This article has been accepted for publication in IEEE Transactions on Intelligent Vehicles. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIV.2023.3269156

State of the Art: Ongoing Research in Assessment Methods for Lane Keeping Assistance Systems

Sijie Wei, Peter E. Pfeffer, and Johannes Edelmann

***Abstract*— A safe, comfortable and intelligent Lane Keeping Assistance System (LKAS) is fundamental on the way to realizing full automated driving. In spite of the current high market penetration of LKAS, vehicle manufacturers, suppliers and researchers are still working on improving the system to reach a higher customer acceptance. To achieve this goal more efficiently, a systematic and ideally objective assessment procedure of the system needs to be established. Till date, a thorough review study of the assessment procedures of LKAS does not exist, especially of those which also take the subjective impression of the driver into consideration. The motivation of this paper is to benchmark the state-of-the-art assessment methods of LKAS and identify where problems and research potentials lie. The paper summarizes the relevant characteristics of LKAS based on representative research in the past decades. It also compares the characterization of LKAS with the field of steering feel and vehicle handling in the aspects of procedure, strategy, as well as their advantages and disadvantages. This paper contributes to transferring the know-hows and valuable experiences from other well-developed fields of vehicle dynamics into the rapid-evolving branch of Advanced Driver Assistance System (ADAS), in order to help standardizing its assessment procedure and give insights on system design from a human-centered perspective in the early development phase, thus shortening the ADAS development cycles.**

***Index Terms*— ADAS, LKAS, subjective assessment, objective assessment, objectification method**

I. INTRODUCTION

ASSISTED and automated driving has been identified as one of the technologies of the future. According to estimates, this market will grow at an annual rate of 12-14% [1]. In addition, companies from the IT industry, which are not traditionally active in the automotive sector, are actively investing in research into technologies for assistance systems and automated driving (AD). Besides, due to the danger of losing touch with technology, a majority of the companies in the automotive industry are currently working on this topic. According to the classification of Automation Levels from Society of Automotive Engineers (SAE) (see Table 1), level 2 (partial automation) systems, typically Lane Keeping Assistance System (LKAS) and Adaptive Cruise Control (ACC), have reached a high market penetration rate. In July 2021, a regulation legalizing vehicles equipped with level 3 (high automation) systems up to 60 km/h came into force in Germany [2]. The German Original Equipment Manufacturer (OEM) Mercedes-Benz has brought the first level 3 system *Drive Pilot* to the market. When upgrading the current technology to higher automation levels, a reliable lateral guidance function plays a central role.

LKAS observe the environment, currently mainly using vision-based technologies, and detect potential unintended departure from the ego lane. They then intervene and help the driver to steer the vehicle back into the lane. The series products on the market are usually named brand-specifically. Example of systems of level 2 are *Active Lane Assist* of Audi, *Autopilot* of Tesla, *Lane Trace Assist* of Toyota etc. For a better clarity, we will use the term "Lane Keeping Assistance System" with the abbreviation "LKAS" in this paper. Here, systems with lateral control functions in all automation levels are included.

TABLE I
LEVELS OF DRIVING AUTOMATION ACCORDING TO [3]

	Level	Name	Example Features
	0	No Driving Automation	Automatic Emergency Braking (AEB)
Driver Support	**1**	Driver Assistance	Lane Centering **OR** ACC
Driver Support	**2**	Partial Driving Automation	Lane Centering **AND** ACC at the same time
Automated Driving	**3**	Conditional Driving Automation	Traffic Jam Chauffeur
Automated Driving	**4**	High Driving Automation	Local driverless taxi
Automated Driving	**5**	Full Driving Automation	Same as level 4, but feature can drive everywhere in all conditions

The current level 2 LKAS on the market, according to the subject studies in [4], still does not fully meet the requirements of the customers. Though several European as well as international standards were introduced in the last two decades to test the LKAS, they are to be considered as a legal guideline. A universally accepted reference for designing and tuning the system has not been systematically established. This means, the assessment process of LKAS is still far from standardized. To speed up the development cycle and reduce the cost while targeting customer needs, more effective and

(Corresponding author: Johannes Edelmann)
Sijie Wei and Johannes Edelmann are with the Institute of Mechanics and Mechatronics, Vienna University of Technology, Getreidemarkt 9, 1060 Vienna, Austria (e-mail: e11946216@student.tuwien.ac.at, johannes.edelmann@tuwien.ac.at).
Peter E. Pfeffer is with the Department of Mechanical Engineering, Munich University of Applied Sciences, Lothstraße 64, 80335 Munich, Germany (e-amail: peter.pfeffer@hm.edu).

efficient development methods are required throughout the whole Product Evolution Process (PEP). Following other branches in the automotive industry, the attempt to supplement or even replace subjective assessments with objective assessments has started in the field of ADAS development.

This paper reviews current research in assessment methods for LKAS and other ADAS for automated driving. It analyzes the requirements for ADAS, mainly LKAS, from a human-centered point of view and provides insight improve future ADAS/AD development to enhance customer acceptance. It also provides a clear view on the heterogeneous research approaches in this field. Furthermore, this paper benchmarks the state of the art in the field of classic vehicle dynamics characterization, especially considering steering feel and handling, with the aim of transferring the experience and know-hows into ADAS development and assessment. Based on the discussion of current research trends and their results, deficiencies and research potentials towards an objective, systematic and cross-domain assessment method are identified. This paper contributes to standardizing the assessment and calibration procedure of LKAS and ADAS in order to meet the complex subjective requirements of the driver.

This paper is structured as follows. The second chapter gives a brief introduction of LKAS, including its theoretical background, classifications and the state-of-the-art technologies. The third chapter summarizes the most relevant characteristics of LKAS based on previous publications. Chapter four gives an overview of different assessment methods in vehicle development. Chapter five to eight introduce and analyze these methods of assessing the system characteristics individually, where literatures from the well-established fields of steering feel and handling assessment are regarded as a reference point for the new territory of ADAS and AD. The discovered results, achievements and limitations in the previous works, as well as potentials for future researches are discussed and summarized in the last chapter.

II. Background of LKAS

Due to the increasing technical possibilities, ADAS and AD have recently gained increased attention. Today's assistance systems can support the driver with information for certain parts of the driving task (information systems), relieve them of certain subtasks (comfort systems) or help them to cope with critical driving situations more safely (safety systems) [5]. In this context, the integration of driver assistance and automated driving functions has significantly increased the system complexity of the vehicle. LKAS is a combination of the three functionalities mentioned above. It supports the driver in keeping in lane and avoiding unintentional lane departure or takes over these tasks during automated driving.

The introduction of LKAS has on one hand reduced the accidents caused by unintentional lane departure. Lane departure crashes are mainly made up of 3 major crash types related to lateral vehicle movement, namely head-on, single-vehicle, and overtaking/lane changing crashes. These have been identified as one of the most common crash types in many countries [6]. The study [7] analyzed the data on fatal passenger car crashes in 2014–2016 in Finland. The results indicated that LKAS is able to deliver notable road safety benefits. This statement could be confirmed by a more current study [8], which estimated that LKAS was effective in reducing the overall number of target population crashes by 60% ± 16%. On the other hand, the system is able to relieve the driver from certain mundane routine tasks in the road traffic, such as keeping in lane in long distance travels or during traffic congestions. This aspect is gaining more relevance starting from Automation level 3.

According to the 3-level driver behavior model from Donges [9], LKAS takes over the driving tasks on the “guidance” and “stabilization” level. A basic lateral controller can be designed based on the concept of this driver behavior model (see Fig. 1). On the “guidance” level, the system derives the target values, normally the target path and/or the target velocity of the vehicle, with respect to the environmental information such as the previewed road curvature. The system then intervenes with an open-loop control in an anticipatory way. On the “stabilization” level, the system corrects the deviation from the reference values (lateral position, heading angle etc.) in a closed-loop manner [5].

The LKAS on the market can be divided into Type I and Type II according to their intervention strategy. Type I system only intervenes when the vehicles are near the lane markings and tend to depart the lane. It leaves a corridor around the center of the lane where the driver can drive freely without any support. It is to be regarded as a lane departure prevention. Type II system, on the contrary, supports the driver in keeping the vehicle near the center of the lane permanently, as long as the vehicle is within the defined Operational Design Domain (ODD). Type II system is to be considered as a lane centering support or an automated lateral guidance, which is more similar to what highly automated driving (HAD) or AD is equipped with.

The state-of-the-art LKAS of commercial vehicles nowadays are vision-based, which rely mainly on a camera for lane and object detection, potentially combined with Radar and Lidar. Research projects also make use of technologies like Laser Scanner and digital maps for higher resolution. However, the use of camera sensors has several advantages in particular in AD applications, e.g. considering the available infrastructure for traffic monitoring that is mainly based on imaging [5].

LKAS collects information of ego vehicle’s relative position to the road, including the distance and relative orientation to lane markings or road edges. Together with the vehicle dynamics signals of the ego vehicle, such as velocity and acceleration, the system detects potential lane departure by means of Time to Line Crossing (TTLC) and/or Distance to Line (D2L). The simplest way to calculate TTLC, denoted t_{LC}, is

This article has been accepted for publication in IEEE Transactions on Intelligent Vehicles. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIV.2023.3269156

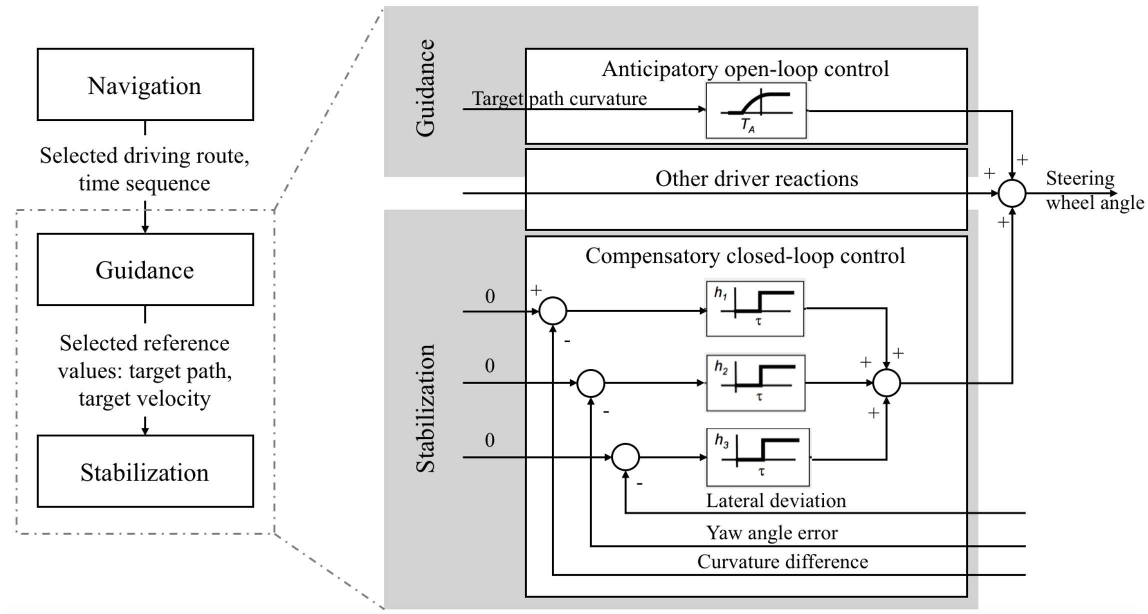

Fig. 1. 3-level driver behavior model according to [5].

$$t_{LC} = d_{LC}/v \cdot sin\,(\psi), \quad (1)$$

where v is the vehicle's velocity and ψ is the vehicle's heading angle relative to the lane marking.

Based on that, the system determines whether and how much of a system intervention is needed. A basic and commonly used control algorithm in commercial vehicles is a PID controller due to its suitability for the low dynamic lateral control problem and real-time integration. Further algorithms, e.g. Model Predictive Control (MPC), state control, Neural Networks, are also introduced to improve the system performance in terms of control accuracy, reaction time, adaptability, etc.

Similar to other industrial sectors, also the development of ADAS/AD is shifting towards the virtual world. It is no longer feasible to test and validate ADAS and AD functions only with real driving tests, due to the complexity of the system, the test cases and the required test scope [5]. For higher level of automation (SAE level 4-5), it is especially challenging to not only thoroughly test hardware such as sensors, control units and vehicle components, but also to train and test respective AI algorithms. For this purpose, an enormous amount of data is required to ensure the function and robustness of the algorithms. In the context of traffic vision for ADAS, a bulk of labeled data is required for tasks like object detection and semantic segmentation. Nowadays, there are many publicly accessible datasets obtained from real scenes, such as ImageNet [10], KITTI [11], MS COCO [12] and PASCAL VOC [13]. Researchers have recently also developed virtual datasets and driving scenes, e.g. the ParallelEye [14] [15], to complement the disadvantages of real-world dataset such as time- and cost-intensity, the lack of flexibility and diversity. Validation has proven that the inclusion of virtual datasets improved the accuracy of object detection and semantic segmentation.

As the automation level increases, there is a need to bring a more systematic approach to ADAS and intelligent vehicle testing. Unlike conventional performance tests of vehicles, where the test scenarios and the desired final state of the test are clearly specified, it is not easy to determine what kind of scenarios an intelligent vehicle may encounter [16]. Li *et al.* [17] designed a semantic diagram to define driving intelligence in four levels – scenarios, tasks, function atoms and functionalities. Based on this, Li *et al.* [16] then defined three kinds of tests: performance test, which emphasizes the specific "tasks"; functionality test, which addresses whether special "function atoms" can be successfully accomplished; intelligence test, which considers all four parts of driving intelligence. Here, "scenarios" become essential. The authors presented a Probably Approximately Correct (PAC) testing framework and showed that the major difficulty of testing lies in fathoming the set of scenarios rather than sample scenarios when the properties of scenarios set are assumed [16]. In addition to system validation, scenarios are also essential in system design and verification etc. Zhang *et al.* [18] did a systematic mapping study on finding the critical scenarios for AD regarding design and Verification and Validation (V&V).

III. Characterization of LKAS

Meanwhile, researches are carried out intensively in assessing ADAS and AD functions with higher quality and efficiency. The most important system characteristics that OEMs and researchers focus on are performance, comfort, safety, interaction with occupants, driving style, etc. The most investigated characteristics are analyzed in the following subsections based on published results respectively.

A. Performance

The first and most basic aspect to examine on an ADAS is its performance, namely whether it fulfills the functional requirements of the system well enough. Taking LKAS as an example, it is to evaluate if a potential lane departure is successfully avoided and if the departing vehicle is guided back into the lane center according to the planned trajectory accurately and efficiently. This is also the focus of the most published standards, e.g. ISO 11270 [19], UN ECE R79 [20], EURO NCAP [21].

It is not difficult to use objective metrics to define that. Typical specifications from the classic control theory can often be used, which are stability, rejection of step disturbances, time-responses such as rise time, peak overshoot and settling time, as well as further requirements such as integrated tracking error. For example, Diab *et al.* [22] has developed an LKA system based on a PID controller to minimize the yaw error. The controller parameters were tuned based on the performance indices of integral square error (ISE), integral absolute error (IAE), integral time square error (ITSE) and integral time absolute error (ITAE). For example, ITAE is calculated by

$$ITAE = \int_0^\infty t|e|dt. \tag{2}$$

Wang *et al.* [23] has selected a similar performance index, namely the integrated squared lateral deviation over time for the human-centered feed-forward control system they developed. System responses in frequency domain were used to compare the closed-loop system performance of systems with various preview times in different frequency ranges in [24]. The most common metrics for LKAS are lateral deviation from lane center, lateral velocity and acceleration, heading angle, yaw rate, etc. In Wang *et al.* [25], a generic and systematic testing and performance evaluation approach for autonomous vehicles is proposed. This approach is composed of three layers – sensing & perception, decision-making & planning, control & execution. For the performance evaluation of the latter two layers, metrics such as "response time", "driving efficiency", "tracking adaptability", "tracking accuracy" and "tracking smoothness" are selected.

As for driver perception, it is noticeable that most of the literatures focus on the sub-aspects of driver perception of comfort and safety. The evaluation of system performance is, on the other hand, mainly undertaken objectively. Nevertheless, some researches strive to create a holistic assessment system for ADAS by covering more dimensions of the system characteristics. In the developed assessment procedures for LKAS [26] [27] [28] [29] [30] the attribute performance was included. Most notably in Ref. [26] [27], lateral guidance performance was further specified and subdivided into "track accuracy", "lane keeping", "goodness of lane guidance" (position deviation), "quality of lane guidance" (angle deviation), "lane usage", "distance to left/right boundary" respectively. In doctoral thesis [30] the author included a performance criterion "lane guidance", which was defined as the capability of the vehicle to follow a straight or a curved road. It is strongly correlated with metrics based on deviation from lane center, lateral acceleration, steering angle, yaw angle error etc.

According to doctoral thesis [31], unlike the criteria comfort and safety, the criterion performance is categorized as the "assessment of perception". This means that the test driver should function like a "sensor" and evaluate if the perceived system performance is "high" or "low", despite the personal preference.

A list of the reviewed literature that investigated driver perception of the performance of LKAS and other ADAS is shown in Table II. The test objects and scenarios of each reference are illustrated, as well as the specific system characteristics that were investigated. As can be seen in the table, the driver perception of the same category "performance" can be divided into more detailed sub-criteria, for example "track accuracy" and "lane guidance". This means that research in this field focused on different levels of detail when categorizing and defining driver perception. The main category of the characteristic "performance" is shown in bold type in the table while the sub-criteria are shown in regular type.

TABLE II
REVIEWED LITERATURES ON DRIVER PERCEPTION OF SYSTEM PERFORMANCE OF LKAS

Literature	System under Test	Test Scenario: Maneuver	Test Scenario: Test Track	Driver Perception – Performance
2014 - Holzinger [27]	LKAS	-	-	track accuracy; lane keeping
2017 - Holzinger [26]	LKAS	-	-	goodness of lane guidance; quality of lane guidance; lane usage; distance to left/right boundary
2018 - Fen [28]	LKAS	-	-	**performance**
2019 - Oschlies [30]	LKAS	lane keeping	test track (straight, curve)	lane guidance
2019 - Schick [29]	LKAS	various according to maneuver catalogue	various according to maneuver catalogue	track guidance quality
2019 - Park [32]	LKAS	following (hands on/off)	highway	lane keeping performance; dynamics behavior performance
2018 - Kidd [33]	LKAS, ACC	daily driving	public roads	performance improvement
2017 - Wang [23]	Human-centered	following	simulator; simulation	**performance**

This article has been accepted for publication in IEEE Transactions on Intelligent Vehicles. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIV.2023.3269156

	steering system			
2018 - Wang [25]	ADAS/AD	-	-	**performance**

B. Ride Comfort

Elbanhawi *et al.* [34] summarized the main factors that influence the ride comfort of autonomous vehicles. In the context of AD, ride comfort basically refers to passenger comfort, since no human driver is involved anymore. Aside from the traditional ergonomic comfort factors such as vibrations, noise, temperature and air quality in the vehicle, the authors listed other factors that play an important role in the AD situations where the driver is out of the control loop. They are summarized as resulting forces, natural paths, motion sickness and apparent safety. Compared to other literatures, the ‘ride comfort’ was given a broader meaning in Ref. [34]. Ref. [34] studied how these factors can be taken into consideration to optimize the autonomous vehicles at the path planning phase. To minimize the resulting forces and jerks on the passengers while planning the trajectory, a smooth path with continuous curvatures is of great relevance. It contributes to reducing the motion sickness in autonomous cars as well. Previous researches on human control behaviors and maneuver characterization can be utilized for ADAS and AD to generate natural paths. Apparent safety which can be interpreted as the feeling of safety perceived by driver and passenger, can be affected, according to the authors, by the following factors. a) distance to obstacles or other vehicles; b) safe maneuvering time, where a balance between fast enough system reactions and driving comfort should be found; c) smooth trajectory tracking, where instability and overshooting should be avoided. The authors pointed out that most prior researches that considered comfort in planning were limited to the minimizing the resulting forces and did not take the other three factors mentioned above into account.

Gallep *et al.* [35] analyzed the path planning of AD regarding motion sickness in depth. The authors used a multidimensional motion sickness model from Kamiji *et al.* [36], which was validated with the results from Donohew *et al.* [37]. Ref. [35] simulated the Motion Sickness Incidence (MSI) with 3 different maneuvers and analyzed the influence of different translational and rotational motions on the MSI. As it turned out, angular velocities have a much lower influence on the MSI in comparison to translational accelerations due to the limited inclination angles. The highest MSI values occurred in coherence with Ref. [37] at 0.2 Hz. They then proposed a path generator based on the model to minimize motion sickness.

The detailed list of the reviewed literature on comfort of LKAS and other ADAS is included in Appendix 2.

C. Safety

Safety of ADAS and AD is categorized into cyber security and road safety by Ref. [34]. Road safety depends on system responsiveness, safe maneuvering and system reliability. The first two characteristics can be perceived directly by the occupants, which affects the customer acceptance directly. Responsiveness can be described as the ratio between the product of the vehicle speed v and response time T and the sensor look ahead distance d (or vehicle gap for urban scenarios) as follows [38]:

$$\text{Response Ratio} = v{\cdot}T/d \tag{3}$$

To achieve a high perceived safety, the system is mainly required to identify dangerous scenarios accurately and to react to them safely.

The perception of safety varies among individuals and can be influenced by physiological factors like personality, gender and age. Ping *et al.* [39] designed an experiment to investigate the visual risk perception of drivers from different age group with different driving experiences on city roads. The participants drove a predefined route and were asked to review the videos taken and rate the level of risk they perceived afterwards. The researchers extracted the environmental features from the videos. Combined with the ego vehicle data, they modeled the relation between the environmental factors and the risk perception of each individual driver using deep learning methods. The different sensitivity to risks were able to be identified among different driver categories. It can be seen that novice drivers were much less sensitive compared to the elderly and the experienced. Furthermore, the strongest influential factors which determined the risk perception also varied among the driver categories. For example, the percentage of important risk factors originating from the outside environment of the more experienced drivers is much higher than that of the novice driver, even though both groups were sensitive to ego speed. The elderly group, however, found insufficient distance to the leading vehicle the riskiest factor. As for lateral dynamics factors, experienced drivers and instructors were less likely to relate lateral acceleration and steering to danger than the novice as well as the elderly drivers.

Heiderich *et al.* [40] studied the influence of vehicle motions caused by external disturbances on the perceived safety on a driving simulator. The authors then objectified the perceived safety by correlating the subjective impression of safety with the vehicle motion data. The objective criteria are defined in the frequency domain due to the complex combination of ride comfort and handling aspects in the evaluation of the feeling of safety. The biggest influence on the perceived safety was found to be caused by roll rate motions in the range of 2-8 Hz and lateral accelerations between 0-2 Hz and 2-8 Hz. Jentsch *et al.* [41] did a validation study on safety evaluation in a static driving simulator. The authors compared the evaluation results of an autonomous emergency braking (AEB) system on a test track and in a driving simulator. The results confirmed that the participants were able to evaluate the AEB similarly in both test environments, thus validating the assessment method of

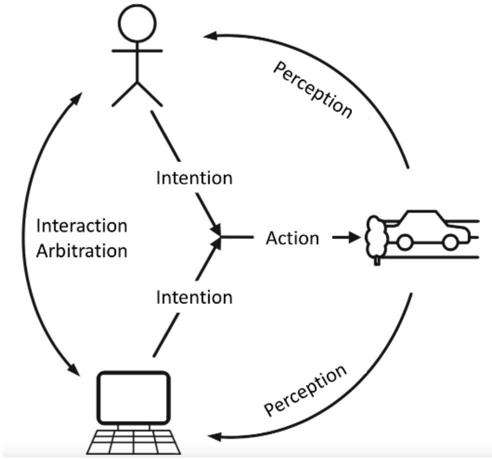

Fig. 2. Shared control loop [5].

perceived safety in a static driving simulator. However, it should be kept in mind that evaluating the safety feeling of a more comfort oriented ADAS that does not only intervene in emergency situations, such as LKAS, might not necessarily yield the same results.

Similar to section B, the detailed list of the reviewed literature considering driver perception of the safety of LKAS and other ADAS is included in Appendix 2.

D. Driver Interaction

For lower-level automation, i.e. level 1 to level 3 according to SAE, the driver is still an active component in the control loop, see Fig. 2. When the system is activated and intervening, the driver should be kept in the control loop (level 1-2) or ready to take over upon request (level 3). It is important to ensure a clear and reproducible driver-vehicle interaction without ambiguity. The Human-Machine-Interface (HMI) is roughly subdivided into display and operation elements. The display elements, from which the driver intake the information, trigger the information processing procedure of the human brain. The information could be provided optically (e.g. Head-up Display), acoustically (e.g. warning tones) or haptically (e.g. vibration of the seat). With the operation elements, the driver can execute the actions after processing the information. [5] For lateral guidance, the steering wheel acts as both display and operation element and is therefore the focus of the following literature study.

Aside from the warning functions of LKAS, the information that is being exchanged through the steering wheel is steering wheel movement, such as steering wheel angle, velocity and acceleration, and steering wheel torque. For level 2 systems, the drivers are supposed to keep their hands on the steering wheel permanently and only be assisted when necessary. There should be therefore no abrupt changes to the above-mentioned variables to guarantee a natural and comfortable interaction between the driver and the automated vehicle. For level 3 systems, the drivers could be freed from the driving task when the system is in control and are supposed to take over when being requested. The "driver interaction" here mainly concerns the take-over scenarios. Furthermore, the movement of the steering wheel should be reproducible under similar situations, for example the timing and intensity of intervention on the roads with similar curvatures. Thus, the driver would know what to expect from the system under certain situations and react accordingly.

The assessment system in Ref. [30] has included a subjective criterion "dynamics", which could be categorized as "driver interaction" according to the definition of this review paper. The criterion "dynamics" is defined as the quickness and the intensity of the steering intervention. Schick *et al.* [29] defined "driver-vehicle interaction" and "HMI" as two relevant attributes in LKAS development. Park *et al.* [32] attempted to model the intrusive feelings of the driver while using LKAS. The authors have specified the intrusive feelings in four types and derived four performance indicators for LKAS based on those. Two of the four indicators, "steering stability performance" and "non-interference performance", could be interpreted as the criteria of "driver interaction". They are defined as "the ability to minimize the vibration of steering due to the high-frequency intervention" and "the ability to minimize the LKAS torque in the opposite direction to the driver's intention" respectively. In the last decades, haptic shared control has gained more and more attention as a solution of improving driver's satisfaction with the interaction with semi-autonomy by sharing the control task [42]. Shared control is described as a continuous spectrum lying between manual control and full automation. By allocating the authority according to driver's state, driver's satisfaction with the system is found to be improved even when engaging in a demanding secondary task [43].

Another way to investigate the "driver interaction" is to conduct Naturalistic Driving Studies (NDS), where large-scale real-world driving data are collected for analyzing the complex interaction between drivers and automation, for example in Ref. [44]. Compared with conventional test drives, NDS has the advantages of the immense amount of data and the minimized effect of observation. On top of that, more sophisticated Machine Learning (ML) algorithms could be applied in the analysis. Thus, this method is believed to have huge potentials in analyzing human-automation-interaction, which could also have a much broader meaning than the definition in this paper.

The detailed list of the reviewed literature on driver interaction with LKAS and other ADAS is included in Appendix 2.

E. Driving Style

After the basic requirements on the aspects such as system safety, performance and comfort have been covered, the optimization and adaptation of the driving style of ADAS and AD drew the researchers' interest. Traditionally, driving style concerns the way individuals choose to drive or driving habits that have become established over a period of years. It includes choice of driving speed, threshold for overtaking, headway, and propensity to commit traffic violations [45].

This article has been accepted for publication in IEEE Transactions on Intelligent Vehicles. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIV.2023.3269156

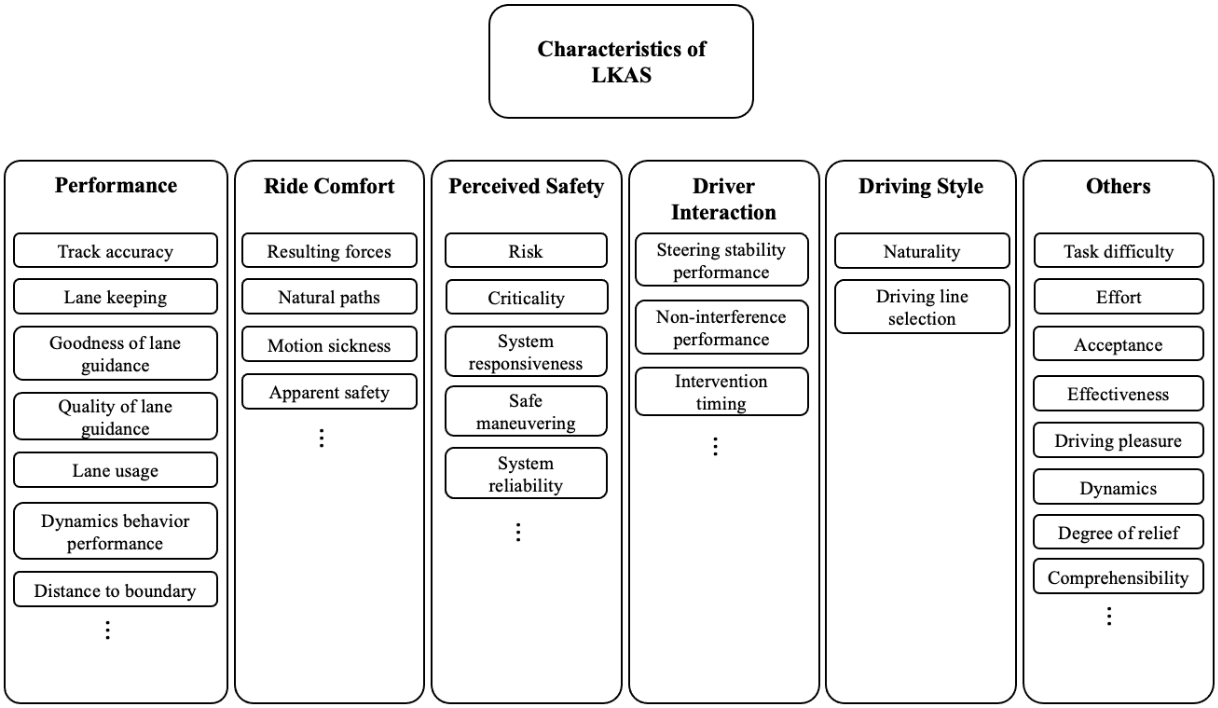

Fig. 3. Categorization of LKAS characteristics

Customers would want the automated vehicle or vehicle equipped with ADAS to still drive naturally with minimized sense of the robotic operation. Ideally, the automated function could even adapt the driving mode to customers' personal preferences. This could be done implicitly by learning the driver preference from manual driving data in advance, or explicitly by learning via the interactions between drivers and automation, or a combination of both [46].

Griesche *et al.* [46] conducted a two-step study on the driver preference of the style of the overtaking maneuver on a highway. In the first step, the participants were asked to overtake the slower leading vehicle in three different situations in a dynamic driving simulator. The driving data were then clustered into four different driving style clusters per situation using a pattern recognition and image processing tool. The vehicle state data used for the identification of driving style are lateral deviation, longitudinal and lateral acceleration, ego velocity, longitudinal and lateral jerk. Three months later, in the second step, the same participants were instructed to evaluate the four driving styles as well as their ego driving style in each situation when being automatically "driven" in the simulator. It is found out that most drivers did prefer their own style or a similar style in automated mode in general. However, the inter-individual difference in preferences was also situation dependent. For example, all drivers showed similar overtaking strategies in the more safety-critical scenario. Also overtaking strategies with smaller safety margins and higher accelerations were disliked by all participants. The authors also mentioned that a perfect matching between manual driving data and the preferred automated driving style is challenging, since not all drivers preferred their own style when being driven automatically in all situations. Therefore, a possibility for the driver to interact and configure the automated functions would be beneficial.

Bellem *et al.* [47] agreed that manual driving data are good indicators for configuring the automated driving styles and took the same approach. The researchers designed a study to distinguish three driving styles – comfortable, dynamic and everyday driving. Participants were instructed to drive a course manually and conduct four main maneuvers with these three styles in a randomized order on rural and urban roads as well as on highways. They focused on the metrics which cannot be manipulated in an automated system but can be perceived by the human vestibular system. For instance, accelerations and jerks were selected instead of velocity, because human vestibular system is not able to perceive speed itself but its changes. The researchers were able to identify objective metrics to distinguish the three driving styles under both rural and urban as well as highway conditions. The results also showed that it was necessary to carry out maneuver-specific analysis, for different maneuvers differ in their primary goals and the interpretation of the driving styles varies accordingly.

The detailed list of the reviewed literature on driving styles of LKAS and other ADAS is included in Appendix 2.

Further specific characteristics other than these five were also mentioned in the literatures, which were not easy to categorize due to the lack of detailed definition of the terminologies. Examples of those are acceptance, driving pleasure, degree of relief and so on. Fig.3 summarizes the five main categories of LKAS characteristics and their sub-aspects

This article has been accepted for publication in IEEE Transactions on Intelligent Vehicles. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIV.2023.3269156

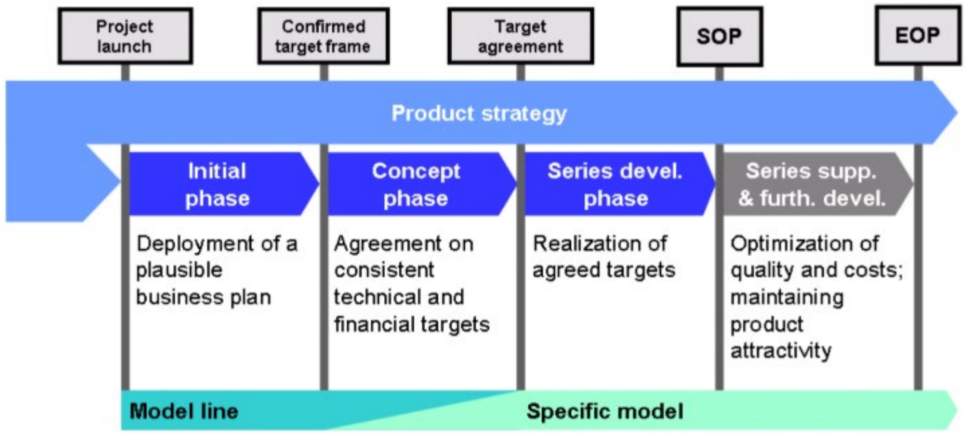

Fig. 4. Main phases of a vehicle development project [48].

reviewed in this chapter as well as further characteristics that do not fall in any of these. This overview shall give an impression of the current heterogeneity in ADAS assessment procedures from the perspective of characteristic definitions. In addition, the test scenarios and conditions are often project-specific, if documented at all, as shown in Table II and Appendix 2. This makes it difficult to reproduce, compare and extrapolate the research results from different projects and to develop further based on these results.

It should be taken into consideration that human's requirements for the vehicle or certain subsystems are multidimensional, subtle and complicated, sometimes even contradictory. As can be seen from the researches cited in this chapter, the investigated system characteristics of ADAS and AD – e.g. comfort, safety, and driving style – have some overlap in definition and they correlate to some extent. For example, the term ride comfort can have a broader meaning and can be certainly influenced, as illustrated in Ref. [34], by the perceived safety or the personal preferred driving style. Therefore, it is recommended that future researches define the system characteristics under investigation as clear and detailed as possible.

It should also be noted that the relations of the identified metrics to the driver perceptions in these researches are not necessarily statistically proven. Moreover, the collected subjective assessment data from the participants cannot necessarily represent the whole population due to the small sample size and the nature of the study. Nevertheless, they build a solid foundation for gaining a holistic understanding of drivers' impression of the system and for developing a systematic objective development method for ADAS and AD.

IV. Assessment Methods

Literatures [49] [50] [51] mentioned that the assessment methods could be classified in different ways:

1) virtual and real testing according to how tests are performed;
2) open-loop and closed-loop testing according to how the vehicle is controlled;
3) subjective and objective assessment according to how the behavior is assessed.

These divisions could be combined: open-/closed-loop maneuvers could be evaluated objectively and/or subjectively [50]. Objective assessment could be conducted in both virtual and real environment, i.e. simulation and driving test. While without a proper objectification tool, subjective assessment is mainly carried out in real world.

For LKAS, however, it is not precise to classify the tests into open- and closed-loop tests. Because the lateral guidance controller is always in the control loop, as shown in the Fig. 2. In this paper, we are going to denote the two categories as "driver-out-of-the-loop" and "driver-in-the-loop" depending on if the upper control loop is closed in the figure.

In the automobile industry, the PEP of vehicles can be roughly divided in 3 phases (see Fig. 4), namely initial phase, concept phase and series development phase. In the initial phase, the project plan including the technical and economical feasibility is clarified. Then the goals and targets are concretized in the concept phase. Finally in the series development phase the concept is realized and the vehicle is prepared for manufacture [48]. Numerical simulation and simulator experiments are applied in the initial and the concept phase, which contribute to a time- and cost-efficient conception in the early phase of development. Driving tests are mainly employed in the series development phase [52]. Generally, it could be stated that increasing objective assessment based on measurement and simulation is being applied in the design and basic tuning stage. However, the final fine tuning is still done by subjective assessment of test drivers, mainly through closed-loop driving tests [51].

The driving tests in the classic vehicle dynamics have reached a certain level of maturity after decades of development. Despite the fact that the functionality of LKAS is rapidly iterating and that emergency LKAS are mandatory in all new vehicles starting from 2022 [53], the assessment process of LKAS is far from standardized. The testing procedure is highly dependent on the manufacturer, which is heterogeneous and not transparent due to confidentiality

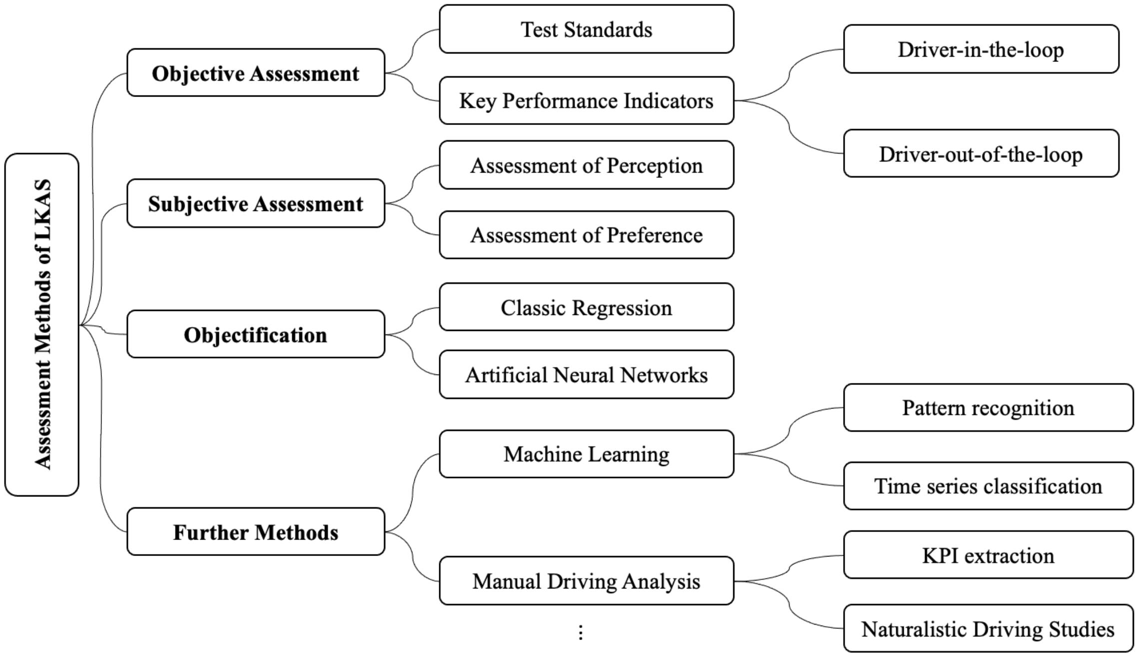

Fig. 5. Classification of the assessment methods of LKAS

agreements. Researchers have been attempting to develop a generic methodology to describe and evaluate the holistic system characteristics of LKAS, mainly regarding the major attributes mentioned in the previous chapter. Since the assessment of classic vehicle dynamics is well established to date, we are also going to review the assessment methods of lateral dynamics without ADAS, especially steering feel and vehicle handling. The experience in those fields will provide valuable insights and could be seen as a reference point for ADAS and AD assessment. In the following chapters, the state-of-the-art assessment methods in the automobile industry as well as the newly developed methods utilized in researches are classified into objective assessment, subjective assessment, objectification methods, Machine Learning methods and manual driving analysis. Fig. 5 gives an overview on the classification of the methods analyzed in this paper.

V. Objective Assessment

Open-loop tests are dominant in the objective assessment procedures because of the minimized human influences. Here, the system under test is “vehicle-environment” and the driver is out of the control loop. Open-loop test maneuvers define only the input to the vehicle or the initial condition and environment of the vehicle. After the initial input, the driver cannot influence the reaction of the vehicle and thus the driver is kept out of the control loop.

During closed-loop test maneuvers, on the contrary, the driver is asked to follow a predefined route/velocity profile or to drive freely. In this case, the complete “driver-vehicle-environment” control loop is being investigated. The driver is involved as part of the control loop, which makes it possible to test the relevant custom-oriented scenarios. However, despite the strict definition of the test procedures, closed-loop tests show a larger variance in the measurements. This could be influenced by driver’s driving strategy, competence, adaptation to a certain vehicle etc. [49] Due to this problem, closed-loop tests are not used in objective assessment alone. As described in ISO 3888-2, a complete and accurate description of the behavior of the road vehicle must necessarily involve information obtained from a number of different tests [54].

Objective assessment is usually carried out according to predefined test procedures on a proving ground. Relevant vehicle dynamics data are collected during the process, with which various predefined characteristic values are then calculated as the test results. However, there is little agreement in how to interpret the results or what to do with them in order to derive the greatest benefit [55].

A. Standards for Objective Assessment

For classic lateral dynamics tests, open-loop tests which are defined in ISO standards are mostly used. For example, Steady-State Circular Test (ISO 4138), Step Input Test (ISO 7401) and Weave Test (ISO 13674-1) are commonly used. One example of standardized closed-loop tests is Double Lane Change (ISO 3888-1). To minimize the environmental influences on the results, these standard tests usually have to be carried out on a proving ground. Using a steering robot can increase the quality of the measurement data noticeably, as proven in doctoral thesis [56]. However, only a few of the previous publications have utilized a steering robot – Ref. [57] [58] [59] [60] [61]. It is also recommended for the objective driving test of lateral guidance functions, if condition allows, to use a steering robot for a better data quality by executing the maneuvers more precisely and reproducibly. It has already been realized in various industrial projects. However, most publications on LKAS assessment either did not document the test maneuvers or logged the data

This article has been accepted for publication in IEEE Transactions on Intelligent Vehicles. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIV.2023.3269156

TABLE III
COMPARISON OF THREE EUROPEAN LKAS TEST STANDARDS

Standard		Test Scenario	Passing Criteria
ISO 11270	Procedure on a Straight	Travel along the straight road with constant speed **20 m/s - 22 m/s** and steer to gently depart from the lane to the left/right at the rate of departure **0.4 ± 0.2 m/s**.	1. LKAS_Offset_max = **0.4m** for light vehicles 2. LKAS_Offset_max = **1.1m** for heavy vehicles
	Procedure in a Curve	Drive along the straight with constant speed **20 m/s - 22 m/s** and set free the steering wheel before entering the left/right curve.	1. LKAS_Offset_max = **0.4 m** for light vehicles 2. LKAS_Offset_max = **1.1 m** for heavy vehicles
EURO NCAP	Road Edge Tests	Start at a constant speed of **72 km/h** and accelerate and steer to the driver/front passenger side to achieve the target lateral velocity of **0.2 - 0.5m/s (0.1 m/s increments)**.	Distance to Lane Edge (DTLE) <= **- 0.1m**
	Dashed Line Tests		DTLE <= **-0.3m**
UN ECE R79 - Automatically Commanded Steering Function (ACSF)	Lane Keeping Functional Test	Follow a curved road with a constant speed within range **V_smin - V_smax; for each speed range***	1. The vehicle **does not cross** any lane marking; 2. The moving average over half a second of the **lateral jerk <= 5 m/s³**.
	Maximum Lateral Acceleration Test		1. The recorded **acceleration** is within the limits specified in paragraph 5.6.2.1.3. of this Regulation. 2. The moving average over half a second of the **lateral jerk <= 5 m/s³**.
	…	…	…

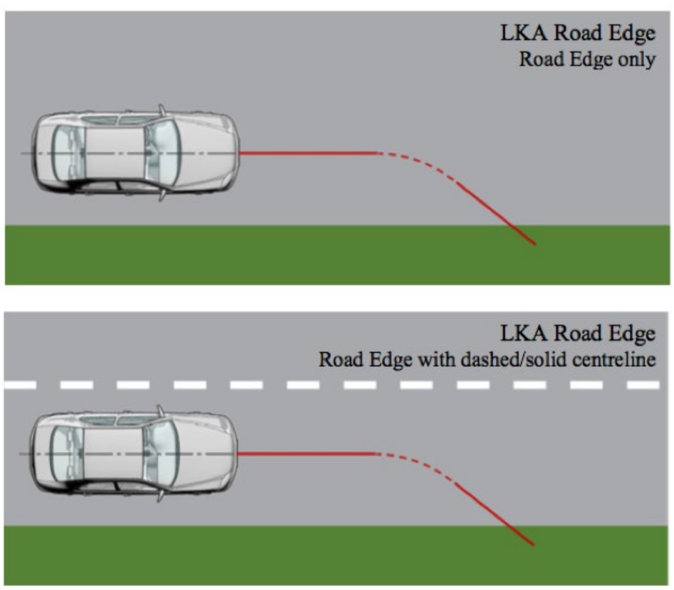

Fig. 6. EURO NCAP Lane Keeping Assist tests [21].

during the subjective test drives simultaneously, for example Ref. [29] and [30].

Since the beginning of this century, several European and international standards have been introduced for testing LKAS. These standards prescribe the system limits, define the standard test maneuvers as well as the fundamental conditions in order to test the functionality concerning safety, driver interaction and robustness. ISO-standards ISO 17361 [62] and ISO 11270 [19] set the minimal requirements on the functionality of Lane Departure Warning (LDW) and LKAS respectively in passenger cars, commercial vehicles and buses. Fundamental Human-Machine-Interface (HMI) elements and test methods were defined for the systems. Within the scope of [19], the two system types LKAS Type I and II are not distinguished. In the regulation UN ECE R79 [20] published by the United Nations, on the contrary, LKA functions are categorized into corrective and automatic steering functions, which correspond to Type II and I respectively. The Test Protocol [21] and Assessment Protocol [63] of EURO NCAP adapted the global NCAP tests to the European standards. The test scenarios and assessment criteria were prescribed for Emergency Lane Keeping (ELK), LKA and LDW. Unlike other standardized tests, different types of lane markings are taken into consideration in the EURO NCAP tests and are also assessed differently. An illustration of the EURO NCAP LKA test scenarios is shown in the following Fig. 6. The test scenarios and passing criteria of the main European standards are summarized in Table III.

Other international organizations have enforced various standardized test procedures under specific road conditions as well. Ministry of Land, Infrastructure, Transport and Tourism (MLIT) in Japan published a Technical Guideline [64] for LKA in motor vehicles. National Highway Traffic Safety Administration (NHTSA) in the USA specified tests for LDW and LKA systems under North American road conditions, while the tests for LKA solely serve the purpose of

This article has been accepted for publication in IEEE Transactions on Intelligent Vehicles. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIV.2023.3269156

information collection [65]. A detailed comparison of the standards is listed in Appendix 1.

However, the test scenarios defined in the standards are limited. The specified KPIs are mostly reaction time, intensity of the system intervention and extreme values of the vehicle dynamics parameters such as lateral position, acceleration and jerk. Although in most cases, there is a lack of theoretical and/or practical justification for the target values. Taking the standards in Table III as an example, the test scenarios are limited to straight and curved road with specified curvatures and the test executions of different standards are heterogeneously defined. The passing criteria for different standards are defined in a seemingly random and inconsistent manner, from the definition of the KPIs (e.g. LKAS_Offset_max and DTLE) to their target values. Due to the lack of theoretical and/or practical justification for the passing criteria and the consideration of aspects such as safety, comfort and complex interaction with the driver, they can only be considered as minimum legal requirements for the systems. Therefore, the target value of the KPIs under realistic conditions with the aim of achieving a well-rounded LKAS often cannot refer to the specified criteria for passing the standardized tests and requires more in-depth research.

B. Selection of Key Performance Indicators (KPI)

The relevant basic channels for analysis of lateral dynamics, such as steering feel and handling, are roughly summarized as vehicle speed and acceleration in longitudinal as well as lateral direction, yaw rate, roll angle and/or roll rate, side slip angle, steering wheel angle and steering wheel torque [55]. For LKAS analysis, the additional measurement channels, namely system activation status, Distance to Line Crossing (DLC), heading angle of the vehicle (with reference to lane markings), lane-keeping controller torque, lateral jerk, steering wheel velocity are also important.

After appropriate pre-processing, such as elimination of outliers and filtering, secondary data can be calculated using descriptive statistics. Secondary data can be derived either from primary data in time domain or those transformed in frequency domain. Typical secondary data to describe the central tendency are mean values, median. Standard deviation, variance, extrema etc. are suitable for describing the dispersion or variability of the measurement data. Relevant secondary data that can describe the system characteristics will be selected as KPIs. Other specific values, for example gradient, overshoot and settling time, are interesting for specific maneuvers and can also be selected.

Fen [28] analyzed the general requirements on the definition of KPIs for evaluating a comfort-orientated ADAS. According to the author, the KPIs should not be vehicle-specific, but should have a high degree of general validity. This means that the variables, from which KPIs are derived, should be at the overall vehicle level. The author listed some possible KPIs to take into consideration based on the general requirements. Ref. [28] also suggested appropriate algorithms, for example Goertzel-Algorithm instead of Discrete Fourier Transformation (DFT) for a shorter calculation time, since in most cases only a specific frequency band is of interest. It is also discussed that the redundancy of KPIs should generally be avoided, while their different physical effects on the driver should be taken into consideration in the meantime. For example, even though lateral acceleration and lateral jerk correlate, they should still both be used as KPIs for their different effects on driver's sense of comfort. The developed KPIs in this paper were categorized into three groups, i.e. performance, comfort and sense of security.

VI. Subjective Assessment

With standardized objective test maneuvers, diverse aspects of the vehicle or subsystem characteristics can be extracted with high accuracy and reproducibility. However, since in the end it is the driving behavior as a whole that has an effect on people, the subjective assessment method cannot be replaced so far [51]. The subjective assessment is based on the human ability to integrally assess dynamic system characteristics, which are composed of complex multidimensional stimuli [56]. The driving test for subjective assessment is carried out exclusively based on the vehicle reaction variables induced and processed by the driver. These are first perceived by the sensory organs of the test engineer and set in relation to internal model conceptions in parallel [31].

Subjective assessment may not be completely reproducible due to the nature of the principle. The difference in the result of subjective assessment may be caused by the characteristics of the vehicle, or by the different competencies and/or personal preference of the drivers. Harrer [56] pointed out that there are two kinds of differences of test drivers that influence the assessment result – inter-individual and intra-individual difference. They refer to differences between individuals and within an individual respectively. Inter-individually, for example, the judgements of test drivers from different manufacturers are not necessarily the same because test drivers from different development areas, such as different model series or different brand philosophies, incorporate additional metrics into the rating model. Intra-individual differences occur because the intrinsic assessment scheme is supported by the experience of the driver, which evolves over time. To keep this inner model up to date, the test drivers have to be permanently in training to keep the comparative databank precise and make reliable judgements. Also, the fluctuation of the personal performance adds to the uncertainty and inconsistency of the judgements [49]. Therefore, the requirements on the competence of the test driver are fairly high. They need to sense the minor differences among models or setups and minimize the personal influence at the same time. That requires the ability to accomplish multiple complicated tasks with a high level of concentration throughout the driving test. This is also the reason why the number of professional test drivers is limited and the cost is fairly expensive.

In addition, the unstandardized test procedure of subjective assessment can lead to procedure-dependent results [55]. For example, there has been a discussion about the use of reference vehicles. There is by far no clear conclusion if the use of reference vehicles improves the quality of subjective assessment. Some argue that it makes it easier for the test driver to build the internal model for assessment. However, this can also falsify the results, as it makes the assessment relative. It is especially problematic when using the subjective assessment for correlation analysis, where the measurement data is absolute. For example,

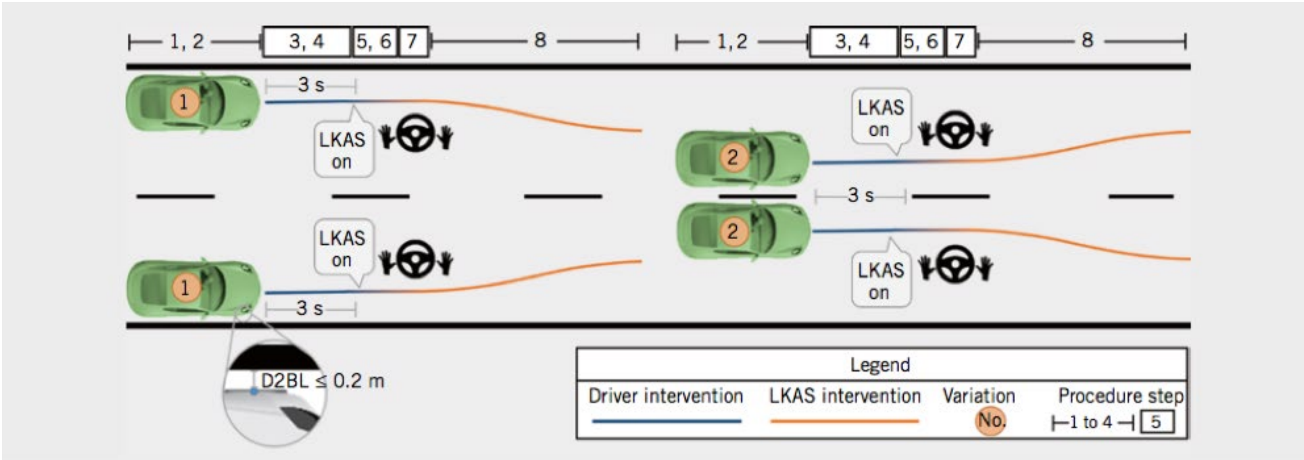

Fig. 7. Example of an LKAS test maneuver in [29].

trainings and pair comparisons were carried out in Ref. [30] before and after the test drives respectively to ensure the quality of the subjective assessments. But many other publications [26] [27] [29] decided against it. On top of that, even in the field of steering feel and handling, where the assessment procedure is relatively mature, the assessment criteria and the terminology are not standardized or universally accepted. Thus, they might be interpreted differently by drivers with bias.

To ensure the validity and reliability of the subjective assessment, researchers have applied various procedures and strategies, from the selection of test maneuvers and the design of questionnaires, the recruiting of test drivers, etc. These procedures and strategies will be discussed in the following subsections.

A. Subjective Driving Maneuvers

In the previous studies on the subjective assessment of conventional lateral dynamics without LKAS, closed-loop tests were mainly carried out. Both predefined maneuvers [66] and free driving [67] [56] could be found. Gómez [55] compared the assessment results of predefined maneuver with free driving. It shows that the predefined maneuver, which is considerably less time-consuming than the free driving, delivered very similar subjective assessments results.

On the contrary, open-loop test maneuvers are usually used in objective assessment procedures. However, for assessing ADAS functions with an SAE automation level 2 or higher, maneuvers without driver interventions could be considered. In this case, test drivers/assessors do not interact with the vehicle through steering wheel and/or pedals and assess the system from the perspective of a passenger. In this way, the uncertain influences of the drivers on the vehicle/system reaction can be eliminated, while the driver could still perceive and evaluate the vehicle reaction. Ref. [29] developed a comprehensive road catalogue as well as maneuver catalogue to cover the test scenarios of LKAS functions. Driver-out-of-the-loop maneuvers such as free ride, step steer, and closed-loop maneuvers such as lane change test are both included. In other researches on subjectively assessing ADAS functions, such as Ref. [26] [27] [28] [30], the test maneuvers are not always documented. But it can be interpreted that for testing LKAS, free ride with and/or without driver intervention on various tracks, including public road traffic and proving ground, is mostly used. An example of the test maneuver for LKAS designed in Ref. [29] is shown in Fig. 7.

B. Various Subjective Assessment Strategies

The subjective assessment procedure in the industry and in published academic works shows a big variety in the strategies. To begin with, the selection of criteria and the design of questionnaires are not standardized. There is by far no common agreement on many design details, for example the number of questions to include, if the term explanation should be included, and the assessment scale, which affects the result interpretation or correlation analysis directly. Harrer [56] classified the assessment scales into open scale and closed scale, both of which could be bi-polar or uni-polar. Other than the metrical scale, some questionnaires also included supplementary descriptive assessment to explain each score given. Ref. [56] claims this to improve the quality of the subjective assessments.

Decker [31] distinguished subjective assessment as the assessment of perception, which is the quasi-objective assessment, and the assessment of preference, which is the subjective assessment. Gómez [55] defined them as ‘evaluation’ assessment (good – bad) and ‘calculation’ assessment (light – heavy). The former one can be used to correlate to the objective assessments to find the preferred range of the characteristics. While the latter one can help identify how well the characteristics can be felt by the drivers.

Another long-lasting discussion is the selection of test driver type. The use of normal drivers has advantages in terms of cost and feasibility, which means it is easier to collect larger quantity of subjective assessment data. Most of the works on assessing LKAS functions used normal drivers because it is more feasible in large amount. Moreover, the test maneuvers regarding LKAS are mostly in low dynamic ranges, so the requirement on driver execution is not as high

This article has been accepted for publication in IEEE Transactions on Intelligent Vehicles. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIV.2023.3269156

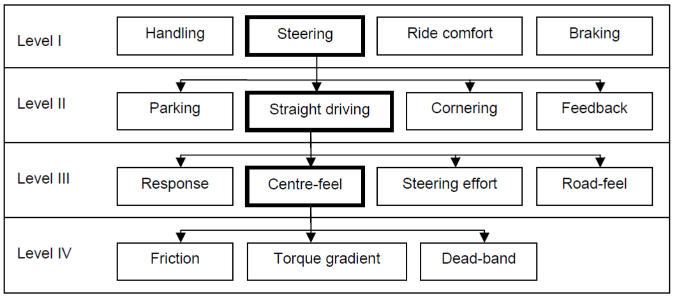

Fig. 8. Multi-Level assessment schema [56].

as for example for handling. However, their ability to perform and assess the tests consistently or to distinguish the fine differences among different setups cannot compete with professional test drivers. For example, they might not be able to interpret the assessment criteria or the assessment scale fully because of the lack of experience in the specific field. While recruiting professional test drivers is usually not easy, they offer in return high quality and more reproducible assessments. Professional test drivers took part in the driving tests in Ref. [29] [56] [68]. The selection of test drivers also has an influence on the design of the questionnaire or even of the whole procedure. For example, different levels of assessment should be carried out for different driver qualification. This is illustrated in Fig. 8 using steering feel as an example.

Schick *et al.* [29] subdivided the assessment criteria into subjective customer level and subjective expert level to evaluate LKAS. The subjective customer level consists of key criteria such as track guidance quality, driver-vehicle interaction, vehicle reaction, sense of safety, HMI and so on. These criteria are broken down into more detailed criteria in the subjective expert level. For example, the criterion “track guidance quality” can be divided into driving line, track precision, drifting and so on. The subjective criteria which Oschlies [30] selected, namely dynamics, lane guidance, driving comfort, security, driving style, comprehensibility and attractivity, are similar to the customer level of the Level-Model in Ref. [29].

Ref. [52] pointed out the importance of proper samples of drivers. Because the goal of the road tests usually is to make general statements, which goes beyond the description of the examined samples and can be applied to the population [52]. The samples should therefore represent either the global or specific attributes of the population as well as possible. An adequate number of test drivers also ensure the representability of the samples. Ref. [55] calculated the minimum number of test drivers needed to represent the whole population of drivers. To achieve a typical confidence level of 95.5% with a margin of error of 5%, at least 400 samples would be needed. However, the number of test drivers involved in the previous researches and projects is far below this theoretical requirement.

From the literature reviewed above, it can be concluded that the subjective assessment procedures are rather diverse and project-dependent. Researchers have used a variety of approaches to guarantee the quality of the subjective assessment data and to minimize potential biases. However, there has been little conclusion on which method is superior to others. The above-mentioned discussion on the use of reference vehicles, the selection of test drivers, and the necessity of predefined maneuvers are all examples. Nevertheless, there are still some general rules that almost all publications follow: 1) ensuring the largest possible number of test drivers; 2) eliminating incomplete ratings; 3) manually going through the questionnaires and comparing the ratings with verbatim comments, if included, to eliminate erroneous or inconsistent data points.

VII. Objectification Methods

It is clear that the state-of-the-art development and testing process in the automotive industry, i.e. the objective and subjective assessment, may be improved. To conclude, objective assessment, in particular open-loop tests, have the advantage of small variance in characteristic values due to the highly reproducible test conditions and maneuvers. Moreover, they are time and cost efficient. However, most predefined test scenarios do not represent regular driving tasks (especially open-loop tests). Most importantly, objective assessment does not take into account complex human perception. Subjective assessment, on the other hand, is a straightforward representation of human perception in customer-relevant scenarios that cannot yet be replaced by objective measurement. However, it is time-consuming, costly, not fully reproducible, non-transparent and, as discussed in the previous chapter, mostly brand-dependent. In addition, subjective assessment methods can only be applied in real test drives, i.e. in the very last stages of the development process.

These problems cannot be ignored in the new field of ADAS evaluation. In addition, in contrast to classical vehicle dynamics tests, there are not many test standards with sophisticated assessment criteria. Results often cannot be extrapolated to other situations/conditions. Furthermore, there are more environmental influences involved in the assessment process when testing in public traffic.

Because of all the above-mentioned drawbacks of the traditional assessment methods, the objectification of the assessment process is helpful and imperative for a more efficient and target-oriented development process. With a proper objectification model, it is possible to derive the target value of parameters and determine the driver preferences more systematically. In particular in the new field of ADAS and AD, the consistent use of a systematic objective evaluation method in various phases of the product development process shortens the development cycle time and reduces costs.

Decker [31] elaborated that the human perception of vehicle dynamics is very complex. When developing an objectification method, several features of human perception need to be taken into consideration: a) variance: the perception varies among different persons, different physical and mental states as well as different duration of the perception (adaption). b) interconnected perception: physically uniquely defined variables are often redundantly perceived. For example, yaw rate is perceived simultaneously via the sense of sight and the vestibular organ. c) reference values: also the relations of stimuli from different sensory channels to each other serve as reference values when processing sensory impressions. d) motion patterns: by exclusively reading the measured time series signals, the connection and the interaction of the individual variables to the perceived complex motion pattern of the vehicle is lost.

In general terms, the objectification of assessments is always about finding a connection in order to link different characteristics of a system and its assessments with a human judgement [69]. The doctoral thesis [49] summarized the general methods for objectification in the field of vehicle dynamics and classified them into three categories – driver-model-based, vehicle-model-based and KPI-based. The three kinds of methods were benchmarked from the perspective of usability as well as fulfillment of the requirements on objectification, which were specified previously in the thesis. Since the first two of the objectification approaches are not commonly used or especially suitable for LKAS assessment, only the KPI-based method will be analyzed in this paper.

KPI-based method is a classic approach and is mostly used in publications. KPIs are calculated based on the measurement of the driving tests, which are mostly open-loop maneuvers. Subsequently, the correlation between KPIs and subjective assessment is investigated. The goal here is to predict the subjective assessment based on KPIs, for example already in the early development phase where only simulation data is available. Furthermore, researchers strived to determine the optimal range of KPIs that are preferred by the drivers or that represent certain brand attributes. This approach theoretically provides a more comprehensive understanding of the relations between subjective and objective assessment and serve as a powerful tool throughout the whole ADAS development process.

The main advantages of this approach are its universal applicability, since the required parameters are available in both simulation and test drives, and the highly reproducible test maneuvers. However, the causality of the found correlations does not necessarily exist. Moreover, this method requires a moderate amount of training data and has no indicators for successful training of algorithm.

A. Data Preparation

As already mentioned before, the small sample size of the subjective assessment data is a general problem of the objectification of vehicle dynamics characteristics. Therefore, the data quality needs to be checked and optimized before the correlation analysis to guarantee statistically reliable results.

In some of the previous publications, the data were first processed with different kinds of data reduction techniques before being used for regression analysis. Käppler *et al.* [70] for example applied Principal Component Analysis (PCA) to the objective assessment to reduce the dimension of the data. The PCA generates synthetical secondary variables, a Principal Component, which correlates with all the primary variables as highly as possible [52]. Thus, the original variables are replaced by one or more independent factors, whose number determines the dimensionality of the data. The drawback is, however, that the physical meaning of the factors becomes difficult to understand. A similar method of data reduction is factor analysis, which Ref. [30] used to reduce the large number of KPIs derived from the measurement. The purpose of factor analysis is to combine many different variables into a few underlying latent variables based on the correlation matrices. In this process, the common issues of multi-collinearity can also be minimized. Other data reduction techniques include Ridge Regression Plots [67] and Correlation Analysis. Correlation Analysis can be applied to both subjective and objective assessment to check the redundancy of the criteria. Thus, only the essential variables will be included in the regression analysis.

In the next step, the generalizability of the results from a relatively small sample size should be checked. Here the inferential statistical methods are usually utilized to examine the reliability of the collected data. The variances and differences between groups of data can be determined by uni- and multi-variate Analysis of Variance (ANOVA). Their statistical significance can be checked by F-tests.

Another frequently used tactics is data normalization. On one hand, normal distribution is the prerequisite of many statistical analysis, for example t-test and ANOVA as well as the following regression analysis. One the other hand, normalization of subjective assessments can eliminate the personal influences of the driver to a certain degree. Ref. [30] normalized the subjective assessments and subsequently performed reverse transformation based on the total variance and the expected value of the samples, so that all the subjective assessments share the identical assessment scale. Harrer [56] designed a special transformation function for his subjective assessment system, which is designed intentionally with positive and negative sign, to prepare the data for further analysis. However, if not performed properly, normalization might cause random changes to the data due to the small sample size.

B. Classic Regression Analysis

Regression analysis is usually utilized to investigate the relationship between dependent and independent variables, which refer to the objective and subjective assessment in our case. It is used to describe and explain the relationships between objective and subjective assessment quantitatively, and/or to estimate as well as to forecast the values of the subjective assessment [71]. With sufficient experience and prior knowledge about vehicle dynamics and the system under test, this concept is straightforward, since no vehicle or driver model is required.

Simple Linear Regression (SLR) is commonly used by many of the researchers, which models the relationship between the dependent variable (subjective assessment) and one independent variable (objective assessment). Harrer [56] pointed out that the sample size has a direct impact on the statistical power and the valid generalization of a regression model. For the limited sample size, which is usually the case in vehicle road tests, it is recommended to choose a single independent variable and utilize simple linear regression to avoid overfitting. The commonly used correlation coefficient r is a dimensionless measure for the strength of the correlation ranging from -1 to +1. Ref. [59] chose a lower threshold of 0.7 and Ref. [56] 0.8054 for |r| to indicate a reliable relationship between the variables.

The simple linear regression model can be extended to multiple independent variables. Multiple Regression Analysis introduces more explanatory variables in the function and reveals the correlation between subjective assessment and multiple objective metrics (KPIs). The subjective assessment can be generally described by the objective metrics as follows:

$$y_i = b_0 + \sum_{j=1}^{k} b_j \cdot x_{ij} + u_i, \tag{4}$$

where y refers to the subjective assessment and x to the objective metrics. Here, b is the regression coefficient, k is the number of objective metrics and u is the stochastic unobservable disturbance.

The regression parameters are obtained by minimizing the sum of the squared residuals. After that, the quality of the regression function needs to be checked, which can be divided into two parts: global checking of the function (goodness of fit) and the checking of the regression coefficients. Global measures for testing the regression function are the coefficient of determination (R^2), F-statistics and the standard error of the estimation. Measures for testing the regression coefficients are the t-value and the beta-value [71]. The coefficient of determination expresses how well the regression function adapts to the observed data [71]. However, for multi regression analysis, each inclusion of an independent variable will increase the coefficient of determination. To avoid excessive model complexity, Ref. [30] used the adjusted coefficient of determination additionally.

Gómez [55] mentioned that problems might arise if the number of regressors excesses the available sample amount of objective assessments, i.e. the number of test vehicles/settings. A general rule is that the ratio should never fall below 5:1, meaning that there should be five observations for each independent variable [72]. Otherwise the results tend to become sample-specific and thus cannot be generalized. Other than the statistical validation of the regression function, the selection of meaningful combination of assessment criterion, objective parameter and test configuration shall be guaranteed [56].

An assessment procedure for steering feel using KPI-based approach was proposed in Ref. [56]. Various strong correlations between the derived KPIs and the subjective assessments were determined using both simple and multi linear regression methods. Moreover, the target ranges of the KPIs were identified. It was noted that some of the found correlations are dependent on vehicle segments, while others show general validity. Other authors who have performed multi regression analysis are [73], [74] and [58] [59].

It is clear that the relationship between subjective and objective assessment is not necessarily linear. Many researchers have taken the non-linearity into account in their studies. However, previous works in vehicle dynamics using second and third order regression functions with one independent variable have not shown further results [56]. For this reason, Ref. [56] handled the non-linear relationship between the subjective and objective assessment using transformation of the subjective assessment (Transformed Assessment Index). Thus, the non-linearity is included in the TAI and the linear regression formulae can therefore be applied, with the TAI representing the independent variable of the SLR.

The KPI-based objectification methodology was transferred to the field of ADAS development in Pawellek *et al.* [75]. The authors investigated the Adaptive Cruise Control (ACC) regarding the column braking maneuver in both city and on highway scenarios, with low and high criticality respectively. Three KPIs derived from the longitudinal deceleration were adopted to characterize the intervention strategies of the ACC. They were correlated with the subjective assessment concerning comfort and dynamics, evaluated by expert and normal drivers, using regression analysis. As a result, the predicted objective ratings calculated by the objectification model showed a very high correlation with the subjective ratings. The series status ACC was optimized based on the objectification model. The improvement was then proven by both expert and normal drivers in a post study on public roads.

Inspired by Ref. [75], a methodology to virtually calibrate LKAS was developed in the doctoral thesis [30]. The models of objective scores were created to illustrate the mathematical relationship between subjective and objective assessment of LKAS. 29 normal drivers assessed the test vehicle with 4 controller variations regarding dynamics, lane guidance, driving comfort, security and further criteria on two specified tracks. The median, the absolute mean, the standard deviation and the mean squared error of the recorded vehicle dynamics data were calculated as universal KPIs. Besides that, specific KPIs were calculated for each vehicle dynamics parameter. The correlation matrices between the KPIs and the subjective assessment criteria were created by means of linear regression methods. Based on that, factor analysis was conducted to reduce the dimension of the model and to eliminate multicollinearity of the KPIs. Lastly, the models of objective scores based on factors were developed with reference to the two specified tracks. It could be determined that the models of objective scores show “strong” up to “very strong” relations to all subjective assessments. Subsequently, a model predictive controller was designed for vehicle lateral control, which was then optimized with the help of the created models of objective scores. The thesis presented a complete process for virtual optimization of the lateral guidance function and showed the potential of applying the model of objective scores in the development phase. However, solely linear correlations were inspected in this work. There is therefore still a great potential for optimization and further development of the models. Furthermore, the created models of objective ratings in Ref. [75] and [30] were limited to the specified test scenarios, for which reason the robustness and generality of the models is still to be proven in further researches.

It has also drawn the researchers' attention that how the level of automation influences the assessment of the driving function. The conference paper [76] concerning objectification of a combined longitudinal and lateral guidance system introduced a tuning method for ADAS and Highly Automated Driving (HAD) system based on a similar objectification approach to Ref. [75]. An ADAS of SAE level 2, which consist of an ACC and an LKAS, and a SAE level 3 automated driving function operated by a distracted driver was investigated on the highway. A variation scheme of the KPIs was defined during the research, according to which the controller behavior could be varied in representative maneuvers. In the following evaluation study, the assistant and the automated driving function of a prototype with different controller calibrations was subjectively evaluated. The vehicle was assessed regarding sense of security, comfort, dynamics and perceived situation criticality. The assessment results in a column braking maneuver of a basic and an optimized controller, which brakes harder with a larger minimum distance to the vehicle in front, were displayed. It was found out that the optimized controller of the ADAS function was assessed better in sense of security, dynamics and maneuver criticality. In contrast, the assessment of the HAD function with these two controllers does not distinguish from each other. This means that the level of automation of the assistance function should be taken into consideration when assessing as well as objectifying the system. Because the driver involves as a different role in assisted and automated driving and therefore perceives the system characteristics differently.

The following papers have proposed many relationships between subjective assessment and KPIs based on professional experience but have not explicitly proved the relationships with mathematical models. These could be seen as preparation works for objectification in ADAS development. In Ref. [28], a Congestion Assistant with SAE level 2 was investigated, which consists of a Full Speed Range ACC with Stop-and-Go function and an LKAS. It was discussed in this work how a comfort orientated ADAS should be evaluated with KPIs. General requirements on the definition of KPIs were analyzed.

Schick *et al.* [29] introduced a generic procedure model for an attribute-based ADAS/HAD development. A Level-Model of the assessment criteria, which is subdivided into subjective customer level, subjective expert level and objective level, was introduced to evaluate the LKAS. Relevant subjective assessment criteria, i.e. desired brand attributes, and KPIs were developed in expert workshops, benchmark tests and measurement campaigns. The subjective customer criteria cover the aspects of system dynamics, precision, safety and comfort, and are then broken down into detailed expert criteria. Finally, the subjective expert criteria are linked to KPIs in the objective level based on expert knowledge. The strengths of the links are rated as high, moderate, low, none or unknown. The desired brand attributes can thus be translated

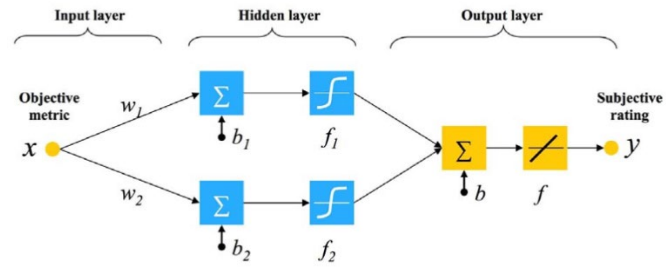

Fig. 9. An example of an NN structure [55].

into objective targets using the Level-Model. A modular simulation environment was built and a maneuver catalogue as well as algorithm for KPI-calculation were implemented to produce comparable results throughout the whole development process. Note that the significance and statistical power of the linkages between subjective expert assessments and KPIs still need to be validated in further researches. Although not yet statistically proven, this Level-Model offered great insights from the experts, which could be used as a roadmap for further researches in exploring the correlations of subjective and objective assessment of LKAS or other ADAS functions.

C. Artificial Neural Networks

Other than the classic regression analysis, Artificial Intelligence (AI) algorithms offer promising solutions for revealing especially the non-linear relationships between subjective and objective assessment. Algorithms such as Artificial Neural Networks (ANN or NN) have been often used in previous research.

Using NN to explore the correlation is of great benefit, since the user does not necessarily have to make a presumption about the relationship between variables in the analysis procedures. Furthermore, NN can also process variables with different scales. The construction principle of NN is based on the processes in the nervous system of humans and animals and tries to reproduce "biological learning" by means of suitable mathematical operations. The basic structure of a Multi-Layer Perceptron (MLP) NN usually consists of one input layer, one or more hidden layers and an output layer. Similar to a biological neuron, a cell can receive a variety of input signals from the preceding cells. It combines these input signals according to their weights into a uniform input value of the neuron. The so-called activation function then determines whether the neuron is activated and sends out a signal or not. In the hidden layer, non-linear transformations are carried out, allowing the neural network to map non-linear relationships between input and output values [77] [71].

In the modelling, objective metrics calculated from measurements are often used as input signals and the subjective assessments are the outputs. The most often used transfer functions are linear, hard-limit and tan-sigmoid. An example of the NN structure used in doctoral thesis [55] is shown in Fig. 9. The data usually need to be preprocessed before being used in NN. For example, multicollinearity of the input data, i.e. the objective metrics, should be checked. All the objective metrics also need to be normalized to a mean of zero and a standard deviation of one.

Ash [67] and King *et al.* [78] recycled the data of Crolla *et al.* [79] and attempted to conduct non-linear analysis using NN. However, it is found out that the requirements for the data suitable for NN is not necessarily the same as those for linear analysis. Large networks were not possible due to the limited subjective and objective data. As a result, they decided on a simple NN with one input and one output. King *et al.* [78] also tried to investigate the combinatory effects of the subjective and objective assessment by applying a dual-input NN with one hidden neuron with a tan-sigmoid transfer function. Because of the limited data available, however, this attempt led to problems such as overfitting. Furthermore, the relationships found were not evident across different drivers and questions. It showed that attempting to fit more than one objective assessment to a set of subjective ratings using non-linear methods only served to cloud relationships already found through single input networks [78]. Nybacka *et al.* [59] and Gómez [55] were based on the same KPI-based approach to objectify subjective assessment of steering feel and broadened the scope of Ref. [56]. They applied ML methods such as NN besides the linear regression methods to analyze the correlations. Ref. [59] chose a single-input NN with 3 hidden neurons to model the correlation of one subjective assessment and one KPI. As in Ref. [78], the authors also chose a partition fraction of 0.75 for the training and test data to ensure adequate training and generalization given the limited nature of the data. An r-value larger than 0.7 and a mean squared error (MSE) lower than 0.5 was selected for adequate fit. In order to determine the full range of subjective driver preference, a sufficient spread range of objective assessment, i.e. a sufficient amount of test vehicles/configurations, is required. Based upon Ref. [59], further AI methods were utilized in the doctoral thesis [55], in order to explore the non-linear relations and tackle the problem of the small sample size. The author utilized a SISO NN with one hidden layer consisting of only two sigmoidal hidden neurons (Fig. 9) to avoid overfitting and the requirement on large quantity of data for training the model. Based on the preliminary results, Ref. [55] later introduced a new method using two NNs built upon each other to model the relationship between 27 objective metrics and one single subjective assessment of 51 vehicles. The authors implemented a Self-Organizing Map (SOM) to classify the tested vehicles based on merely objective assessment. A General Regression Neural Network (GRNN) of subjective assessment was then implemented on top to extend the SOM. The map was finally complemented by word-clouds from interviewing the test drivers, which describe the vehicle attributes in natural language. The advantage of this method is that it is completely based on objective assessments, which do not change over time. The superimposed subjective assessment surface is on the other hand highly flexible, which means it would be possible to extend the map by connecting to new subjective assessment surfaces.

Monsma [50] applied three methods to explore the vehicle handling characteristics preferred by the driver under the influence of different tires – KPI-based objectification, workload measures and driver model method. By using the objectification method, the author built a model for each of the two test drivers and an average of the two to predict their subjective assessment in 9 aspects based on objective metrics. To bypass the common drawbacks of regression analysis, such as the risk of choosing an inappropriate model based on limited knowledge, the author chose GRNN to model the relationships between subjective assessment and objective metrics. GRNN is a radial basis neural network. The value of the radial basis activation function is dependent on the distance r of a data point $\boldsymbol{x}$ to a certain center point $\boldsymbol{c}$, which is defined as

$$r(\boldsymbol{x}, \boldsymbol{c}) = \|\boldsymbol{x} - \boldsymbol{c}\|. \tag{5}$$

Here, $\boldsymbol{x}$ and $\boldsymbol{c}$ can be one or multidimensional. The activation of the neuron is then given as follows:

$$a = \varphi(r \cdot b), \tag{6}$$

with the radial basis activation function φ and bias value b.

To construct a GRNN, no a-priori knowledge about the model structure is required. Thus, one does not have to presume the relationships between predictors and targets when they are unknown. Furthermore, it is suitable for prediction based on very limited datasets. GRNN has a one-pass learning phase, which means no iterations are done. And no division between the training and validation set is required. Unlike the multilayer feedforward NN with backpropagation, who needs large amount of data for proper learning. [50] Since limited datasets and insufficient knowledge about the relationships between the metrics and the subjective impressions remain two major problems for LKAS assessment, GRNN could have good potential in this research field.

In Ref. [27] and [26], ACC, LKAS and Lane Change Assist (LCA) were investigated regarding objectifying the subjective assessment using NN. The systems were evaluated from the point of view of the driver as well as of the front passenger. The subjective assessment of LKAS regarding track accuracy, lane keeping and driving line selection was correlated with KPIs lateral deviation, angle deviation, lane utilization and distance to lane boundaries. The center line of the lane was assumed to be the target trajectory, to which the calculation of lateral deviation and angle deviation refers. It was mentioned that NN were utilized in order to create the correlation model. Yet no detailed information about the algorithm was documented. It was for example determined in these papers that the sense of security of the driver depends on the difference between the expected and the driven trajectory.

This article has been accepted for publication in IEEE Transactions on Intelligent Vehicles. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIV.2023.3269156

VIII. Further Assessment Methods

A. Other Machine Learning Algorithms

There were also researchers who took a different path and explored other ML algorithms like pattern recognition. Instead of using characteristic objective metrics, a clustering method was used in Ref. [46] to cluster the driving data into different styles. Moser *et al.* [68] pointed out the limitations of the state-of-the-art correlation and regression methods, which includes the variance of the subjective perceptions, high impact of the system designers, and the information loss due to the reduction of time series data to a characteristic value. The application of NN like in Ref. [27] can also cause problems because the model is often a black box and hard to interpret. The authors therefore implemented a new assessment method based on time series classification to overcome the limitations of the conventional methods. A lane keeping and a lane change driving scenario was conducted on a test ground and on a highway respectively. The ride comfort was rated by the drivers on a scale of 1-7, which were used as the training dataset. After each test, a small expert driving test was conducted to create a test dataset. The vehicle data were analyzed using the proposed k-Nearest Neighbors Classifier and Kernel-Density-Estimation Classifier based on Uni-/Multivariate Dynamic Time Warping (MDTW) and compared with the estimation results from a Multiple Linear Regression (MLR) and a simple NN model. As a result, in the lane keeping test, the time series classifier method outperformed both MLR and NN. In the lane change test, the proposed approach lacked the robustness of MLR, especially with sparse data. However, the MDTW approach recognized boarder/extreme cases much better than MLR (limitation of linearity). The authors summarized that the MDTW method has the advantage of easier data generation since no hypothesis needs to be made before the test, and the better ability to recognize extreme cases compared to MLR. It is recommended to apply the method in combination with MLR for the ADAS assessment to complement each other. The test set in the study, however, was relatively small. With increasing database, the proposed method would have a great potential.

B. Manual Driving Data Analysis

Some researches explore the optimal or natural driving characteristics by drivers first, then tried to describe the ideal system setting using objective metrics. These researches gave a great insight into drivers' requirements for ADAS and AD. Ref. [46] and [47] analyzed manual driving data on driving simulator and in the real traffic as a reference for future ADAS and AD. They both objectively characterized different driving styles in specific driving scenarios. Griesche *et al.* [46] implemented a pattern recognition method on the transformed images of vehicle state data snippets to cluster the overtaking maneuver into different styles. The participants were then asked to evaluate those styles using Best-Worst-Scaling. Bellem *et al.* [47] identified maneuver specific KPIs which can characterize different driving styles in four common driving scenarios. The manual driving data were analyzed using ANOVA and Bonferroni-corrected post-hoc tests, in order to identify if the selected KPIs distinguish between different styles. The above-mentioned publication [68] also made use of data collected from synthetic oscillation in lane keeping scenario and manual lane change maneuver.

Gordon and Srinivasan [80] used two sets of Naturalistic Driving Data (NDD) to validate a well-known linear steering model under low-workload conditions. It was found that the closed-loop behavior of the fitted model is instable. Additionally, it does not represent the steering rate reversals in realistic steering behaviors. The authors proposed an alternative Pulse Control Model to represent the intermittent pulse-like steering in real-world scenarios. Based on this concept, Zhang *et al.* [81] analyzed and decomposed the steering angle measurement into a combination of trend, integrated sine components and sine components. The study validated the Pulse Control Model for vehicle lane keeping proposed in Ref. [80], which benchmarked the normal driver behaviors and could be used for identifying abnormal driver conditions such as tiredness and distraction. The pulse behavior on the other hand could be used for evaluating the lane keeping naturality of ADAS and AD. Benderius and Markkula [82] analyzed a large number of manual driving data exceeding a total length of 1000 hours, collected from various projects. The dataset includes three driving simulator studies, one test track study and a field operational test (FOT), some of which were not designed originally for the purpose of this research. According to the authors, it is believed that the nature of this work is general and independent of any specific data properties or purposes [82]. The results showed that the steering corrections can be generally explained by bell-shaped steering wheel rate profiles. Since a strong relationship between the maximum steering wheel rate and steering wheel deflection was found, a constant correction duration could be expected. For corrections that could not be described by a single bell-shaped rate profile, it was found that they could be described by a superposition of at most four motor primitives. These findings are expected to offer a fresh perspective for the driver modeling and LKAS development, because the work indicated that natural steering might rather be composed of intermittent open-loop bursts than pure closed-loop continuous control action [82].

Naturalistic Driving Studies (NDS) also gained a lot of interest in the last decades as a promising approach to study natural driver behavior. Video cameras and a suite of sensors are installed on participants own vehicles and are used to continuously record the driver, the vehicle, and the environment over an extended period of time [83]. NDS makes it possible to observe the drivers in naturalistic settings (during their regular, everyday driving) in an unobtrusive way [84]. It was mainly utilized for investigation in driver behavior that leads to safety-critical events to begin with, mainly related to crashes. However, NDSs are also an effective tool for the design, testing, and evaluation of driver assistance systems [83]. For example, a running project conducted by Massachusetts Institute of Technology (MIT) aims to gain a holistic understanding of how human beings interact with vehicle automation. Fridman *et al.* [44] outlined the methodology and underlying principles governing the design and operation of the study vehicle instrumentation, data collection, and the use of deep learning methods for automated analysis.

IX. Conclusion and Discussion

This paper reviewed the assessment of LKAS regardless of the automation levels from the publications in the last decades. We firstly summarized the most relevant characteristics of the system, as well as of HAD and AD, namely performance, comfort, safety, driver interaction and driving style. The exact scope of the individual characteristics in different literatures and the discovered results are analyzed. It is noticeable that the definitions of the attributes have not reached a common agreement in the field (see Appendix 2). The levels of assessment criteria are heterogeneous, for example "quality of lane guidance" is more detailed and could be included in "performance". Many characteristics overlap to a certain degree and influence each other, for example "perceived safety" is hugely influenced and more or less implicitly implied by many other factors such as "performance", "comfort" and "driver interaction". These ambiguities will directly influence the quality of the objectification based on the fundamental definition. Therefore, it is recommended for future research to subdivide the assessment criteria into different detail levels according to the competence of the test drivers in the subjective studies and define the characteristics as clear as possible to avoid divergent interpretations.

We then discussed how the relevant characteristics can be evaluated – objectively, subjectively, using objectification tools or other methodologies. The assessment of classic lateral dynamics without ADAS, more specifically steering feel and vehicle handling, is also benchmarked here. Rich experiences in this field can be taken as a reference for the development of a systematic evaluation tool for LKAS or other ADAS functions such as ACC. Each method is introduced and analyzed in detail with examples from representative literatures. In the new territory of ADAS assessment, the major drawbacks in the conventional vehicle dynamics assessment still exist. For example, the limited number of test drivers, the unstandardized test maneuvers and subjective assessment strategies, the weak causality in the objectification. Other than these, there are some other specific problems and potentials regarding LKAS assessment within the scope of each assessment method such as the following points: 1) objective – the available international standards are not as well-established as in vehicle dynamics. It can be seen that the bare minimal safety and functional requirements are not suitable for characterizing an LKAS. The research projects by far are not able to deliver robust target values for KPIs for LKAS design due to various reasons. For example, the test scenarios were limited, or the concrete results were not made public due to confidentiality agreement. 2) subjective – due to the ODD (Operational Design Domain) of the system and the more limited availability of suitable proving ground for lateral guidance assessment, most of the test drives are conducted on public roads. This on one hand represents customer-oriented scenarios, on the other hand introduces more uncertain environmental factors into the assessment. Another influence factor that may arise is the automation level. Human driver/passenger might have very different requirements on the system of different automation levels. 3) objectification with regression – the explanatory power of the single objective metrics is not strong enough to correlate to a customer-leveled subjective assessment. For example, we cannot simply correlate the characteristic values of lateral deviation/velocity/acceleration with "perceived safety" and expect a robust causal relationship of the model. This could be solved with a) more complicated models with more predictors, individually or using factor analysis, which might however be hard to interpret as a result; b) using more detailed subjective criteria on expert-level as targets of the model; c) using other ML algorithms such as pattern recognition [46] or time series classification [68] which preserves more information in the objective data than just the characteristic values. 4) other ML algorithms and manual driving analysis – these models normally require more data for training, but with an extensive database they still have great potentials. Manual driving analysis serves as a meaningful reference for developing ADAS and AD. For example, examining human steering characteristics offers a great insight into the naturality of lateral guidance systems. However, it cannot be assumed that human would like to be driven in the same way as they would drive, especially in higher level automations, as reported in previous literatures.

It can be concluded that, despite the considerable progress that has been made, there is still a lot of potential for further development of LKAS or ADAS/AD in general, particularly with a focus on human needs. With more advanced AI algorithms, it may be expected that the future LKAS will be able to adapt to human habits and preferences. With the rapid development of 5G, V2X technologies and other infrastructure, the information available for intelligent vehicles and traffic, and the ability to process this information, will increase further.

However, attention may be drawn to the current drawbacks of AI algorithms, e.g. Ref. [85]. Researchers have suggested that the current AI needs to evolve from feature engineering to scenario engineering. The theoretical framework of scenario engineering, which aims to address the challenges of interpretability and reliability of AI, and thus to build more trustworthy AI systems as well as foundation models in metaverse in general, is presented in Ref. [86] [87].

This article has been accepted for publication in IEEE Transactions on Intelligent Vehicles. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIV.2023.3269156

Appendix

Appendix 1 – Comparison of international standards on LKAS

Standard		Test Scenario				Test Execution	Passing Criteria
		Test Track	Lane Markings	Test Speed			
				Longitudinal	Lateral		
ISO 11270	Procedure on a Straight	Straight	depending on the applicable regulations for highway like roads	20 m/s - 22 m/s	0.4 ± 0.2 m/s	Travel along the straight road and steer to gently depart from the lane to the left/right at the rate of departure.	the outer edges of the tires do not exceed the lane boundary more than LKAS_Offset_max during the test, where 1. LKAS_Offset_max = 0.4m for light vehicles 2. LKAS_Offset_max = 1.1m for heavy vehicles
	Procedure in a Curve	Straight entering a curve*			-	Drive along the straight and set free the steering wheel before entering the left/right curve.	the outer edges of the tires do not exceed the lane boundary more than LKAS_Offset_max during the test, where 1. LKAS_Offset_max = 0.4 m for light vehicles 2. LKAS_Offset_max = 1.1 m for heavy vehicles
EURO NCAP	Road Edge Tests	Straight	Road edge only	72 km/h	0.2 - 0.5m/s (0.1 m/s increments)	Accelerate and steer to the driver/front passenger side to achieve the lateral velocity.	Distance to Lane Edge (DTLE) <= - 0.1m
			Road edge with dashed centerline				
			Road edge with solid centerline				
	Dashed Line Tests	Straight	Dashed line only				DTLE <= -0.3m
			Dashed line with solid centerline				
	Solid Line Tests	Straight	Solid line only				
			Dashed line with solid centerline				
NHSTA (for information collection only)	Low Lateral Velocity Tests	Straight	Solid white lines	72 km/h	0.1 to 0.6 m/s with a target lateral velocity of 0.5 m/s	Pass the entrance gate and steer to depart to the left/right with the target lateral velocity until the vehicle has crossed at least 0.5m over the lane edge boundary.	No performance requirements, results are for informative purposes only.

This article has been accepted for publication in IEEE Transactions on Intelligent Vehicles. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIV.2023.3269156

		Threshold Determination Tests	Straight	Solid white lines		iteratively increased by 0.5 – 0.6 m/s starting from 0.5 m/s	Pass the entrance gate and steer to depart to the left/right with the target lateral velocity until the vehicle has crossed at least 0.5m over the lane edge boundary. Increase the lateral velocity by 0.5 - 0.6 m/s until the LKA system can no longer prevent a lane departure.	
UN ECE R79	**Corrective Steering Function (CSF)**	Warning Test	-	Lane markings on both sides	Within the operating range of the system	-	attempt to leave the lane and cause CSF intervention to be maintained for a period longer than 10s (for M1, N1) or 30s (for M2, M3, N2, N3) without distinguishing the directions.	The acoustic warning is provided no later than 10 s (for M1, N1) or 30 s (for M2, M3, N2, N3) after the beginning of the intervention…
		Overriding Force Test	-	Lane markings on both sides		-	1. attempt to leave the lane and cause CSF intervention2. the driver shall apply a force on the steering control to override the intervention	Force applied by the driver on the steering control to override the intervention <= 50 N.
	Automatically Commanded Steering Function (ACSF)	Lane Keeping Functional Test	Curved track	Lane markings on both sides	Constant speed within range V_smin - V_smax; for each speed range*	-	follow road	1. The vehicle does not cross any lane marking; 2. The moving average over half a second of the lateral jerk <= 5 m/s³.
		Maximum Lateral Acceleration Test	Curved track	Lane markings on both sides		-	follow road	1. The recorded acceleration is within the limits specified in paragraph 5.6.2.1.3. of this Regulation. 2. The moving average over half a second of the lateral jerk <= 5 m/s³.
		Overriding Force Test	Curved track	Lane markings on both sides	Constant speed within range V_smin - V_smax	-	Follow the curve without driver intervention, and the driver shall then apply a force on the steering wheel to override the system intervention and leave the lane.	The force applied by the driver < 50N
		Transition Test; Hands-on Test	Curved track	Lane markings on both sides	between V_smin + 10 km/h and V_smin + 20 km/h; between Vsmax - 20 km/h and Vsmax - 10 km/h or 130	-	The driver shall release the steering control and continue to drive until the ACSF is deactivated by the system.	* Acoustic warning assessment criteria omitted here

This article has been accepted for publication in IEEE Transactions on Intelligent Vehicles. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIV.2023.3269156

					km/h whichever is lower			

This article has been accepted for publication in IEEE Transactions on Intelligent Vehicles. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIV.2023.3269156

Appendix 2 – List of literature on ADAS/AD assessment

TABLE A
LITERATURE ON DRIVER PERCEPTION OF RIDE COMFORT OF LKAS

		Test Scenario		
Literature	**System under Test**	**Maneuver**	**Test Track**	**Driver Perception – Ride Comfort**
2014 – Siebert [88]	ACC	following	simulator	**comfort**
2015 - Elbanhawi [34]	ADAS/AD	-	-	resulting forces; natural paths; motion sickness; apparent safety
2016 - Scherer [89]	highly automated driving functions in longitudinal direction	startup, braking	barricaded area of public road	**comfort**
2016 - Pawellek [75]	ACC	column braking	proving ground	**comfort**
2017 - Krauns [76]	LKAS and ACC (ADAS vs HAD)			**comfort**
2018 - Gallep [35]	AD	sinus steering, sinus accelerating	simulation	motion sickness
2018 - Fen [28]	LKAS	-	-	**comfort**
2018 - Kidd [33]	ACC, LKAS	daily driving	public roads	**comfort**
2019 - Oschlies [30]	LKAS	lane keeping	test track (straight, curve)	**comfort**
2019 - Moser [68]	manual driving	synthetic lane keeping with oscillations; synthetic lane change	lane keeping - proving ground; lane change - highway	**comfort**
2019 - Schöggl [90]	ADAS	-	-	**comfort**

TABLE B
LITERATURE ON DRIVER PERCEPTION OF PERCEIVED SAFETY OF LKAS

		Test Scenario		
Literature	**System under Test**	**Maneuver**	**Test Track**	**Driver Perception – Perceived Safety**
2014 - Jentsch [41]	EBS	-	simulator; test track	**perceived safety**
2014 – Siebert [88]	ACC	following	simulator	risk; criticality
2015 - Elbanhawi [34]	ADAS/AD	-	-	system responsiveness; safe maneuvering; system reliability
2016 - Pawellek [75]	ACC	column braking	proving ground	**perceived safety** (implicit)
2017 - Krauns [76]	LKAS and ACC (ADAS vs HAD)	-	-	**perceived safety**
2018 - Heiderich [40]	manual driving	road surface excitations	simulator	**perceived safety**
2018 - Fen [28]	LKAS	-	-	**perceived safety**

This article has been accepted for publication in IEEE Transactions on Intelligent Vehicles. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIV.2023.3269156

2019 - Oschlies [30]	LKAS	lane keeping	test track (straight, curve)	**perceived safety**
2019 - Schick [29]	LKAS	various according to maneuver catalogue	various according to maneuver catalogue	**perceived safety**
2019 - Schöggl [90]	ADAS	-	-	**perceived safety**
2018 – Ping [28]	manual driving	-	city roads	risk perception

TABLE C
LITERATURE ON DRIVER PERCEPTION OF INTERACTION OF LKAS

		Test Scenario		
Literature	**System under Test**	**Maneuver**	**Test Track**	**Driver Perception – Interaction**
2019 - Schick [29]	LKAS	various according to maneuver catalogue	various according to maneuver catalogue	**interaction**
2019 - Fridman [44]	ADAS	NDS	NDS	**interaction**
2019 - Park [32]	LKAS	following (hands on/off)	highway	steering stability performance; non-interference performance
2018 - Ran [91]	LKAS	-	simulator	intervention timing
2018 - Naujoks [92]	AD/HAD	take over	-	**interaction**

TABLE D
LITERATURE ON DRIVER PERCEPTION OF DRIVING STYLE OF LKAS

		test scenario		
Literature	**system under test**	**maneuver**	**test track**	**driver perception – driving style**
2014 - Benderius [82]	manual driving		driving simulator; test tracks; field operational test	naturality
2014 - Gordon [80]	manual driving	lane keeping while steady driving	highway	naturality
2014 - Holzinger [27]	LKA	-	-	driving line selection
2016 - Griesche [46]	AD	overtaking	simulator	**driving style**
2016 - Bellem [47]	manual driving	accelerating; decelerating; following; lane change	rural road/urban highway	**driving style**
2018 - Zhang [81]	manual driving	lane keeping		naturality
2019 - Oschlies [30]	LKAS	lane keeping	test track (straight, curve)	**driving style**

TABLE E
LITERATURE ON DRIVER PERCEPTION OF OTHER CRITERIA OF LKAS

		test scenario		
Literature	**system under test**	**maneuver**	**test track**	**driver perception – others**
2014 – Siebert [88]	ACC	following	simulator	task difficulty; effort

This article has been accepted for publication in IEEE Transactions on Intelligent Vehicles. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIV.2023.3269156

2015 - Huang [93]	LDW		fixed based driving simulator	acceptance; effectiveness
2016 - Scherer [89]	highly automated driving functions in longitudinal direction	startup, braking	barricaded area of public road	driving pleasure
2016 - Pawellek [75]	ACC	column braking	proving ground	dynamics
2017 - Krauns [76]	LKAS and ACC (ADAS vs HAD)			dynamics
2017 - Wang [23]	Human-centered steering system	following	simulator; simulation	degree of relief
2018 - Kidd [33]	ACC, LKAS	daily driving	public roads	acceptance
2019 - Oschlies [30]	LKAS	lane keeping	test track (straight, curve)	dynamics; comprehensibility; attractivity
2019 - Schick [29]	LKAS	various according to maneuver catalogue	various according to maneuver catalogue	vehicle reaction; availability; degree of relief
2019 - Park [32]	LKAS	following (hands on/off)	highway	summarized as intrusive feeling

REFERENCES

[1] H. J. Heider, M. Schallehn, C. Schlegel and K. Stricker, „Autonomes Fahren: Zukunftsstrategien für Automobilzulieferer - was führende Zulieferer jetzt angehen müssen," [Online]. Available: https://www.bain.com/de/insights/an-autonomous-car-roadmap-for-suppliers/. [Accessed 24 08 2019].

[2] "Gesetz zum autonomen Fahren tritt in Kraft," Bundesministerium für Digitales und Verkehr, 27 July 2021. [Online]. Available: https://www.bmvi.de/SharedDocs/DE/Artikel/DG/gesetz-zum-autonomen-fahren.html. [Accessed 10 September 2022].

[3] "SAE J3016 - Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," SAE International, 2021.

[4] B. Schick, C. Seidler, S. Aydogdu and Y.-J. Kuo, "Driving experience vs. mental stress with automated lateral guidance from the customer's point of view," in *9th International Munich Chassis Symposium*, Munich, Germany, 2018.

[5] H. Winner, S. Hakuli, F. Lotz und C. Singer, Handbuch Fahrerassistenzsysteme - Grundlagen, Komponenten und Systeme für aktive Sicherheit und Komfort, Springer Fachmedien Wiesbaden, 2015.

[6] S. Sternlund, "The safety potential of lane departure warning systems—A descriptive real-world study of fatal lane departure passenger car crashes in Sweden," *Traffic Injury Prevention,* vol. 18, no. S1, pp. 18-23, 2017.

[7] R. Utriainen, M. Pöllänen and H. Liimatainen, "The Safety Potential of Lane Keeping Assistance and Possible Actions to Improve the Potential," *IEEE Transactions on Intelligent Vehicles,* vol. 5, no. 4, pp. 556-564, December 2020.

[8] M. E. Dean and L. E. Riexinger, "Estimating the Real-World Benefits of Lane Departure Warning and Lane Keeping Assist," SAE Technical Paper, 2022.

[9] E. Donges, "Aspekte der Aktiven Sicherheit bei der Führung von Personenkraftwagen," *Automobil-Industrie,* vol. 27, no. 2, pp. 183-190, 1982.

[10] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and F.-F. Li, "ImageNet: A Large-Scale Hierarchical Image Database," in *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, 2009.

[11] A. Geiger, P. Lenz and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *IEEE Conference on Computer Vision and Pattern Recognition*, Providence, 2012.

[12] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár and C. L. Zitnick, "Microsoft COCO: Common Objects in Context," in *European Conference Computer Vision*, Zurich, 2014.

[13] M. Everingham, L. Van Gool, C. K. Williams, J. Winn and A. Zisserman, "The PASCAL Visual Object Classes (VOC) Challenge," *International Journal of Computer Vision,* vol. 88, no. 2, pp. 303-338, September 2009.

[14] X. Li, K. Wang, Y. Tian, L. Yan, F. Deng and F.-Y. Wang, "The ParallelEye Dataset: A Large Collection of Virtual Images for Traffic Vision Research," *IEEE Transactions on Intelligent Transportation Systems,* vol. 20, no. 6, pp. 2072-2084, June 2019.

[15] X. Li, Y. Wang, L. Yan, K. Wang, F. Deng and F.-Y. Wang, "ParallelEye-CS: A New Dataset of Synthetic Images for Testing the Visual Intelligence of Intelligent Vehicles," *IEEE Transactions on Vehicular Technology,* vol. 68, no. 10, pp. 9619-9631, October 2019.

[16] L. Li, N. Zheng and F.-Y. Wang, "A Theoretical Foundation of Intelligence Testing and Its Application for Intelligent Vehicles," *IEEE Transactions on Intelligent Transportation Systems,* vol. 22, no. 10, pp. 6297-6306, 2021.

[17] L. Li, W.-L. Huang, Y. Liu, N.-N. Zheng and F.-Y. Wang, "Intelligence Testing for Autonomous Vehicles: A New Approach," *IEEE Transactions on Intelligent Vehicles,* vol. 1, no. 2, pp. 158-166, 2016.

[18] X. Zhang, J. Tao, K. Tan, M. Törngren, J. M. G. Sánchez, M. R. Ramli, X. Tao, M. Gyllenhammar, F. Wotawa, N. Mohan, M. Nica and H. Felbinger, "Finding Critical Scenarios for Automated Driving Systems: A Systematic Mapping Study," *IEEE Transactions on Software Engineering,* 2022.

[19] "ISO 11270:2014(E) Intelligent Transport Systems - Lane Keeping Assistance Systems (LKAS) - Performance Requirements and Test Procedures," International Organization for Standardization, 2014.

[20] "UN Regulation No.79: Uniform Provisions Concerning the Approval of Vehicles with Regard to Steering Equipment," United Nations, 2017.

[21] "Euro NCAP: Test Protocol - Lane Support Systems," European New Car Assessment Programme, 2017.

[22] M. K. Diab, A. N. Abbas, H. H. Ammar and R. Shalaby, "Experimental Lane Keeping Assist for an Autonomous Vehicle Based on Optimal PID Controller," in *2nd Novel Intelligent and Leading Emerging Sciences Conference*, 2020.

[23] W. Wang, J. Xi, C. Liu and X. Li, "Human-Centered Feed-Forward Control of a Vehicle Steering System Based on a Driver's Path-Following Characteristics," *IEEE Transactions on Intelligent Transportation Systems,* vol. 18, no. 6, pp. 1440-1453, June 2017.

[24] S. Xu and H. Peng, "Design, Analysis, and Experiments of Preview Path Tracking Control for Autonomous Vehicles," *IEEE Transactions on Intelligent Transportation Systems,* vol. 21, no. 1, pp. 48-58, January 2020.

[25] G. Wang, W. Deng and S. Zhang, "A Comprehensive Testing and Evaluation Approach for Autonomous Vehicles," *SAE Technical Paper 2018-01-0124,* 3 April 2018.

[26] J. Holzinger und E. Bogner, „Objektvierte Bewertung von Fahrerassistenzfunktionen," *ATZ,* pp. 16-20, 2017.

[27] J. Holzinger, P. Schöggl, M. Schrauf und E. Bogner, „Objektivierung der Fahrbarkeit beim automatisierten Fahren," *ATZ,* pp. 48-53, 2014.

[28] A. Fen, „Simulative Ermittlung von Key-Performance-Indikatoren komfortorientierter Fahrerassistenzsysteme," *ATZ,* pp. 96-99, 2018.

[29] B. Schick, F. Fuhr, M. Höfer and P. E. Pfeffer, "Attribute-based Development of Advanced Driver Assistance Systems," *ATZ Worldwide,* pp. 64-69, 04 2019.

[30] H. Oschlies, Komfortorientierte Regelung für die automatisierte Fahrzeugquerführung, Springer Fachmedien Wiesbaden GmbH, 2019.

[31] M. Decker, Zur Beurteilung der Querdynamik von Personenkraftwagen, Technische Universität München, 2009.

[32] K. Park, J. Kwahk, S. H. Han, M. Song, D. G. Choi, H. Jang, I. S. Jeong, Y. D. Won and I. S. Jeong, "Modelling the Intrusive Feelings of Advanced Driver Assistance Systems Based on Vehicle Activity Log Data: Case Study for the Lane Keeping Assistance System," *International Journal of Automotive Technology,* vol. 20, no. 3, pp. 455-463, 2019.

[33] D. G. Kidd and I. J. Reagan, "System Attributes that Influence Reported Improvement in Drivers' Experiences with Adaptive Cruise Control and Active Lane Keeping after Daily Use in Five Production Vehicles," *International Journal of Human–Computer Interaction,* vol. 35, no. 11, pp. 972-979, 2019.

[34] M. Elbanhawi, M. Simic and R. Jazar, "In the Passenger Seat: Investigating Ride Comfort Measures in Autonomous Cars," *IEEE Intelligent Transportation Systems Magazine,* pp. 4-17, 24 July 2015.

[35] J. Gallep and S. Müller, "Vehicle Guidance Control for Automated Driving Considering Motion Sickness," in *AVEC*, Beijing, 2018.

[36] N. Kamiji, Y. Kurata, T. Wada and S. Doi, "Modeling and Validation of Carsickness Mechanism," in *SICE Annual Conference*, Kagawa, Japan, 2007.

[37] B. E. Donohew and M. J. Griffin, "Motion sickness: Effect of the frequency of lateral oscillation," *Aviation Space and Environmental Medicine,* pp. 649-656, 2004.

[38] A. Kelly and A. Stentz, "Rough Terrain Autonomous Mobility—Part 1: A Theoretical Analysis of Requirements," *Autonomous Robots,* pp. 129-161, May 1998.

[39] P. Ping, Y. Sheng, W. Qin, C. Miyajima and K. Takeda, "Modeling

Driver Risk Perception on City Roads Using Deep Learning," *IEEE Access,* pp. 68850-68866, November 2018.

[40] M. Heiderich, S. Leonhardt, W. Krantz, J. Neubeck and W. Jochen, "Method for analysing the feeling of safety at high speed using virtual test drives," in *18. Internationales Stuttgarter Symposium*, Stuttgart, Germany, 2018.

[41] M. Jentsch and A. C. Bullinger, "Is simulation (not) enough? Results of a validation study of an autonomous emergency braking system on a test track and in a static driving simulator," in *Human Factors and Ergonomics Society Europe Chapter 2014 Annual Conference*, Lisbon, Portugal, 2014.

[42] K. Park, S. H. Han and J. Kwahk, "Toward Trustworthy and Comfortable Lane Keeping Assistance System: An Empirical Study of the Level of Haptic Authority," *International Journal of Human-Computer Interaction,* pp. 1347-1363, 18 Feburary 2021.

[43] M. Benloucif, C. Sentouh, J. Floris, P. Simon and J.-C. Popieul, "Online adaptation of the Level of Haptic Authority in a lane keeping system considering the driver's state," *Transportation Research Part F,* vol. 61, pp. 107-119, 12 September 2019.

[44] L. Fridman, D. E. Brown, M. Glazer, W. Angell, S. Dodd, B. Jenik, J. Terwilliger, A. Patsekin, J. Kindelsberger, L. Ding, S. Seaman, A. Mehler, A. Sipperley, A. Pettinato and B. D. Seppelt, "MIT Advanced Vehicle Technology Study: Large-Scale Naturalistic Driving Study of Driver Behavior and Interaction With Automation," *IEEE Access,* pp. 102021-102038, 1 July 2019.

[45] J. Elander, R. West and D. French, "Behavioral Correlates of Individual Differences in Road-Traffic Crash Risk: An Examination of Methods and Findings," *Psychological Bulletin,* pp. 279-294, April 1993.

[46] S. Griesche, M. Dotzauer and D. Käthner, "Should my car drive as I do? What kind of driving style do drivers prefer for the design of automated driving functions?," in *17. Braunschweiger Symposium Automatisierungssysteme, Assistenzsysteme und eingebettete Systeme für Transportmittel (AAET)*, Braunschweig, 2016.

[47] H. Bellem, T. Schönenberg, J. F. Krems and S. Michael, "Objective metrics of comfort: Developing a driving style for highly automated vehicles," *Transportation Research Part F,* pp. 45-54, 5 July 2016.

[48] J. Weber, Automotive Development Processes, Springer-Verlag Berlin Heidelberg, 2009.

[49] C. Schimmel, „Entwicklung eines fahrerbasierten Werkzeugs zur Objektivierung subjektiver Fahreindrücke," Ph.D. dissertation, Technische Universität München, Munich, Germany, 2010.

[50] S. Monsma, "Feel the Tire - Tire Influence on Driver's Handling Assessment," Ph.D. dissertation, Aalto University, Helsinki, Finland, 2015.

[51] M. Ersoy and S. Gies, Fahrwerkhandbuch, Wiesbaden: Springer Vieweg, 2017.

[52] W. D. Käppler, Smart Vehicle Handling, Springer-Verlag Berlin Heidelberg, 2015.

[53] "Fahrassistenzsysteme - Pflicht ab dem 6. Juli 2022," TÜV Nord, [Online]. Available: https://www.tuev-nord.de/de/privatkunden/ratgeber-und-tipps/technik/fahrassistenzsysteme/. [Accessed 10 September 2022].

[54] ISO3888-2:2011(E), "Passenger cars — Test track for a severe lane-change manoeuvre — Part 2: Obstacle avoidance," International Organization for Standardization.

[55] G. G. Gómez, "Towards efficient vehicle dynamics: From subjective assessments to objective metrics, from physical to virtual testing," Ph.D. dissertation, KTH Royal Institute of Technology, Stockholm, Sweden, 2017.

[56] M. Harrer, "Characterisation of Steering Feel," Ph.D. dissertation, University of Bath, Bath, UK, 2007.

[57] Higuchi, Akira; Sakai, Hideki;, "Objective Evaluation Method of On-Center Handling Characteristics," in *SAE 2001 World Congress*, Detroit, Michigan, 2001.

[58] M. Nybacka, X. He, G. G. Gómez, E. Bakker and L. Drugge, "Links Between Subjective Assessments and Objective Metrics for Steering," *International Journal of Automotive Technology,* pp. 893-907, 2014.

[59] M. Nybacka, X. He, Z. Su, L. Drugge and E. Bakker, "Links between subjective assessments and objective metrics for steering, and evaluation of driver ratings," *Vehicle System Dynamics: International Journal of Vehicle Mechanics and Mobility,* 2014.

[60] G. L. Gil Gómez, E. Bakker, M. Nybacka and L. Drugge, "Analysing vehicle dynamics objective and subjective testing in winter conditions," in *24th International Symposium on Dynamics of Vehicles on Roads and Tracks*, Graz, Austria, 2015.

[61] G. L. Gil Gómez, M. Nybacka, E. Bakker and L. Drugge, "Findings from subjective evaluations and driver ratings of vehicle dynamics: steering and handling," *Vehicle System Dynamics,* p. 1416–1438, 2015.

[62] "ISO 17361:2017(E) Intelligent Transport Systems - Lane Departure Warning Systems - Performance Requirements and Test Procedures," International Organization for Standardization, 2017.

[63] "Euro NCAP: Assessment Protocol - Safety Assist," European New Car Assessment Programme, 2018.

[64] "Japanese Technical Guideline to Lane-Keeping Assist Devices of Motor Vehicles," Ministry of Land, Infrastructure, Transport and Tourism (MLIT), 2013.

[65] "Lane Departure Warning System Confirmation Test and Lane Keeping Support Performance Documentation," U.S. Department of Transportation National Highway Traffic Safety Administration (NHTSA), 2013.

[66] Z. Xin, X. Zhang, G. Shi and Y. Lin, "Steering Feel Study on the Performance of EPS," in *IEEE Vehicle Power and Propulsion Conference (VPPC)*, Harbin, China, 2008.

[67] H. A. S. Ash, "Correlation of Subjective and Objective Handling of Vehicle Behaviour," Ph.D. dissertation, the University of Leeds, Leeds, UK, 2002.

[68] U. Moser, N. Harmening and D. Schramm, "A new method for the objective assessment of ADAS based on multivariate time series classification," in *19. Internationales Stuttgarter Symposium*, Stuttgart, 2019.

[69] M. Fritzsche, Ein Beitrag zur Objektivierung des Sicherheitsempfindens mit Fahrermodellen bei Lenkstörungen, Universität Duisburg-Essen, 2016.

[70] W. D. Käppler, P. A. Ruijs, M. J. van Randwijk and J. Godthelp, "Methodology for predicting car and truck handling assessments," in *XXIV FISITA Congress*, London, UK, 1992.

[71] K. Backhaus, B. Erichson, W. Plinke und R. Weiber, Multivariate Analysemethoden, Springer-Verlag Berlin Heidelberg, 2018.

[72] J. F. Hair, W. C. Black, B. J. Babin and R. E. Anderson, Multivariate Data Analysis, Pearson Prentice Hall, 2009.

[73] D. C. Chen and D. A. Crolla, "Subjective and objective measures of vehicle handling: drivers & experiments," *Vehicle System Dynamics,* p. 576–597, 1998.

[74] M. Rothhämel, J. IJkema and L. Drugge, "A method to find correlations between steering feel and vehicle handling properties using a moving base driving simulator," *Vehicle System Dynamics,* 08 2011.

[75] T. Pawellek, L. Liesner, R. Henze und F. Kücükay, „Objektivierungsverfahren für eine kundenoptimale ACC-Systemauslegung," *ATZ,* pp. 74-79, 04 2016.

[76] F. Krauns, „Objektivierung kombinierter Längs- und Querführung," in *AAET*, Braunschweig, 2017.

[77] K. Backhaus, B. Erichson und R. Weiber, Fortgeschrittene Multivariate Analysemethoden, Springer-Verlag Berlin Heidelberg, 2015.

[78] R. P. King, D. A. Crolla, H. A. S. Ash and J. Whitehead, "Identification of Subjective-Objective Vehicle Handling Links Using Neural Networks for the Foresight Vehicle," in *SAE World Congress*, Detroit, Michigan, 2002.

[79] D. A. Crolla, D. C. Chen and J. P. Whitehead, "Vehicle handling assessment using a combined subjective-objective approach," in *International Congress and Exposition*, Detroit, Michigan, 1998.

This article has been accepted for publication in IEEE Transactions on Intelligent Vehicles. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIV.2023.3269156

[80] T. Gordon and K. Srinivasan, "Modeling Human Lane Keeping Control in Highway Driving with Validation by Naturalistic Data," in *IEEE International Conference on Systems, Man, and Cybernetics*, San Diego, CA, USA, 2014.

[81] Y. Zhang, T. Gordon, M. Martinez-Garcia and C. Bingham, "Steering measurement decomposition for vehicle lane keeping – A study of driver behaviour," *Measurement,* pp. 26-38, 21 March 2018.

[82] O. Benderius and G. Markkula, "Evidence for a fundamental property of steering," in *Human Factors and Ergonomics Society 58th Annual Meeting*, 2014.

[83] A. Eskandarian, B. Van Arem, E. Pucher, A. Pruckner, R. Stroph, P. Pfeffer, P. Handel, W. Huber, A. Broggi, K. Kompass, S. Andrews, C. Laugier and M. Parent, Handbook of Intelligent Vehicles, London: Springer-Verlag, 2012.

[84] H. Lietz, T. Petzoldt, M. Henning, J. Haupt, G. Wanielik, J. Krems, H. Mosebach, J. Schomerus, M. Baumann and U. Noyer, Methodische und technische Aspekte einer Naturalistic Driving Study, Verband der Automobilindustrie, 2008.

[85] D. Heaven, "Deep Trouble for Deep Learning," *Nature,* vol. 574, pp. 163-166, October 2019.

[86] X. Li, P. Ye, Z. Liu, L. Cao and F.-Y. Wang, "From Features Engineering to Scenarios Engineering for Trustworthy AI: I&I, C&C, and V&V," *IEEE Intelligent Systems,* vol. 37, no. 4, pp. 18-26, 2022.

[87] X. Li, Y. Tian, P. Ye, H. Duan and F.-Y. Wang, "A Novel Scenarios Engineering Methodology for Foundation Models in Metaverse," *IEEE Transactions on Systems, Man, and Cybernetics: Systems,* 2022.

[88] F. W. Siebert, M. Oehl and H.-R. Pfister, "The influence of time headway on subjective driver states in adaptive cruise control," *Transportation Research Part F: Traffic Psychology and Behaviour,* vol. 25, pp. 65-73, July 2014.

[89] S. Scherer, D. Schubert, A. Dettmann, F. Hartwich and A. Bullinger-Hoffmann, "Wie will der „Fahrer“ automatisiert gefahren werden? Überprüfung verschiedener Fahrstile hinsichtlich des Komforterlebens," in *VDI/VW-Gemeinschaftstagung Fahrerassistenzsysteme und automatisiertes Fahren*, Wolfsburg, Germany, 2016.

[90] P. Schöggl, E. Bogner, E. Ramschak and L. Welsersheimb, "Der Mensch im Mittelpunkt der ADAS-Entwicklung und -Objektivierung," *ATZ,* pp. 40-45, 2019.

[91] W. Ran, H. Chen, J. Chen, X. Lan, Y. Nishimura, K. Kitahara and K. Ando, "Study on Driver’s Preferred Lane-keeping Assistance System Intervention Timing Based on Subjective and Objective Correlation," in *AVEC*, Beijing, 2018.

[92] F. Naujoks, K. Wiedemann, N. Schömig, O. Jarosch and C. Gold, "Expert-based controllability assessment of control transitions from automated to manual driving," *MethodsX,* vol. 5, pp. 579-592, 2018.

[93] Z. Huang, Y. Wu and J. Liu, "Research on effects of pattern, amplitude and frequency of pulse steering torque warnings for lane departure," *Transportation Research Part F: Traffic Psychology and Behaviour,* vol. 31, pp. 67-76, May 2015.

Sijie Wei received the B.S. degree in automotive engineering from Tongji University, Shanghai, China, in 2014. She received the M.S. degree in automotive mechatronics from Munich University of Applied Sciences, Germany, in 2017. She is currently Ph.D. student at TU Wien. Her research field is vehicle dynamics and human-centered development of Advanced Driver Assistance Systems.

Peter E. Pfeffer received the B.S. and the M.S. degree from the Vienna University of Technology, Austria. From 2002 to 2007 he was lecturer at the Mechanical Engineering Department of the University of Bath, where he researched the interaction of vehicle and steering system regarding handling and received his Ph.D. degree in 2007. He is currently a Professor of automotive engineering at the Munich University of Applied Sciences, Germany, since 2007 and is cofounder and CEO of MdynamiX AG since 2014. He is the Scientific Director of the International Munich Chassis Symposium and editor of the *Steering Handbook* (Springer Cham, 2017). He is a member of the scientific board of the VDI (the Association of German Engineers) and is leading the regional section South Bavaria. His research interests are automated driving, vehicle dynamics, steering systems, steering feel, elastomer and hydro mounts.

Johannes Edelmann received the Ph.D. degree in 2008 and his venia docendi for Technical Dynamics in 2011 at TU Wien, Austria. From 2018 to 2022 he was an Associated Professor at the Institute of Mechanics and Mechatronics, TU Wien. Since October 2021 he is a Professor for Technical Dynamics at the Institute of Mechanics and Mechatronics. His research interests are applied multi-body dynamics, stability of dynamical systems, vehicle dynamics, driver modelling, dynamic behavior of bicycles, scooters, motorcycles, and automobiles, development of driver assistance systems and vehicle dynamics control systems.