Towards self-aware vehicle automation for improved usability and safer automation mediation

De Campos, Gabriel Rodrigues; Knauss, Alessia; Tanov, Nikita; Mano, David; Bakker, Bram; Farah, Haneen; Yuan, Yufei; Andersson, Stefan

DOI
10.1109/IV55156.2024.10588774

Publication date
2024

Document Version
Final published version

Published in
35th IEEE Intelligent Vehicles Symposium, IV 2024

2024 IEEE Intelligent Vehicles Symposium (IV)
June 2-5, 2024. Jeju Island, Korea

Towards self-aware vehicle automation for improved usability and safer automation mediation

Gabriel Rodrigues de Campos, Alessia Knauss, Nikita Tanov, David Mano, Bram Bakker, Haneen Farah, Yufei Yuan, Stefan Andersson

2024 IEEE Intelligent Vehicle Symposium (IV) | 979-8-3503-4881-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/IV55156.2024.10588774

Abstract— **This paper investigates the development of self-aware mechanisms for automated vehicles, introducing the notion of an automation state estimation system. This system is capable to understand its capabilities in a given context, and can leverage that knowledge to estimate the current and near-future automation performance based on internal metrics, as well as external, static (e.g. lane geometry) and dynamic environmental elements (e.g. traffic and weather information). From an application perspective, we consider automation state estimation in the scope of automation mediation, as part of a broader and holistic mediation system, with the goal to tackle challenging aspects related to transitions of control, mode confusion, and driver engagement. We used real-world data for system design, and implemented the proposed automation estimation system in a prototype vehicle. Based on 70 hours of real-world driving, we also validated the performance of the automation state estimation for automation mediation purposes.**

I. INTRODUCTION

Partial automation of the driving task can induce a state of cognitive underload, mind-wandering, and sleepiness [1], [2]. It is therefore crucial to acknowledge that increased automation in automotive systems means that the role of the human driver will change from an active to a more passive and supervisory role which, naturally, will lead to a reduced workload with respect to the primary control task and can create problems with respect to vigilance [3], for instance. To tackle some of the aforementioned challenges, some works investigated how to design the transitions from an Human-Machine-Interface (HMI) perspective. For example, the range of take-over time values depending on the driving context and driver condition has been studied in [4], while [5] approached the effects of communicating potentially valuable information through HMI design. Also, recent research indicates that the safety of the driver's responses, in terms of the visual behavior, and the timing and quality of the driver's actions, depends on the type of vehicle automation [6].

While the above mentioned contributions are important steps within the autonomous driving domain, this paper takes a step further, by proposing an automated system that is (increasingly) *aware* of its own capabilities. This would add significant value in both automation-to-driver and driver-to-automation transitions, and in reducing mode confusion, overreliance, and misuse of automation. This would be similar and analogous to a driver state estimation system (which assesses the driver's fitness), but providing instead an assessment of automation's current and near-future fitness. In addition, we add a very important *predictive* element, determining where and when in the near future automation will likely have good vs. low performance. The idea is that this can be used to clearly and proactively communicate the system boundaries to the driver, therefore creating increased awareness. Furthermore, it can serve as a foundation for "meta-level" monitoring and prediction systems that are able to increase safety by assessing who, among the human driver and the automation, is the fittest to drive, and intelligently select transitions by combining the strengths of automated system's and human driver's capabilities for each situation. While there has been plenty of research on assessing the state of the driver through driver monitoring systems [7], [8], there is yet a gap when it comes to similarly determining the automated driving system's capabilities.

II. RESEARCH SCOPE AND CONTRIBUTIONS

In this paper we investigate the notion of real-time automation monitoring and performance prediction, building upon the concept of automation "self-awareness" [9], already described informally above. This work is structured around the following Research Question (RQ):

- **RQ**: How to design an automation monitoring system that allows to reliably estimate the current and future performance of automation systems at run time, including the assessment of static (e.g, road characteristics) as well as dynamic elements (e.g, traffic, accident, roadworks, and weather information, as well as internal sensor and world model information)?

To answer this exploratory, proof-of-concept research question, we design, implement, and validate a data-driven fitness assessment algorithm for an ADS, which we call here *Automation State Estimation (ASE)* system. Furthermore, we conduct a real-world experiment (including 70 hours with seven test drivers) to assess the performance and utility of the proposed ASE proof-of-concept system.

The focus and contributions of this paper are as follows:

- The development of the ASE concepts, estimation and prediction models, for an automation system's behavior based on both static and dynamic properties of the

* This work was supported by the Mediator project, part of the European Union's Horizon 2020 programme under grant agreement 814735.

G. Rodrigues de Campos, A. Knauss, N. Tanov and D. Mano are with Zenseact, Sweden. E-mail: name.surname@zenseact.com
B. Bakker is with Cygnify BV, Netherlands. E-mail: bram.bakker@cygnify.ai
H. Farah and Y. Yuan are with TU Delft, Netherlands. E-mail: H.Farah@tudelft.nl, y.yuan@tudelft.nl
Stefan Andersson is with Autoliv, Sweden. E-mail: stefan.andersson@autoliv.com

driving environment, as well as the automation system's internal performance parameters.

- The implementation of a prototype system for a real-life experiment on public roads, for both system design and validation purposes.
- A validation of the proof-of-concept ASE system in the context of the automation mediation concept through a real-world experiment.

While the notion of safe-aware automation is not new, the novelty of this work relies on the predictive nature of the algorithm, as well as the development and application of the concept in a real vehicle and driving conditions.

III. Background and related work

This section provides some background elements as well as a brief overview of related literature.

A. The Mediator system and ASE within it

This work was performed in the context of the Mediator project (www.mediatorproject.eu). Broadly speaking, the project focused on the various issues related to automation usage and mediation between driver and automation for vehicle automation systems up to SAE level 4, and considered a broad range of aspects such as human factors, automation performance, and HMI aspects. The core idea of Mediator is to have a central, "meta-level" system that simultaneously and in real time monitors both the driver and the automation system and compares their relative performance, also denoted here as *fitness*. Based on that assessment, the Mediator system recommends transitions of control, and provides relevant information to HMI elements in order to help the driver. Therefore, the ASE system described later in this paper is just one, but yet a crucial and novel, subsystem of this whole Mediator concept, see Fig. 1. For more information on the Mediator concepts, see [10], [11].

B. Related work

This work is based on the notion of self-awareness, often referred to as the system's capability to recognize its own state, possible actions and the result of these actions on the system itself and on its environment [9]. While researched in various scientific domains such as, for example, system design, systems-of-systems or cloud-based systems/services [12], [13], [14], the idea of self-awareness has not gained yet much attention in the automotive domain, most of the approaches being limited to anomalies detection [13],

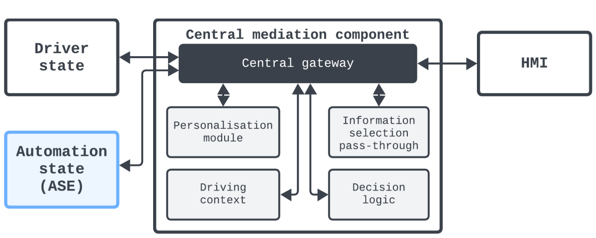


Fig. 1. The Mediator architecture.

[15], or cross-layer designs [9], [16]. To the best of the authors' knowledge, no other research works have approached, from a holistic perspective, self-awareness for an ADS, by providing it with an application context (e.g. automation mediation) and evaluating it in real world conditions.

In this paper we focus on a holistic approach, laying the groundwork for a type of self-awareness for ADS. We present a system capable of comprehensive and predictive automation state estimation capabilities, able to recognize the system's own state and predict future states, based on observed and predicted environmental conditions and the system's own capabilities for these predicted conditions. In other words, self-awareness can be seen as having accurate albeit probabilistic knowledge regarding good/bad driving performance (by the ADS), based on conditions that are "known" to affect the automation capabilities. Since this is a novel and exploratory work, the goal is not to have a perfect and all-encompassing system but rather to create a valid proof-of-concept system and validate it in a real vehicle and in real-world driving conditions. Hence, certain state-of-the-art vehicle automation aspects were excluded such as higher level automation systems, sensor models, and aspects related to high-definition mapping solutions.

IV. Research Methodology

The research methodology considered for this work is summarized in Fig. 2. First, we focus on the design, implementation, and validation of the ASE system on its own. In a second stage, we integrate it within an automation mediation context, and in particular within the Mediator system, together with the remaining Mediator components, see Fig. 1. This step is meant to understand and evaluate whether such self-awareness mechanisms can potentially be integrated within an automation mediation context and to evaluate its performance.

A. Data Collection Setup

1) Test drivers: The research approach of this work leverages data collection with test drivers for both: i) the design and evaluation of the ASE system and ii) the validation of the Mediator system principles on public roads. Ethical approval was obtained from the Mediator project's ethical review board, exploiting also the vehicle partner's (Veoneer) special Swedish testing permission. The study included seven test drivers, which were employees of either Veoneer, Zenseact, or Autoliv and that could drive the prototype vehicle on public roads. The test drivers had a history of working with safety system development and are used to driving prototype vehicles. Drivers were between 27 and 58 years old (average age is 42), one driver was female and the other six were male.

2) Prototype vehicle: The vehicular platform used in this work consists of the LIV (Learning Intelligent Vehicle) research vehicle from Veoneer, previously presented at CES. LIV is based on a Volvo XC90 production vehicle and is equipped with a basic level ADS, a real (Level 2, "Continuous Mediation", "Pilot Assist") automation system. For the sake of the design and evaluation of the ASE, the prototype

Fig. 2. Design and validation methodology of the ASE system on its own and in the context of the Mediator automation mediation system.

platform had to include all major concepts and components of the Mediator system, see Fig. 1.

All components communicated to each other using clearly defined interfaces and a real-time LAN messaging protocol, allowing to evaluate if the ASE would be suitable to be embedded in such a real-world system, where the ASE is embedded in a Mediator system in conjunction with the driver state and driving context modules. This allowed us to demonstrate and evaluate the proposed ASE algorithms as well as the broader Mediator logic. In the context of the Mediator project, this vehicle was one of the main prototypes, of particular importance for testing the automation state component running in real time, in a real vehicle and driving conditions.

Data relevant for the driving task and the driver are displayed on the center stack, see Fig. 3, such as the driving mode, time budgets (explained below), instructions for the driver (e.g., keep hands on steering wheel), as well as context information (e.g., roadworks zones). Additional LED strips on the dashboard and the steering wheel, as well as ambient lighting, indicate the current driving mode, the remaining time, and the intended or necessary changes in the driving mode (e.g., by different colours, brightness, and pulsation frequencies). Furthermore, a sound system was also implemented for audio alerts.

B. Design, Implementation, and Validation of ASE System

The system design of the ASE system is based on real-world data that was used to determine the algorithm logic. Hence, as a first step, data from real public roads was collected on a first route, denoted as Route 1. Based on this data, the ASE algorithm was designed and revised to consider the ADS system capabilities and ensure good performance against this initial data. After the implementation of the algorithm, the ASE was validated on a complementary data set, from other public roads where the ASE system was additionally run on, denoted later as Route 2. By using, for validation purposes, a different route from the one used for the initial data collection, we were able to observe how well the system is able to adapt to various road conditions and different combinations of those.

Fig. 3. HMI example highlighting the Time To Automation Fitness (TTAF) [left figure] and the Time To Automation Unfitness (TTAU) [right figure].

1) ASE System - Data Collection (on Route 1): We consider the ADS available on the prototype vehicle as a black box system, without a detailed understanding of its capabilities and performance in different context conditions. To predict its own performance and capabilities (i.e., to create self-awareness), a small real-world dataset was gathered and analysed. This data was retrieved on Route 1, corresponding to about 60 km and one hour of driving, and comprising the segment Vårgårda-Vara (through road E20) and the segment Vara-Herrljunga-Vårgårda (through road 181). It is composed of a variety of road types such as city, small to large rural roads, as well as highway with a wall divider. There were construction zones, intersections, roundabouts, and road links along the route, as well as segments with either degraded or missing lane markings. A majority of the data collection was performed in dry weather and light traffic conditions, even though a small part of the data set contained rainy weather (or wet road surface), as well as dense traffic (targeting rush hours), in order to understand how the system behavior varies and generalizes with respect to different dynamic conditions, for the same static parameters.

The main goal of this data collection was to expose the ADS to diverse enough situations such that the ASE prediction model is able to generalise, as much as possible, to other routes. The logged data in the prototype vehicle consisted of information provided by the ADS itself, such as automation availability and activation states, detected objects and road geometry, vehicle information such as vehicle speed and location and, lastly, a front looking camera used for offline annotations of both static and dynamic conditions. This dataset corresponds to 20 hours of driving, and is mainly aimed at providing availability, perception, and environmental information (e.g. detected objects, lane markings, etc).

2) ASE System - Algorithm Design: In order to analyse the relation between the automation availability (retrieved on the dataset mentioned before) and the different environmental cues, a correlation analysis was performed. This process was done through semi-automatic annotation of the dataset in terms of the presence/absence of specific elements, such as, for example, lane markings, road dividers, intersection or roundabouts, or the presence of leading vehicles, for instance.

The annotated dataset was used to define the baseline used later for benchmark purposes. First, the minimally limiting definition of the baseline was taken by restricting the automation availability to a subset of conditions in which the ADS reported L2 automation as available at least once. Secondly, it was then further constrained to limit the automation availability to highway and rural roads only, as urban roads were deemed (by this research group) as too complex and an unsafe environment for this specific ADS system to drive, and was therefore deemed not suitable for automation driving. As a result, the following subset of the environment conditions has been chosen as the initial baseline, for which the automation may be considered available/suitable:

- **Road type:** highway, major or minor rural road
- **Lane markers:** center lane marker visible
- **Curvature:** straight and lightly curved roads
- **Turns:** no turning between two roads on the planned route and no roundabouts

As the tests and data collection were performed, the model was further refined to improve the estimations' accuracy against the actual availability of the ADS. But while the aforementioned initial limitations imposed on the automation availability haven't changed, the variable part of the system, represented by the variations in ADS behaviour dictated by technical limitations of sensors and other internal components, allowed to further improve the estimation model.

3) ASE System - Algorithm Implementation: Once the automation estimation logic was developed and the algorithm design concluded, the algorithms were implemented as software units on the prototype vehicle platform. Within the Mediator framework described before, the ASE system was developed in parallel to the related driving context module, that feeds the ASE system, see Fig. 1. The driving context module stores and provides runtime information regarding the route, combined, through map matching, with static environmental information previously stored on the route.

4) ASE System - Design Validation: After we have finalized the algorithm's implementation and determined how the ASE system is expected to work, it was essential to understand and quantify the performance and accuracy of the implemented ASE module. Such an analysis was articulated around the following guiding question:

- **Guiding question 1:** How accurate are the automation fitness and unfitness predictions when compared to the actual automation system's (un)availability?

A new data collection was performed, consisting of 16 hours of driving on a new route, denoted as Route 2, in preparation for the validation of ASE as part of the Mediator setup. Leveraging this data, the algorithm was integrated in a simulation environment for the purpose of iteration of the system design. This analysis was necessary to determine the proper design of the ASE system and to identify ways to improve it to suitable standards. It is worth noting that weather and traffic information was not available at runtime. However, their pertinence and usability for the ASE system has been evaluated in the simulation environment.

C. ASE in the Context of the Mediator System

The ASE was also implemented into the prototype vehicle in order to study the automation mediation aspects. The ASE was evaluated under different driving contexts in order to validate the ASE system design, as well as for assessing and validate how well the ASE system can predict automation performance. A real-world system validation of the ASE concepts was conducted based on 70 hours driving data.

1) Mediator System - Integration Testing: For this purpose, the ASE was integrated with the multiple remaining modules in the aforementioned prototype vehicle. As further detailed in [11], the automation mediation task is allocated to several main components, which communicate to each other using a clearly defined interface. Such components include the decision logic, driver state, driving context, as well as HMI manager, among others, see Fig. 1.

The central mediation component plays a central integration, decision, and communication role, and interacts with each of the other main components, in particular the driving context and decision logic components. All components were running in real time and in real conditions to demonstrate the overall functionality of the Mediator system and the ASE module, as a crucial part of the design, and the correct interaction with the other modules and vehicle automation systems. All the main hardware (apart from the driver state cameras and the HMI components) were placed in the vehicle's trunk, consisting of the power supply, the computer network equipment, as well as several computers running the different components.

2) Mediator System - Data Collection (on Route 2): In order to validate and analyze both the ASE system (at the sub-system level), as well as within the Mediator complete system (at the "meta" decision making level), a real-life test campaign and a quantitative analysis was performed. The ASE system's outputs, transformed to usable driver information (see below), were communicated to the driver via a HMI setup as illustrated in Fig. 3.

This study used seven test drivers, for a total of 70 hours of driving. The participants drove the prototype vehicle on a specified route (Route 2) that included a mix of road types and driving conditions. Route 2 included the segment Vårgårda-Ingared (through road $E20$), as well as the segment Ingared-Alingsås (through road 180). Each cycle consisted of 1 hour driving, and each drivers performed two driving cycles in a row (i.e, 2 h driving). The route contains certain stretches known to affect automation performance. In those stretches, the driver might wish to deactivate assisted driving, or the automation system may reach its (not precisely known) design limitations and deactivate by itself.

The choice of the route was made in order to ensure sufficiently long stretches of automated driving that fulfilled the use cases/scenario requirements for the Mediator study. The route included city road stretches, highways, as well as rural road stretches. To ensure the safety of the drivers, the chosen route did not include many road work segments.

3) Mediator System - Quantitative analysis: In order to evaluate the design choices and the performance of the proposed algorithms, we performed a quantitative analysis. This analysis was articulated around the following two questions:

- **Guiding Question 2:** How accurate are the automation state predictions in terms of the time to (un)availability for the next route section?
- **Guiding Question 3:** Based on the ASE estimations, do drivers switch on/off automation when automation becomes (un)available?

While Guiding Questions 1 and 2 mainly concern the ASE system' performance, Guiding Question 3 tackles the behavior of the ASE within the Mediator system context.

V. Self-aware automation: ASE system

This section presents the proposed Automation State Estimation (ASE) system. Without leveraging design knowledge of the underlying logic of the ADS system itself, as such insights are often not possible nor properly documented, we consider it as a black-box system. The focus of this work is therefore to identify, contextualize, and predict the limiting dimensions and context cues of the ADS performance.

A. Principles

This work is based on the notion of self-awareness [9], as a means for a system to cope with complexity and to recognize its own state and possible actions. As groundwork for self-awareness for ADS, we present here a comprehensive and predictive automation state estimation system, able to recognize the system's own state and predict the future driving capabilities (in contrast to existing HMI elements that solely report the current availability of certain features, such as lane keeping assistance).

In the context of this paper, good/safe automation performance is defined when the system is available (and would not make the human driver uncomfortable if activated) under traffic conditions that were intended for by the system designers. For instance, the considered L2 automation "Pilot Assist" might be available for activation in city environments, though such systems are likely not intended to be used in cities but rather highways or large rural roads. While the system may not be technically prevented to be activated, it is argued here that such an activation is deemed not safe due to the expected low driving capabilities. Factors known to affect (safe) automation performance can be both internal conditions and external and environmental conditions:

- **External/environmental conditions,** such as adverse weather, dense traffic or roadworks, and even road types for which ADS either is not suitable (upon the definition of the Operational Design Domain (ODD)). All of those aspects can be measured and determined in this work as part of the driving context.
- **Internal conditions,** such as sensor shortcomings, computation limitations, and hardware failures.

Leveraging these notions, we propose here an approach to quantify how good/safe the automation system is expected to perform, by taking into account i) knowledge from the system itself (i.e., expected automation availability), ii) detection of driving context cues and iii) a definition of safe ODD, see Fig. 4 for an illustration. We also introduce the fitness score metric, an internal variable aiming at quantifying the automation performance on a scale from 0 to 5, given a set of conditions and self-assessment performance measures. Here, 0 denotes no automation availability (i.e., no automation fitness) and 5 full availability (i.e., high automation fitness). The underlying idea is that measurable and predictable internal (i.e., performance self-assessment) and external (i.e., driving context) conditions can be used as indications for upcoming availability or degraded performance, see Fig. 5 for an illustration.

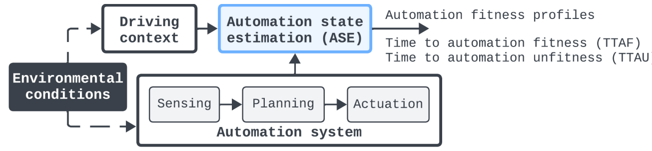


Fig. 4. Overview of the interplay between the environmental conditions, driving context, and the ASE system (in blue).

B. Time-To Fitness and Unfitness and Time Budgets

A central Mediator concept, crucial to make the driver fitness and automation fitness assessment *comparable*, is to express "fitness" (performance) in terms of *time*. In essence, the idea is to estimate the following two values for both human and automation driving:

- **Time To Driver/Automation Unfitness** (when Driver/Automation is *active*). It represents the estimated amount of time until Driver/Automation can no longer function safely. For the driver, this is based on estimated distraction and fatigue, determined by a driver monitoring system. For automation, this corresponds to the dynamically estimated remaining time, performed by the newly proposed ASE system, for which the automation can operate safely.
- **Time To Driver/Automation Fitness** (when Driver/Automation is *inactive*). It represents the estimated amount of time until Driver/Automation can function safely (again). For the driver, this is based on estimated distraction and fatigue, determined by a driver monitoring system. For automation, this corresponds to the dynamically estimated time (given the current planned or predicted route) until the automation reaches the ODD limits and is predicted to be able to function safely, as assessed by the newly proposed ASE system.

Each of the above metrics can, potentially, be extended by upper and lower bound ("confidence") values.By bringing both driver and automation fitness into a "shared space" of *time*, one makes them comparable for the central Mediator decision making logic. This Time-To concept has the important benefit of including, by design, a limited near-future forecast of predicted fitness into the estimation algorithms; and allowing "time budget" values to be communicated to the driver through the vehicle's HMI (see Fig. 3), which inform the driver about how much longer automation is

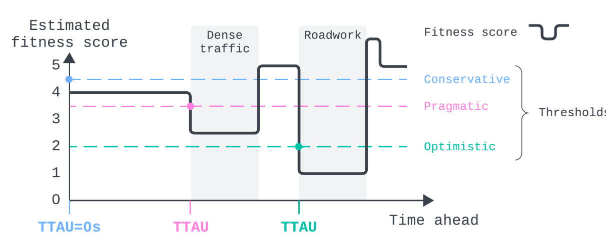


Fig. 5. Illustration of the automation fitness estimation functionality.

operational or when, in the near future, automation can likely be turned on. Thus, the estimation of the current and upcoming automation fitness score is used to derive the main output of the automation state estimation system, namely the Time-To-Automation-Fitness (TTAF) and Time-To-Automation-Unfitness (TTAU), corresponding to the predicted time until the automation system is deemed to be fit and unfit to drive, respectively (see Fig. 4).

From an implementation perspective, the TTAF and TTAU metrics are calculated for the current and the next automation mode stretch, and computed whenever an automation mode switch is possible. The following elements are considered: i) distance from current position to the first encountered automation switch; ii) distance from the first encountered automation switch to the next switch. These distances are then converted to time spans using the estimated speed and fed to the HMI, as illustrated in Fig. 3 and Fig. 6.

C. Driving Context and Self-Assessment Measures

Automation availability is predicted based on the static and dynamic road conditions. Static conditions are strongly linked to the road surface and topography and are therefore not subject to changes during the drive. Dynamic conditions are related to real-time state of the road environment (e.g., traffic, weather or roadworks information) as well as internal automation sensor performance and detection information, and are expected to be dynamically updated during the drive.

From an implementation perspective, the static elements considered in this research were mainly road class, permanent lane markings, and existence of roundabouts and intersections. Dynamic aspects such as roadworks (present on the test route during the trials), weather and traffic information were also implemented. Road-related information was linked to 10 m road segments over the different routes based on high-frequency map-matching.

Internal automation sensor performance was considered as well, by measuring whether the automation system had automation availability based on accurate lane detection. Within the Mediator system implementation on the prototype vehicle, all the above information was gathered as a look-up table, updated in real time and provided to ASE system for fitness score calculations, and in essence providing all the information for all the different points of the driving route. Such a table or similarly dynamically updated data structure could, for future use and implementations, be based on any high-definition mapping solution (for the static aspects) and, for the dynamic aspects, on any underlying vehicle automation system, infrastructure elements, or weather stations.

Fig. 6. HMI implemented in the prototype vehicle, showing the relevant time budget information derived from the proposed ASE, the LED bars show driving mode, as well as the remaining time budget by gradually shrinking.

D. Coefficients and Automation Scores

Upon the information available on the static and dynamic elements for each road segment, the goal of ASE is the calculation of the automation availability estimates, as well as the degradation reason that indicates the principal cause behind the loss of automation availability. To determine the overall automation availability score, each of the different road conditions is assigned a coefficient, determined based on a mix of engineering judgement and real-world automation performance data for the conditions in question. The coefficients, determined based on data from Route 1, are provided in Table I. The values considered multiple automation fitness profiles and were determined by maximizing the accuracy of the model in terms of the baseline performance. The degradation reason is picked based on the lowest coefficient that degrades the automation availability, at a given moment. Note that more complex (AI-) and data-driven approaches could be used for this task in future research efforts.

In this work, the automation availability for each road edge is determined by a decimal score in the range from 0.0 to 5.0, as illustrated in Fig. 5. The score is determined based on the driving context and road conditions. Each value of each road condition is assigned a coefficient ranging from 0.0 (critical condition disabling the automation) to 1.0 (perfect condition, does not affect the automation). The automation score is calculated as follows:

$$As = 5\ \Pi_1^n\ x\ , \tag{1}$$

where n represents the number of all the road conditions and x the coefficient assigned to the value of given condition.

E. Thresholds and Automation Profiles

In the scope of this paper, automated driving is considered to be available (to the driver) if the score is greater than or equal to a given threshold. Such threshold levels are aimed to capture the variations in the automation performance, and to provide granularity to the situation assessment, rather that

TABLE I
AVAILABILITY COEFFICIENTS

Condition	Coefficient
Road class	
Highway-main	1.0
Highway-connect.-link	0.0
Rural	1.0
Urban	0.0
Roadwork presence	
True	0.75
False	1.0
Left lane marker	
True	1.0
False	0.98
Center lane marker	
True	1.0
False	0.0
Right lane marker	
True	1.0
False	0.98
Ambiguous lane markers	
True	0.99
False	1.0

Condition	Coefficient
Curvature	
No/low curvature	1.0
High-curvature	0.5
Intersection	
True	0.98
False	1.0
Road change/Junction	
True	0.0
False	1.0
Roundabout	
True	0.0
False	1.0
Weather	
Clear weather	1.0
Light-rain	0.83
Heavy-rain	0.75
Traffic	
Free-flow	1.0
Congested busy	0.83
Incident slow	0.83

leveraging a binary classification paradigm. In order to also incorporate the notions of uncertain estimations and different confidence levels, we consider three availability profiles:

- **Pragmatic**: Using reported automation availability of the prototype vehicle as the baseline.
- **Conservative**: Assuming that automation will perform worse than usual.
- **Optimistic**: Assuming that automation will perform better than usual.

An illustration of the interplay between the different automation profiles is given in Fig. 5.

VI. Evaluation and Validation Results

This section presents the evaluation and validation results for i) the Automation State Estimation module and ii) the performance of the ASE within the Mediator system context. That analysis and discussion is articulated around the guiding questions and research methodology presented in Sec. IV. The automation availability thresholds are defined in Table II.

TABLE II

AUTOMATION AVAILABILITY THRESHOLDS

Automation profile	Threshold
Conservative	4.95
Pragmatic	3.76
Optimistic	1.99

TABLE III

ASE PREDICTIONS PERFORMANCE

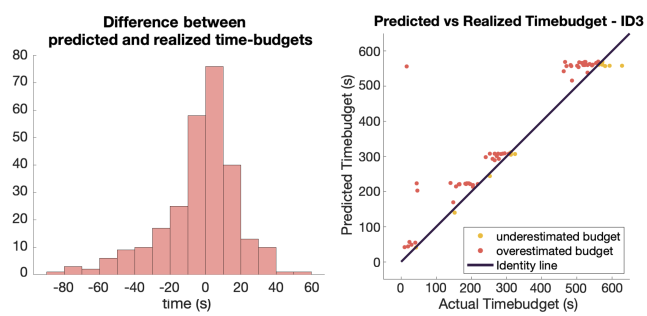
Automation availability	Precision	Recall	F1 score
Available	0.992	0.996	0.994
Unavailable	0.988	0.975	0.981

Guiding Question 1: How accurate are the automation fitness and unfitness predictions when compared to the automation system's actual (un)availability? Accuracy numbers were calculated to assess and quantify the overall performance of the ASE system. The accuracy assessment was done by calculating how much the ASE predictions differ from the ground truth, as defined in Section IV. Table III presents the metrics of the automation estimations in the scope of the Guiding Question 1. There, the F1 score shows that both segments with available and unavailable automation are predicted with high accuracy. It is worth noting that these results may be due to the iterative improvements made to the route data, as an attempt to overcome map-matching and route annotation issues, aspects to which the ASE is sensitive. We emphasize, though, that the purpose of this work was not demonstrating high accuracy out of sample generalization, but rather an early proof-of-concept to be evaluated by test drivers.

Guiding Question 2: How accurate are the automation state predictions in terms of the time of (un)availability for the next route section? To investigate the time budget accuracy as formulated in Guiding Question 2, a comparison was conducted between the elapsed times between the moments the first TTAU and TTAF time budgets were computed and the actual time until that zone is actually reached by the vehicle. The first TTAU or TTAF time budget values were computed after each change in the automation availability zone, recall Section V for a definition of road segments and availability zones. Fig. 7a illustrates that the time-budget mismatch falls within the range of 60s, with most of the values within the range of -20s and $+20$s. Similar to before, these accuracy results should be viewed as an indication of what could be achieved in practice, in the context of this proof-of-concept work. The mismatch is computed over the time budgets for all road segments, knowing that each segment has different lengths, going from a few seconds to up to 10 minutes (the actual time budget in Fig. 7a correlating to the length of the segment). While not visible from the figure, it has also been observed that the longer the road segment, the more uncertain the estimations become. Hence, for shorter stretches, the mismatch is small, and since the ASE system updates the estimates on a regular basis, longer mismatches are not perceived by the drivers, since the ASE estimates are updated as the automation switch gets closer. This was confirmed by the drivers' feedback, that perceived the time budgets as very accurate.

For a deeper look into this aspect, the results from a particular driver are also shown in Fig. 7b. The results of these comparisons mostly align around the identity line, therefore indicating a generally good match between the predicted and actual time budgets.

Guiding Question 3: Based on the ASE estimations, do drivers switch on/off automation when automation becomes (un)available? This evaluation is aimed at determining whether and how drivers use the proposed ASE module as a self-awareness provider for automation mediation, in the context of the Mediator system. The moment when drivers switched off the automation, when approaching the end of an automation zone/period, was measured both for the conditions of "end of ODD" and for "roadworks". Fig. 8 shows that the mean remaining time in "roadworks" conditions is comparable to the "end of ODD", and is typically (median values) approx. 20 to 25 seconds before automation is really no longer available. Importantly, the Mediator system always gave its first proactive notification to prepare for manual



(a) (b)

Fig. 7. Comparison between predicted and actual time budgets: (a) Histogram showing the magnitude of the time-budget mismatch. (b) Predicted time budget versus actual time budget for the ID3 driver.

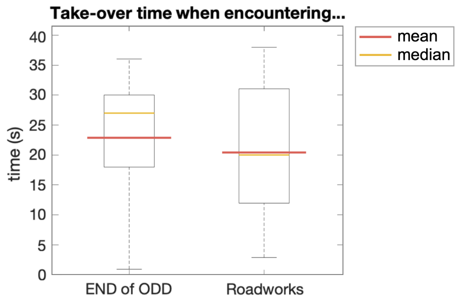


Fig. 8. Remaining time to unfitness (at the instant of human deactivation).

driving 30 seconds before the predicted end of automation. Thus, the actual median (typical) take-over moments are in line with those first, proactive suggestions. And since there is still so much time left, these results suggest that there is sufficient time for a safe and smooth transition of control. Albeit the preliminary results (these were, after all, only seven internal participating company test drivers instructed to test the ASE and Mediator systems), this indicates that such proactive notifications can in practice be useful.

We have also analysed the percentage of the highway section in which drivers used Pilot-Assist, normalized to the total length in which Pilot-Assist was available. The data shows that the automation system was available and the driver activated it in 87% of the highway environments, which indicates that the Mediator system encourages the usage of automation and that the ASE's time budget estimations were used and followed by the drivers.

VII. Conclusions and Future Research

This paper presents an automation state estimation system aimed at providing self-awareness and run-time performance prediction capabilities for vehicle automation systems. We present the underlying principles and mechanisms, and contextualize this development within a broader, holistic automation mediation system (i.e. the Mediator). We show that it is possible to design an automation monitoring system that allows to reliably estimate the current and future performance of automation systems at run-time, including the assessment of static as well as dynamic elements. In order to validate the proposed design and its merits, we performed real-traffic evaluations of the ASE system, showing real-world feasibility and potential in the context of the Mediator system.

While this paper focuses on self-aware automation systems, it is not claimed nor aimed to be a complete study, in particular considering the small number of participants in the evaluation. It should therefore be seen as an initial work on such concepts and its merits. This work is a successful proof-of-concept of the proposed ASE system, implemented and tested in a real vehicle and public roads. Future research should focus on the validation of the proposed ASE system in larger datasets, higher automation levels, richer runtime inputs to the ASE including more external and internal dynamic elements that affect automation availability, as well as a deeper study of the hand-over procedures and the inherent human factors (timings, HMI, etc).

VIII. Acknowledgements

The authors would like to thank all the institutions and individuals that contributed to this work, and in particular J. Thor, C. J. Pålsson, J. Lang, D. Nylen, J. Holmsten, M. Sebu, P. Thalya, E. Kotelnikov, M. Brännström, N. Berger, N. Rauh, A. Larsson, E. van Grondelle, and I. van Zeumeren.

References

[1] M. Körber, A. Cingel, M. Zimmermann, and K. Bengler, “Vigilance decrement and passive fatigue caused by monotony in automated driving,” *Procedia Manufacturing*, vol. 3, pp. 2403–2409, 2015.

[2] D. Miller, A. Sun, M. Johns, H. Ive, D. Sirkin, S. Aich, and W. Ju, “Distraction becomes engagement in automated driving,” *Human Factors and Ergonomics Society Annual Meeting*, vol. 59, no. 1, pp. 1676–1680, 2015.

[3] V. A. Banks and N. A. Stanton, “Analysis of driver roles: modelling the changing role of the driver in automated driving systems using east,” *Theoretical Issues in Ergonomics Science*, vol. 20, no. 3, pp. 284–300, 2019.

[4] 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,” *Transp. Research Part F: Traffic Psychology and Behaviour*, vol. 64, pp. 285–307, 2019.

[5] A. M. Tinga, D. Cleij, R. J. Jansen, S. van der Kint, and N. van Nes, “Human machine interface design for continuous support of mode awareness during automated driving: An online simulation,” *Transp. Research Part F: Traffic Psychology and Behaviour*, vol. 87, pp. 102–119, 2022.

[6] L. Pipkorn, “The driver response process in assisted and automated driving,” PhD thesis, Chalmers University of Technology, 2022.

[7] M.-H. Sigari, M.-R. Pourshahabi, M. Soryani, and M. Fathy, “A review on driver face monitoring systems for fatigue and distraction detection,” *International Journal of Advanced Science and Technology*, vol. 64, pp. 73–100, 2014.

[8] A. Kashevnik, R. Shchedrin, C. Kaiser, and A. Stocker, “Driver distraction detection methods: A literature review and framework,” *IEEE Access*, vol. 9, pp. 60 063–60 076, 2021.

[9] J. Schlatow, M. Moostl, R. Ernst, M. Nolte, I. Jatzkowski, M. Maurer, C. Herber, and A. Herkersdorf, “Self-awareness in autonomous automotive systems,” in *Design, Automation Test in Europe Conference Exhibition (DATE), 2017*, 2017, pp. 1050–1055.

[10] A. Fiorentino, B. Bakker, D. Brase, J. Karlsson, A. Knauss, T. Santhià, P. Thalya, A. Toffetti, E. van Grondelle, and D. Wang, “Final in-vehicle prototypes, deliverable D2.10 of the H2020 project mediator,” Tech. Rep., 2022.

[11] B. Bakker, A. Knauss, G. Rodrigues de Campos, N. Tanov, S. Andersson, D. Mano, E. Van Grondelle, M. Spaan, and S. Enhuber, “Functional and technical specifications of the mediator system, deliverable D2.11 of the H2020 project mediator,” Tech. Rep., 2022.

[12] F. Faniyi, P. R. Lewis, R. Bahsoon, and X. Yao, “Architecting self-aware software systems,” in *IEEE/IFIP Conference on Software Architecture*, 2014, pp. 91–94.

[13] D. T. Kanapram, P. Marin-Plaza, L. Marcenaro, D. Martin, A. de la Escalera, and C. Regazzoni, “Self-awareness in intelligent vehicles: Feature based dynamic bayesian models for abnormality detection,” *Robotics and Autonomous Systems*, vol. 134, p. 103652, 2020.

[14] M. Ravanbakhsh, M. Baydoun, D. Campo, P. Marin, D. Martin, L. Marcenaro, and C. Regazzoni, “Learning self-awareness for autonomous vehicles: Exploring multisensory incremental models,” *Trans. on Intelligent Transportation Systems*, vol. 22, no. 6, pp. 3372–3386, 2021.

[15] M. Ravanbakhsh, M. Baydoun, D. Campo, P. Marin, D. Martin, L. Marcenaro, and C. S. Regazzoni, “Learning multi-modal self-awareness models for autonomous vehicles from human driving,” in *International Conference on Information Fusion (FUSION)*, 2018.

[16] A. Sadighi, B. Donyanavard, T. Kadeed, K. Moazzemi, T. Mück, A. Nassar, A. M. Rahmani, T. Wild, N. Dutt, R. Ernst, A. Herkersdorf, and F. Kurdahi, “Design methodologies for enabling self-awareness in autonomous systems,” in *Design, Automation Test in Europe Conference Exhibition (DATE)*, 2018.