Validating SuperHuman Automated Driving Performance

Zlatan Ajanovic[1], Matthijs Klomp[2], Bakir Lacevic[3], Barys Shyrokau[4], Paolo Pretto[1], Hassaan Islam[1], Georg Stettinger[1] and Martin Horn[5]

***Abstract*— Closed-loop validation of autonomous vehicles is an open problem, significantly influencing development and adoption of this technology. The main contribution of this paper is a novel approach to reproducible, scenario-based validation that decouples the problem into several sub-problems, while avoiding to brake the crucial couplings. First, a realistic scenario is generated from the real urban traffic. Second, human participants, drive in a virtual scenario (in a driving simulator), based on the real traffic. Third, human and automated driving trajectories are reproduced and compared in the real vehicle on an empty track without traffic. Thus, benefits of automation with respect to safety, efficiency and comfort can be clearly benchmarked in a reproducible manner. Presented approach is used to benchmark performance of SBOMP planner in one scenario and validate SuperHuman driving performance.**

***Index Terms*— automated driving, validation, lane change, multi-lane driving, traffic lights, urban driving, planning, SuperHuman.**

I. Introduction

Autonomous Vehicles (AV) (vehicles with SAE Level 5 [1]), are promising to enable Automated Driving (AD) functions with SuperHuman performance in terms of safety, efficiency, and comfort [2]. However, besides challenges related to development (involving multi-objective goals), validation and benchmarking of various AD solutions on real vehicles, in realistic traffic situations, is one of the critical technical issues. Current traffic safety sets a very high bar with about 210 million km driven between two fatal accidents (about 230 years of non-stop driving with velocity 100 km/h) [3]. Therefore, a large efforts in testing with diverse participants is necessary to test and ensure AD acceptance. In this respect, besides in-vehicle testing, the usage of software simulators and driving simulators can reduce testing costs and enhance consistency, thus offering controlled and repeatable driving conditions.

Expected mainstream approach to validation is by providing the statistical performance comparison of Autonomous

*https://youtu.be

[1]Z. Ajanovic, Paolo Pretto, Hassaan Islam and Georg Stettinger are with Virtual Vehicle Research GmbH, Graz, Austria. zlatan.ajanovic@v2c2.at

[2]Matthijs Klomp is with the Volvo Car Group, Gothenburg, Sweden. matthijs.klomp@volvocars.com

[3]B. Lacevic is with the Faculty of Electrical Engineering, University of Sarajevo, 7100 Sarajevo, Bosnia and Herzegovina. bakir.lacevic@etf.unsa.ba

[4]B. Shyrokau is with the Department of Cognitive Robotics, Delft University of Technology, The Netherlands, Mekelweg 2, 2628 CD Delft, The Netherlands. b.shyrokau@tudelft.nl

[5]M. Horn is head of Institute of Automation and Control at Graz University of Technology, Graz, Austria. martin.horn@tugraz.at

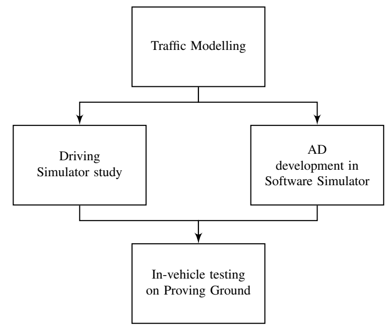

Fig. 1: Validation methodology overview.

Vehicle fleet with regular human driving statistics [4]. However, this requires a large fleet of automated vehicles (Level 4-5). Such fleet is not yet available on the road, and requires some kind of validation in advance, rendering it chicken-and-egg kind of a problem. The strategy of many players in the field is to avoid real world testing by utilizing software simulators [5]. However, software simulators still lack realism and real vehicle testing cannot be fully avoided [6].

Besides the statistical validation on a fleet of the vehicles and larger scenario distribution, other approach is to validate benefits in a particular scenario. The interest from society regarding autonomous vehicles is usually reflected with scenario related quations, ethical dilemmas (variants of Trolley Problem), where only bad options remain as possible and the vehicle has to choose between one bad action over another [7]. To quantify benefits of AD in a particular scenario, one approach could be to reconstruct a particular driving scenario and measure the performance of human drivers and AVs. The main challenge (which is practically unsolvable) in such approach is to accurately reconstruct the real driving situations involving multiple traffic participants.

The main contribution of this paper is a novel approach to reproducible, scenario-based validation by decoupling the problem into several sub-problems, without breaking the crucial couplings. Furthermore, the proposed approach was used to benchmark SBOMP planner (presented in [8]) in one scenario and validate SuperHuman driving performance.

The rest of the paper is structured as follows. In the section II, the validation problem is presented. Section III summarizes the proposed methodology. Sections IV-VII present individual components of the methodology. Finally, Section VIII concludes the paper.

II. Autonomous Vehicle Validation Problem

Validation and homologation of automated driving remains as a challenge and is an active research topic. In [9], Beglerovic et. al provide an overview of the challenges for Autonomous Vehicle validation. For validation of AD decision making and control system in particular, we can state the following major challenges:

- It is practically impossible to realistically recreate the same traffic scenario with many participants, for different experiment runs.
- The traffic is a multi-agent problem with closed loops, as actions of ego driver influence other traffic. Therefore, replaying previously recorded real traffic scenarios is not realistic for validation of AD decision making and control systems (as compared to perception systems).
- Testing in-vehicle might be expensive, complicated for legal approval or dangerous.

A methodology presented here, overcomes these challenges by decoupling the problem and using advantages of multiple validation modalities.

III. Methodology

In this work, a novel approach to reproducible, scenario-based validation is presented. Decoupling of the problem is performed by carefully choosing decoupling points, such that the crucial couplings (closed-loops) are not broken, as illustrated in Figure 1. First, a realistic scenario is generated from the real urban traffic. Second, human participants, drive in a virtual scenario (in a driving simulator), based on the real traffic. Third, human and automated driving trajectories are reproduced and compared in the real vehicle on an empty track without traffic. Thus, benefits of automation with respect to safety, efficiency and comfort can be clearly benchmarked in a reproducible manner. Such decoupling makes the approach technically and economically feasible. In this work, decoupling enabled experiments to be performed at several different locations across Europe, including Graz (Austria), Delft (The Netherlands), Sarajevo (Bosnia and Herzegovina) and Gothenburg/Sandhult (Sweden), effectively, utilizing available resources. In fact, in [9], Beglerovic et. al state that “A clever combination of methods and validation environments (SiL, HiL, test-track, public road etc.) is necessary”. This concurs with the approach presented here.

The assumptions made here are as follows.

- Traffic can be realistically simulated (this holds if scenario is not highly interactive, as it is the case in this scenario).
- Simulated vehicle dynamics represents real vehicle dynamics.
- Driving simulator can be used to gather realistic human driving behavior.
- Vehicle dynamics and passenger comfort can be tested without other vehicles on an empty proving ground.

Table I summarizes advantages provided by each validation modality, demonstrated in a realism of different aspects of the overall validation problem. Decoupling of the problem is performed such that each modality supports the overall validation by providing some aspects as ”real” (marked in bold). All other crucial aspects, neccessary to realistically validate that aspect, are either real or high fidelity models. Eventually, all aspects are real and complete reproducibility of the validation procedure is provided as well.

The realistic traffic scenario is generated from real traffic. The focus in this stage is to capture the road geometry, environment (traffic lights position and timing) and traffic density. As the situation is highly dynamic but not highly interactive, simplified driver models can be used, without loss of realism. Therefore, from public roads, infrastructure and other traffic participants aspects were covered. Traffic scenarios on a public road are not reproducible. However, once traffic scenario is captured, it can be reproduced in the simulators.

Vehicle dynamics and traffic simulator were used for the development process of AD function, as well as for the initial validation. Simulator provided high fidelity model for vehicle dynamics, other traffic participants as well as complete reproducibility.

To obtain naturalistic driver behavior, multiple human participants were tested in driving simulator, in the same traffic scenario. This modality satisfies coupling of human driver and dynamic driving environment, with an appropriate vehicle model. It provides the overall validation with realistic human driving behavior and complete reproducibility of the procedure.

Finally, the driving trajectories from driving simulator and AD system were executed using the real vehicle on an empty track on the proving ground. This modality satisfies the coupling of vehicle dynamics and the occupant comfort. It supports the overall validation by testing both vehicle dynamics and driving comfort and provides full reproducibility.

IV. The real traffic

For validation scenario, the traffic from a segment of the street “Zmaja od Bosne” in Sarajevo, Bosnia and Herzegovina, was chosen. In particular, the $750\ \mathrm{m}$ long segment from “Trg Bosne i Hercegovine” to the Campus of the University of Sarajevo was used. Figure 2 shows the environment from the considered street. The segment contains three traffic lights on a short distance and is relatively straight. The timings of the traffic lights were experimentally obtained based on a recorded video. Traffic lights are located at $188\ \mathrm{m}$, $361\ \mathrm{m}$ and $682\ \mathrm{m}$ after the segment starting point. They turn yellow $12.7\ \mathrm{s}$, $25.7\ \mathrm{s}$ and $47.7\ \mathrm{s}$ after the scenario start, respectively. They turn red $3\ \mathrm{s}$ after yellow. They have a phase of red light of $45\ \mathrm{s}$, $45\ \mathrm{s}$ and $24\ \mathrm{s}$, respectively.

Artificial traffic was created with the density of $30\ \mathrm{veh/km/lane}$ and the average velocity of $12\ \mathrm{m/s}$. All traffic participants are implemented with driver model that satisfies traffic light signals, keeps the current lane and keeps appropriate spacing to other vehicles in front.

V. Validation in software simulator

Motion Planning algorithms can be efficiently developed and preliminary validated in software simulator, following basic principles of Test-driven development (TDD). In this

TABLE I: Comparison of validation modalities.

	Public roads	Vehicle & Traffic simulator	Driving simulator	Proving ground
Human driver	Real	N/A	**Real**	Real
AI driver	N/A	**Real**	N/A (Real)	N/A (Real)
Passenger	Real	N/A	High fidelity	**Real**
Vehicle dynamics	Real	High fidelity	High fidelity	**Real**
Infrastructure	**Real**	High fidelity	High fidelity	Real/High fidelity
Other traffic participants	**Real**	High fidelity	High fidelity	Not realistic
Reproducible	No	Yes	Yes	Yes

Fig. 2: Real traffic scenario.

Fig. 3: Prescan Simulator.

work, particular Motion Planning algorithm under test is SBOMP planner [8]. It is a search-based motion planning algorithm, providing lateral and longitudinal desired motion in urban traffic, aiming to achieve energy-optimal driving. To speed up planning it may utilize different heuristics. One based on Dynamic Programming (DP) solution from the relaxed problem [10], and the second one analytical uderestimate based on the model (MB) [11].

Two kinds of software simulator are used in this work. Custom developed traffic simulator is used for the development and initial validation and high fidelity vehicle dynamics simulator is used for adapting lateral vehicle motion behavior and (soft) real-time implementation.

The lane change feasibility within a sufficiently large time interval is validated using a higher fidelity vehicle model, Figure 3. The used vehicle model has 10 degrees of freedom (DoF) covering 6 DoF of the vehicle body and 4 DoF of vertical motion of unsprung masses. The vehicle body motion in space has longitudinal, lateral, vertical, roll, pitch and yaw motions. Assuming a smooth driving in high friction conditions results in small wheel slip, the wheel rotational dynamics can be neglected. We assume the linear cornering stiffness for the considered vehicle operational conditions, with the relaxation behavior included. The longitudinal motion of the vehicle body is modelled taking into account the applied wheel torques (both traction and brake torques), air drag, road resistance and slope forces.

SBOMP planner was developed in custom developed traffic simulator. As scenarios of interest are not highly interactive, other traffic participants are modelled based on IDM [12]. The simulator enables convenient and script based adjustment of scenarios, including street geometry, other traffic participants, traffic light, etc. Test scenario was created based on a real traffic scenario as presented in section IV. Figure 4 presents one situation from the scenario. In this situation, the ego vehicle plans a lane change in order to pass the red vehicle in front and catch the green light. To make the clearance for lane change, the ego vehicle speeds up to get close to the preceding red vehicle (where the gap is), slows down during the lane change (to provide enough time for lane change) and again speeds up (to pass the red vehicle) to catch the green light. This situation truly demonstrates the importance of integrated planning for longitudinal and lateral motion. The blue tree represents searched trajectories, and the red trajectory represents the final solution. Projections of trajectories on the $s \times l$ and $s \times t$ plane are shown on the left and middle plot respectively. Vehicles shown on the left plot represent polygonal obstacles in the middle plot. The rightmost image is the screenshot from PreScan software showing the part of the real street used in the study. The resulting red trajectory shows the vehicle reaching just behind the red vehicle, slowing down to provide enough time for changing to the left lane and speeding up to passing the red vehicle, while catching the green light.

To demonstrate the robustness, stochastic variations of the scenario are created by introducing randomized perturbations of initial positions and velocities. The results indicate that the proposed approach is robust to variations in the scenario and without significant deviations from the initial solution regarding the cost and time of travel. These results serve as a reference for comparison with human driving performance.

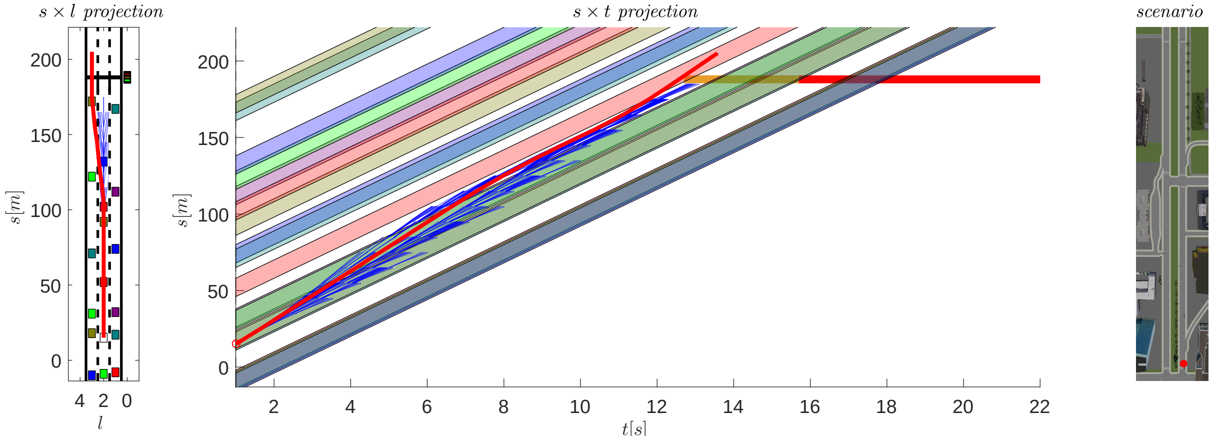

Fig. 4: Urban driving in the presence of traffic light.

VI. Driving Simulator

Driving simulators are cost-effective tools for objective evaluation of the human driving behavior in a controlled environment, that enable reproducibility of scenarios [13]. In this work, driving simulator was used to collect data for a comparison of AD system with actual human driving performance in the same scenario.

Human participants were asked to drive along the road as they would normally do, in a virtual environment created within the driving simulator to emulate traffic scenario as presented in section IV. The scenario consisted of a straight road with multiple lanes, passing through a city with relatively congested traffic. Two sessions of repetition were performed for each participant, to acquire both “normal driving” and “energy-efficient driving”. The complete study, for each participant, lasted approximately 30 minutes, including small breaks after the familiarization and between the 2 driving sessions.

A. DriveLab driving simulator

For human driver study, DriveLab driving simulator located at Virtual Vehicle Research, Graz (Austria) is used. DriveLab driving simulator is state-of-the-art 3-DoF driving simulator. It has a mock-up vehicle cockpit that is fabricated from a small sedan car body, resembling the real vehicle interior with some added features. Software system is based on VI-DriveSim[1] (VI-grade, Italy) and SCANeR Studio[2] (AV Simulation, France) running on Concurrent Real-time Workbench[3]. It enables vehicle dynamics and environment simulations essential for this study. With the use of three integrated visual projectors with circular screen, a horizontal field of view of $215°$ is achieved. Together with three rear mirror screens, it gives a fully immersive feeling to the driver. For acoustic support, a 7.1 surround sound system is used to replicate the acoustics of the real vehicle. A digital dashboard provides the information to the driver about the vehicle speed and other attributes, including engine rpm and engaged gear.

[1]https://www.vi-grade.com/en/products/static-simulator/
[2]https://www.avsimulation.fr/solutions/
[3]https://www.concurrent-rt.com/products/redhawk-linux/

Fig. 5: Driving simulator.

The vehicle with automatic transmission is considered in the study, so only two pedals (throttle and brake) are used, without gear shifting.

B. Virtual Environment

The virtual environment resembling an urban street was created within driving simulator software, based on the real traffic scenario. A three-lane road with traffic lights and lane-markings as described in section IV is generated. To provide full urban environment feeling, architecture and urban design is also modelled. The traffic was created by placing 17 vehicles on predefined positions. Vehicles are programmed to drive with the target velocity of 12 $\mathrm{m/s}$, but they react to the environment and slow down for the traffic light and other vehicles. They follow their initial lane and do not change the lane. The length of the road of interest is $1000\ \mathrm{m}$, ending with the finish line. To provide the same scenario for all participants and enable further benchmarking of the results, scenario trigger is used when participant reaches the segment of interest. They drive on an empty road and approach traffic that is frozen. When they reach the trigger point, the scenario starts, including other vehicles and traffic lights timing. Participants are instructed to approach other traffic in the middle lane with approximately $50\ \mathrm{km/h}$, without slowing down. The real-time data was collected, including the driven trajectories with a specific time stamp for later analysis.

C. Procedure

Each participant was assigned dedicated time-slot for the study. First, participants received a safety briefing and oral driving instructions about the driving task, and completed an intake questionnaire.

Next, participants completed a 5-minute training session to get used to the driving simulator. Five minutes was regarded as sufficient for becoming accustomed to driving in a simulator according to various studies. During the training session, the participants drove in the environment which was different then the environment used in the study. It consisted of more turns, intersections and interactions with other vehicles. It was intentionally set as such to expose participants in order to gain more experience and develop the feeling of how the virtual vehicle was responding to the input command and how the visual scene in driving simulator behaves.

After the training session, the actual experiment started, consisting of two sessions, with each session having three repetitions. First, a driving task was briefly rehearsed. Participants were then instructed to drive naturally as they would drive in reality, and keep an appropriate driving behavior during the whole experiment, respecting the following safety and legal considerations:

- respect the speed limit of 60 km/h,
- avoid collision with other vehicles,
- respect the traffic lights,
- it is possible to to use both sides to overtake other vehicles.

They were particularly advised to approach the trigger point with 50 km/h, so that all participants have the same scenario start. When scenario is triggered they could drive as they would normally drive in the traffic. Participant would start the scenario from standstill, drive into the traffic (which triggers the traffic) and drive the full scenario until the finish line. This was repeated three times for the first session. For the second session and three more repetitions, they were advised to drive more energy efficiently, to try to catch the green light but still keep smooth driving.

As a conclusion, a talk with the participants was used to inform them about future steps of the study and to get feedback about the realism of the driving simulator study. There were no objections of the participants. The complete procedure for each participant lasted about 30 minutes.

D. Results

In total, 28 volunteers (23 males, 5 females) participated in the study in the months of May and August 2019. All of them had normal or corrected-to-normal vision. They all had driving experience and were in a possession of valid driving license. The average participant age was 28,5 years (std 4,3 years).

From 28 participants, 25 participant completed the study. Each participant drove maximum 6 driving trajectories. In total, 113 driving trajectories were collected. From 113 trajectories, 103 trajectories were within $50 \pm 10 \text{ km/h}$ initial velocity and were used for comparison. From 103 trajectories, 55 trajectories fall into category of normal driving while 48 trajectories correspond to energy-efficient driving.

TABLE II: Instances of braking traffic rules by human drivers.

variant	number of trajectories	speed limit	speed limit $+5 \text{ km/h}$	speed limit $+10 \text{ km/h}$	yellow TL	red TL	collision
normal	55	28	15	7	38	0	2
normal [%]	100	50.91	27.27	12.73	69.09	0	3.64
energy-efficient	48	27	9	5	35	0	0
energy-efficient[%]	100	56.25	18.75	10.42	72.92	0	0

Different driving trajectories are presented on the Figure 6 and 7. From $t \times s$ projection, it can be observed that there are four main clusters of trajectories. The three clusters of vehicles that stop on some traffic light, and one cluster of trajectories that manage to get the green wave. It can be observed that many drivers pass during yellow and some even during red light. From $l \times s$ projection, it can be observed that the lateral motion is not so distinct among drivers. The stopping on three traffic lights can also be seen on $s \times v$ diagram from Figure 7. From the same diagram, it can be seen that many drivers violated the speed limit of 60 km/h.

Table II presents the results of the analysis of traffic rule violation in this scenario. Clearly, drivers violated the speed limit in more than 50% of driving runs and more than 10% of driving runs with a margin of 10 km/h. In more than 69% of driving runs, drivers passed during the yellow light. Moreover, drivers caused 2 critical situations which can be categorized as collisions. Table III presents the results of the driving behavior based on energy efficiency and travel time. First, results from "normal driving" session are presented, followed by "energy-efficient driving session". Additionally, the single results of the best and the worst run based on energy-consumption and travel time are presented. It can be observed that drivers generally improve performance in terms of energy efficiency and travel time in the second session (energy-efficient driving). This might be caused by the advice, but might also be due to gained experience about the scenario. Still, the average performance of human drivers is more than 60% worse than AD. In fact, AD is about 20% **more energy-efficient** than the best energy-efficient human driver, and **faster than the fastest human driver** (without violating traffic rules). Presented results here include human-driven trajectories which violate traffic rules, therefore the results would be even more in AD advantage.

VII. Proving Ground

Once all other stages are successfully completed, experimental evaluation on the proving ground with the real vehicle

TABLE III: Driving performance of human drivers based on energy and travel time.

driving	energy used [kJ]	diff. [%]	travel time [s]	diff. [%]
normal driving	815.42 ± 128.41	+95.87	106.11 ± 24.31	+84.53
energy-efficient driving	670.94 ± 104.09	+62.44	82.14 ± 20.47	+42.85
the most energy-efficient	510.79	+22.69	68.99	+19.98
the fastest	571.59	+37.3	60.0	+4.34
Automated Driving	416.3	0	57.5	0

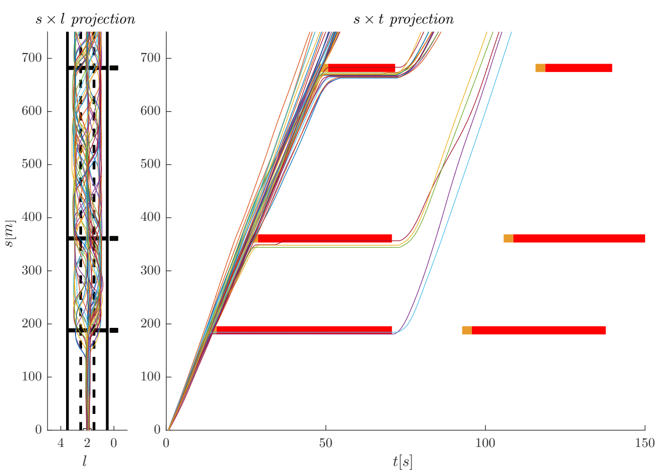

Fig. 6: Results from human driver study in driving simulator.

can be executed, as the final stage. Testing on the proving ground provides reproducibility by the use of driving robots. Testing is performed with a goal to verify simulation results and to validate occupant driving comfort. Verification of the simulation should assess how well the models represent the real world. The expected outcome is that the feasibility and user acceptance of the planned trajectories are confirmed. Tests in a real vehicle are also important for passenger driving comfort validation, as it cannot be validated otherwise. Driving simulators generally do not provide the full range of driving dynamics experience.

In this work, testing on the proving ground is performed in cooperation with Volvo Cars and AstaZero[4]. Human-driven trajectories from driving simulator and automated driving trajectories are executed on the empty test track by the vehicle instrumented with a steering robot. Similar approach to testing on empty field was presented in [14]. There, the authors presented the test of an autonomous vehicle on an empty track with simulated dynamic environment. However, the results were not benchmarked with human drivers and in realistic urban traffic situation with traffic lights. On the other hand, the presented work in this paper has covered this drawback.

[4]http://www.astazero.com/

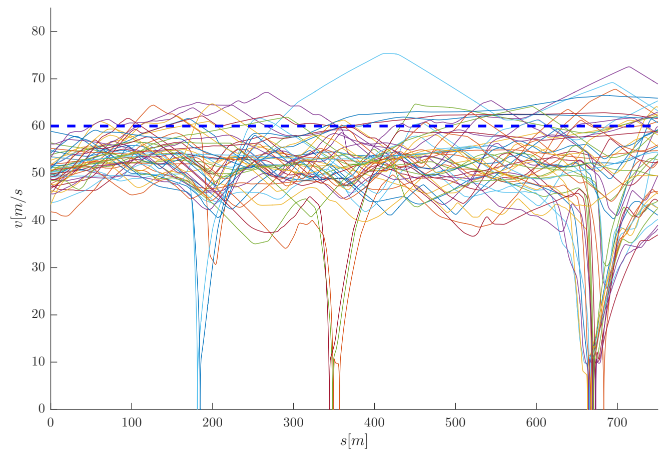

Fig. 7: Results from human driver study in driving simulator.

Fig. 8: Test vehicle Volvo S90.

A. Demo Vehicle

Volvo S90 T5 was used as a demo vehicle. Figure 8 presents the demo vehicle used for this work on the test-track. Demo vehicle is instrumented with the steering robot ABDynamics. Figure 9 shows the steering robot setup in the vehicle, actuating the steering wheel, gas and brake pedals. It has differential GPS positioning with RTK (real-time kinematics). To execute maneuvers, steering robot uses dedicated file format, so-called PMC files, which can be recorded by driving to reproduce some tests. For this work, a software tool was created that interprets human-driven trajectories from driving simulator and generates appropriate PMC files, that can be further executed on the real vehicle using steering robot. During experiments, data from the steering robot and vehicle CAN signals were collected.

B. Proving ground

As mentioned earlier, experiments were executed at AstaZero Hällered, Sandhult, (Sweden) proving ground. In particular, multilane test track was used, as shown in Figure 10. Some experiments were also performed on a rural road, but they are not reported here. Multilane road is 700 m long. As some room is needed to accelerate the vehicle in order to

Fig. 9: Steering robot used for executing driving trajectories.

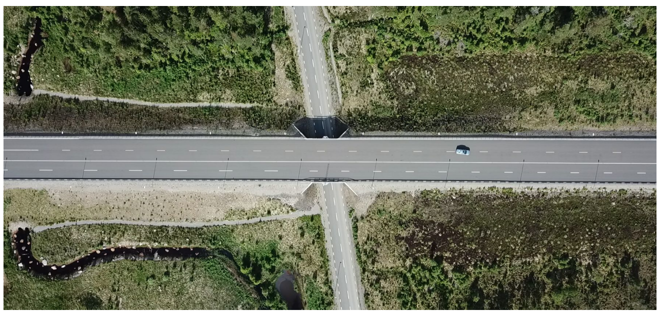

Fig. 10: AstaZero proving ground, multilane road.

start scenario with $50~\mathrm{km/h}$, the segment of $550~\mathrm{m}$ was used for reproducing the driving trajectories. That was enough to reproduce the trajectories, including first two traffic lights. Multilane road has four lanes, from which the left three lanes are used in experiments. The cones were placed to mark the position of traffic lights

C. Experiments

Experiments were performed for a duration of several days in the months of June and August 2019. Human driven trajectories and automated driving trajectories were executed for several times each, while the robot and vehicle data were collected. The developed tools for interpreting trajectories for a robot showed to be very robust, requiring minimal additional manual work by test engineers. In total, 67 trials were executed and recorded.

The trajectories were successfully reproduced with a tracking errors for lateral deviation and longitudinal speed less than $0.01~\mathrm{m}$ and $2~\mathrm{m/s}$, as shown in Figure 11. It is worth pointing out that some trajectories were very extreme (emergency full braking), with longitudinal accelerations exceeding $11~\mathrm{m/s^2}$. Trajectory execution was robust, so it was possible to reproduce same trajectories several months after the initial tests.

To validate the perceived human safety and comfort, 8 participants were driven as passengers in the vehicle. Vehicle executed three driving trajectories. Two driving trajectories were human-driven from the driving simulator, and one corresponded to AD. Participants were asked to rate the driving

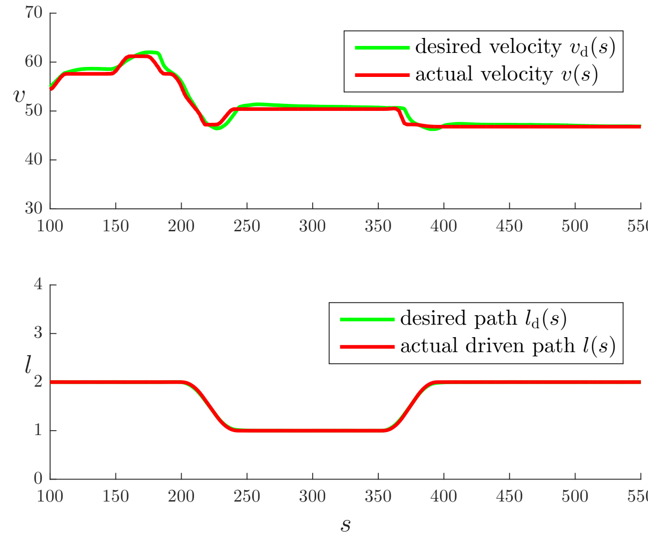

Fig. 11: Results from executing Automated Driving trajectory by test vehicle.

on the scale of 1 to 5 based on their perceived safety and comfort. The ratings from 1 to 5 indicate driving as: "very bad", "bad", "ok", "good" and "very good" respectively. Table IV presents the results of the study. Average rating is presented along with the standard deviation.

Although the sample is not sufficiently large to make a general conclusion, the presented results demonstrate that automated driving was rated higher in average than two human-driven trajectories. One human trajectory in particular was rated bad, as it included a sudden stop on the second traffic light.

TABLE IV: Passenger ratings of driving trajectories (2 human and one automated driving).

	Comfort			Safety		
	1st run	2nd run	AD run	1st run	2nd run	AD run
Rating	1.75 ± 0.66	3.125 ± 1.05	3.625 ± 0.86	2.25 ± 1.20	3.25 ± 0.66	3.875 ± 0.60

Successful execution of human-driven and AD trajectories on the proving ground provided the confirmation of the simulation study. The reproducibility of the testing on the proving ground is demonstrated also by the fact that scenarios are reproduced even a few months after their initial setup.

VIII. Conclusion

The presented approach to scenario-based validation of AD functions demonstrates that the complex problem of validation can be effectively decoupled. Modelling the realistic traffic helped to obtain useful traffic scenarios. Simulation tools helped significantly during the development and the initial validation. Driving simulator was an effective solution to get a wide variety of human responses to the same driving scenario. The study confirmed that the scenario is challenging, as even good drivers had to focus to consistently reproduce the good performance every time. Even with the experience in the scenario (in 5. or 6. iteration), it would happen that drivers do not succeed to reproduce good performance. Additionally, accident rates obtained in the study are orders of magnitude larger than current safety standards from everyday driving, as distribution of scenarios is different in everyday driving, containing easier and harder scenarios [3]. The experiments on the proving ground confirmed the simulation results, and enabled to acquire useful feedback from participants about the perceived safety and comfort.

Validation effort demonstrated that the developed AD function based on SBOMP planner achieves SuperHuman driving performance in terms of safety, efficiency and comfort, in this scenario. While many drivers violate traffic rules (56% of drivers violate speed limit, 10% even with margin of $10\ \mathrm{km/h}$) and cause accidents, AD system does not. AD system has demonstrated better energy efficiency, in particular 22% better than the best human driven trajectory (from more than 100 trials and almost 30 participants). Finally, passengers rated AD better than other two human-driven trajectories in terms of perceived safety and comfort.

As for now, this validation represents a proof of concept and is based on a single scenario. For more general conclusions a deeper study is necessary, including more participants and more scenarios. This methodology is not designed for validation of interactive scenarios, but could be potentially extended by coupling more driving simulators.

ACKNOWLEDGMENT

The project leading to this study has received funding from the European Unions Horizon 2020 research and innovation programme under the Marie Skodowska-Curie grant agreement No 675999, ITEAM project.

The publication was written at Virtual Vehicle Research GmbH in Graz and partially funded within the COMET K2 Competence Centers for Excellent Technologies from the Austrian Federal Ministry for Climate Action (BMK), the Austrian Federal Ministry for Digital and Economic Affairs (BMDW), the Province of Styria (Dept. 12) and the Styrian Business Promotion Agency (SFG). The Austrian Research Promotion Agency (FFG) has been authorised for the programme management.

The validation work on proving ground was one of the Open Research projects at AstaZero, partially financed by RISE and Chalmers and led by SAFER.

REFERENCES

[1] On-Road Automated Driving (ORAD) committee, “Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles,” 400 Commonwealth Drive, Warrendale, PA, United States.

[2] Z. Ajanović, “Towards SuperHuman Autonomous Vehicles,” PhD thesis, Graz University of Technology, Graz, Austria, 2019.

[3] W. Wachenfeld and H. Winner, “The Release of Autonomous Vehicles,” in *Autonomous driving*, M. Maurer, J. C. Gerdes, B. Lenz, and H. Winner, Eds. New York NY: Springer Berlin Heidelberg, 2016, pp. 425–449.

[4] 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, B. D. Seppelt, L. Angell, B. Mehler, and B. Reimer, “MIT Advanced Vehicle Technology Study: Large-Scale Naturalistic Driving Study of Driver Behavior and Interaction With Automation,” *IEEE Access*, vol. 7, pp. 102 021–102 038, 2019.

[5] Darrell Etherington, “Waymo has now driven 10 billion autonomous miles in simulation,” 2019. [Online]. Available: https://techcrunch.com/2019/07/10/waymo-has-now-driven-10-billion-autonomous-miles-in-simulation/

[6] Q. Chao, H. Bi, W. Li, T. Mao, Z. Wang, M. C. Lin, and Z. Deng, “A Survey on Visual Traffic Simulation: Models, Evaluations, and Applications in Autonomous Driving,” *Computer Graphics Forum*, vol. 60, no. 3, p. 1083, 2019.

[7] J.-F. Bonnefon, A. Shariff, and I. Rahwan, “The social dilemma of autonomous vehicles,” *Science*, vol. 352, no. 6293, pp. 1573–1576, 2016.

[8] Z. Ajanović, B. Lacevic, B. Shyrokau, M. Stolz, and M. Horn, “Search-Based Optimal Motion Planning for Automated Driving,” in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 10/1/2018 - 10/5/2018, pp. 4523–4530.

[9] H. Beglerovic, S. Metzner, and M. Horn, “Challenges for the Validation and Testing of Automated Driving Functions,” in *Advanced microsystems for automotive applications 2017*, ser. Lecture Notes in Mobility, C. Zachäus, B. Müller, and G. Meyer, Eds. Cham, Switzerland: Springer, 2018, pp. 179–187.

[10] Z. Ajanović, M. Stolz, and M. Horn, “Energy Efficient Driving in Dynamic Environment: Considering Other Traffic Participants and Overtaking Possibility,” in *Comprehensive energy management*, ser. SpringerBriefs in applied sciences and technology, Automotive engineering : simulation and validation methods, 2191-530X, D. Watzenig and B. Brandstätter, Eds. Cham, Switzerland: Springer, 2017, vol. 6, pp. 61–80.

[11] Z. Ajanović, M. Stolz, and M. Horn, “A novel model-based heuristic for energy-optimal motion planning for automated driving,” *IFAC-PapersOnLine*, vol. 51, no. 9, pp. 255–260, 2018.

[12] M. Treiber, A. Hennecke, and D. Helbing, “Congested traffic states in empirical observations and microscopic simulations,” *Physical Review E*, vol. 62, no. 2, pp. 1805–1824, 2000.

[13] B. Shyrokau, J. de Winter, O. Stroosma, C. Dijksterhuis, J. Loof, R. van Paassen, and R. Happee, “The effect of steering-system linearity, simulator motion, and truck driving experience on steering of an articulated tractor-semitrailer combination,” *Applied ergonomics*, vol. 71, pp. 17–28, 2018.

[14] M. Werling, S. Kammel, J. Ziegler, and L. Gröll, “Optimal trajectories for time-critical street scenarios using discretized terminal manifolds,” *The International Journal of Robotics Research*, vol. 31, no. 3, pp. 346–359, 2012.