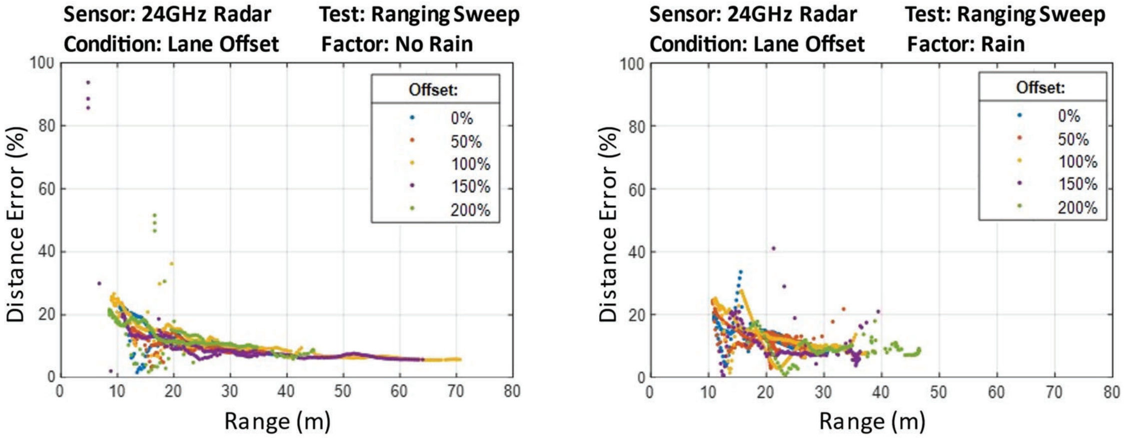


Figure 11.SR 24 GHz radar detection accuracy plots without rain (left) and with rain (right).

Optical degradations can affect LiDAR intensity values and number of points. Adding the sandblasted degradations (which impacts transmission and scattering of light) to the solid-state LiDAR in the ranging sweep trials decreased the maximum and average intensity values (Figure 12). The high-level sandblasted degradation resulted in very few returned points. While this level of surface wear may not be realized, it does provide a limit condition for characterization testing.

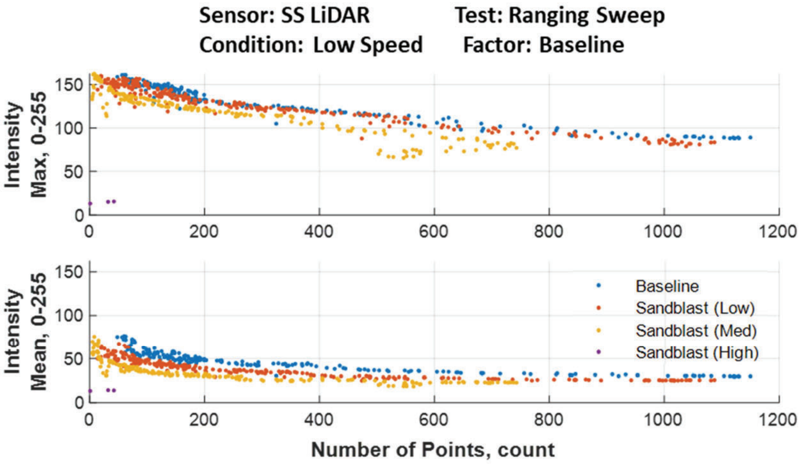


Figure 12. Intensity values for the solid-state LiDAR with three levels of sandblasted degradations.

DISCUSSION AND LIMITATIONS

Standardized tests are designed to minimize noise and subsequently test in what could be considered best-case conditions with ideal targets. However, this may not provide a good indication of the performance in the driving environment. The approach presented can be applied across a variety of sensors to provide metrics to compare different sensors and sensor types for a common set of conditions. The method developed provides a framework to guide testing and analysis. The use of the method will require the practitioner to select factors and variables that are relevant to the sensors and applications of interest. This could include additional test conditions, external factors, and metrics.

As noted, this research focused on sensor evaluation rather than perception software. The methods used for extracting object information is not indicative of how it would be accomplished in system integrated into a vehicle. The method employed used the DGPS signal and knowledge regarding the size of the target to identify points associated with the target object. Consequently, the analysis could identify individual returns associated with the object of interest. Since the DGPS is doing the job of the perception system in identifying the target, it is unlikely that the system, when integrated into a vehicle, will be able to detect objects with this limited of data. Without an understanding of the constraints of the object detection algorithm to be used, this could lead to an unrealistic expectation of system performance based on the characterization results. For example, if an algorithm needs a minimum number of points distributed across the vehicle to perceive it in the point cloud, the maximum detection range may be lower than what the characterization data shows. Therefore, it is important to understand the method and its assumptions in the context of how the sensor may be used and the associated data processing. Though not discussed, this also holds true for evaluating lane line detection performance for sensors that do not natively output this information.

Another consideration in the use of this method is that it, by design, is unlike a standardized test which is tightly defined to help ensure repeatability and reproducibility. This method requires careful documentation and a means to account for external factors that could affect repeatability (e.g., sunlight) or that may vary and effect reproducibility (e.g., the use of different target vehicles). The quality of the ground truth measurement system could also influence the results. The DGPS system used for this study could measure the location to within 10 cm, or approximately the width of a lane line. This may be adequate for many applications. The performance of a different ground truth system may or may not be sufficient for another party's system requirements.

CONCLUSIONS

The research team developed a methodology that provides a structured way to characterize the performance of sensors and sensor systems. The method defines a set of generalized test templates, a means to select the most relevant test factors, and an approach for analysis that includes multi-factor metric fitting to facilitate sensor performance characterization in the expected operational conditions. This approach targets representing the final performance rather than the current sensor data provided based on laboratory tests.

However, the test space can become large when considering the different parameters and combinations. The method outlined provides a framework to aid in the identification and selection of key test configurations. The flexibility associated with the method also makes it applicable to sensor systems that provide object level data. This flexibility does come at a cost. Unlike a standardized test where the steps and conditions are tightly defined, successful application of developed method benefits from knowledge and understanding of the sensors, their potential applications, operational conditions, and relevant measures and metrics.

It is also important to apply the results appropriately. While the results from this method may not necessarily reflect the final performance of the integrated perception, it does provide a means to generate results that more closely reflect how a sensor would likely perform in the automotive environment.

REFERENCES

[1] Wu, M., & Coifman, B. (2019, 10/01). Quantifying what goes unseen in instrumented and autonomous vehicle perception sensor data – A case study. *Transportation Research Part C: Emerging Technologies, 107*, 105-119. https://doi.org/10.1016/j.trc.2019.07.024.

Evaluation of Different ADS Material Concepts using Various Safety Metrics

William Thomas Hollowell, Ph.D.
WTH Consulting LLC

Rudolf Reichert
Steve Kan, Ph.D.
Chung-Kyu Park, Ph.D.
George Mason University

United States

Paper Number 23-0251

ABSTRACT

Research Question/Objective: New vehicle concepts of occupied and un-occupied Automated Driving Systems [(U)ADS] are fast evolving. Their design, materials used, and energy absorbing structures can significantly differ from traditional vehicles. Appropriate analysis methods and safety metrics can help to evaluate their crashworthiness and compatibility when colliding with other vehicles or road-side hardware. This paper explores the effect different material concepts of an ADS vehicle's outer body has on self-protection and partner protection. The research is considered an example to demonstrate how various impact configurations and simulation analysis tools and metrics can be used to assess structural and occupant aspects for this new type of vehicle class.

Methods and Data Sources: Previously developed Finite Element (FE) models of generic ADS vehicles in combination with validated road-side hardware, crash barrier, and occupants were used to understand the effect different material concepts can have on self- and partner-protection. Partner-protection was analyzed using EuroNCAP's mobile progressive deformable barrier (MPDB) and its respective compatibility metric Occupant Load Criterion (OLC), where lower values represent better compatibility. Self-protection was studied using occupant injury metrics recorded during a run-off road impact scenario, where the ADS vehicles impacted a New Jersey Barrier (NJB).

Results: Differences in crash compatibility were observed depending on the material concepts used. The impact of a mid-size ADS vehicle using thermoplastic material for select components with the MPDB resulted in an OLC value of 18. The same vehicle with a composite material concept showed an OLC value of 19, while an OLC value of 22 was recorded for the baseline vehicle with a steel material concept. Differences in occupant metrics HIC, BRIC, chest deflection, and femur loads were small when comparing the three material concepts in a 35-mph oblique impact into a NJB.

Discussion and Limitations: The use of different material concepts resulted in different total vehicle mass. The vehicle using thermoplastic material for select components had a mass of 3,653 kg. The same vehicle with composite material concept had a mass of 3,718 kg, while the baseline vehicle using steel had a mass of 4,273 kg. Lower vehicle mass correlated with better partner-protection based on OLC metrics. Occupant metrics were mainly affected by the interior concept, which was identical for all three vehicles. Differences in occupant load was therefore small. The same vehicle design and underlying structure was used during this study and no optimization towards the respective material concept was performed.

Conclusions and Relevance to Session Submitted: The research is relevant to demonstrate how simulation tools can contribute to assessing this new type of ADS vehicle class. Material concepts that resulted in a smaller vehicle mass tended to show better partner protection. The interior concept, which was the same for all three ADS vehicle variations, was the main factor for producing similar occupant injury metrics for the evaluated impact scenario.

INTRODUCTION

In perpetuation of successful research collaboration for more than a decade [1,2,3], the American Chemistry Council (ACC) and the George Mason University (GMU) continued to conduct research to understand opportunities for using plastics and composite materials for future automated vehicles.

Automated Driving Systems (ADS) have the potential of significantly reducing fatalities and serious injuries by reducing the number crashes on US roadways. These new technologies, however, may present unique challenges for protecting occupants in the remaining crashes that still occur. New ADS are expected to include new vehicle types that are configured to carry cargo or occupants or both. Current vehicles are designed to have their crash response in accordance with the Federal Motor Vehicle Safety Standard (FMVSS) occupant protection standards. Crash configuration of fully automated vehicles can be expected to use advanced sensor technology. However, these new ADS would potentially encounter crashes with existing vehicles and road-side hardware due to sensor malfunction, for example. Little research exists that explores crash scenarios with the existing road-side hardware that go beyond impact conditions evaluated today.

The GMU-Team has previously developed generic Finite Element (FE) models of different size ADS using traditional steel materials, funded by the Department of Transportation (DOT) [4]. The models are used to understand the performance of these new type of vehicles in run-off road crashes, i.e., when impacting road-side hardware devices. Subsequently, these models were used to study compatibility aspects when an ADS collides with a traditional vehicle. [5]

NHTSA recently published a Notice of Proposed Rulemaking (NPRM) “Occupant Protection for Automated Driving Systems” that addresses potential rulemaking changes to vehicles with and without ADS functionality, such as the definition of driver and protections required when there is not a steering wheel or steering column in a motor vehicle.

OBJECTIVE

Use generic ADS vehicle models that can also be used for occupant transportation to study the effect of using plastics and composite materials instead of steel for select components. Crash configurations included run-off road conditions and frontal impact compatibility crash test scenarios.

Apply commonly used metrics, such as OLC, to estimate differences in occupant loads. Conduct integrated occupant vehicle simulation using a generic sled model and adequate interior component FE models.

---

[1] Park C-K, Kan C-D, Hollowell W T, and S.I. Hill, “Investigation of opportunities for lightweight vehicles using advanced plastics and composites,” Report No. DOT HS 811 692, NHTSA, Washington, DC, 2012 http://www.nhtsa.gov/DOT/NHTSA/NVS/Crashworthiness/Plastics/811692.pdf

[2] Park C-K, Achstetter T, Kan C-D, Hollowell W T, “Understanding of Numerical Polymer/Composite Material Models and Their CAE applications,” George Mason University Final Report, 2017

[3] Hollowell W T, Kan C-D, Park C-K, Reichert R, Evaluation of the Safety Performance and Weight Reduction Using CFRP Modified Automotive Structures in NHTSA's Frontal Oblique Impact Test, ESV Conference, Eindhoven, Netherlands, 2019

[4] Reichert R, Marzougui D, Kan C-D, “Simulations Between Non-Occupied Automated Driving Systems and Road-side Hardware”, Report Number : DOT HS 812 871, NHTSA, Washington, DC, 2020 URL : https://rosap.ntl.bts.gov/view/dot/54288

[5] Reichert R, Kan C-D, Park C-K, Crash Compatibility for Unoccupied Automated Driving Systems, NHTSA, 2022

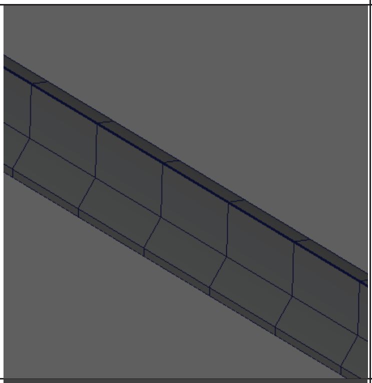
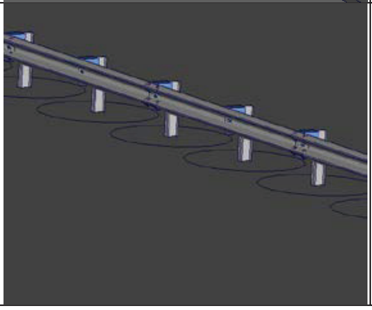
In addition, an advanced composite material model for shell elements is being developed and validated in cooperation with the Ohio State University (OSU) and Honda. The “MAT_213” material model will be made publicly available in LS-DYNA, once completed.

METHODS

Road-side Hardware

Current road-side testing practices are described in the Manual for Assessing Safety Hardware (MASH). [6] Two representative road-side devices, i.e., (1) a rigid “New Jersey Barrier and (2) a W-beam guardrail, are shown in Table 1. Run-off-road crashes, which usually involve only a single vehicle, contribute to a large portion of fatalities and serious injuries to motor vehicle occupants.[7] The NJB was selected to study the effect of different material concepts for ADS vehicles. Additional studies with the W-Beam guardrail and other road-side devices were conducted in previous research efforts.[4]

Table 1.
Representative Road-side Hardware

	Road-side Device	FE Model	Picture of similar Physical Device	Relevant References
1	New Jersey Barrier			Related studies and validation [8]
2	W-Beam Guardrail			Related studies and validation [9]

---

[6] Manual for Assessing Safety Hardware (MASH) Transition, DOT Federal Highway Administration, Federal Register Volume 80, Issue 219, 80 FR 70288, Docket No. FHWA-2015-0008

[7] Cejun Liu, Ph.D., and Tony Jianqiang Ye, “Run-Off-Road Crashes: An On-Scene Perspective,” Report No. DOT HS 811 500, NHTSA, Washington, DC, 2011

[8] Marzougui et al, Crash Test & Simulation Comparisons of a Pickup Truck & a Small Car Oblique Impact into a Concrete Barrier, GMU, Fairfax (2014).

[9] Marzougui et al, Evaluation of Rail Height Effects of the Safety Performance of W-Beam Barriers, NCAC, Ashburn (2007).

Traditional and ADS Vehicles

Vehicles recommended for testing under MASH are the 1100C small car, 1500A mid-size sedan, and the 2270P pick-up truck. FE models of a Toyota Yaris small car and a Chevrolet Silverado have been developed, validated, and used in previous research.[4] Pictures of two conventional vehicles are shown in Figure 1. They were used as references in related research.[4]

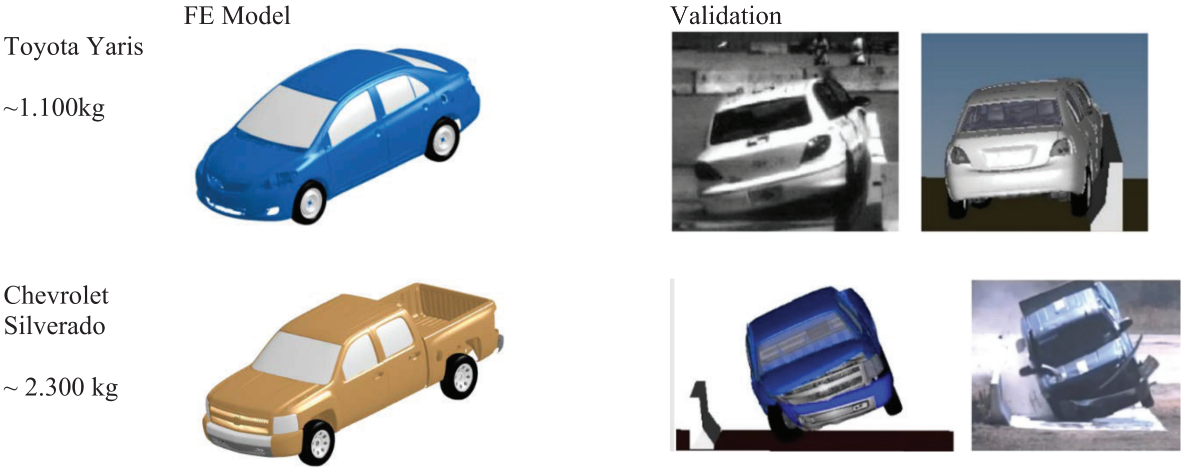


Figure 1. Conventional Vehicles

Other validated FE models that can be used as traditional vehicle references for material concept and other research studies include the 2014 Honda Accord [10], 2015 Toyota Camry[11], and 2020 Nissan Rogue. [12] The 2015 Toyota Camry representing the sedan vehicle class and the 2020 Nissan Rogue representing the SUV vehicle class have been validated against existing pedestrian safety impact configurations, in addition to validation against frontal and side impact scenarios. ADS vehicles for this study are based on the two concepts shown in Figure 2. The mid-size ADS could be used to carry cargo or up to 5 occupants comparable to a sedan vehicle. The large ADS could be used to carry cargo or up to 10 occupants comparable to an airport shuttle, for example.

Mid-size ADS "Mercedes Vision" [13]

Large ASD "Enride T-Pod" [14]

Figure 2. ADS Concepts

---

[10] Singh H, "Vehicle Interior and Restraints Modeling", EDAG Inc., Washington DC, 2017, https://www.nhtsa.gov/sites/nhtsa.gov/files/documents/812545_edagvehicleinteriorandrestraintsmodelingreport.pdf,

[11] Reichert R, Kan C-D, (2017). "Development of a 2015 Mid-Size Sedan Vehicle Model". 11th European LS-DYNA Conference

[12] Reichert R, Mahadevaiah U, Fuchs L, Kan C-D, "Development of a 2020 SUV vehicle FE model", 16th LS-DYNA Forum 2022, Bamberg, Germany, 2022. https://www.ccsa.gmu.edu/models

[13] https://www.theverge.com/mercedes-benz-vision-urbanetic-self-driving-electric-concept-design, accessed June 2019

[14] https://venturebeat.com/swedens-einride-debuts-prototype-t-pod-an-autonomous-electric-truck-that-can-also-be-controlled-remotely, accessed June 2019

Previously developed methods to develop generic ADS FE models included the transformation into an electric drive and the use of skateboard-type chassis. See Figure 3.

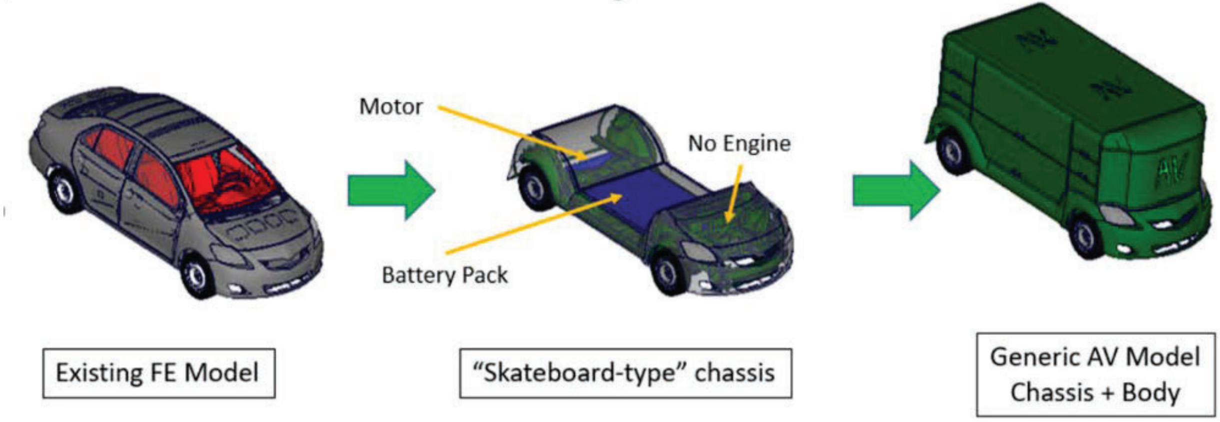


Figure 3. Development of Generic ADS Models

Existing ADS vehicle concepts served as a reference for the development of generic mid-size and large ADS vehicle models. For examples, see Figure 4.

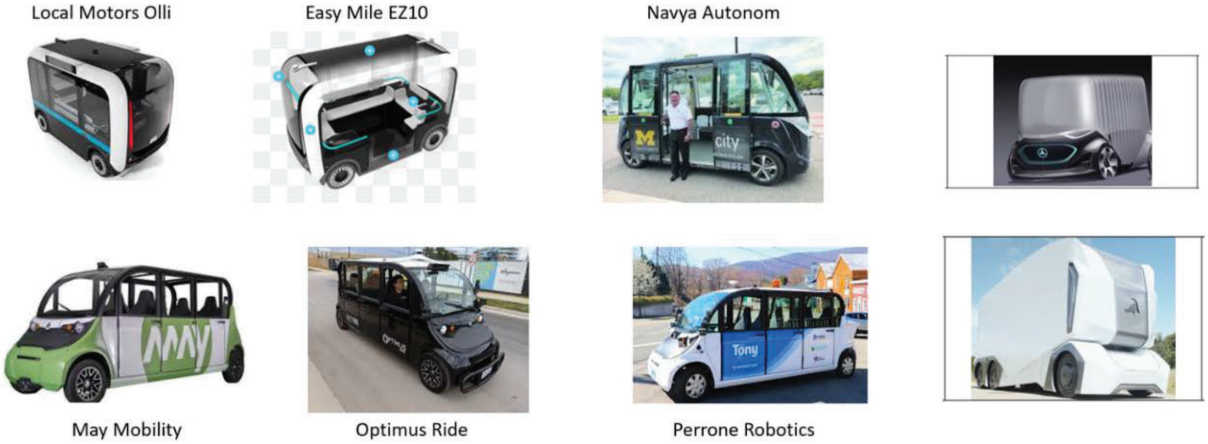
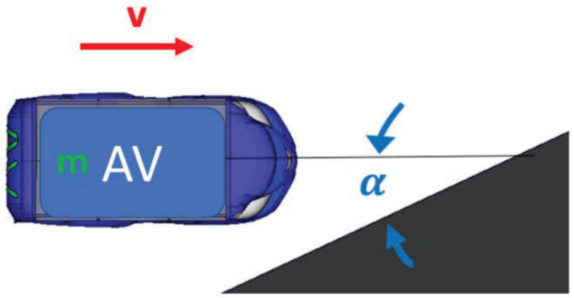


Figure 4. ADS Vehicle Concept Examples

Impact Configurations and Parameters
A standard impact into a NJB guardrail was conducted with vehicles of specified mass, at 100km/h impact speed and a 25° angle. Since impact scenarios with ADS vehicles are expected to differ in impact angle and impact speed, a wider range of parameters was studied. The three parameters are illustrated in Figure 5.



v: impact velocity

α: impact angle

m: vehicle mass

Figure 5. Impact Configuration and Parameters

Material Concepts
Steel, composite, and thermoplastic material concepts for the ADS vehicle body were considered, as shown in Figure 6.

Figure 6. Generic Vehicle Model and Material Concepts

A braided carbon-fiber thermoset composite material has been used in previous projects. [15] In addition, GMU has developed a material model of MAT 213 Shell Element version for implementation into LS-DYNA[16]. The development of the material model including failures has been completed and was integrated into the code.

Development of Component Models for Occupant Analysis
ADS vehicles are anticipated to allow unconventional seating configurations, such as rotated orientations. Therefore, the seat-belt D-ring will have to be integrated into the seat, rather than mounted at the vehicle structure's B-Pillar, for example. Consequently, a FE model of a seat with integrated seatbelt has been developed, as shown in Figure 7, and used in previous research efforts. [17]

---

[15] Park C-K, Kan C-D, Hollowell W T, and S.I. Hill, "Investigation of opportunities for lightweight vehicles using advanced plastics and composites," Report No. DOT HS 811 692, NHTSA, Washington, DC, 2012 http://www.nhtsa.gov/DOT/NHTSA/NVS/Crashworthiness/Plastics/811692.pdf

[16] Achstetter et al, "Development of a Composite Material Shell-Element Model for Impact Applications", LS-DYNA Conference, 2020

[17] Reichert, R. and Kan, C.-D., "Effect of Reclined and Rotated Seating for Automated Driving Systems," SAE Technical Paper 2022-01-5048, 2022, doi:10.4271/2022-01-5048

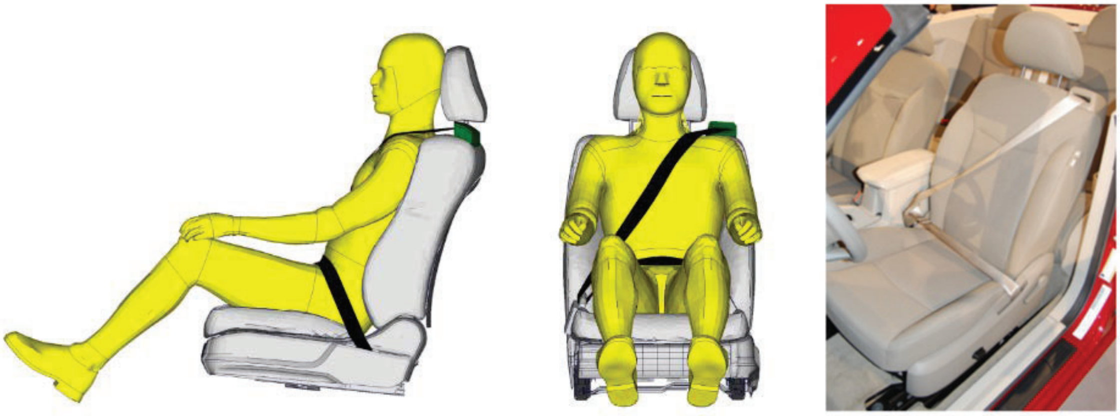


Figure 7. Seat with Integrated Seatbelt

“Seat-squash” and “seat-belt fitting” procedures are used to position respective occupants in the seat and to realistically place the seatbelt on the occupant.

Development of Component Models for Occupant Analysis
In preparation for subsequent studies using the developed generic ADS vehicle models with different material concepts in combination with interior and occupants, generic seat models have been developed. See Figure 8. They allow the study of a variety of unconventional seating arrangements including front-, side-, and rear-facing orientations.

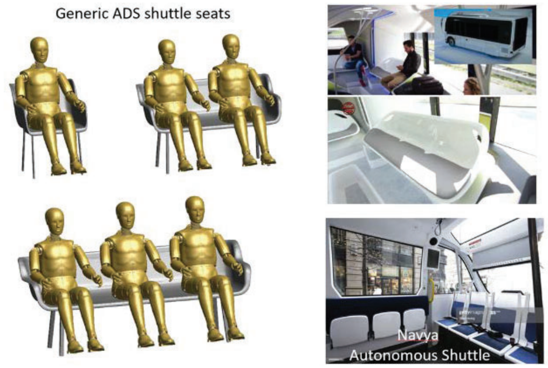


Figure 8. Generic Shuttle Seats

Development of FMVSS No. 201 Compliant Interior
Similarly, in preparation for subsequent studies using the developed generic ADS vehicle models with different material concepts in combination with interior and occupants, generic interiors were developed. FMVSS No. 201 requires vehicle interiors to be designed to produce a Head Injury Criterion, HIC(d) < 1000 when impacted by a Free Motion Head-form (FMH) at defined locations, angles, and impact speeds. Energy absorbing generic interior models with thermoplastic material characteristics were developed, that produced a HIC(d) value below 800, as shown in Figure 9. Their effect in comparison to non-energy absorbing “rigid” interior concepts were studied in previous research. [18]

---

[18] Reichert R, Kan C-D, Park C-K (2022, October). Crash safety considerations for speed-limited ADS shuttles (Report No. DOT HS 813 354). National Highway Traffic Safety Administration.

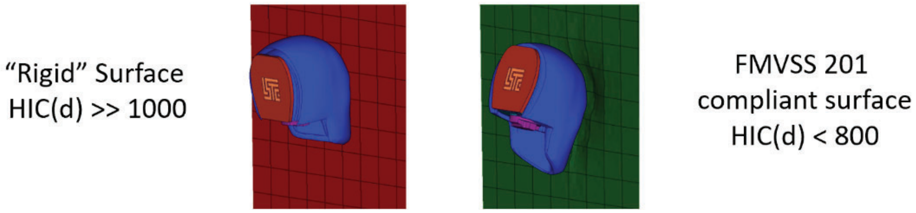


Figure 9. Development of FMVSS No. 201 Compliant Interior

Compatibility Assessment using EURONCAP's MPDB
EURONCAP has introduced an frontal offset test procedure with a Mobile Progressive Deformable Barrier (MPDB)[19] to assess the partner protection of a vehicle, as shown in Figure 10.

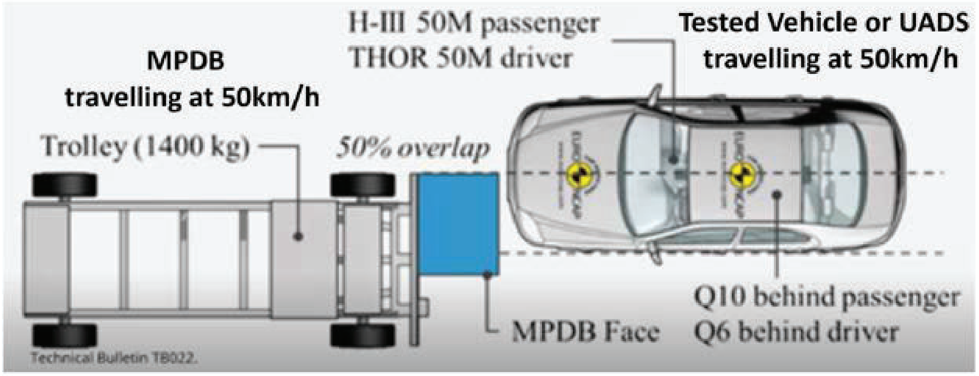


Figure 10. Vehicle Compatibility Rating Test Configuration

Respective MPDB FE models to assess the compatibility of traditional and ADS vehicles exist.[20] They were used to evaluate effect of different material concepts for ADS crash compatibility in frontal impact configurations.
`

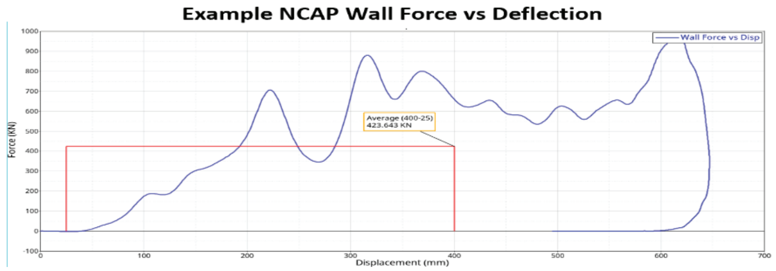
Different compatibility metrics to assess partner protection exist. They include (1) bumper height assessment, which aims to enhance partner protection primarily through geometric matching of front structural components of cars and light trucks and vans; (2) Average Height of Force (AHoF), which is calculated from NCAP load cell measurements to quantify vertical geometric alignment of a vehicle; (3) Crush Work Stiffness (Kw400), a metric to quantify the front-end stiffness related to the crush energy absorbed by a vehicle in the first 400 mm of crush when impacting a rigid wall; (4) OLC criteria used by EuroNCAP, calculated from the velocity pulse of the MPDB.

The crush work stiffness (Kw400) was used in related research, to study baseline ADS vehicles and to develop variations with different compatibility characteristics. Details are outlined in the discussion section. The crush work stiffness is determined by calculating the area under the Force – Deflection (F-D) curve between 25 and 400 mm of front-end crush in a NCAP full frontal impact configuration, as shown in Figure 11.

---

[19] Reichert R, Kan C-D, Park C-K, Crash Compatibility for Unoccupied Automated Driving Systems, NHTSA, 2022

[20] https://lsdyna.ansys.com/lstc-barrier-models. Accessed November 2022



$$\int_{25}^{400} F\,dx = \frac{1}{2} Kw400\,[(400)^2 - (25)^2]$$

$$Kw400 = \frac{2F}{425}$$

Figure 11. Crush Work Stiffness Compatibility Metric; (a) Force-Displacement Example; (b) Kw400 Equation

The resulting stiffness value ‘K’ is termed Kw400, based on the equation outlined in Figure 11 (b), where F is the average of the total force on the barrier between 25 and 400 mm of vehicle crush. The first 25 mm of crush is ignored to account for soft materials and noise in the measured data. The maximum crush is limited to 400 mm to isolate the high inertial forces on the load cell wall due to engine contact.

EuroNCAP’s OLC metric was selected as the main metric to evaluate the energy absorbing characteristics of ADS vehicles. The OLC metric is derived from the virtual dummy responses estimated from a governing equation involving an assumed restraint system and a given vehicle crash pulse. This metric is independent of the actual dummy response. It assumes a virtual and uniform restraint system and that a virtual dummy will be in free-flight-phase along a displacement of 65 mm. In the restraining-phase an ideal restraint is assumed that would decelerate the occupant until the relative velocity between the occupant and the vehicle becomes zero. It is assumed that the distance between the vehicle and the occupant at point B is an additional 235 mm, as shown in Figure 12. For EuroNCAP’s compatibility assessment, the OLC is evaluated using a sliding scale between 25 g and 40 g. OLC values below 25 g result in four points and values above 40 g result in zero points.

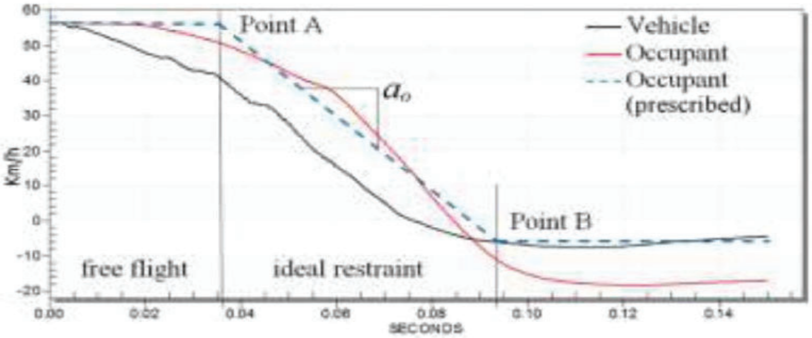


Figure 12. Occupant Load Criterion (OLC)

Bottoming out is defined as an area of the barrier that is 40 mm x 40 mm in height and width that has been penetrated by 630 mm or more. It is determined from a physical examination of the barrier face and vehicle.

RESULTS

Simulation Study 1 - Mid-size ADS impacting New Jersey Barrier (NJB)

Results using the mid-size ADS impacting the NJB are presented in this study. Figure 13 shows a top and side view of the developed mid-size ADS vehicle and impact configuration. Studied impact parameters and vehicle characteristics are listed below:

- Range of impact angles: 20°, 25°, and 30°
- Range of impact speeds: 25 mph, 35 mph, and 45 mph
- Material concepts for vehicle body (orange): steel, composite, and thermoplastic

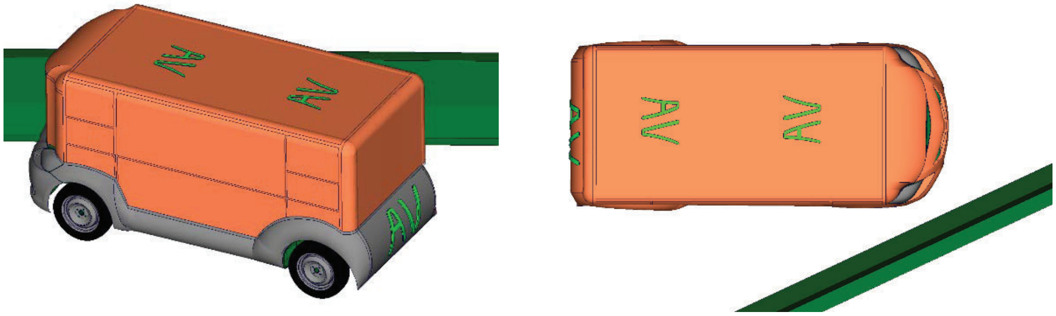


Figure 13. Mid-size ADS into NJB Impact Configuration

Simulations were conducted for all impact parameters and vehicle characteristics. For reference, an existing FE model representing a traditional sedan vehicle was evaluated, as shown in Figure 14. Vehicle mass and the height of the Center of Gravity (CG-z) are listed for the respective vehicles and material concepts.

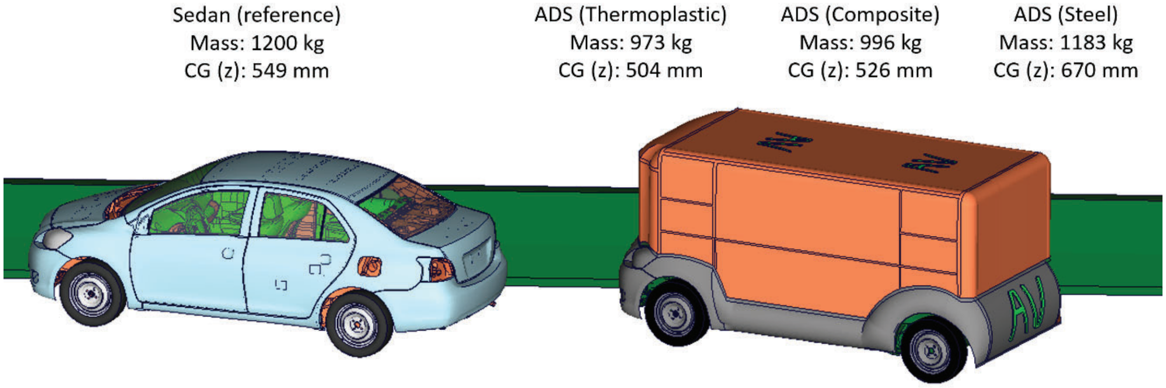


Figure 14. Comparison of traditional and mid-size ADS vehicle

Studies were conducted for impact velocities of 25 mph, 35 mph, and 45 mph of the mid-size ADSs into the NJB. For impacts at 25 mph, maximum roll angles well below the defined critical value of 40 degrees were observed for all cases. ADS with plastic or composite vehicle body tended to show marginally smaller roll angles compared to steel body for 25 mph impacts. See Figure 15.

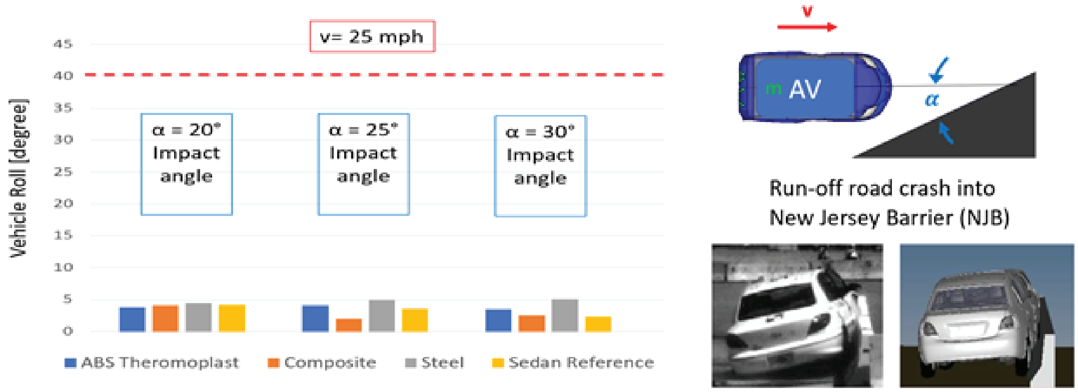


Figure 15. Mid-size ADS into NJB at 25 mph

For impacts at 35 mph, maximum roll angles below the defined criteria of 40 degrees were observed for all material concepts. ADS with a plastic or composite vehicle body showed clearly less critical roll angles compared to a steel body for all impact angles, especially for the 30° impact angle. See Figure 16.

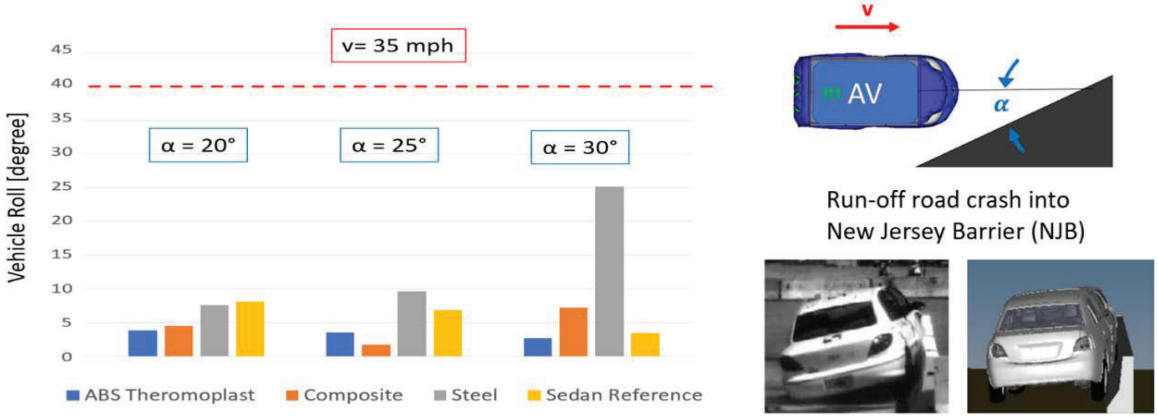


Figure 16. Simulation Study 1 – Mid-size ADS into NJB at 35 mph

For impacts at 45 mph, the ADS with a plastic or composite vehicle body again showed clearly less critical roll angles compared to a steel body for all impact angles, especially for the 30° impact angle. See Figure 17.

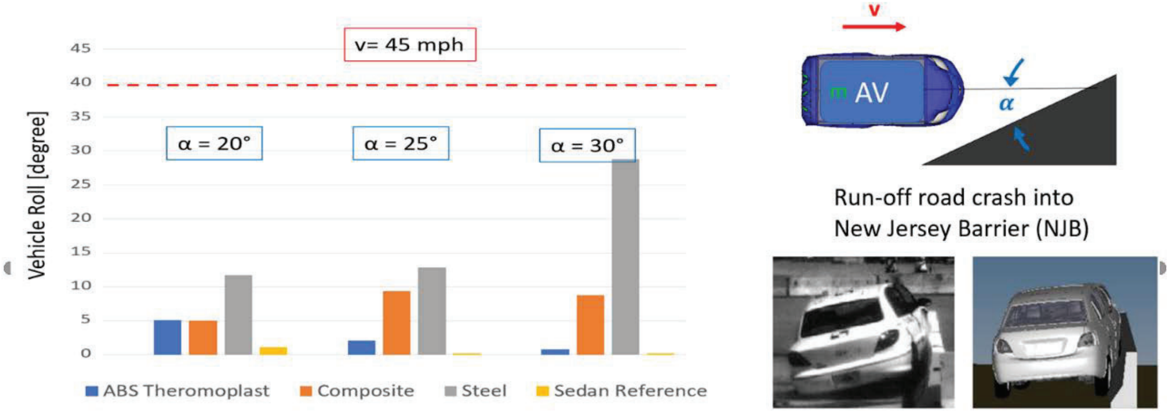


Figure 17. Simulation Study 1 – Mid-size ADS into NJB at 45 mph

Simulation Study 2 - Large-size ADS impacting New Jersey Barrier (NJB)
Figure 18 shows a front view of a Chevrolet Silverado reference vehicle impacting a NJB in simulation and full-scale test.

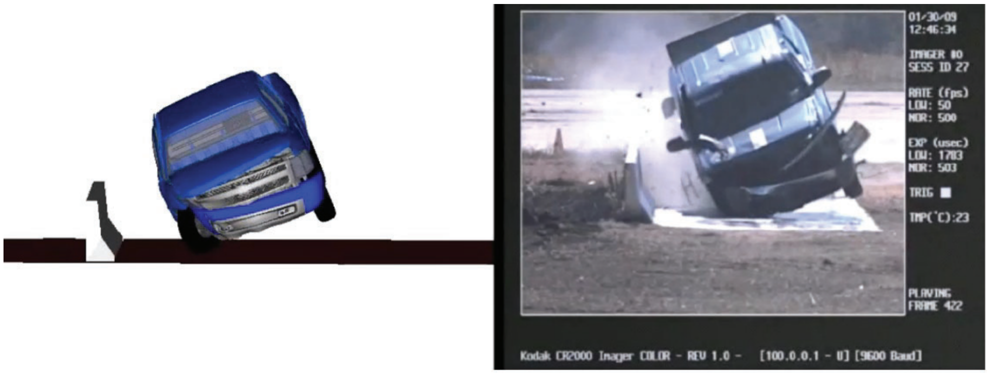


Figure 18. Large ADS Reference Vehicle impacting NJB (a) Simulation; (b) Test

Figure 19 illustrates studied impact parameters and vehicle characteristics used for the large ADS study.

- Range of impact angles: 20°, 25°, and 30°
- Impact speeds: 35 mph
- Material concepts for vehicle body: steel, composite, and thermoplastic

Figure 19. Large ADS with different impact angles

Table 2 summarizes the mass and height of the Center of Gravity (CG) for the SUV reference and the large ADS vehicles using different material concepts. Note that all ADS vehicles have a higher mass than the SUV reference vehicle due to their size with the steel version being the highest. Similarly, the CGs of the large ADS vehicles were higher than the CG of the Silverado reference SUV, with the steel version being the highest.

Table 2.
Large ADS: Mass and CG Comparison

	SUV (Reference)	**Large ADS Thermoplastic**	**Large ADS Composite**	**Large ADS Steel**
Mass [kg]	2,271	3,653	3,718	4,273
CG-z [mm]	732	764	789	975

The results for the simulation study 2 for impacts of the large ADS travelling at 35 mph into the New Jersey Barrier are shown in Figure 20. Maximum roll angles well below the defined criteria of 40 degrees were observed for all cases. ADS a with plastic or composite vehicle body showed significantly smaller roll angles compared to steel body.

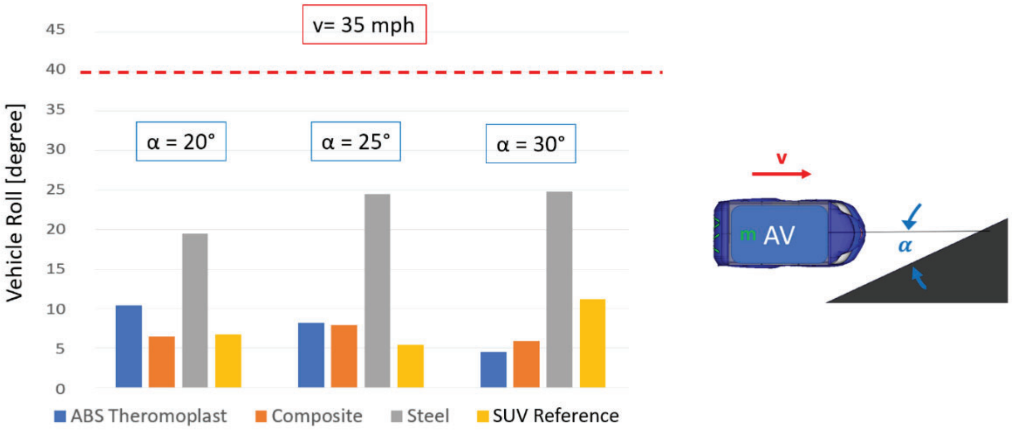


Figure 20. Simulation Study 2 – Large ADS into NJB

Simulation Study 3 – Mid-size ADS impacting NJB – Occupant Analysis
Vehicle pulses from the mid-size ADS vehicles with different material concepts impacting the NJB were recorded. The recorded vehicle pulses were then used to assess occupant injury risk with a previously developed generic sled model.[16] The generic sled model includes relevant interiors and restraints and allows the analysis of occupant injury risk, as shown in Figure 21. An ADS vehicle allowing manual or automated driving mode was assumed. The study was conducted for an occupant seated on the driver seat with a steering wheel and driver airbag present.

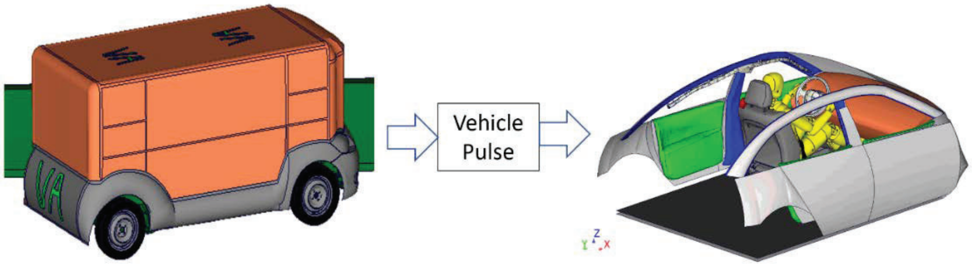


Figure 21. Evaluation Process of Occupant Responses using a Generic Sled Model

The generic sled model was developed to evaluate the effect of different vehicle pulses, seating orientations, and postures. It therefore assumes a seat-integrated restraints system. Dimensions, interior geometry, and package were adopted from a detailed finite element (FE) model of a 2014 Honda Accord[21]. The generic sled model exterior is represented by rigid parts, depicted in gray. Interior components with elastoplastic material and energy absorbing characteristics were used to allow for realistic occupant-to-vehicle interactions. This included a deformable floor, instrument panel, windshield, B-Pillar, and door trim components.

For the front-facing scenarios in the frontal impact scenario, interaction of the occupant with the seat, seatbelt, and airbag are dominant. In addition, interaction of the feet with the floor and the knees with the instrument panel affect occupant kinematics and loads.

---

[21] H. Singh, "Vehicle Interior and Restraints Modeling", EDAG Inc., Washington DC, 2017

https://www.nhtsa.gov/sites/nhtsa.gov/files/documents/812545_edagvehicleinteriorandrestraintsmodelingreport.pdf, accessed July 7, 2021

The developed generic occupied sled model was used to evaluate the effect of different vehicle pulses recorded from impacts with the mid-size ADS with different material concepts. The 35mph impacts of the mid-size ADS and the reference Toyota Yaris structural vehicles into the NJB at a 25-degree angle were selected, as highlighted by the green frame in Figure 22.

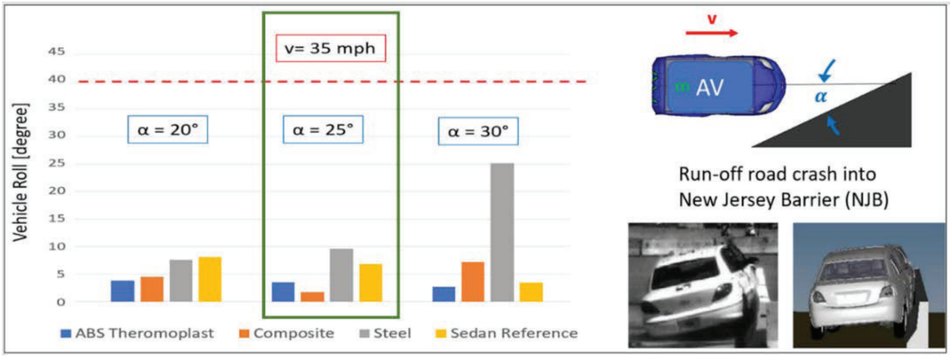


Figure 22. Simulation Study 2 – Mid-size ADS into NJB

In addition to the vehicle roll angle; vehicle pitch, vehicle yaw, and x-, y-, and z-pulses were recorded. The recorded motion was then applied to the previously developed generic sled model with a Hybrid 3 ATD on the driver seat, as shown in Figure 23.

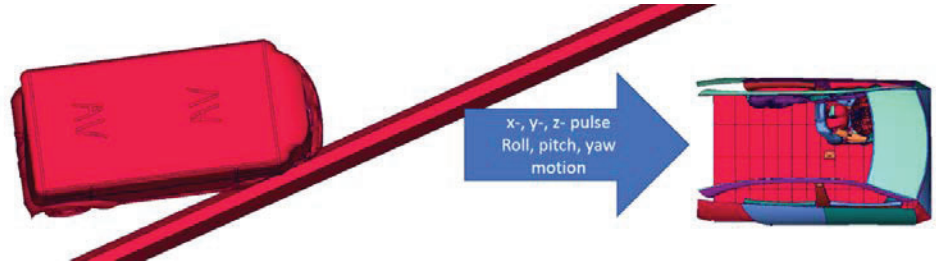


Figure 23. Mid-size occupant study: recorded motion from structural simulations applied to generic sled model

Figure 24 shows ATD and sled model kinematics 200ms after initial impact with the NJB for vehicle motion from ADS with steel body in gray and for vehicle motion from ADS with composite body shown in orange. Note that the lateral head trajectory tends to be marginally higher for the steel ADS motion compared to the composite ADS motion. Lateral head trajectory is relevant for far-side lateral impacts and is rated by EuroNCAP for side pole impact configurations, for example.

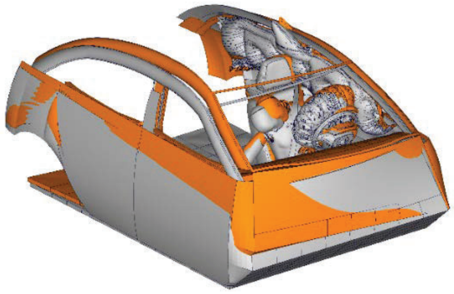


Figure 24. ATD and Sled Model Kinematics 200ms after Initial Impact with the NJB Example: Vehicle Motion from ADS with Steel Body (Gray) and Vehicle Motion from ADS with Composite Body (Orange)

Table 3 summarizes characteristic values of the Hybrid 3 ATD in the identical generic sled model interior environment with respective vehicle motions. Color coding is used to highlight the different material concepts studied. Reference values for the evaluated injury metrics, HIC, BRIC, chest deflection, and femur loads are shown in the second column. Results for the mid-size ADS motion with an ABS, composite, and steel body are highlighted in blue, orange, and gray, respectively. Results when applying pulses from the Toyota Yaris sedan vehicle are highlighted in yellow.

Table 3.
Hybrid 3 ATD characteristic values

	Reference	ABS	Composite	Steel	Sedan
HIC	1000	68	68	70	69
BRIC	1.05	0.68	0.67	0.67	0.69
Chest [mm]	22	8	9	9	9
Femur left [N]	8558	822	691	656	804
Femur right [N]	8558	1231	1116	1267	1273

Note that all injury values are well below the documented reference values. Only small differences for the ATD metrics were observed when applying the vehicle motions from the respective ADS and sedan reference vehicle impacts into the NJB.

Simulation Study 4 - Compatibility Assessment

Compatibility metrics for the mid-size ADS with different material concepts were evaluated. Figure 25 shows the results for the EuroNCAP MPDB compatibility configuration. The impact of the vehicle using thermoplastic material for select components and a mass 3,653 kg with the MPDB resulted in an OLC value of 18. The same vehicle with composite material for selected components and a mass of 3,718 kg showed an OLC value of 19, while an OLC value of 22 was recorded for the baseline vehicle using steel with a mass of 4,273 kg. Lower OLC values indicate better compatibility. Hence, the use of thermoplastic and composite material resulted in lower vehicle mass and better compatibility metrics.

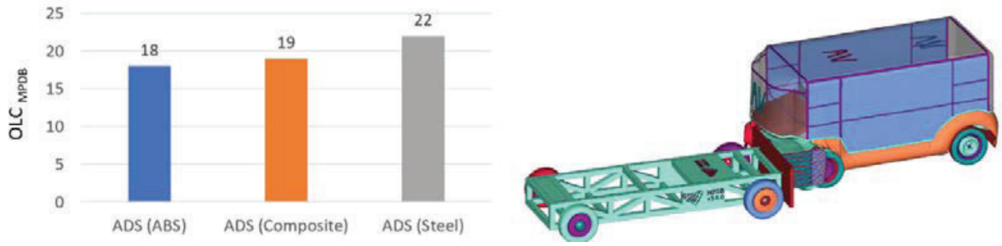


Figure 25. OLC compatibility metric for midsize ADS with different material concepts

ANCILLARY RESEARCH

In collaboration with Honda and OSU, material coupon specimen tests for a select composite material have been conducted. The validation and verification process using a previously developed LS-Dyna material model[22] and the generated test data is ongoing. Specimen test series included tension, compression, and shear loading conditions. Tests were conducted quasi-statically, at different rates, different hysteresis load patterns, and different temperatures. The material model will be made publicly available, once completed. Additional information regarding the material model development process is documented in Appendix 1.

In addition to adequate material, state-of-the-art vehicle, occupant, and restraint modelling, realistic representation of battery modules will be essential to ensure crashworthiness of electric vehicles. Hence, a lithium-ion battery pack capable of capturing electrochemical, electromagnetic, and thermal-mechanical effects is currently being developed. Additional detail can be found in Appendix 2. The planned application of using the developed battery pack in context of the conducted research is discussed in the next section.

---

[22] Tobias Achstetter, "Development of a Composite Material Shell-Element Model for Impact Applications," 2019, Dissertation, George Mason University

DISCUSSION

Effect of different material concepts

The use of different material concepts resulted in different total vehicle mass and CG location, which affected the vehicles'compatibility characteristics. For example, the large ADS vehicle using thermoplastic material for select components had a mass of 3,653 kg and a OLC of 18. The same vehicle with composite material concept had a mass of 3,718 kg and a OLC of 19, while the baseline vehicle using steel had a mass of 4,273 kg and an OLC of 22. Lower vehicle mass correlated with better partner-protection based on OLC metrics. Differences in OLC correlated with the difference in vehicle mass, which is considered small and can potentially be compensated by optimizing frontal vehicle structures. This is demonstrated by comparing differences in OLC and KW400 metrics for UADS vehicles with different structural characteristics from previous research[23], as shown in Figure 26.

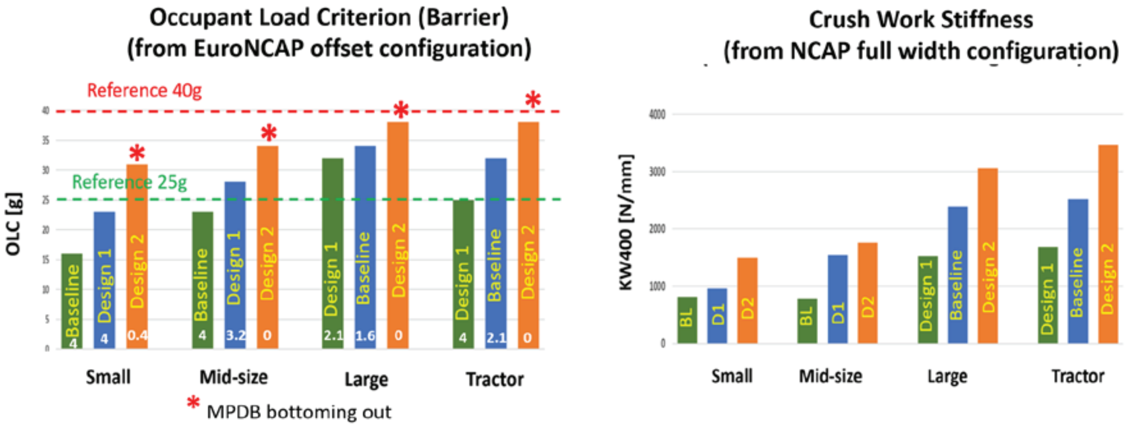


Figure 26. Summary of compatibility metrics for 4 UADS categories (a) OLC; (b) KW400

Two variations documented in the aforementioned previous research in addition to the baseline model for each of four UADS categories were developed by modifying the material strength and material thickness of relevant frontal structural components. Figure 26 (a) summarizes the compatibility characteristics for small, mid-size, large, and tractor UADS categories, based on the OLC, calculated from the MPDB barrier pulse in EuroNCAP's offset impact configuration. Reference lines representing an OLC of 25 g and 40 g are shown in green and red, respectively. In addition to the OLC, barrier bottoming out was evaluated. If bottoming out occurs, a 2-point penalty modifier applies. For example, if a vehicle generates an OLC of 32.5 g and barrier face bottoming out is observed, zero points would be given for the overall rating. The red asterisk on top of the design variations of the four UADS categories with the highest OLC values indicates that bottoming out was observed. The respective EuroNCAP score based on the OLC value and the bottoming out penalties is documented by the numbers, shown in white at the bottom of each bar. Figure 26 (b) summarizes the crush work stiffness (KW400) values for the respective baseline models and UADS variations, as calculated from the force versus deformation characteristics in the NCAP full overlap impact. Similar trends can be observed for the OLC and KW400 metrics.

Effect of Material Concepts on Risk of Roll-Over in NJB Impact

Lower vehicle mass and lower CG correlated with reduced risk of rollover, based on run-off road impact scenarios. Differenced in risk of rollover during a 30° NJB impact was observed for the large and mid-size ADS vehicles. Using thermoplastic or composite materials for select vehicle components of the large ADS resulted in maximum roll angles of 2° and 7°, respectively, versus 25° for the baseline steel version. Similarly, the reduction of maximum roll angle for the mid-size ADS is considered significant. Maximum recorded roll angle during a 45mph NJB impact was reduced from 28° to 1° and 9°, respectively, when using thermoplastic or composite materials compared to the

---

[23] Reichert R, Kan C-D, Park C-K, Crash Compatibility for Unoccupied Automated Driving Systems, NHTSA, 2022

baseline version using steel materials. In addition to using lightweight materials, optimizing suspension characteristics and CG location are considered important parameters influencing risk of rollover during run-off road crashes.

Use of Interior Sled Models in Combination with Recorded Vehicle Pulses

The documented methodology of using a generic sled model with interiors and restraints in combination with previously recorded vehicle pulses allowed for the study of the effect of different vehicle material concepts on occupant injury risk. Occupant metrics were mainly affected by the interior concept, which was identical for all three vehicles during the studied run-off road impact. Differences in occupant load was therefore small for all body regions and well below select reference criteria. HIC and BRIC head injury metrics were found to be similar when using different material concepts. Similarly, chest and lower extremity femur loads were considered to be of similar magnitude when using different material concepts. During the conducted simulation study, different maximum roll angles were observed, while no roll over occurred for any of the considered vehicles. Relevant differences in occupant injury risk can be expected if a roll over does occur.

The methodology of recording vehicle kinematics and applying relevant pulses to a generic interior sled model was found to be advantageous not only with respect to numerical efficiency. The technique allowed to understand the effect of different full vehicle structural kinematics while keeping all other occupant, interior, and restraint characteristics identical. Similar techniques were used for related research[24], where different seating concepts, seating arrangements, and interior concepts were studied for a variety of speed limited ADS shuttles for frontal, side, and rear impact crash scenarios. Examples are depicted in Figure 28.

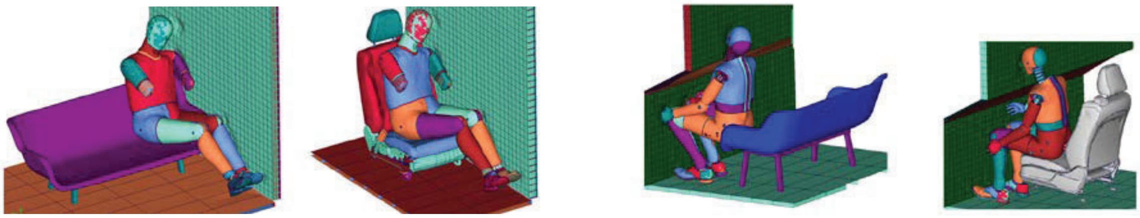


Figure 28. Examples of Generic Sled Application using Different Seating and Interior Concepts for (a) Side; (b) Frontal Impact Scenarios

Integrated Approach to Study Vehicle Material Concepts, Occupant Risk, and EV Safety

Ongoing research includes the combination of vehicle, material, occupant, and detailed battery pack FE models in an integrated approach. A FE model based on a Waymo self-driving vehicle is currently being equipped with relevant interiors, restraints, and battery pack components. The developed model allowed for the evaluation the effect of using different material concepts on structural deformation characteristics, occupant injury risk, and thermal run-away due to mechanical loads during a side pole impact, for example, as shown in Figure 29. Initial studies with a conventional vehicle were conducted to demonstrate the mass effect on vehicle intrusion, as shown in Figure 29 (d). Increased vehicle mass resulted in increased intrusion. Similar effects are expected when studying different material concepts in combination with a ADS vehicle. Consequently, the increased risk of battery pack damage can be anticipated, for increased structural intrusion, depending on the material concept used, for example.

---

[24] Reichert R, Kan C-D, Park C-K (2022, October). Crash safety considerations for speed-limited ADS shuttles (Report No. DOT HS 813 354). National Highway Traffic Safety Administration.

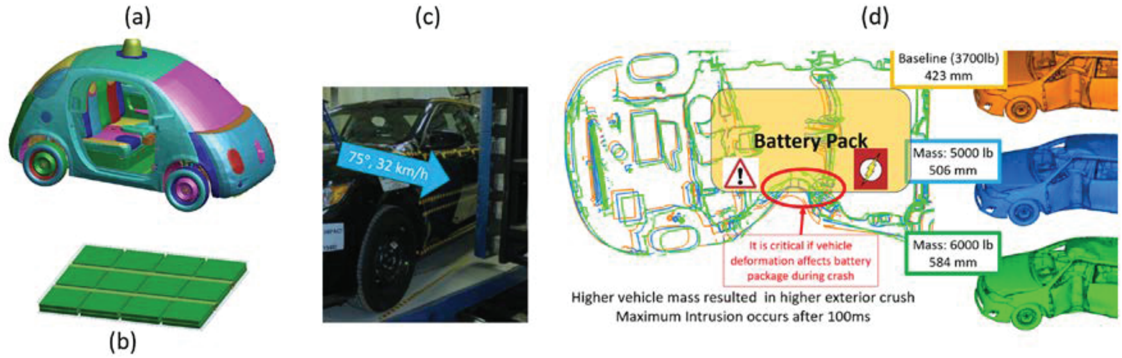


Figure 29. Waymo Self-driving Vehicle; (b) Multi-physics Battery Pack; (c) Side pole Impact; (d) Initial Study to Determine Mass Effect on Intrusion

Limitations

The same vehicle design and underlying structures were used during the study where standard steel materials were replaced with thermoplastic and composite materials, i.e. no optimization towards the respective material concept was performed.

CONCLUSION

The conducted research and the methods used demonstrate how simulation tools can contribute to assessing new type of ADS vehicle designs. Material concepts that resulted in a smaller vehicle mass tended to show better partner protection. Reduced vehicle mass and the corresponding lowered CG tended to reduce the risk of roll over. The interior concept, which was the same for three ADS vehicle variations studied, was the main factor for producing similar occupant injury metrics for the evaluated impact scenario.

ACKNOWLEDGEMENT

The authors wish to acknowledge the American Chemistry Council (ACC) for providing the main funding for this research. We also wish to acknowledge ACC’s Gina Oliver and Barb Robertson for their technical and administrative support for this project. We also wish to acknowledge the National Highway Traffic Safety Administration (NHTSA) for the funding of their related research mentioned in this paper. Finally, we wish to acknowledge Honda which provided the main funding for the finite element material model development.

APPENDIX 1

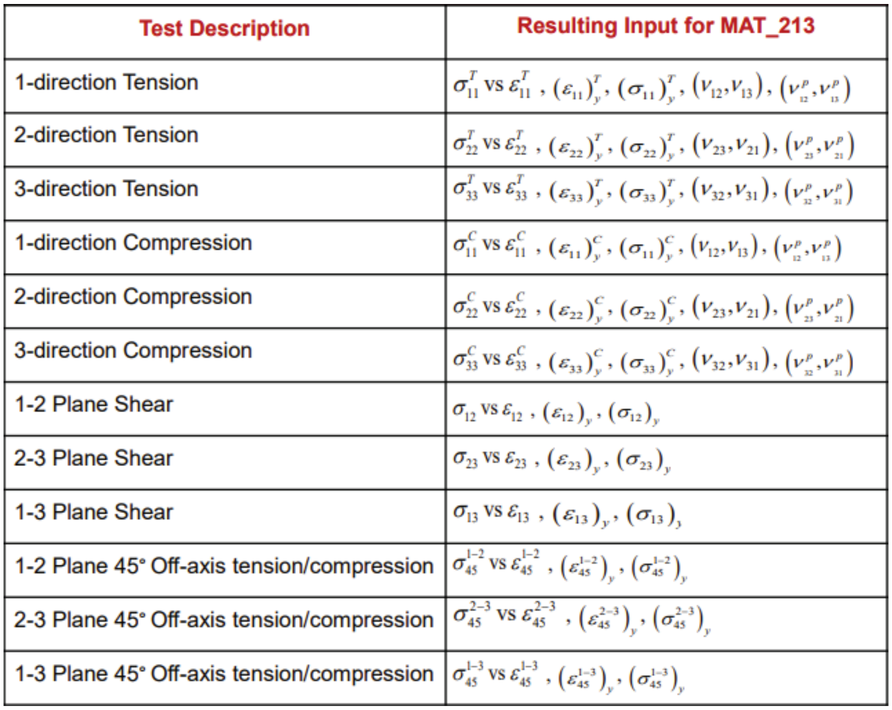
COMPOSITE MATERIAL MODEL DEVELOPMENT

Introduction

Since the predictability of the existing analytical composite material models was not satisfactory in the application of high-velocity impact simulations, a team of George Mason University, Arizona State University, Ohio State University, the National Aeronautics and Space Administration, and the Federal Aviation Administration have collaborated to develop a new composite material model in LS-DYNA for high-velocity impact simulations.

As a result, *MAT_213 in LS-DYNA has been developed. *MAT_213 is an orthotropic constitutive model developed with strain hardening, non-associated plasticity, Tsai-Wu yield surface, and the second Tsai-Wu surface as flow surface. It uses tabulated hardening curves based on experimental material test data. It has three failure modes, such as Tsai-Wu Failure Criterion, Puck Failure Criterion, and Generalized Tabulated Failure Criterion. A damage model accounts for decreasing strength/stiffness. It could simulate pure linear material behavior with tension-compression asymmetry. In addition, the model includes a stochastic analysis function. Initially, it was developed for solid elements. Currently, *MAT_213 (V1.3.5) is available in a developing version to AWG (Aerospace Working Group) and sponsors (Honda). Soon, it will be available in future version of LS-DYNA R13 to the public.

After developing and implementing *MAT_213 into LS-Dyna, a dataset for T800/F3900 composite was developed. A tabulated experimental input dataset was created by conducting a series of material tests and simulations to satisfy the input requirements for the three sub-models' deformation, damage, and failure. For the deformation model, twelve (12) material coupon tests at various strain-rate and temperature combinations are needed, as shown in Figure 30. For the damage and failure models, additional test series are needed.



Test Description	Resulting Input for MAT_213
1-direction Tension	σ_{11}^{T} vs ε_{11}^{T} , $(\varepsilon_{11})_{y}^{T}$, $(\sigma_{11})_{y}^{T}$, (ν_{12},ν_{13}), $(\nu_{12}^{p},\nu_{13}^{p})$
2-direction Tension	σ_{22}^{T} vs ε_{22}^{T} , $(\varepsilon_{22})_{y}^{T}$, $(\sigma_{22})_{y}^{T}$, (ν_{23},ν_{21}), $(\nu_{23}^{p},\nu_{21}^{p})$
3-direction Tension	σ_{33}^{T} vs ε_{33}^{T} , $(\varepsilon_{33})_{y}^{T}$, $(\sigma_{33})_{y}^{T}$, (ν_{32},ν_{31}), $(\nu_{32}^{p},\nu_{31}^{p})$
1-direction Compression	σ_{11}^{C} vs ε_{11}^{C} , $(\varepsilon_{11})_{y}^{C}$, $(\sigma_{11})_{y}^{C}$, (ν_{12},ν_{13}), $(\nu_{12}^{p},\nu_{13}^{p})$
2-direction Compression	σ_{22}^{C} vs ε_{22}^{C} , $(\varepsilon_{22})_{y}^{C}$, $(\sigma_{22})_{y}^{C}$, (ν_{23},ν_{21}), $(\nu_{23}^{p},\nu_{21}^{p})$
3-direction Compression	σ_{33}^{C} vs ε_{33}^{C} , $(\varepsilon_{33})_{y}^{C}$, $(\sigma_{33})_{y}^{C}$, (ν_{32},ν_{31}), $(\nu_{32}^{p},\nu_{31}^{p})$
1-2 Plane Shear	σ_{12} vs ε_{12} , $(\varepsilon_{12})_{y}$, $(\sigma_{12})_{y}$
2-3 Plane Shear	σ_{23} vs ε_{23} , $(\varepsilon_{23})_{y}$, $(\sigma_{23})_{y}$
1-3 Plane Shear	σ_{13} vs ε_{13} , $(\varepsilon_{13})_{y}$, $(\sigma_{13})_{y}$
1-2 Plane 45° Off-axis tension/compression	σ_{45}^{1-2} vs ε_{45}^{1-2} , $(\varepsilon_{45}^{1-2})_{y}$, $(\sigma_{45}^{1-2})_{y}$
2-3 Plane 45° Off-axis tension/compression	σ_{45}^{2-3} vs ε_{45}^{2-3} , $(\varepsilon_{45}^{2-3})_{y}$, $(\sigma_{45}^{2-3})_{y}$
1-3 Plane 45° Off-axis tension/compression	σ_{45}^{1-3} vs ε_{45}^{1-3} , $(\varepsilon_{45}^{1-3})_{y}$, $(\sigma_{45}^{1-3})_{y}$

Figure 30. Material coupon test series for MAT_213

Figure 31 shows an example of a material coupon specimen.

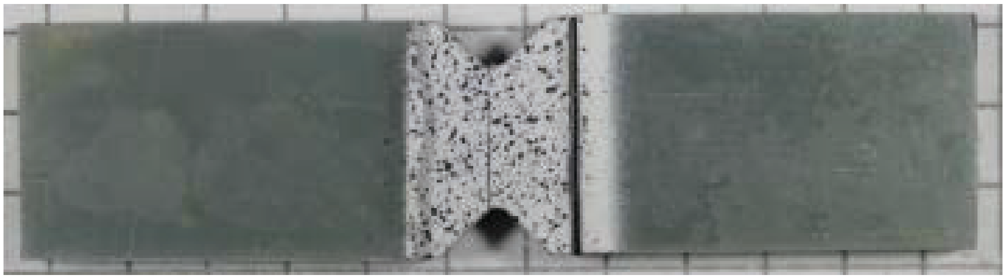


Figure 31. Material coupon specimen example

Honda R&D funded GMU to develop and validate the shell element version of *MAT_213 in LS- DYNA for automotive crash application in four phases. The first phase is to develop verify the code has been completed. Ongoing work during Phase 2 includes the characterization of a material law for a composite material based on coupon testing. During Phase 3 a comparison of different discretization techniques when used in conjunction with *MAT_213. Finally, validation of the material model based on component testing will be conducted during phase 4.

Composite material model validation

Ballistic impact conditions were selected for quality of material data and to validate robustness of the material model. Physical tests were conducted at NASA-GRC using a 50g projectile and unidirectional T800/F3900. Sixteen (16) fully integrated elements through the thickness and cohesive elements between all layers were modeled. The model consisted of ~400,000 shell elements and ~300,000 solid cohesive elements.

LS-DYNA material model *MAT_213 shell routine development work was previously conducted by (Achstetter, 2019) [25] , which included:

Develop and implement plane stress plasticity algorithm by stress projection
Develop and implement orthotropic VE-VP algorithm with non-linear visco-elasticity.
Develop and implement strain rate smoothing algorithm
Update tabulated failure model
Develop and implement stochastic option
Improve robustness of plasticity algorithm (radial return 'backup' solution)

The developed material model was used to simulate three cases from a test series with different impact and rebound/exit velocities, as shown in Figure 32.

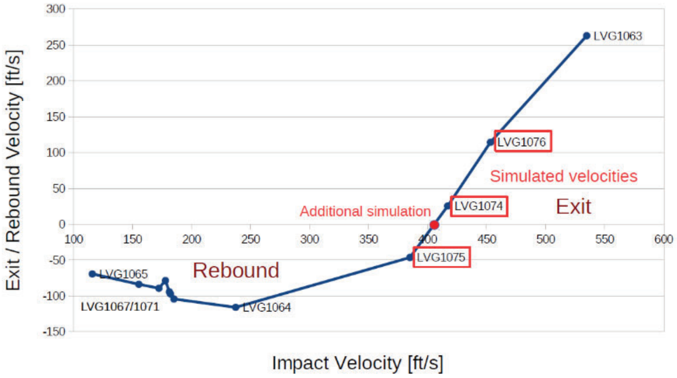


Figure 32. Exit / rebound velocity versus impact velocity test series and select simulations

Delamination characteristics were well captured as shown in Figure 33.

---

[25] Tobias Achstetter, "Development of a Composite Material Shell-Element Model for Impact Applications," 2019, Dissertation, George Mason University

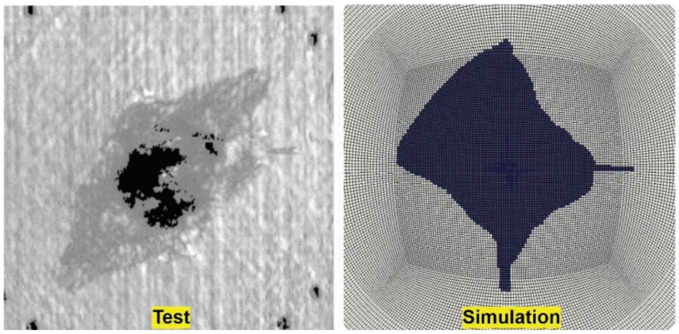


Figure 33. Comparison of delamination in test (LVG1075) and respective simulation

Overall good correlation when comparing the composite fracture and delamination between test and simulation was observed, as shown in Figure 34.

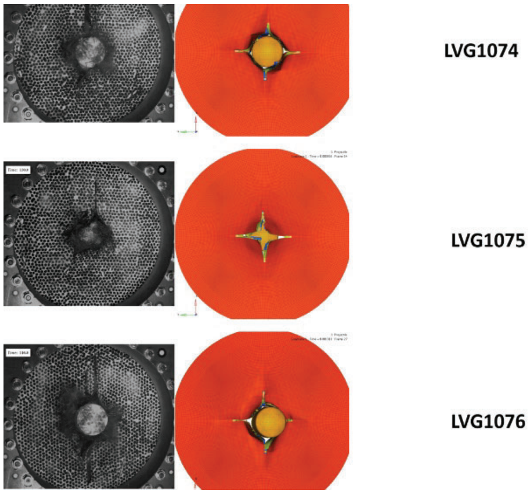


Figure 34. Comparison of test and simulation for different impact velocities

C-Channel Crush

Honda conducted C-Channel crush tests using composite layup [45/-45/0/90]. Two modeling approaches were applied. In approach 1, one element through the thickness and 8 integration points were defined using *MAT_58. In approach 2, one element per ply and 2 integration points were specified using *MAT_213. It was found that MAT_213 compares better with the test results showing closer failure behavior and more localized stress concentration at the crush area, as shown in Figure 35.

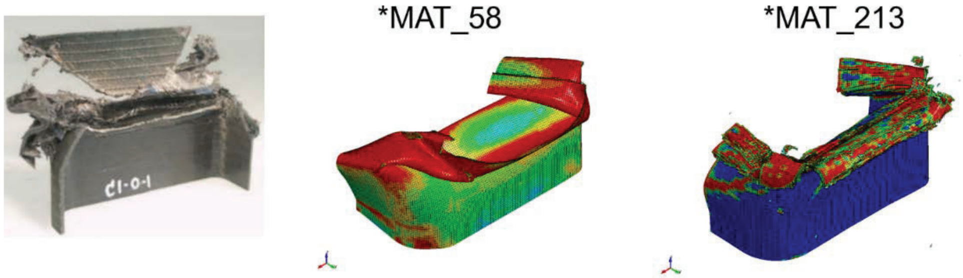


Figure 35. C-Channel Crush (a) Test; (b) Mat_58; (c) Mat_213

APPENDIX 2

BATTERY MODELLING

Lithium Battery Background

The Nobel Prize in Chemistry 2019 was awarded jointly to John B. Goodenough, M. Stanley Whittingham, and Akira Yoshino for the development of lithium-ion batteries. The history of lithium-ion battery can be dated back to 1970s. Whittingham created the first working, rechargeable metallic Lithium battery in 1976. In 1979/1980, John B. Goodenough discovered that Lithium cobalt oxide could serve as a cathode material in a “ion transfer cell” configuration. In 1985, Akira Yoshino identified that certain qualities of petroleum coke can be used as anode material and created the first commercialized lithium-ion battery, which is the Sony rechargeable battery. Since then, lithium-ion battery technology has rapidly progressed. For example, modem lithium-ion batteries are composed by multiple layers of thin membranes. A typical cell consists of a layer of aluminum coated with Lithium manganese oxide, a polyethylene separator, and a layer of copper coated with graphite. All these membranes are very thin, and they are rolled up as to form either a cylinder shape or a porch shape. Additional protective circuit is often necessary for safe operation.

Lithium-ion battery’s advantages are high energy density, high open-circuit voltage, low internal resistance, long cycle life, and quick charging. The disadvantage is the safety issues. Fire or explosion may occur when overcharge occurs, when exposed to high temperature, or when short circuit or puncture occurs. Examples include fire in some Tesla electric cars, initiated from the battery after impact. Lithium-ion battery fires are also reported on cellphones, drones, and electric bicycles and scooters.

The development of reliable and robust battery models for simulating their behavior in crashes involving electric vehicles (EV) is needed. To address safety related issues, GMU has started to develop a lithium-ion battery FE model that can predict battery failure in EV Crash impacts and other high velocity impact applications. The battery model method adopted is aimed to allow computationally efficient simulations for engineering applications.

Lithium Battery Modeling Approach

To capture the physics of a lithium-ion battery cell in a fully coupled analysis, three aspects must be addressed:

Electrochemical
Electromagnetic
Thermal-mechanical

Different battery cell verification tests allow validation of a battery cell FE model. They may include the three-point bending test to validate mechanical behavior, a battery circuit test to verify electrical performance, and a punch test at high and low temperatures to validate combined mechanical and thermal behavior. Figure 36 shows promising initial results of a hemispherical punch test and a simulation. The force versus displacement comparison depicts reasonably good correlation using material properties and test results from literature.

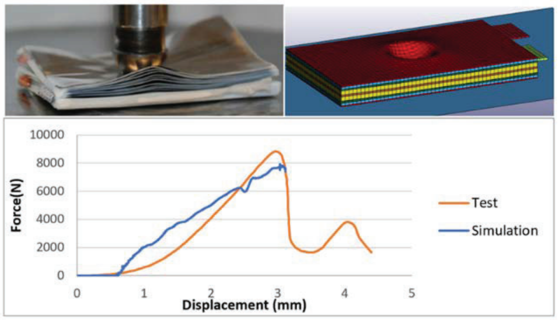


Figure 36. Physical punch test (upper left); simulation result (upper right); and test versus simulation force-displacement comparison (bottom)

To verify combined thermal and electrical predictability capabilities of a FE model, the external short circuit test can be used. Figure 37 shows an example of the electrical current density distribution for the preliminary FE model.

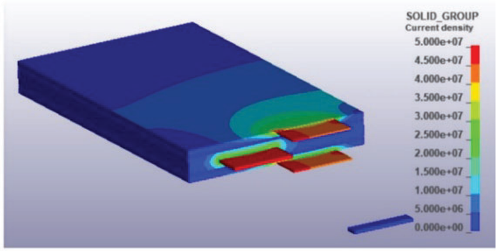


Figure 37. Electric current after external short circuit

Figure 38 shows the volumetric heat power during an external short circuit using the same FE model.

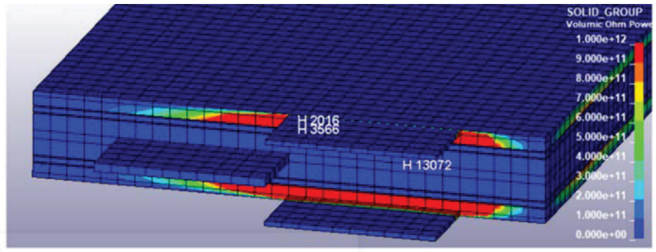


Figure 38. Volumetric heat power during an external short circuit

Figure 39 shows a so-called coated cathode material test, where the material coupon is placed in a thermal chamber to generate data for a separator test with elevated temperatures. This setup is currently being evaluated.

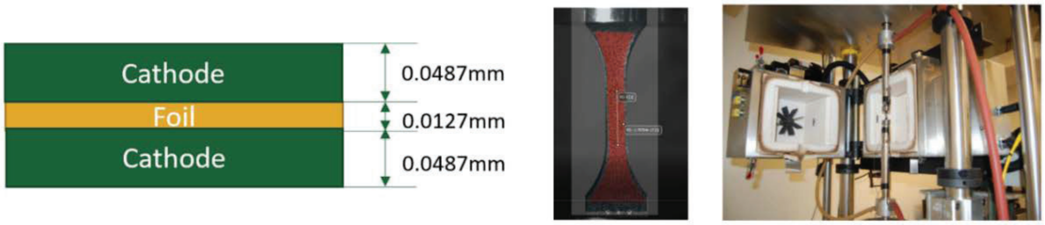
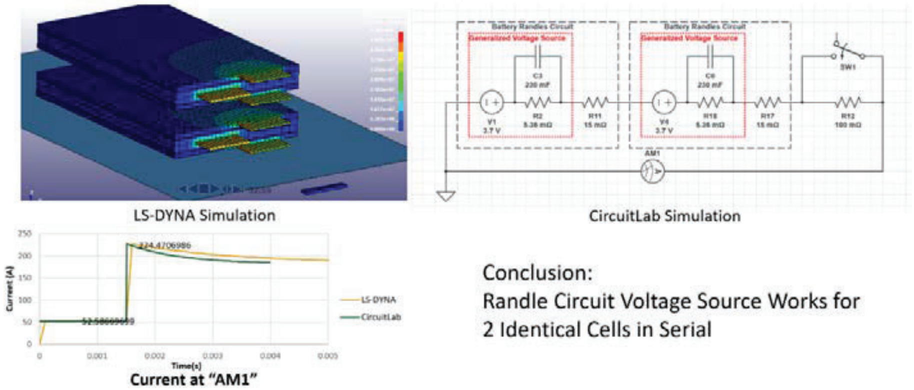


Figure 39. Volumetric heat power during an external short circuit

Lithium battery modeling from cell to module level

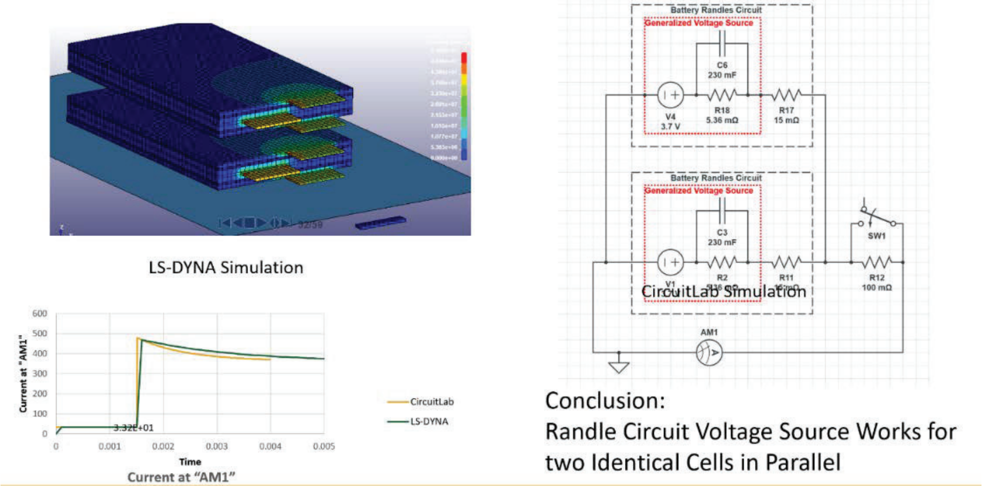
Modeling multi physics of an individual lithium battery cell is well understood and can be considered state-of-the-art. Significant progress of modeling two battery cells in series has been made, as shown in Figure 40. Promising results when comparing a two cell "battery pack" with respective test results were achieved. A so-called Randle's circuit voltage source was used.



Conclusion:
Randle Circuit Voltage Source Works for
2 Identical Cells in Serial

Figure 40. Randle's circuit voltage source works for 2 DIFFERENT cells in serial

Similar to the two cells in serial evaluation, a two cell in parallel configuration was studied, as shown in Figure 41.



Conclusion:
Randle Circuit Voltage Source Works for two Identical Cells in Parallel

Figure 41. Randle's circuit voltage source works for 2 DIFFERENT cells in parallel

Being able to combine two cells makes scalability and therefore modeling of a detailed battery pack with multiple individual cells realistic. The findings have been published in a journal paper by (Wang et. al., 2022).[26]

Lithium Battery Modeling Outlook

The development and validation of a lithium battery pack is ongoing. The selected modelling technique will allow to evaluate electric vehicles during a crash event and provide a step-by-step approach to capture the thermal runaway condition, for example.

---

[26] Leyu Wang, Chenxi Ling, Cing-Dao Kan & Chi Yang. 2022. "A coupled thermal electrical mechanical analysis for lithium-ion battery." Journal of Micromechanics and Molecular Physics. https://www.worldscientific.com/doi/abs/10.1142/S2424913021420108?journalCode=jmmp

SAFETY GRADING SCHEME FOR HIGHWAY ASSISTED DRIVING TECHNOLOGY

Dominic Tough
Colin Grover
Matthew Avery
Thatcham Research
United Kingdom

Paper Number 23-0274

ABSTRACT

There is increasing availability of assisted driving technology on vehicles and manufacturers are developing innovative features and functionality. When the driver engages assisted driving technology, the vehicle supports the driver with the steering and speed control, however the driver retains the responsibility for the safe driving. Assisted driving offers the potential safety benefits of improved speed and headway regulation and lane guidance, addressing the most common front-to-rear crash type and lane drifting and run-off road crashes. Because the systems relieve some of the driving workload, fatigue is also addressed. However, system implementation must be carefully considered to ensure the driver remains engaged with the driving task.

Assisted driving systems are implemented differently by individual vehicle manufactures. The objective of this research was to identify the key features that lead to safe implementation of assisted driving technology, enabling the development of a consumer safety grading scheme to guide vehicle manufacturers to safe implementation and provide an independent, objective means of assessing systems.

Vehicle Assistance and Driver Engagement were identified as the two critical aspects. The level of assistance provided must be matched by the perception of the driver and the ability of the system to keep the driver engaged. Vehicle Assistance assesses the steering support technology and the selection and application of appropriate speed control. Driver Engagement assesses driving collaboration, driver monitoring and system status in use, and also the consumer information including how the system is named, marketed and its appropriate usage described.

A third key area identified for safe implementation was Safety Backup, namely the advanced emergency support the system provides in case of an imminent collision beyond the capability of the assistance, in case of an unresponsive driver who becomes disengaged for a long period, or a system failure.

The research was implemented by developing test and assessment protocols in association with Euro NCAP acknowledging the results of broad range of research vehicle testing. A four-tier grading scheme was developed (Entry, Moderate, Good and Very Good) ranking vehicles on the sum of Assistance Competence (balancing Vehicle Assistance and Driver Engagement) and Safety Backup.

To date, 21 vehicles have been assessed and a range of results have been achieved that span across the four grades, indicating the relevance of the scheme and its ability to differentiate systems. The scheme has identified an apparent imbalance between Vehicle Assistance and Driver Engagement in one case. In another, a vehicle has been reassessed and gained an improved grading after an over-the-air update.

A limitation of the grading scheme is it is currently focused on highway functionality, whereas assisted driving technology can be utilised by the driver wherever the system deems it is capable of operating. In this first iteration of the grading scheme, only interaction on highway-like roads with other restricted vehicle types has been considered. Expanding the assessment beyond highway usage will necessarily involve assisted driving relevant interactions with a broader range of road types and features, traffic control and road users etc.

INTRODUCTION

Assisted driving technology can provide safety and comfort benefits to the driver by ensuring safe driving practice is observed. This leads to a reduction in the frequency of potential critical collision avoidance situations, therefore minimising the risk of common car to car crash types. For safe driving to be achieved, a balance is required between the level of assistance that the vehicle provides and keeping the driver engaged with the driving task through the systems interaction with the driver. "*Automation needs to be designed either so that it does not rely on the driver or so that the driver unmistakably understands that it is an assistance system that needs an active driver to lead and share control*". [1]

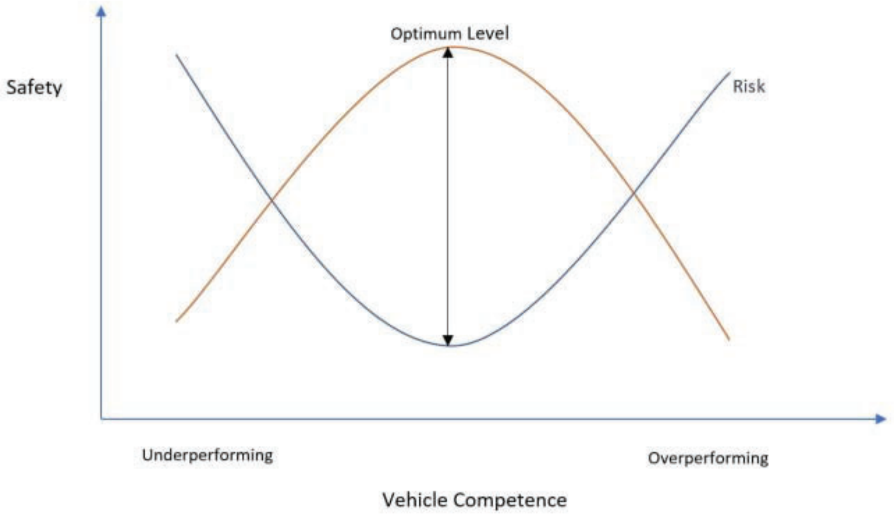


Figure 1 Assisted Driving - safety against vehicle competence.

Each Vehicle Manufacture has their own philosophy in both the way they implement their assisted driving system, which in turn affects the level of driver engagement, but also the technical capability of the system. To give the consumer a greater understanding of the capability and safety of the system, an independent grading system is needed to objectively measure each individual system accurately and fairly.

RESEARCH QUESTION/OBJECTIVE

There is an increasing number of vehicles giving the option of fitting assisted driving technology and manufacturers are developing innovative features and functionality to support the driver with steering and speed control.

Assisted driving offers the potential safety benefits of improved speed and headway regulation and lane guidance, addressing the most common crash types. Because systems relieve some of the driving workload, fatigue is also addressed on longer journeys which could potentially keep the driver from entering a critical accident scenario. However, system implementation must be carefully considered to ensure the driver remains engaged with the driving task. A balance should be achieved between the amount of vehicle competence and the level of driver engagement

The objective of this research was to identify the key features that lead to safe implementation of assisted driving technology, enabling the development of a consumer safety grading scheme to guide vehicle manufacturers to safe implementation and provide an independent, objective means of assessing systems.

METHODS AND DATA SOURCES

Building on previous work completed in 2018 [2] Vehicle Assistance and Driver Engagement were identified as two critical aspects for the grading scheme. In order for the system to be deemed balanced, the level of assistance provided must be matched by the driver perception and the ability of the system to keep the driver engaged. Vehicle Assistance assesses the steering support and the selection and application of appropriate speed control. Driver Engagement assesses driving collaboration, driver monitoring and system status, and also the consumer information including how the system is named, marketed and its appropriate usage described.

Assistance Competence - Vehicle Assistance

Vehicle Assistance focuses on the technical capability of the system as a whole. This section involves test scenarios to assess both the longitudinal (Adaptive Cruise Control) and lateral capability (lane centering) of the system. The Vehicle Assistance assessment consists of three elements:

- Speed Assistance
- Adaptive Cruise Control (ACC) Performance
- Steering Assistance

The Speed Assistance builds on the already well-established Speed Assist Systems (SAS) [3] assessment from the Euro NCAP 5-star safety rating scheme. Additional points are awarded to ACC systems which respond to the road environment, through the GPS navigation and windscreen mounted camera. Points are awarded for automatically reducing the vehicle speed prior to a slip road, sharp curve, roundabout, and adjusting the set ACC speed to both fixed and temporary speed signs.

Adaptive Cruise Control (ACC) Performance uses highway based car-to-car scenarios identified in previous work [2] with scoring based on the vehicle performance at each test speed. Maximum points awarded for full avoidance, half points awarded for mitigating impact, reducing impact by more than 5km/h, quarter points awarded for producing a Forward Collision Warning (FCW) 1.5s prior to impact. This part of the assessment takes into consideration only the capability of the ACC system, this is defined where braking levels stay below approximately $-5 m/s^2$ or where it is confirmed that Autonomous Emergency Braking (AEB) did not intervene.

Approaching a stationary car

Approaching a slower moving car

Approaching a braking car

Car cutting-in in front

Car cutting-out in front

Figure 2 ACC performance car-to-car scenarios.

Steering Assistance assess the competency of the lateral support part of the system. A steering assistance function should support the driver to keep the vehicle in lane, not only on straight roads. If a car departs from its lane there is an increased risk of collision. It is not expected that vehicles are able to stay in the centre of the lane in all road corners, but expects the vehicle to always support the driver by directing the vehicle to the correct heading. Tests for the Steering Assistance are conducted in a so-called S-Bend at three different speeds, 80km/h, 100km/h & 120km/h.

Assistance Competence - Driver Engagement
Driver Engagement is the opposing component of the Assistance Competence balancing concept. To ensure safe driving is maintained throughout the use of the assisted driving system, the driver needs have a sufficient level of engagement with the road environment ahead. The level of Vehicle Assistance needs to equal the level of Driver Engagement. The level of Driver Engagement can be affected by both the pre-determined perception of the systems capability and the feedback from the vehicle whilst driving with the assisted system activated. The Driver Engagement assessment consists of four elements:

- Consumer Information
- System Status
- Driver Monitoring
- Driving Collaboration

Consumer Information looks at the consumers expectations of how much assistance the system will provide them, this expectation will be influenced by information they are subjected to before operating the system. An example of this is marketing on the vehicle manufacturers website and detailed information within the operations handbook. It should always be clear to any potential consumer that the system is an assistance system only and that driver oversight is always required. This assessment is designed to examine the information supplied to the consumer relating to the assistance system.

The System Status assessment is designed to evaluate the information supplied to the driver on a continuous basis, confirming the level of driving assistance being provided by the system. This is anticipated to be visual information only. This assessment is also designed to evaluate the information supplied to the driver in case the level of assistance by the system changes. This is commonly provided as visual, audible and/or haptic information or warnings.

Figure 3 shows an example of the System Status, indicating the active lateral support via a green steering wheel and the longitudinal support via the display of a lead vehicle.

Figure 3 Volvo's Pilot Assist system status [4]

Driver Monitoring technology at the time of developing the grading scheme was limited to indirect monitoring. The most frequently adopted form on indirect monitoring is a detected torque threshold in the steering column to indicate through steering inputs that the driver's hand(s) were on the wheel. Another technology used is touch capacitive sensors within the steering wheel to detect physical pressure on the wheel. Due to the current limited advances within this field and the industry knowledge of Direct Driver Monitoring systems, such as in cabin infra-red cameras to measure the drivers' eyes gaze, the maximum points awarded for driver monitoring was limited to 15/30 until future iterations of the grading scheme.

Driver Collaboration was found to be a key dynamic indicator to the driver of the systems capability whilst the system was active. It evaluates the resistive torque of the steering support system as the driver is interacting with the steering wheel. A full sine wave of steering angle to the vehicle steering wheel, with an amplitude of 5 degrees and frequency of 0.25Hz is applied, this allows the measurement of the resistive torque both with the system on and off whilst keeping the vehicle in the lane. Points are awarded based on the percentage difference in torque measurements with the system on and off. Using the difference instead of the absolute value allows manufactures the freedom to implement heavy or light steering when the system is not active.

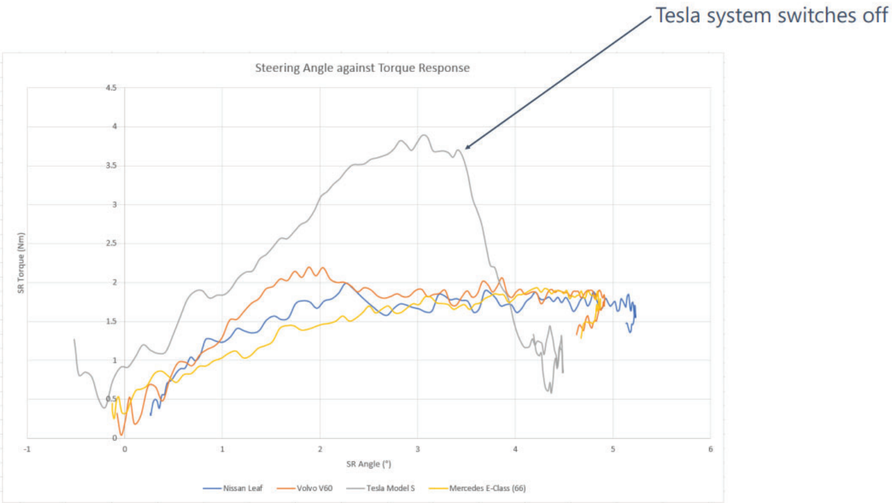


Figure 4 shows early research into steering support torque measurements of different vehicles.

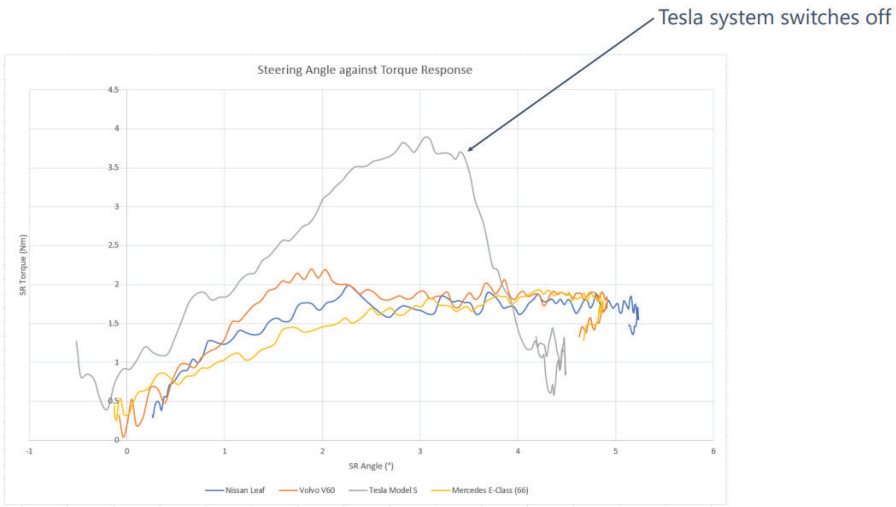


Figure 4 Assisted driving steering torque measurements

Safety Backup

A third key area identified for safe implementation was Safety Backup, namely the advanced emergency support provided in case of an imminent collision beyond the assistance capability, in case of an unresponsive driver or a system failure. The Safety Backup assessment consists of three elements:

- System Failure
- Unresponsive Driver Intervention
- Collision Avoidance

System Failure corresponds to the vehicles response to blockage of the sensors responsible for the lateral and longitudinal control of the systems. Typically, a forward-facing radar for the longitudinal aspect and a windscreen mounted camera for the lateral aspect. A common example is the front radar becoming blocked with snow whilst parked overnight. For a vehicle to score highly it will de-active the relevant system within two minutes of becoming blocked. The driver needs to be informed with a visual warning within the instrument cluster within five minutes to score additional points. It is known that complete blockage of these sensors is not a common scenario and therefore future research will investigate partial sensor blockage, increase blockage overtime, and sensor degradation.

Unresponsive Driver Intervention builds on the basic UN ECE regulation no. 79 [4] for the vehicle escalation when "hands off" the steering wheel is detected. The grading scheme awards points if after the detection of an unresponsive driver the steering support continues whilst safely bringing the vehicle to a controlled stop. Current regulation of steering support systems requires the vehicle stay within its lane however, future amendments will allow the vehicle to automatically (if deemed safe) move out of the lane and stop on the hard shoulder or inside lane of a multi-laned highway. This will be rewarded in future developments of the Assisted Driving grading scheme.

Collision Avoidance uses the same car-to-car scenarios found in the ACC performance section. At this stage, the system is only being assessed for performance when driving on a highway, therefore only car-to-car performance is assessed. In this assessment "Collision Avoidance" the capability of the vehicle to avoid a collision using both assisted driving systems and emergency systems combined is assessed.

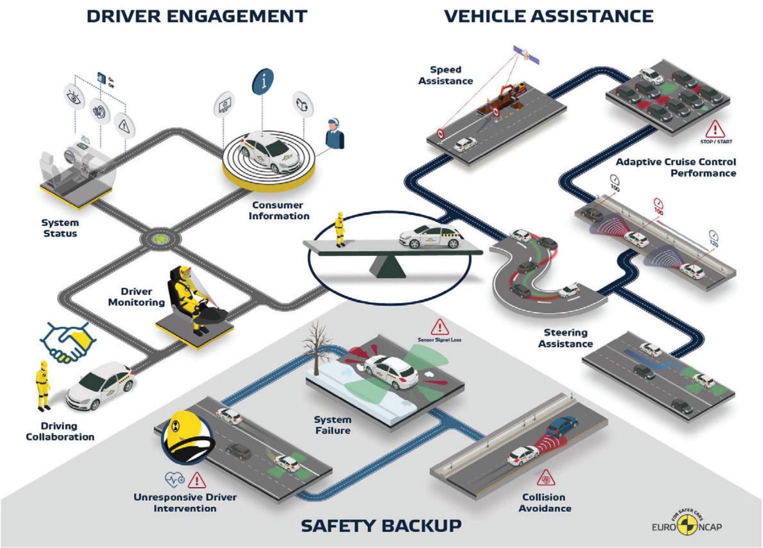


Figure 5 Euro NCAP Assisted driving grading scheme breakdown.

Grading Scheme

The research was implemented by developing test and assessment protocols in association with Euro NCAP acknowledging the results of a broad range of research vehicle testing. A four-tier grading scheme was developed (Entry, Moderate, Good and Very Good) ranking vehicles on the sum of Assistance Competence (balancing Vehicle Assistance and Driver Engagement) and Safety Backup. The naming for the four tiers of grades was agreed through Euro NCAP and industry members such that the wording was applicable throughout different European languages. Entry is given to the most basic of systems in acknowledgement that assisted driving systems are an optional extra and the additional safety benefit given to the consumer by these basic systems is not to be discouraged.

The Assistance Competence score is the balance between Vehicle Assistance and Driver Engagement. The higher the level of assistance, the more the driver must be engaged by the system. In principle, the Assistance Competence score equals the Vehicle Assistance score, but only when the Driver Engagement score (at least) matches Vehicle Assistance. Where Vehicle Assistance outscores Driver Engagement, the Assistance Competence score is limited to the Driver Engagement performance.

A total of 200 points is available to each system with a maximum of 100 points available for Assistance Competence and a maximum of 100 for Safety Backup. The sum of Assistance Competence and Safety backup determines the Grading.

The breakdown of the grading boundaries are set to the following:

- ≥ 100 Entry
- ≥ 120 Moderate
- ≥ 140 Good
- ≥ 160 Very Good

The complete assisted driving assessment and detailed classification of points allocation can be found on the Euro NCAP website: Euro NCAP AD Test and Assessment Protocol v1.1 [6]

RESULTS

Table 1. Euro NCAP Assisted Driving Results to Date

Vehicle	Year Tested	Score /200	Grading
Mercedes-EQ EQE	2022	185	Very Good
Mercedes-Benz GLE	2020	174	Very Good
BMW 3 Series	2020	172	Very Good
BMW iX3	2021	169	Very Good
Nissan Qashqai	2022	167	Very Good
Audi Q8	2020	162	Very Good
VW ID.5	2022	161	Very Good
Ford Kuga	2020	152	Good
Ford Mustang Mach-E	2021	152	Good
Cupra Formentor	2021	144	Good
Polestar 2 (OTA Update)	2022	141	Good
VW Passat	2020	137	Good
Hyundai IONIQ 5	2021	137	Moderate
Tesla Model 3	2020	135	Moderate
Nissan Juke	2020	134	Moderate
Volvo V60	2020	121	Moderate
Jaguar I-Pace	2022	112	Entry
Toyota Yaris	2021	109	Entry
Renault Clio	2020	105	Entry
Opel Mokka-e	2021	101	Entry
Peugeot 2008	2020	101	Entry

To date 21 vehicles have been assessed and a range of results achieved that span across the four gradings as shown in Table 1. Euro NCAP Assisted Driving Results to Date. Most systems have been appropriately balanced albeit of differing assistance capability.

Overall, there is a range in performance between each system, showing both the qualities and limitations of each assisted driving system. The results show that “very good” gradings are not limited to high end expensive vehicles but can be achieved in affordable family vehicles such as the Nissan Qashqai and the Ford Kuga.

In the short time the assisted driving protocol has been in place, it is evident that there is an incentive for the vehicle manufacturers to improve their system to achieve a higher score and ultimately improve the grading of the vehicle. One example is the Nissan Juke which was tested as part of the first ever set of vehicles to be assessed using this protocol in 2020 where it achieved a 134/200 score This gave it a Moderate rating due to having a balanced system but was limited in some technical capability. From the latest set of results in 2022 Nissan improved the technical capability of the system for example, adding ACC functionality for junctions and roundabouts, whilst still retaining a high level of driver engagement. This improvement resulted in an increased score of 167/200 and a Very Good rating, a similar level to high end luxury vehicles.

Other examples of improvement to vehicle performance have been achieved through the use of Over The Air (OTA) updates. One manufacture improved the steering component of their assisted driving system to allow the vehicle to cope with complex road layouts which upgraded their rating from and Moderate to Good.

The scheme identified an apparent imbalance between Vehicle Assistance and Driver Engagement in one case.

Detailed datasheets of every vehicle can be found on the Euro NCAP website: Euro NCAP | Assisted Driving Gradings

DISCUSSION AND LIMITATIONS

The scheme yielded a range of safety grading results which indicates its relevance and ability to differentiate systems. The grading scheme highlights the capability and limitations of each system to the consumer to give them greater understanding of the functionality of the system such that they do not overestimate the system's ability on the road.

During the limited time the grading scheme has been implemented, it has been evident that vehicle manufactures are developing their systems to further improve both the level of Assistance Competency but also balancing with appropriate levels of Driver Engagement. This should bring about the additional safety benefits which assisted driving systems can provide and subsequently reduce the frequency or severity of accidents on roads.

A limitation of the grading scheme is that it is currently focused on highway functionality only, whereas assisted driving technology can be utilised by the driver wherever the system deems it is capable of operating. In this first iteration of the grading scheme, only interaction on highway-like roads with other restricted vehicle types has been considered. Expanding the assessment beyond highway usage will potentially involve assisted driving relevant interactions with a broader range of road types and features, traffic control and road users etc. Therefore, further research is required to identify future test scenarios and the use of additional road users to further expand the grading scheme beyond highway usage only.

CONCLUSIONS

Assisted driving systems can provide road safety benefits and reduce the number of vehicle accidents on the road. Each system is implemented differently by individual vehicle manufacturers and the independent grading scheme successfully differentiates those systems offering the essential elements from those incorporating more advanced features, acknowledging the necessary balance to achieve safe adoption. This drives vehicle manufacturers towards implementing safe systems and supports consumers making safer choices.

Whilst the grading scheme is currently only limited to highway type scenarios it is acknowledged that these systems can be used on non-highway road types and also provide a safety benefit. Future research work will develop the grading scheme to assess the system performance on non-highway road types to further differentiate those systems incorporated higher functionality.

REFERENCES

[1] Tivesten, E., Victor, T. W., Gustavsson, P., Johansson, J., Sangberg, F., & Aust, M. L. (2018). Automation Expectation Mismatch: Incorrect Prediction Despite Eyes on Threat and Hands on Wheel. *Sage Journals*, 1081-1243.

[2] Grover, C., Avery, M., & Tough, D. (2018). *AN ASSESSEMENT APPROACH TO ASSISTED DRIVING SYSTEMS.* NHTSA.

[3] Euro NCAP. (2018, January). Retrieved from Euro NCAP: https://cdn.euroncap.com/media/32290/euro-ncap-sas-test-protocol-v20.pdf

[4] Volvo Cars. (2022, June 27). *Tips for using Pilot Assist*. Retrieved from Volvo Cars: https://www.volvocars.com/uk/support/topics/use-your-car/car-functions/tips-for-using-pilot-assist

[5] United Nations Economic Commission for Europe . (2018). Regulation No 79 of the Economic Commission for Europe of the United Nations (UN/ECE) — . *Official Journal of the European Union.*

[6] Euro NCAP. (2023, January). *Euro NCAP*. Retrieved from Euro NCAP: https://cdn.euroncap.com/media/75441/euro-ncap-ad-test-and-assessment-protocol-v11.pdf

PROPOSAL FOR AN IN-USE SAFETY AND SECURITY MONITORING FRAMEWORK FOR AUTOMATED VEHICLES (AV)

William Perren, Craig Arnold, Dr Kostas Kourantidis, Ben Simpson
Transport Research Laboratory (TRL Limited)
United Kingdom

Bryn Balcombe
ADA Innovation Lab Ltd.
United Kingdom

Nick Reed
Reed Mobility Ltd.
United Kingdom

Sam Chapman
The Floow Ltd.
United Kingdom

Paper Number 23-0280

ABSTRACT

The UK government are committed to bringing forward legislation to allow the safe and secure deployment of self-driving vehicles, as set out in the recent policy paper, “Connected and automated mobility 2025: realising the benefits of self-driving vehicles”. As part of the Connected and Automated Vehicle Process for Assuring Safety and Security (CAVPASS) programme, TRL was commissioned to propose a concept for assuring the safety of Automated Vehicles (AVs) throughout their operational life.

The work involved developing technical, procedural and administrative approaches for safety incident identification, investigation and reporting based on an evidence review of current and proposed in-vehicle datasets, safety metrics and collision investigation methodologies and supported by expert judgment. A Hazard Analysis and Risk Assessment (HARA) and an analysis of domestic traffic rules was conducted to assess the monitoring coverage of relevant risk events. Based on these activities, an overall framework for in-use safety and security monitoring has been proposed.

The study identified the need to monitor compliance against the behavioural competencies and safety arguments stated prior to deployment in order to continually assess the performance of the AV and the validity of the safety case during operation. A taxonomy for event classification has been developed to specify events to monitor safety and rules compliance. The study proposes that event-based data capture is the most feasible method of capturing data required to understand event context and causation to enable investigation. A minimum dataset specification has been developed which specifies a set of data metrics and thresholds for event detection as well as the data to be recalled supporting incident investigation. The HARA found that the proposed measures could not cover all safety relevant events and data sources external to data processed by the AV are required. Therefore, a set of operational processes for monitoring have been proposed.

A concept for monitoring traffic rules compliance has been introduced whereby AV perception data is processed independently. Analysis of domestic traffic rules identified requirements to record relevant dynamic objects, static objects and AV behaviours to enable monitoring of rules compliance. Processes for in-depth investigation and data analysis have been developed to enable the identification of compliance issues, produce learnings to be shared across the industry, and continuously improve the safety scheme. In-use monitoring data was found to be vital in ensuring accountability of AV safety performance by the manufacturer and contributes to an open and transparent safety culture by enabling just and proportionate regulatory sanctions to be applied.

Due to their paucity, data from AV collisions could not be used to base monitoring approaches on. The approach taken in this work was to identify safety monitoring protocols based on known approaches from conventional driving and other transport domains. A principle of continuous improvement was proposed such that the accuracy, quality and relevance of the monitoring framework can be assessed through AV deployment. This independent study proposes a framework for the safety performance assessment of AVs during operation to provide regulatory oversight, accountability and improve public trust in the technology.

INTRODUCTION

The United Kingdom government are committed to bringing forward legislation to allow the safe and secure deployment of self-driving vehicles. As part of the CAVPASS (Connected and Automated Vehicles Process for Assuring Safety and Security) programme, a partnership led by TRL was commissioned to develop a concept for a framework for monitoring the safety and security of automated vehicles during their use (i.e. post-deployment). The study initially focussed on Low-Speed Automated Vehicles (LSAVs), i.e. fully electric vehicles with a maximum speed not greater than 20 miles per hour (approx. 32 kilometres per hour) to be used in mixed traffic on roads with a speed limit not higher than 30 miles per hour (approx. 48 kilometres per hour). Table 1 provides additional details of the use cases and vehicle designs in scope. Further, it was envisaged that these vehicles would be owned and operated as part of a fleet providing goods or passenger services, rather than privately owned.

Table 1.
Scope of use cases and vehicle designs considered

Characteristic	Scope
Body shape	To include novel vehicle designs which do not conform with legacy design conventions such as windscreens, long bonnets, driver controls, etc.
Purpose	Carriage of goods or passengers (seated, standing or mixed)
Powertrain	Fully electric
Maximum speed	20 mph
Maximum mass (gross vehicle weight)	5,000 kg for passenger-carrying vehicles 3,500 kg for goods vehicles
Operating environment	Roads with a speed limit up to 30 mph with mixed traffic (including VRUs), or Dedicated roadways (which may or may not have segregation barriers) Areas which may include high density of pedestrians Operating on a fixed route or within a fixed geographical area

The initial focus on LSAVs was because they are expected to be an early use case for Connected and Automated Mobility (CAM) technology, and because there are no conventionally (i.e. human) driven vehicles in widespread use today that are comparable. While LSAVs were the primary focus, the work considered how a framework could be scaled and adapted to other CAM technology and use cases in line with the UK government's CAM policy [1]. This work considered an in-use monitoring framework to support a safety assurance process. As such, the purpose of the work was to develop a framework that provides continued validation of the safety and security of the automated vehicle during its deployment lifetime which allows for oversight and accountability for manufacturers and operators. This paper presents the work performed to develop an in-use monitoring framework.

METHODS

Evidence Review

The scope of this work set the basic requirements which an in-use monitoring framework must meet. These requirements were:

- The framework must specify requirements for manufacturers and operators to capture data to support evaluation of safety performance throughout deployment. Safety performance should be assessed and validated against the safety performance claimed or expected prior to deployment.

- Data requirements should be set out which support Automated Vehicle (AV) collision investigations as well as enable monitoring of any trends in safety performance during normal operation. The data requirements and monitoring processes should allow for the identification of causal and contributory factors associated with an event which can support safety learning and continuous improvement within the nascent AV industry.

- The framework should be aligned to and interoperable with AV assurance approaches in development by other nations and at the international level.

- The framework should enable non-compliance with, for example, regulations, standards or best-practice to be identified and reported and allow for intervention if necessary.

Based on these requirements, an evidence review was conducted which reviewed current and proposed monitoring approaches for road vehicles including proposals made for AV regulatory schemes such as the United Nations Economic Commission for Europe (UNECE) in-service monitoring and reporting [2] as well as current conventional vehicle data recording methods such as Event Data Recorders (EDR) [3], telematics for usage-based insurance, and fleet monitoring. Monitoring approaches from other transport industries such as rail, aerospace and marine transport were also reviewed. It was noted that safety assurance and governance approaches for these industries have a greater focus on safety learning and continuous improvement, which would provide a good basis for a safety assurance scheme for AVs. Approaches for investigating road collisions were also reviewed to identify the data necessary to support investigation. This primarily focused on approaches from national road safety investigation branches which are focused on independent blame-free investigations for safety learning and continuous improvement. Based on the review, in-vehicle data elements and safety performance metrics relevant to in-use safety monitoring were collated. This was evaluated against the objectives of this scheme to identify a rationalised set of data elements which forms the basis of an in-use monitoring framework using data collected from the automated vehicle.

Hazard Analysis and Risk Assessment (HARA)

It was necessary for the monitoring approach to be able to adequately identify realised risk events such as collisions and near miss events as well as capture data to support investigation and analysis. In order to understand whether proposed monitoring approaches adequately cover all likely risk scenarios encountered by an LSAV, a hazard analysis and risk assessment was conducted. Hazards were identified and structured using a deductive logic approach starting with a top-level hazard and then broken down into the sequential causes. An example is shown in Figure 1.

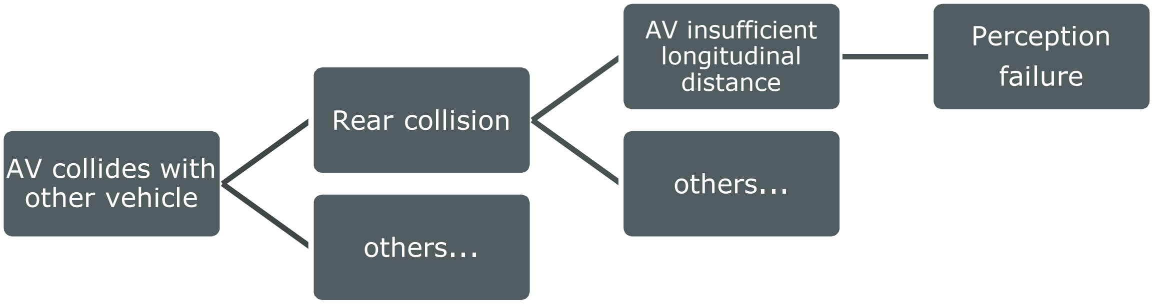


Figure 1: Hazard analysis structure

Each identified hazard or hazardous scenario was mapped to the data elements identified during the evidence review where there was a credible scenario where said data element could identify either the hazard or its causes. This allowed for evaluation of the coverage of the proposed data metrics and identified where in-use monitoring using vehicle data could not detect them.

Analysis of Domestic Traffic Rules

Compliance with road traffic rules (both mandatory and good practice) was found to be an important measure of safety performance and a key requirement to assure consistent, predictable and safe behaviour of AVs when operating alongside other road users. As such, proposals for an in-use monitoring framework include methods for identification of noncompliance with traffic rules by AVs. Other work in the CAVPASS programme set out a refined list of traffic rules relevant to the scope of operation which were collated from the Great Britain (GB) Highway Code [4]. For each of the GB domestic traffic rules relevant to LSAV operation, monitoring requirements were specified in this work, which identified:

- Elements of the rule that define relevant Operational Design Domain (ODD) attributes
- Elements of the rule that define Object Event Definition and Response (OEDR) requirements and driving behaviour considered part of the Dynamic Driving Task (DDT)
- If the rule specifies or requires the use of metrics and thresholds to determine compliance
- The type of measures, metrics, thresholds that are specified or required by the rule, where relevant
- Whether data regarding the vehicle's environment is required to determine rule compliance/violation

The findings from this analysis were used to develop a proposal for rule compliance monitoring using in-vehicle data.

Framework Development

Based on the above work, a concept for a holistic in-use monitoring framework has been proposed the enables oversight of LSAV deployments in the UK which is intended to be scalable to further CAM deployments. For context, it is assumed that the monitoring framework is a part of a wider framework which includes pre-deployment assessment, which would involve the development of a safety case. Monitoring would be in support of validation of the requirements of the safety case. Additionally, the framework was required to define the requirments of 4 types of entities that would be involved in conducting in-use monitoring. Their roles are summarized below:

Manufacturers Responsible for producing the LSAV and putting it forward for pre-deployment assessment. They will have responsibilities to collect data around safety performance and share that data with the regulator to provide evidence of ongoing safety performance and to support investigation.

Service operators Responsible for the safe operation of the vehicle while the Automated Driving System (ADS) is engaged. Their responsibilities include oversight of the vehicle (including remote supervision or operation) and will have an operator Safety Management System that will reference operational monitoring.

In-use regulator The agency (or agencies) with the responsibility to ensure the continuing safety and legal compliance of self-driving vehicles while they are in-use by learning from mistakes and preventing their re-occurrence. They will monitor AV safety performance, investigate safety events and road traffic infractions, and intervene to maintain safety.

Independent investigator This independent body will select safety events involving AVs to investigate for the purposes continual learning, but not attributing blame or liability. This role in the UK may, for example, be carried out by the newly proposed Road Safety Investigation Branch [5].

In this work, we use these terms without prejudice to terms identified elsewhere (for example in the policy paper Connected & Automated Mobility 2025: Realising the benefits of self-driving vehicles in the UK [1]). A 'manufacturer' may assume the role of an Authorised Self-Driving Entity (ASDE) and an 'operator' may be assume the role of a No User-in-Charge Operator (NUiCO).

RESULTS

Monitoring Using Vehicle Data

In order to fulfil the requirements of in-use monitoring and reporting, data needs to be collected that identifies potential occurrences where the AV was not safe (i.e. collisions and unsafe occurrences), not compliant with traffic rules or, acted in a way not in line with the safety claims made by manufacturer. Data then needs to be collected around these occurrences to provide an explanation of the event. The starting point for this work was to consider how best to utilise the wealth of data collected and computed by AVs during normal operation to develop an effective method for collecting the data needed for in-use monitoring.

Event based data capture Event-based data capture involves monitoring for road scenarios and occurrences that act as trigger events for more comprehensive data capture. Trigger events are specified such that they are indicative of a risk event, or potential risk event. In this way, comprehensive data can be collected around events that are likely to be of interest for in-use monitoring, thus minimising the amount of data collected in nominal situations.

In this work, measures for triggering events have been expressed as having 3 components; the data, the metric, and the threshold. Data is collected on the vehicle to calculate a metric in real-time. That metric correlates to increased safety risk. When the measured data exceeds some threshold value, a safety event is identified, and further data is then recorded by the vehicle which supports investigation and analysis. Real-time evaluation of these measures mitigates the need for continuous data capture which is likely to be resource intensive and infeasible. Instead, event-based data capture allows for comprehensive data to only be collected where relevant to specific safety events. This framework is required to detect collisions as well as detect potential safety issues prior to harm arising. This gives rise to leading and lagging measures.

- Lagging measures strongly correlate to a realised risk outcome. They are highly precise in that they are very likely to identify an event of interest, such as a collision. However, they are likely to only cover events where a risk outcome has already occurred.

- Leading measures on the other hand are proxies, they indicate a potential increase in risk of a collision occurring but does not necessarily mean a risk event has occurred. Because of this they are considered less precise, meaning they have a higher propensity to capture events that are not relevant (i.e., false positives). However, they also cover a wider range of possible risk events and thus provide a wider data set for analysis and the ability to predict safety performance in the absence of real risk events.

It is proposed that LSAV manufacturers should be required to collect a set of both leading and lagging measures. Based on this, a minimum set of lagging and leading measures have been proposed, summarized below in Table 2 and Table 3 respectively.

Table 2.
List of proposed lagging measures and the benefits and usefulness of the measures to identify collisions and other risk events.

Trigger considered	Description
Vulnerable Road Users (VRU) impact detection system activation triggers	The operating environment for the most common LSAV use cases is likely to include a high proportion of VRUs such as pedestrians. Many collision scenarios involving VRUs do not typically include high delta velocity changes. The activation of any of the following technologies should act as a trigger for data collection: • body panel VRU impact detection • continuously running pedestrian protection control algorithms • non-reversible deployable pedestrian protection device
"wake up" of occupant roll over protection systems	Crash scenarios in lower speed urban environments do not commonly include roll-over events. However, "wake up" of occupant roll over protection systems, if equipped, should also trigger data collection.
Minimum Risk Manoeuvre (MRM) activation	Execution of an MRM may not result in risk realisation but remains a risk scenario of high importance given that exit from an approved operating design domain is likely to correlate to a risk event.
System triggers of "wake-up" occupant protection systems	Activation of the Airbag control unit/module (ACU/ACM) or other "wake-up" occupant protection systems likely correlate to a high-risk scenario.
Battery / under vehicle impact protection	Presents insight to risk realisation from vehicle grounding or object non avoidance.
Vehicle door release when in motion	Representative of a realised passenger risk if it occurs.

Safety Envelope close proximity detected	Extremely close proximity to objects while in motion is indicative of a collision scenario. Proximity may be measured in different methods and may be an absolute measurement of distance or may be a time-based algorithm for collision risk such as Time To Collision (TTC). Many different proximity measures exist. A summary of those reviewed and considered is provided in Appendix A. Rather than specify which one to be used, it was found that manufacturers could define one that best suited their technology and use case which would need to be justified prior to deployment (i.e. within a safety case).
Passenger emergency or operator control override mechanism	Indicator of emergency disengagement event or passenger emergency requiring data capture.
Vehicle dynamics beyond expected ranges (e.g. over max speed, or harsh events beyond design range)	Extremely harsh events such as harsh braking or acceleration which are outside of operational parameters could be indicative of a realised risk scenario or a collision avoidance manoeuvre. In either instance, data collection should be triggered.
Unavailable or disabled automated system sensor or control, fault triggers	Considers degraded functionality of automated systems (e.g loss of perception system) which constitutes an unintended deviation from the operational design safety parameters.

Table 3.
List of proposed leading measures and the benefits and usefulness of the measures to identify near miss events and assess safety performance.

Trigger considered	**Description**
Infraction Measurement – excess speed (Limit)	A measure of compliance with enforceable speed limits. Not strictly relevant to safety performance but a necessary measure of compliance with road traffic laws.
Infraction Measurement – excess speed (Safe)	Relative to context dependent safe speed such as reducing speed in roads with reduced visibility or in icy conditions. Manufacturers are expected to define their strategy for selecting a safe speed for the driving context and monitor for compliance.
Safety Envelope – proximity	Proximity may be measured in the same way as in Table 2 but with a larger threshold representing a breach of safe operating parameters and potential operation at increased risk.
Driving style – longitudinal and lateral jerk	Jerk is often used as an indicator of risky driving as it correlates to an increased rate of collision involvement. Longitudinal jerk measures the rate of change of acceleration/deceleration in the axis of forward motion of the vehicle and indicates instances of harsh braking or rapid acceleration. Lateral jerk is measured perpendicular to the direction of motion and is indicative of unsafe turning. Both measure undesired and unsafe driving performance.
Operational Design Domain (ODD) exit	Outside of operational design safety parameters that may represent operation at risk.
Hazard Identification, reaction, and risk perception	It is anticipated that some automated vehicle control systems may implement a model for dynamic risk assessment of current and foreseen situations for motion planning. If utilised, outputs of such a system which predict imminent high-risk scenarios should trigger data capture. Manufacturers would be expected to define how their control system is utilised to monitor for risk scenarios and identify the thresholds which would trigger data capture.
Safety pre trigger events – e.g., Anti-lock Braking System (ABS) pre-charge, Forward Collision warning	Pre-activation of safety systems are indicative of higher risk scenarios which may develop into realised risk events or represent narrowly avoided “near-miss” scenarios.

The proposed set of measures listed above represent a minimum set of measures thought to be necessary for safety event detection and assessment of safety performance. It is noted that some LSAV manufacturers may wish to use a broader set of measures for event-based data capture if required for adequate safety oversight of their system. If so, those additional measures would be expected to be defined and shared within a safety case prior to deployment.

It is noted that some measures are dependent on whether safety systems are equipped to the vehicle (e.g. ABS). It was outside of the scope of this work to determine whether such systems are to be mandated on LSAVs, however it was recommended that if they are equipped, they should be considered as inputs for data collection for in-use monitoring. It was found that many of the proposed measures are expected to be defined relative to the operating parameters of the system, such that data capture would be triggered should the defined operating parameters be violated. This implicitly requires the operating parameters that are monitored against to be clearly defined prior to deployment.

The HARA found that the context behind the event is crucial in determining whether a proposed measure could credibly detect the event. The activation of data triggers may be dependent on the severity of the hazardous event. For example, a high-speed differential collision may trigger the wake-up of occupant protection systems, whereas a lower speed differential collision may not. Detection of other events, based upon proximity or acceleration, may be dependent on the value of the detection threshold. The HARA also found that safety envelope proximity triggers were found to be key for both leading and lagging measures due to its high coverage of safety relevant events. By selecting different thresholds, it could also serve as a measure for near miss scenarios and realised risk events, serving as both a leading and a lagging measure. However, proximity triggers are reliant on correct object event detection by the Automated Driving System (ADS). The HARA found that the following causes of a hazardous event would not be possible to identify by relying on vehicle data alone:

- Detection of object too late
- Failure to detect object
- Incorrect classification of object
- Failure to identify ODD exit
- Failure to predict object trajectory
- Detection of object which does not exist

As such, operational monitoring approaches that are not reliant on vehicle data have been proposed as redundancies and are discussed later in this paper. It is inevitable that there will be instances where that collision avoidance fails and other situations where a collision is truly unavoidable. In these instances, the HARA found that the residual risk for not detecting a collision is high. If the AV fails to detect a collision it will not initiate the appropriate response (MRM, or E-stop, etc.) which could result in increased consequence severity and potential for secondary collisions before intervention. As a result of this finding, it is recommended that collision detection should itself become a safety goal which must be argued prior to deployment.

Threshold selection For most of the above measures, the threshold for data capture may be discrete, i.e. it can take a fixed number of values. This corresponds to a state change in the system, such as safety system activation, deactivation, or fault code. In these cases, thresholds should correspond to a discrete value where a system change indicates a risk event has occurred. For proximity based and driving style metrics, data may be continuous and so the threshold may be set at any value. This work found that for these measures, the threshold selected will delineate between what is considered safe and unsafe driving performance.

Thresholds for acceptable performance may relate to current standards and rules for driving in the nation/region. For example, the GB Highway Code recommends that drivers maintain a 1.5 metre clearance when overtaking cyclists at 30 mph or less. This sets a clear threshold for safe performance. However, this threshold is only applicable during an overtaking manoeuvre for a cyclist at 30 mph. For overtaking horse riders, the recommended passing distance is 2 metres. As such, this study found that thresholds must be dependent on the ODD elements (in this case speed and the type of road user) as well as the driving context (in this case overtaking) in order to adequately monitor for contextually safe behaviour. It is not appropriate to develop universal thresholds and they need to be context dependent. For other measures, such as longitudinal jerk, the GB Highway Code does not provide clear threshold values. In this case it is proposed that manufacturers develop their own thresholds that define acceptably safe limits of performance for their vehicle in its defined ODD. Manufacturers may wish to use simulation to define these thresholds prior to deployment, however it is expected that manufacturers will evaluate their thresholds and refine them throughout operation. Manufacturer-defined thresholds should be shared with the regulator prior to deployment (potentially as part of a safety case). The regulators would be expected to assess the manufacturer's processes for defining thresholds and determine their suitability for the proposed deployment. They would also assess whether any manufacturer defined thresholds conflict with the relevant driving rules and whether that conflict is acceptable or not.

Monitoring compliance with traffic rules There is a clear need to be able to establish when road rules have been breached and gather evidence on these incidents in order to apply the appropriate level of corrective action. The analysis of 165 LSAV relevant Highway Code rules identified that a majority of them are context specific and refer to both elements of the driving situation and the environment. Therefore, for an LSAV to be able to comply with a traffic rule, it must have awareness of both the ODD and DDT (OEDR) elements that are specified in that rule. In order to monitor for compliance with the rule, a metric must also be defined which assesses the relevant OEDR performance. An example of the results is shown in Table 4 for rule 212 of the GB Highway Code.

Table 4.
Summary of the traffic rules analysis for rule 212 of the GB Highway Code.

Rule 212 excerpt: Give motorcyclists, cyclists, horse riders, horse drawn vehicles and pedestrians walking in the road (for example, where there is no pavement), at least as much room as you would when overtaking a car. Drivers should take extra care and give more space when overtaking motorcyclists, cyclists, horse riders, horse drawn vehicles and pedestrians in bad weather (including high winds) and at night.	
ODD attributes relevant to rule	• Weather (ice/snow/rain/high winds) • Time of day • Presence of motorcyclist, cyclist, horse rider, horse drawn vehicles and pedestrians in lane
DDT (OEDR) performance relevant to rule	• Passing distance - "giving more space" in bad weather • "taking extra care"
Performance metric and threshold requirements	• Proximity to object during passing manoeuvre • Different thresholds for operation during bad weather

Table 5 summarises the results of the analysis for the 165 identified LSAV relevant rules within the GB Highway Code. It shows that compliance with 97% of the rules requires the AV to have knowledge about either the DDT performance or the ODD attributes (or both). It also shows that compliance for 90.3 % of the rules can be monitored with a performance metric calculable by an AV.

Table 5.
Summary results of traffic rules analysis for 165 LSAV relevant UK Highway Code rules identifying which rules require DDT elements, ODD attributes and performance metrics to assess rule compliance.

Rule attributes	**No of rules**	**Percentage %**
Total LSAV relevant rules	165	100 %
Specifies DDT elements only	41	24.8 %
Specifies DDT and ODD attributes	119	72.2 %
Does not specify any DDT or ODD elements	5	3 %
Monitored via OEDR performance metric and threshold	149	90.3 %

OEDR requirements and ODD attributes are both datasets collected by an AVs perception module as part of normal operation and used as an input into the planning module. Since this data must be collected for functional operation of the AV, it is proposed that this data also be made available for monitoring compliance with traffic rules. As this data is extracted from the AV prior to OEDR planning, it provides an output of how the AV perceives the world, and is the information used by the AV for its decision making and planning OEDR execution. By processing this data independently from the ADS, it is possible to establish the desired OEDR performance which can be compared to the actual performance. By doing this it is possible to determine potential non-compliance as a result of improper OEDR performance. This processing can feasibly be conducted onboard the vehicle in real-time so that safety critical rule infractions can be detected as they occur and used as trigger events for more comprehensive data recording. It is recommended that the manufacturer's safety case should outline what data is collected and how it is used for assessing OEDR performance.

An ADS does not need to explicitly classify rule relevant ODD and DDT elements in order to drive safely and comply with traffic rules and it is known that some ADS solutions do not have this capability. For example, an AV may overtake all other road users with a clearance of 2.5 metres. Rather than classifying the object and selecting a more specific clearance, broad compliance with this element of the Highway Code rules has still been achieved.

Manufacturer Defined Monitoring Processes

It is common good practice with safety case development to ensure a process by which the arguments and assumptions made in the safety case are monitored. In line with this practice, it is proposed that manufacturers develop a monitoring plan that is evidenced within the safety case. The monitoring plan should evidence that there are processes in place to collect and investigate data in line with the minimum dataset specification as well as any additional processes required to monitor continued compliance with the safety case. A key part of monitoring for safety arguments will be to test the assumptions made in the safety case around the presence of hazards, and the effectiveness of the proposed mitigations. To this end, manufacturer defined monitoring processes should seek to understand:

- The occurrence of unmitigated hazards and partially mitigated hazards.
- The occurrence of hazards that have been accepted without mitigation.
- Violations of assumptions, design goals, and conclusions made based on an evaluation of evidence made in the safety case.
- Manufacturers should be encouraged to identify further measures and other monitoring approaches to ensure tolerable coverage and evidence this as part of their safety case.

Operational Monitoring Approaches

The HARA showed that both leading and lagging measures and monitoring of traffic/safety rules accounted for a considerable amount of safety relevant events but identified certain hazards which could not be monitored. Notably risks caused by a failure to detect objects were identified as being impossible to monitor using the monitoring methods proposed, as they all rely on data collected by the vehicle. In order to account for this, it is proposed that operational monitoring mechanisms should be integrated into in-use monitoring and reporting processes in order to widen the coverage of possible risk events. Three mechanisms are proposed below and would fall under the responsibility of the LSAV operator.

Maintenance and inspection Operators of current fleets have a responsibility to develop and operate processes for walkaround checks, maintenance and inspection of their vehicles. The purpose of these processes is to identify and remedy any vehicle faults prior to operation to minimise risk. It was found that reports made through these processes could be incorporated into a monitoring system. These processes could provide context on issues with the vehicle that may not have been internally detected, such as sensor that appears to function but is not calibrated accurately to detect objects and so provide context to events that could have previously happened but weren't detected, or reasons as to why an event happened that were not considered in internal monitoring. It could also uncover previously unseen damage that triggers a review of data to investigate the time during operation when the damage occurred.

Public feedback It is proposed that service operators maintain a mechanism for public feedback and outline their processes. Oversight for this could be achieved through a mechanism of deployment licensing and the regulator would be expected to sample public reports made to operators to ensure they are handled appropriately. This work specifies that a public feedback mechanism should be in place that allows eyewitnesses and passengers to report unsafe events to the operator. Since the scope of this work considers vehicles that have no driver who is usually legally responsible for reporting a collision, witness or passenger reporting is thought to be an effective redundancy for detecting events that supports vehicle data led approaches. Limitations are foreseen with this system, however. Members of the public may report significant amounts of irrelevant events since they do not know what constitutes a safety relevant event. This could lead to unnecessary burdens on the operator. It is recommended that public reports be used to trigger a review of operational data that then may trigger further investigation and reporting. It would be expected that the processes for handling public reports, disregarding false reports, and investigating genuine events be defined and evidenced by the operator.

Police and enforcement reports In the immediate term, traffic events are expected to be reported to the police as with conventional vehicles, including through witness testimony and collision reports. Furthermore, speed cameras and other enforcement infrastructure may also identify an issue. It is recommended that once the police identify that an AV was involved, then the responsible party for the AV would be notified who would have a duty to report to the regulator. We expect operators would use the reports as a trigger to conduct their own investigation in collaboration with the AV manufacturer as required.

Recall Data

The purpose of data recall is to provide data that supports investigation of an event to determine its causes as well as any corrective action required to prevent future occurrences. This work proposes primarily event-based data capture whereby a comprehensive set of data is collected by the vehicle when an event is detected. This method means that only data relevant to the event in question is captured. However, it was found that some degree of continuous data collection is required.

Continuously transmitted data This data is not associated with any event identified by the vehicle. As a result, it can provide data for a basic risk evaluation if a risk event is reported where an ADS may not be aware of it. However, because data needs to be continuously transmitted, there are limits on the amount of data that can collected before its capture and storage becomes economically and practically unfeasible. As a result, it cannot support detailed investigation of safety events to understand their causes. It is proposed that a minimal dataset be collected continuously. This may be stored on the vehicle or transmitted elsewhere. The primary purpose of this dataset will be to provide basic risk analysis and liability determination in situations where no other data is recorded. A continuously transmitted dataset is proposed in Table 6, which is based on commercial telematics systems.

Table 6.
List of data elements proposed for continuous collection during LSAV operation to enable basic risk and liability determination.

Data elements	Values
Continuous transmission	
Vehicle telemetry	GPS, speed, gyroscopes, accelerometers, telemetry accuracy and quality measurement
Values transmitted upon state change	
Autonomous systems	Operating status change and override events
Door, boot, window and hood status	Open/closed/locked/position/status
Horn and light operations	On/off/low beam/high beam/flash/fog/hazard, accuracy and quality measurement
Vehicle dynamics and safety systems	ABS pre-charge, forward collision warning, stability and traction control, etc.
Crash restraint and seat sensors	Status, occupancy, accuracy and quality measurements
Wipers	Speed/state/front/rear/accuracy and quality measurement
Trailer / wheelchair ramp / assistive systems	Trailer/wheelchair ramp/assistive systems - status/ detection
Ignition control	Interaction and operation of ignition and auto/start-stop technologies or in the case of EVs engine on and off.

Event data recall The purpose of comprehensive data recall is to provide data that supports investigation of the event to determine its cause as well as any corrective action required to prevent future occurrences. Broadly, it is proposed that a requirement is placed on the manufacturer to store and share all data necessary to the regulator to enable investigation upon a trigger event. This work found that it is not possible to specify this data completely as the data required would be manufacturer-specific and dependent upon the configuration of sensors, compute and software used. Rather than specifying the exact sensors, data rates and formats for this, the manufacturer should define what data is recalled as part of a monitoring plan that is submitted prior to deployment. A minimum dataset has also been defined that is to be recalled following a trigger of leading and lagging measure triggers. This minimum dataset is defined in Appendix B. This is to create a basic standardised data set for collection within the framework that is technology agnostic and enables fundamental event investigation. This also ensures similar data is being captured across manufacturers to enable identification of safety themes. Manufacturers would be expected to assure regulators that the minimum dataset is being captured in addition to any further data defined by the manufacturer as necessary.

Reporting to Regulator

It is proposed that the scope of reportable occurrences should align with best practice from aviation. In addition, this framework aims to support continued validation of the safety performance declared pre-deployment and monitor ongoing compliance. As such, manufacturers and operators should be required to report:

- Any occurrence which endangers or which, if not corrected, would endanger the AV, its occupants or any other person.
- Any occurrences or set of occurrences that indicates an actual or potential violation of the safety case or safety performance declared at prior to deployment
- Any occurrence which violates road traffic or other laws (such as data privacy)

Sub definitions for events in scope of this scheme have been developed under this project and are summarized in Table 7. These event classifications are intended to allow compliance monitoring of the AV as well as code events to generate data for comparative assessment of safety performance across an AV fleet.

Table 7.
Definitions for proposed event classifications used for trend analysis.

Event	Definition
Collision	An incident in which the LSAV makes contact with an object, either moving or fixed, at any speed, in which kinetic energy is measurably transferred or dissipated. Includes other vehicles, roadside barriers, objects on or off the roadway, pedestrians, cyclists, or animals.
Near-collision	Any circumstance that requires a rapid, evasive manoeuvre by the LSAV (or any other vehicle, pedestrian, cyclist, or animal) to avoid a collision. A rapid, evasive manoeuvre is defined as steering, braking, accelerating, or any combination of control that is significantly greater than that expected in normal operation.
Safety critical event	Any circumstance that requires a collision avoidance response on the part of the LSAV or any other vehicle, pedestrian, cyclist, or animal that is less severe than a rapid evasive manoeuvre but greater in severity than a normal operation to avoid a crash. A collision avoidance response can include braking, steering, accelerating, or any combination of control inputs.
Proximity conflict	Any circumstance resulting in extraordinarily close proximity of the LSAV to any other vehicle, pedestrian, cyclist, animal, or fixed object when, due to apparent unawareness on the part of the vehicle, driver, pedestrians, cyclists or animals, there is no avoidance manoeuvre or response. Extraordinarily close proximity is defined as a clear case in which the absence of an avoidance manoeuvre or response is inappropriate for the driving circumstances (e.g. speed, sight, distance, etc.).
Non-conflict critical incident	Any event that increases the level of risk associated with driving but does not result in any of the events as defined above.
Safety-relevant infraction	Road rule violations that have direct safety implications even if another event type (e.g. collision, near collision etc.) does not occur.
Traffic infraction	Road rule violations not directly related to safety but that negatively impact the flow of traffic or safe movement of other road users.

Individual event reporting The reporting of safety critical incidents such as collisions is required in regulatory schemes in other transport modes and is common practice in the aviation industry. The purpose in aviation is to ensure relevant information on safety is reported, collected, stored, protected and disseminated with the prevention of accidents and incidents as the sole priority. It is recommended that a similar approach be adopted for reporting be adopted for an in-use monitoring framework. It is proposed that individual reports be made for:

- Lagging measures (see Table 2) – their low frequency would inhibit statistical evaluation, but their high data availability allows in depth analysis.
- Police reported events such as traffic infractions – where the regulator would require the manufacturer to investigate the event.
- Events not recorded by the event data capture system but detected through other processes – indicating there is a potential failure
- Any other event requested by the regulator – For instance, where the regulator receives significant public complaints.

Individual event reports should be underpinned by case study analysis with the approach for investigating and analysing events detailed within a manufacturer's Safety Management System. The broad requirement set is that the event reports should display sufficient information required for the regulator to understand the nature and cause of the event. It is also recommended that collisions be reported to the regulator immediately after identification in order to enable the regulator to manage crisis communications and media reporting as this was found to be required to maintain public confidence in the regulator.

Aggregated data reports The purpose of collecting aggregated datasets is to enable statistical evaluation in order to evaluate safety performance over time. This is to identify trends in safety performance that may be inconsistent with the safety performance expected prior to deployment. It is proposed that aggregated data sets of the rate of occurrence of leading and lagging measures be reported to the regulator. This includes the measures defined by the data set as well as any others required by the manufacturer for safety case compliance monitoring.

Aggregated data should detail:

- The event partner and the category of event
- The number of false positives and confirmed events to determine the suitability of selected thresholds
- The number of confirmed events investigated to root cause and their conformance with the safety arguments stated within the safety case
- The relevant crash characteristics that could inform blame determination. For example: The ego vehicle is hit from behind at a stop sign would by default be blamed on the trailing vehicle. An elevated rate of not-blamed loss events could still be indicative of safety issues, such as the AV behaving in a manner that provokes mistakes by human drivers.
- The demographic of involved parties. This is intended to identify patterns is safety performance for different user groups. For example, this may identify biases in machine learning training sets or defects in ODD construction so that they can be corrected.

Reporting frequency It is proposed that reporting occur in two ways; periodically, and immediately following certain events. Periodic reporting would be at a fixed interval (e.g., every six months) and its purpose would be to report the aggregated data. Reactive reporting would take place as soon as collected data (either aggregated data or case studies) provides evidence of an inconsistent ADS behaviour compared to the safety performance claims made prior to market introduction, or when collected data provides evidence of degradation of safety performance). It is also recommended that collisions be reported to the regulator immediately in order to enable the regulator to manage crisis communications and media reporting to ensure public confidence in the regulator.

Regulator Assessment and Sanctions

Once in-use monitoring detects a safety issue, it will be necessary for the regulator to determine the most appropriate corrective action. It is proposed that rather than punitive criminal action, the regulator should have access to a gradated system of enforcement actions, similar to aviation authorities. It was found in this work that the proposed in-use monitoring data is a desirable data source for regulators to be able to understand the extent and nature of the failure and the potential risk of recurrence which would enable them to take an evidence-based decision on the fairest and most proportionate level of corrective action to apply. It is proposed that any regulatory system for enforcement be based on evidence gathered through in-use monitoring. Fair and appropriate sanctions, rather than cautious and severe ones, are expected to reinforce an open and transparent safety culture that promotes proactive reporting to the regulator. Manufacturers and operators both have responsibilities for ensuring safety. As such, it is proposed that sanctions can be applied to both parties if necessary for the interest of public safety.

DISCUSSION AND LIMITATIONS

This study proposes a framework for in-use monitoring to assess the safety performance of automated vehicles. This novel framework proposal is intended to be used to support ongoing safety assurance on behalf of a regulator responsible for ensuring the safety of AVs throughout their deployment lifetime. However the overall approach is consistent with best practice safety management and is applicable to AV manufacturers and operators outside of a regulatory scheme. The proposals made in this framework are novel in that there is no currently comparable mechanism of in-use monitoring for conventionally driven road vehicles. However, this study found that other transport sectors, notably aviation, had mature safety assurance processes that promotes a culture of shared learning and prioritising safety above all else. The principles of knowledge sharing, open and transparent data collection, reporting practices and evidence based regulatory enforcement are all good practice approaches used in aviation have been adapted for this framework. While based in good practice of other sectors, there is limited practical evidence to determine whether it remains feasible for automated vehicles.

The scope of this framework was primarily (although not solely) focused on LSAV use cases as these were likely to be some of the earliest use cases adopted. As a result, it remains to be seen as to whether the proposed approaches are scalable and adaptable to other use cases. It is noted however that the proposed framework is largely independent on specific ODD attributes, use cases, and operating contexts that would limit its applicability to other use cases. One limitation noted is that the proposed framework is reliant on the existence of a fleet service operator who has responsibilities to cooperate with manufacturers, and to collect and report data. There is no equivalent entity for privately owned Automated Vehicles, instead having to pass some or all of this responsibility to the driver. Nevertheless, much of the data elements needed for assurance are likely to be transferable.

This study proposes that the wealth of data collected by automated vehicles during operation can and should be utilised for the purposes of ongoing assurance. It is anticipated that event-based data capture be the primary method of collecting data as continuous capture of all the data required to support event investigation would be inefficient and largely unfeasible. Much of the data required by the proposed scheme is already being collected and utilised for existing processes such as current event data recorders, telematics-based insurance models and fleet monitoring. However, due the paucity of automated vehicle collision data, it is not possible to determine the value and usefulness of collecting the proposed datasets. A principle of continuous improvement is proposed such that the accuracy, quality and relevance of the monitoring framework is assessed through AV deployment.

This work has shown that monitoring for compliance with traffic rules requires data to be collected on ODD attributes, DDT (OEDR) requirements and performance metrics relevant to the rule. As such a method of continuous rule compliance assessment is recommended by independently processing the data and comparing against real-time AV performance. As this relies solely on datasets only available to automated vehicles, there is limited evidence to suggest the ease and practicality of collecting and processing such data. However, the data required for this is the same dataset required for safe AV operation.

CONCLUSION

This independent study proposes a framework for the safety performance assessment of AVs during operation to provide oversight, accountability and improve public trust in the technology. The proposed framework is based on the principle of leveraging the data collected by automated vehicles during normal operation to assess and validate the safety performance of the AV against the performance expected prior to deployment. The study outlines the data elements necessary to support ongoing assurance and investigation of incidents as well as the administrative and technical processes and procedures necessary for sharing and reporting safety learnings to drive improvements in the nascent automated vehicle industry.

ACKNOWLEDGMENTS

The authors wish to acknowledge the United Kingdom's Department for Transport for funding this study. The findings and recommendations do not necessarily reflect the views of the UK Department for Transport and will not necessarily be implemented in future AV safety assurance requirements.

REFERENCES

[1] HM Government, *Connected and automated mobility 2025: Realising the benefits of self-driving vehicles in the UK* (2022). London. Retrieved December 11, 2022, from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1099173/cam-2025-realising-benefits-self-driving-vehicles.pdf.

[2] Informal Working Group (IWG) on Validation Methods for Automated Driving (VMAD) 2022, *New Assessment/Test Method for Automated Driving (NATM) Guidelines for Validating Automated Driving System (ADS) – amendments to ECE/TRANS/WP.29/2022/58.* Available at: https://unece.org/sites/default/files/2022-05/GRVA-13-35e.pdf

[3] United Nations, *E/ECE/TRANS/505/Rev.3/Add.159, UN Regulation No.160 Uniform provisions concerning the approval of motor vehicles with regard to the Event Data Recorder* (n.d.). Retrieved December 11, 2022, from https://unece.org/sites/default/files/2021-10/R160e%20_0.pdf.

[4] HM Government. (2022). *The Highway Code*. Retrieved December 11, 2022, from https://www.highwaycodeuk.co.uk/

[5] HM Government Department for Transport. (2022, June 28). Government launches country's first ever investigation branch focused on Road Safety. GOV.UK. Retrieved December 11, 2022, from https://www.gov.uk/government/news/government-launches-countrys-first-ever-investigation-branch-focused-on-road-safety

APPENDIX A

A summary of the reviewed proximity and safety-envelope calculation metrics is given below:

- **Time to Collision (TTC).** A calculated time to collision between objects if each object continued on current trajectory and speed.
- **Time Exposed Time-to-Collison (TET**). A summation of TTC values (above) over a windowed time period – used to smooth uncertainty in TTC distance and speed estimations.
- **Time Integrated Time-to Collision (TIT).** An integral of TTC values when below a threshold – used in microscopic level traffic simulations.
- **Modified Time-to-Collision (MTTC).** This approach considers possible acceleration changes in objects to present a worst case scenario TTC where speed can increase.
- **Crash Index (CI).** This is a severity index measured by pairs of moving objects kinetic energy differences – used to understand potential crash severity but can be used to minimise strong differential speed risk scenarios.
- **Headway (H).** The elapsed time between following vehicles passing reference locations – used in lane following to associate risk from unsafe stopping distances in following traffic.
- **Time to Accident (TA).** The time until a vehicle would have had an accident had either it or another vehicle not taken evasive speed or direction change already occurred – a what if scenario for risk proximity if no action had been taken to calculate proximity to a realised accident.
- **Post Encroachment Time (PET).** The time between when one road users leaves a potential collision risk area and another enters it – used typically in junction safety understanding.
- **Potential Index for Collision with Urgent Deceleration (PICUD**). The distance between two vehicles if both undergo urgent deceleration – used in some lane changing and merge safety algorithms.
- **Margin to Collision (MTC).** A ratio of the ego and following vehicles stopping distances when following a lead vehicle - used in close following deceleration understanding not just forward but also rear collision potential.
- **Difference of Space Distance and Stopping Distance (DSS).** A difference between stopping distance and actual distance – used to understand degrees of safe operation in following traffic.
- **Time Integrated DSS (TIDSS).** A time integrated DSS (above) approach that factors in duration of risk exposure into its formulae.
- **Deceleration Rate to Avoid the Crash (DRAC).** A declaration indicator looking at differential speeds and closing distance ratios to look for unsafe deceleration when more is required – used in some ADS safety systems.
- **Crash Potential Index (CPI).** An extension of DRAC (above) that considers future time events and potential to exceed a vehicle maximum deceleration rate – used in some ADS safety systems.
- **Criticality Index Function (CIF).** A potential risk severity measure combining vehicle speed with required deceleration – used to indicate a potential severity for a speed and needed deceleration for any impact at a point in time, used in some ADS safety systems.

APPENDIX B

Table 8.
Recommended data elements for recall following trigger of lagging measures including reasoning for inclusion in the minimum dataset

Data element	**Recording interval/time (relative to time zero)**	**Data sample rate (per second)**	**Minimum range**	**Accuracy**	**Resolution**	**Event (s) recorded for**
Delta-V, longitudinal	0 to 250 ms or 0 to End of Event Time plus 30 ms, whichever is shorter.	500	-100 km/h to + 100 km/h.	±10%	1 km/h.	Planar VRU Rollover
				REASONING: This data field provides fine grained velocity change data allowing reconstruction of kinetic energy exchange in a longitudinal direction. Not required if longitudinal acceleration recorded at ≥500 Hz with sufficient range and resolution to calculate delta-v with required accuracy		
Maximum delta-V, longitudinal	0–300 ms or 0 to End of Event Time plus 30 ms, whichever is shorter.	NA	-100 km/h to + 100 km/h.	±10%	1 km/h.	Planar VRU Rollover
				REASONING: This data field provides a single peak value (helping to inform any incident severity) from fine grained velocity change data allowing severity estimation in a longitudinal direction. Not required if longitudinal acceleration recorded at ≥500 Hz.		
Time, maximum delta-V, longitudinal	0–300 ms or 0 to End of Event Time plus 30 ms, whichever is shorter.	NA	0–300 ms, or 0-End of Event Time plus 30 ms, whichever is shorter.	±3 ms	2.5 ms.	Planar VRU Rollover
				REASONING: This data field provides a single timestamp value helping to indicate when in the sample the maximum severity impact occurred in reference to time zero trigger. Not required if longitudinal acceleration recorded at ≥500 Hz.		
Speed, vehicle	-5.0 to 0 sec	2	0 km/h to 250 km/h	±1 km/h	1 km/h.	Planar VRU Rollover
				REASONING: Providing operating speed to enable understanding of overall kinetic energy in precursor to time zero trigger events.		
Motor Transition Demand	-5.0 to 0 sec	2 (or more frequent as possible to record)	0 to 100%	±5%	1%	Planar VRU Rollover
				REASONING: To determine precursor motor transition changes and vehicle motion intention prior to the event.		
Service brake Demand	-5.0 to 0 sec	2 (or more frequent as	0 to 100%	±5%	1%	Planar VRU Rollover
				REASONING: To determine precursor braking		

		possible to record)		operation of the vehicle prior to the trigger event.		
Ignition/start cycle, crash	-1.0 sec	N/A	0 to 60,000	±1 cycle	1 cycle.	Planar VRU Rollover
				REASONING: To determine recorded trigger events by journey cycles to understand power/ignition on/off cycles.		
Ignition/start cycle, download	At time of download	N/A	0 to 60,000	±1 cycle	1 cycle.	Planar VRU Rollover
				REASONING: To determine additional vehicle usage following a trigger event.		
Occupant protection system deployment, time to deploy, in the case of a single stage air bag, or time to first stage deployment, in the case of a multi-stage air bag(s)	Event	N/A	0 to 250 ms	±2ms	1 ms.	Planar VRU Rollover
				REASONING: To detail deployment times for safety systems fitted. Needed to determine effectiveness of mitigations vs. injury in the event of a trigger		
Multi-event crash, number of events	Event	N/A	1 or more	N/A	1 or more.	Planar VRU Rollover
				REASONING: To detail the potential of multiple trigger events in temporal proximity, each adding insight about incidents with multiple impacts or triggers occurring.		
Time from event 1 to 2	As needed	N/A	0 to 5.0 sec	±0.1 sec	0.1 sec.	Planar VRU Rollover
				REASONING: To detail the potential of multiple trigger events in temporal proximity, each adding insight about incidents with multiple impacts or triggers occurring.		
Complete file recorded	Following other data	N/A	Yes or No	N/A	Yes or No.	Planar VRU Rollover
				REASONING: To detail the potential of incomplete recording due to device or sensor damage making expected data unavailable. Indicates mechanical failure of incident recording means in an incident.		
Lateral acceleration (post-crash)	0–250 ms or 0 to End of Event Time plus 30 ms, whichever is	500	-50 to +50g	± 10%	1 g	Planar VRU Rollover
				REASONING: To allow forensic reconstruction post trigger of any side impact.		

	shorter.					
Longitudinal acceleration (post-crash)	0–250 ms or 0 to End of Event Time plus 30 ms, whichever is shorter.	500	-50 to +50g	± 10%	1 g	Planar VRU Rollover
				REASONING: To allow forensic reconstruction post trigger of any front/rear impact.		
Normal acceleration (post-crash)	0–300 ms or 0 to End of Event Time plus 30 ms, whichever is shorter. (This is still under debate and subject to change)	10 Hz	-5 g to +5 g	± 10%	0.5 g	Planar VRU Rollover
				REASONING: Details the downward acceleration (typically gravity) of a vehicle. Is used to determine in any trigger any up-down acceleration of a vehicle which helps forensic reconstruction.		
Delta-V, lateral	0–250 ms or 0 to End of Event Time plus 30 ms, whichever is shorter.	100	-100 km/h to + 100 km/h.	±10%	1 km/h.	Planar VRU Rollover
				REASONING: The cumulative change in velocity in lateral direction that helps to understand kinetic energy transfer in any side impact. Not required if lateral acceleration recorded at ≥500 Hz and with sufficient range and resolution to calculate delta-v with required accuracy.		
Maximum delta-V, lateral	0–300 ms or 0 to End of Event Time plus 30 ms, whichever is shorter.	N/A	-100 km/h to + 100 km/h.	±10%	1 km/h.	Planar VRU Rollover
				REASONING: The highest change value in side velocity during the trigger data capture period. Allows to understand peak severity of side impacts. Not required if lateral acceleration recorded at ≥500 Hz.		
Time maximum delta-V, lateral	0–300 ms or 0 to End of Event Time plus 30 ms, whichever is shorter.	N/A	0–300 ms, or 0-End of Event Time plus 30 ms, whichever is shorter.	±3 ms	2.5 ms.	Planar VRU Rollover
				REASONING: The time point of the highest side velocity change in the monitoring trigger window. Not required if lateral acceleration recorded at ≥500 Hz.		
Time for maximum delta-V, resultant.	0–300 ms or 0 to End of Event Time plus 30 ms, whichever is shorter.	N/A	0–300 ms, or 0-End of Event Time plus 30 ms, whichever is shorter.	±3 ms	2.5 ms.	Planar VRU Rollover
				REASONING: The time from the trigger point (time zero) to the maximum velocity change recorded. Used to understand the point of highest severity in relation to the trigger point aiding forensic reconstruction. Not required if relevant acceleration recorded at ≥500 Hz		
Engine/Motor rpm	-5.0 to 0 sec	2	0 to 10,000 rpm (or high maximum rpm	±100 rpm10	100 rpm.	Planar VRU Rollover

			as needed for the vehicle type)	REASONING: details the number of revolutions per minute of the engine/motor output (in fuel driven vehicles via the crankshaft, in electric vehicles the output rotations of the device applying motive power). This details the engine/motor operating speed in the approach to the trigger event.		
Vehicle roll angle	-5.0 to 5.0 sec	10	-1080 deg to + 1080 deg.	±10%	10 deg.	Rollover
				REASONING: vehicle rollover events being considered this indicates the degree of roll observed in the trigger window. These values can be used in crash reconstruction.		
Anti-lock braking system ABS activity	-5.0 to 0 sec	2	Faulted, Non-Engaged, Engaged Active, Intervening	N/A	Faulted, Non-Engaged, Engaged Active, Intervening	Planar VRU Rollover
				REASONING: If fitted, the status of anti-lock braking pre trigger can help to understand anti-lock braking behaviour in any rapid velocity change before the trigger event.		
Stability control	-5.0 to 0 sec	2	Faulted, On, Off, Engaged Intervening	N/A	Faulted, On, Off, Engaged Intervening	Planar VRU Rollover
				REASONING: If fitted, the status of stability control pre trigger can help to understand stability control status in any rapid velocity change before a trigger event.		
Digital requested Steering input	-5.0 to 0 sec	2	-250 deg CW to + 250 deg CCW.	±5%	±1%.	Planar VRU Rollover
				REASONING: Requested steering input prior to the trigger helps to determine any potential collision avoidance activity or swerving behaviour.		
Safety belt status	-1.0 sec	N/A	Fastened, not fastened	N/A	Fastened, not fastened	Planar VRU Rollover
				REASONING: If fitted, the status of any passenger restraint system has impacts in any trigger events resulting in physical injury.		
Occupant protection systems deployment, time to nth stage,	Event	N/A	0 to 250 ms	±2 ms	1 ms.	Planar VRU Rollover
				REASONING: If fitted the deployment time of multi-stage occupant protection system deployments required to understand the relationship to the ability to mitigate injury in a realised risk incident.		
Occupant size classification, any passenger	-1.0 sec	N/A	6yr old HIII US ATD or Q6 ATD or smaller	N/A	Yes or No.	Planar Rollover
				REASONING: If monitored, seat weight sensors help to understand impact injuries and effectiveness of any fitted restraint systems.		
Automated Driving System Status	[-30.0] to +30.0 second relative to time zero	2	N/A	N/A	On, Off - Manually Deactivated, Off-Automatically	Planar VRU Rollover

					Deactivated Faulted	
				REASONING: To detail the operating status of any automated driving system (one for each possible in the vehicle) to understand the status in connection to an incident as automated vs non automated will require differing handling and statistical aggregation of events.		
Automated Driving System - Minimal Risk Manoeuvre	[-30.0] to +30.0 second relative to time zero	2	N/A	N/A	Yes or No	Planar VRU Rollover
				REASONING: To detail any activation of MRMs in relation throughout a trigger window.		
Automated Driving System - Override	[-30.0] to +30.0 second relative to time zero	2	N/A	N/A	List of possible overrides	Planar VRU Rollover
				REASONING: To detail any listed override events halting automated driving each record gives the reason behind any unplanned disengagement activity that can have safety impacts.		
Latitude	[-30.0] to +30.0 seconds relative to timezero	1 or higher	World Geodetic System 1984 (WGS84)	WGS84 standard error ranges	WGS84 standard ranges	Planar VRU Rollover
				REASONING: Geopositioning may present GDPR challenges for allowable processing however it is vital to understand locational risk and relation to external factors. The course and trajectory understood also have high value in understanding risk scenarios. This collation is recommended within law commission consultations as well as the Insurance Industry to enable liability determination		
Longitude	[-30.0] to +30.0 seconds relative to timezero	1 or higher	WGS84	WGS84 standard error ranges	WGS84 standard ranges	Planar VRU Rollover
				REASONING: Geopositioning may present GDPR challenges for allowable processing however it is vital to understand locational risk and relation to external factors. The course and trajectory understood also have high value in understanding risk scenarios. This collation is recommended within law commission consultations as well as the Insurance Industry to enable liability determination		
All trigger status	[-30.0] to +30.0 seconds relative to timezero	2	N/A	N/A	List of possible trigger types	Planar VRU Rollover
				REASONING: To capture trigger events timing and type throughout the trigger capture period.		
Operating environment static and mobile objects, relative position,	[-10.0] to +10.0 seconds relative to timezero	10	[-50.0m] – [+50.0m] relative position to centre of LSAV (nearest objects)	Relative position	Position used in LSAV decision making	Planar VRU Rollover
				REASONING: To record observed relative object positions that the vehicle detects in near environment to enable reconstruction of third party object relative movements and positions		

longitudinal						
Operating environment static and mobile objects, relative position, lateral	[-10.0] to +10.0 seconds relative to timezero	10	[-50.0m] – [+50.0m] relative position to centre of LSAV (nearest objects)	Sensor estimate position	Position used in LSAV decision making	Planar VRU Rollover
				REASONING: To record observed object positions that the vehicle detects in near environment to enable reconstruction of third party object relative movements and positions		
Operating environment static and mobile objects, speeds, (nearest ‘x’ objects)	[-10.0] to +10.0 seconds relative to timezero	2	0 km/h to 250 km/h	As per accuracy of observed speeds	As per accuracy of observed speeds	Planar VRU Rollover
				REASONING: To record observed object relative speeds that the vehicle detects in near environment to enable reconstruction of third party object relative movements and positions		
Operating environment static and mobile objects, trajectory	[-10.0] to +10.0 seconds relative to timezero	2	[-180.0] to +180.	As per accuracy of observed trajectory	As per accuracy of observed trajectory	Planar VRU Rollover
				REASONING: To record observed object relative bearing that the vehicle detects in near environment to enable reconstruction of third party object relative movements and positions		
Operating environment static and mobile objects, classification	[-10.0] to +10.0 seconds relative to timezero	2	[static, vehicle, VRU, Moving unknown]	As per accuracy of observed object classification	As per accuracy of observed object classification	Planar VRU Rollover
				REASONING: To record observed object types that the vehicle detects in near environment to enable reconstruction of third party object relative movements and positions.		

Table 9.
Recommended data elements for recall following trigger of leading measures

Data element	**Condition for requirement**	**Recording interval/time (relative to time zero)**	**Data sample rate (per second)**	**Minimum range**	**Accuracy**	**Resolution**
Delta-V, longitudinal	Mandatory	-100 to 200 ms	50	-100 km/h to + 100 km/h.	±10%	1 km/h.
Speed	Mandatory	-10.0 to 10.0 sec	50	0 km/h to 250 km/h	±1 km/h	1 km/h.
Delta-V, lateral	Mandatory	-100 to 200 ms	50	-100 km/h to + 100 km/h.	±10%	1 km/h.
Automated Driving System Status	Mandatory	-10.0 to +10.0 second relative to time zero	2 (or event based upon change)	N/A	N/A	On, Off - Manually Deactivated, Off-Automatically

						Deactivated Faulted
Automated Driving System - Minimal Risk Maneuver	Mandatory	-10.0 to +10.0 second relative to time zero	2 (or event based upon change)	N/A	N/A	Yes or No
Automated Driving System - Override	Mandatory	-10.0 to +10.0 second relative to time zero	2 (or event based upon change)	N/A	N/A	List of possible overrides
Latitude	Mandatory	-10.0 to +10.0 second relative to time zero	1 or higher as supported by LSAV and GPS update frequency	WGS84	WGS84 standard error ranges	WGS84 standard ranges
Longitude	Mandatory	-10.0 to +10.0 second relative to time zero	1 or higher as supported by LSAV and GPS update frequency	WGS84	WGS84 standard error ranges	WGS84 standard ranges
Satellite UTC time	Mandatory	Unsigned long – milliseconds since 1970	1	N/A	N/A	N/A
Operating environment static and mobile objects, relative position, longitudinal	If using proximity leading measures	[-10.0] to +10.0 seconds relative to timezero	10	[-50.0m] – [+50.0m] relative position to centre of LSAV (nearest objects)	Sensor estimate position	Position used in LSAV decision making
Bearing (gyroscope)	Mandatory	-10.0 to +10.0 second relative to time zero	1	[0.0 – 360.0]	+- 10 degrees	N/A
Operating environment static and mobile objects, relative position, lateral	If using proximity leading measures	[-10.0] to +10.0 seconds relative to timezero	10	[-50.0m] – [+50.0m] relative position to centre of LSAV (nearest objects)	Sensor estimate position	Position used in LSAV decision making
Operating environment static and mobile objects, speeds,	If using proximity leading measures	[-10.0] to +10.0 seconds relative to timezero	2	0 km/h to 250 km/h	As per accuracy of observed object speeds	As per accuracy of observed object speeds
Operating environment static and	If using proximity leading	[-10.0] to +10.0 seconds relative to	2	[-180.0] to +180.	As per accuracy of observed	As per accuracy of observed object

mobile objects, trajectory,	measures	timezero			object trajectories	trajectories
Operating environment static and mobile objects, classification,	If using proximity leading measures	[-10.0] to +10.0 seconds relative to timezero	2	[static, vehicle, VRU, Moving unknown]	As per accuracy of observed object classification	As per accuracy of observed object classification