Holistic Evaluation Methodology of the User Experience of Driver Assistance Systems in the Combination of Human-Machine-Interface and Functional Performance

Seda Aydogdu[1], Dr. Thomas Kersten[2] [0009-0005-8386-8925], Kevin Schuler[3], Prof. Bernhard Schick[3] and Gioele Micheli[1] [0009-0003-6747-1050]

[1] MdynamiX AG, Junkerstraße 4, 87734 Benningen, Germany
[2]Volkswagen do Brasil, Via Anchieta km 23,5, 09823-901 São Bernardo do Campo, Brasil
[3] Institut für Fahrerassistenz und automatisiertes Fahren, Junkerstraße 1A, 87734 Benningen, Germany
seda.aydogdu@mdynamix.de, bernhard.schick@hs-kempten.de, thomas.kersten@volkswagen.com.br

Abstract. Driver assistance systems aim to make driving safer and more comfortable while increasingly used in all vehicle classes and almost all markets. However, current research results continue to show low customer satisfaction and user acceptance. This is mainly due to the fact that the user experience and the holistic consideration of driver assistance functions as well as the human-machine interface (HMI) were neglected during development. A consequence of this is a low level of utilization, resulting in unused safety potential.

In this paper, we present a comprehensive evaluation method for driver assistance systems in terms of the combination of HMI and functional performance, especially for the Lane Keeping Assist (LKA) and Adaptive Cruise Control (ACC) systems. For this purpose, a test person study was conducted in close cooperation with Volkswagen Brazil and MdynamiX. The aim of this approach is to identify weak points of driver assistance functions in combination with display and operation (HMI). Based on these findings, concrete measures are derived to increase customer satisfaction, to increase acceptance during use and thus also to increase the degree of use.

In the context of this test person study, the complete user experience (before, during, after the journey) of drivers during a real journey with two different vehicles was surveyed by means of questionnaires. In addition, the functional performance of LKA and ACC was recorded by means of objective data collection. Thus, this work represents a holistic methodology for user-centered development.

The work highlights the need for advanced development of function, operation, display and warnings with a strong focus on the needs of customers to enable clear communication regarding system state and boundaries. In this way, functions can be developed that increase both the level of use and safety, as well as driving comfort.

First and second tier suppliers, but also service providers for "user experience", "usability" and "human-centered development", benefit from this work. The findings are crucial for optimizing driver assistance systems and better meeting customer expectations.

The study is characterized by the holistic view of HMI and function from the customer's perspective. Through this approach, targeted improvements can be made to meet customer requirements and expectations.

Keywords: ADAS, Human Factors, Adaptive Cruise Control, Lane Keeping Assistance System, HMI.

1 Introduction

ADAS functions are getting more and more important in the South American Car market. This is driven on the one hand by South American legislation like Latin NCAP or MoVer (Mobilidade Verde – Green Mobility) which is an energy efficiency law. On the other hand the customers can be convinced to use ADAS functions due to safety and comfort aspects. Taking over platform solutions (e.g. from Europe or Asia) to South America as many OEMs are doing and just doing some market application in a very cost sensitive market does not match the expectations of the majority of typical customers. The following approach leads to customer oriented and tailored ADAS solutions for the South American region and can be easily adapted to other regions with different cultures. Taking the Voice of the customer as a focal point and essential pillar into the development process and using objective values requires also a change in the mindset of development engineers and optimization of the development process from OEM point of view.

2 State of the art

2.1 SAM Region in numbers

Looking to some numbers for the SAM region, we are talking about 500 million people in 29 countries, Central America included. Regarding the amount of sold cars per year, the South American market sells about 4 million cars, which is about 5% of the world-wide car market. Taking the world-wide biggest car market, namely China as a reference for innovation and technology growth, the SAM is on a similar level: With 1.12 mobile connections [1] (China 1.18) per inhabitant, 80.6% Internet penetration [2] (China 73.7%) and 71.3% social media penetration [2] (China 72%). The observed increasingly high technological affinity might indicate a positive reception of well-working ADAS – equipped vehicles in the SAM market.

2.2 ADAS Development at SAM - Overview

In general, the South American Passenger Car market is mainly a market of small cars (A0 and A segment) due to the cost sensitivity if usual passenger cars are considered and not cars like e.g. Pickups. Various ADAS Systems from L0 to L2 have been established in the last years like ACC, AEB, LDW, LKA, Travel assist, RTCA or Park Assist. Hence for example, the region is gathering for camera based instead of Radar-based systems for the functions mentioned above. Looking to the future, the task for development engineers is to climb the mountain of L2+ over even L2++ systems which are cloud-based and usually very expensive under highly cost restrictive circumstances.

2.3 Drivers and traffic conditions at SAM

Before talking about the customers' needs, we have to consider the environmental boundary conditions which are very particular in South America. Among them there are bad road conditions, hard traffic, long distances, high rate of speed bumpers speed control, and high altitudes. Travelling long distances by car is a main driver for ACC, ACC stop and go, Travel assists and LKAS systems. Taking part of huge traffic jams in megacities like Sao Paulo with 20 million people such systems like ACC, stop and go, FCW, and other collision warning systems are warmly welcome due to crossing motorbikes at high speed. Summarizing this situation, the task to the OEM is to generate customer benefit in terms of safety and comfort while taking cultural aspects into account.

Finally, the legislation like Latin NCAP and MoVer in Brazil which generates incentives to the OEMS in case of equipping the cars with ADAS functions are helping to penetrate the South American Automotive industry with ADAS systems.

2.4 Development Process

Nevertheless, assuming to have the best ADAS system in terms of functions won't be accepted in case the usage is bad and complicated and the HMI has no clear communication. Therefore, it is essential to create a new development process which takes care of all aspects regarding functionality, usability, HMI communication and cultural backgrounds. Furthermore, it is of utmost interest to relate the subject evaluation to objective KPIs. Based on that it is planned to an objective measurement of KPIs identify the correlation to the subjective impressions. This task will be done in the next month in a further project.

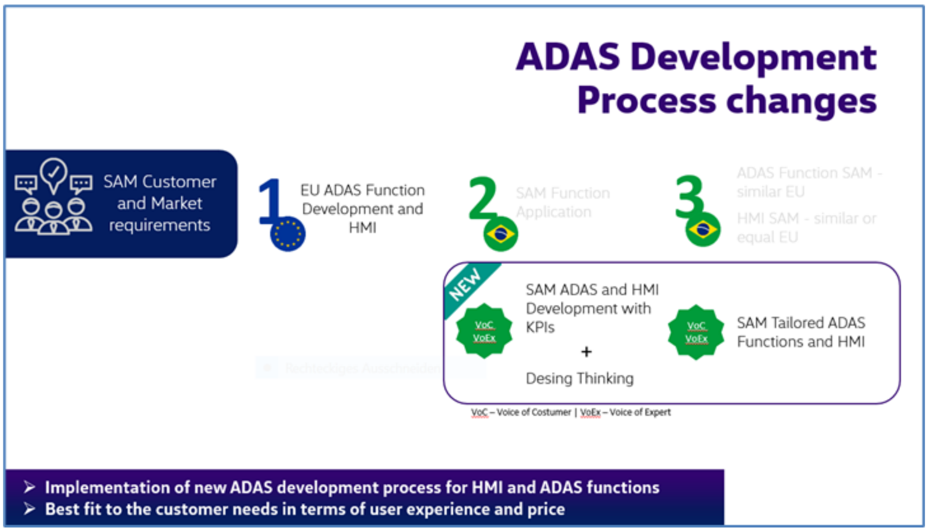

Fig. 1. Improvement of ADAS development Process at Volkswagen do Brasil

2.5 Research questions

Aim of the research is creating a methodology to develop vehicles holistically (HMI's UX and usability, functional performance) and testing whether the approach yields better results in improving driver's satisfaction. This led us to consider the following questions:

RQ1: Can a methodology be developed to compare vehicles holistically considering both Functions and HMI?

RQ2: Does the benchmarking vehicle perform better or worse compared to the Taos in with respect to ADAS and HMI?

RQ3: Which optimization tasks can be derived from participants?

3 Methods

3.1 Experiment design

In this experiment, the ADAS (LKA, ACC) of a VW Taos (model year: 2024) and of a benchmarking vehicle (model year: 2024) are compared in terms of functional and of HMI performance with 43 (scheduled: 50) participants. The benchmarking vehicle was selected based on similarity of class and segment, of availability of relevant ADAS, and of being perceived as a vehicle with capabilities to learn from. Of the 43 participants, five were considered as experts. Results recorded include both quantitative subjective questionnaires as well as qualitative, unguided interviews. Observed ADAS functions include a LKA and an ACC.

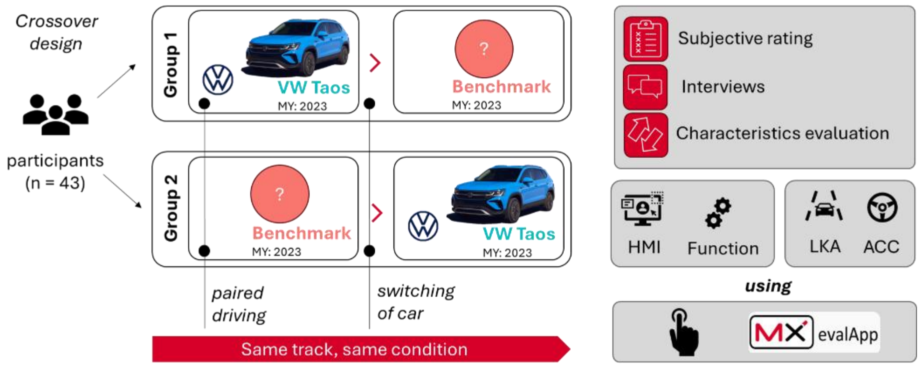

Fig. 2. Overview on the experiment design.

The trial is staged as a cross-over (AB / BA) design (compare with Fig. 2) [3]. The first group (AB) tested the VW Taos first (Period 1) and then the benchmarking vehicle (Period 2), the other group (BA) vice versa. Both participants start driving their assigned cars simultaneously to reduce between-subject variance due to varying traffic conditions. After completing the test sequence, a washout period of 30 minutes is scheduled. After the washout period, participants switch seats and perform the second round of tests with the other vehicle respectively.

Participants were sampled using a friends-and-family approach. A public recruiting was avoided for safety concerns. Participants were not actively balanced for age or gender, although both demographics were recorded. Additionally, participant's employment occupation and education levels were requested, to match participants with customer target group. Participants' prior knowledge was recorded with a pre-evaluation questionnaire. Among the information recorded, vehicle ownership, driving experience in km/year and years, ownership of ADAS, and usage of ADAS systems was recorded.

For each vehicle, there is one experiment lead respectively. The task of the experiment lead is to navigate the participating driver, instruct to initiate maneuvers, record results, and conduct the interview.

3.2 Track and maneuvers

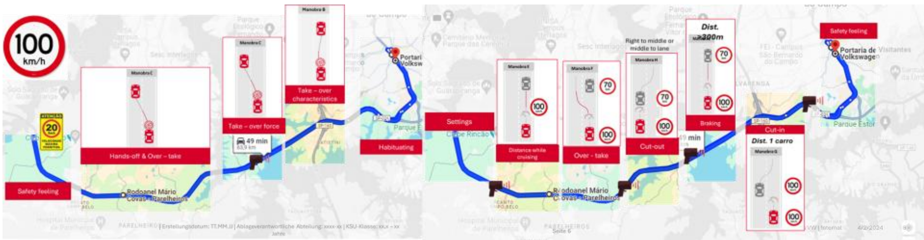

Fig. 3. Map of experiment with maneuvers. Left: LKA maneuver map (moving clockwise / right – to – left). Right: ACC maneuver map (moving counterclockwise / left – to – right).

The track the experiment takes place in starts in Portaria Anchieta (São Bernardo do Campo) following the Rodoanel Mário Covas (SP-021) until in Exit 23 (LKA, Fig 3 left), where the route follows back to the beginning (ACC, Fig. 3 right). It was selected considering the boundary conditions: Safety concerns, route with low probability of traffic jam, optimal length for number of maneuvers, minimum amount of tall collection stations, allowed speed required for ADAS system activation. Experiment leads described weather (subjectively) as *Bad* in 8% (n = 4) of cases, and in 12% (n = 6) of cases traffic was subjectively described as *Heavy*. In total, five (four) maneuvers were considered for evaluating the ACC (LKA). Figure 4 gives an overview on maneuvers considered. A written explanation can be found in Tables 2 and 3.

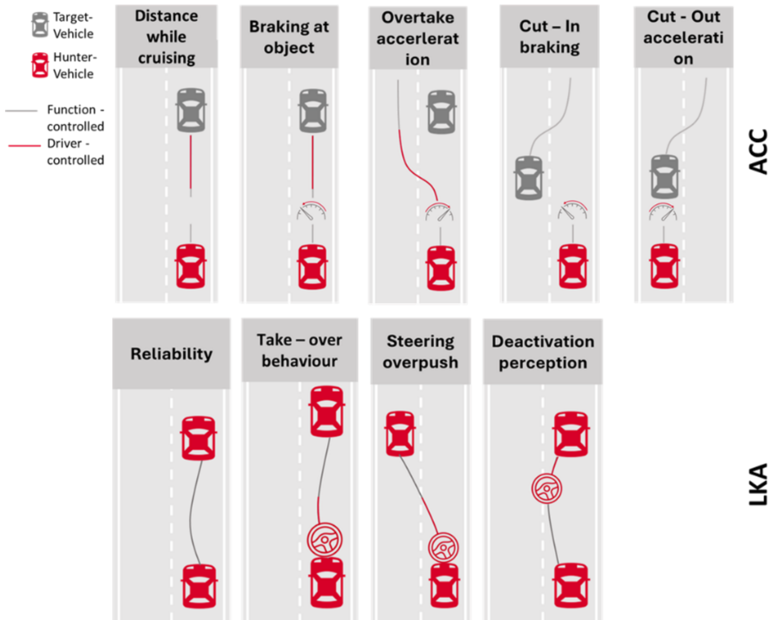

Fig. 4. Overview of maneuvers considered for evaluation.

3.3 Measurements and Criteria

Two types of data were recorded in this experiment using the MXeval App: Qualitative and quantitative data. Quantitative data includes demographic data, subjective ratings of both function and HMI, and evaluations of characteristic information. All subjective are based on a 1 – item 10 – point Likert scales. Additionally, for certain criteria based on maneuvers, a subjective characterization based on the following scales is used:

- **Timing**: How well the ADAS maneuver was timed, rated from very early to very late.
- **Strength**: How well the acceleration force was perceived, rated subjectively from very weak to very early.
- **Distance:** How well was the distance perceived that the vehicle kept from an approaching (also referred to as *target*) vehicle in a maneuver, rated subjectively from very far to very close.

These dimensions can describe parameters of a maneuver that allow creating tradeoff charts. Interviews are the only type of qualitative data recorded. Data was recorded in three phases: Before test drive (Demographic and Pre-Questionnaire), after maneuver (ACC & LKA criteria), and after experiment (HMI, LKA reliability). Criteria are derived from [4] [5] [6].

HMI criteria

The HMI serves as interface with the automated car. Criteria in this section are mainly limited to functions pertaining to the ADAS functions, therefore mainly belonging to the aHMI (automation HMI, according to the taxonomy in [7]). Table 1 provides an overview of the criteria used to rate both the LKA's and the ACC's HMI.

Table 1. Overview of HMI criteria used. (AT = After test drive, LS = Likert scale)

Criterion	**Explanation**	**Measured**
Activation usability (only ACC)	Rating general impression and usability of activation user interface	AT, LS
Adjustment usability (only ACC)	Rating general impression and usability of adjustment user interface	AT, LS
Perceived safety	Subjectively perceived degree of safety when system is active	AT, LS
Display	Subjective rating of usability and attractiveness of the display concept	AT, LS
Supervision workload	Describes workload required to supervise (monitor) the system	AT, LS
Warning system	Describes design of acoustic, optical, and haptic warnings of the ADAS	AT, LS
Ability to self-explain	Describes system usability without explanations	AT, LS
Quality to relieve	Describes subjective degree of relief from the active system	AT, LS

Single Ease Questionnaire (SEQ) [8]	Rating of overall usability in one question	AT, LS
System Usability Score (SUS) [9]	Ten-item question score to assess general usability	AT, LS

Functional criteria ACC

The ACC is an ADAS that allows "following a forward vehicle ahead at an appropriate distance" (ISO 15622 [10]). For evaluating the ACC, only criteria based on maneuvers performed are considered. Table 2 outlines the maneuvers evaluated for ACC.

Table 2 Overview of ACC criteria used. (AM = After maneuver, LS = Likert scale, FT = Force timing characteristics, D = Distance characteristics)

Criterion	**Explanation**	**Measured**
Breaking at object	Strength and intensity of breaking to avoid collision with approaching vehicle	AM, LS + FT
Cut – out acceleration	Strength of acceleration when target vehicle cut-out with ACC active	AM, LS + FT
Cut – In Braking	Timing and acceleration of braking to avoid collision with cutting in target vehicle	AM, LS + FT
Overtake acceleration	Timing and acceleration of the noticeable acceleration with ACC active	AM, LS + FT
Cruising distance	Describes Distance between target and hunter that can be set while driving	AM, LS + D

Functional Criteria LKAS

The LKA (cf. ISO 11270 [11]) is a different ADAS that can control the vehicle longitudinally and laterally to ensure it stays within the lane boundaries. Table 3 gives an overview of criteria used to rate the LKA.

Table 3 Overview of LKA criteria used. (AM = After maneuver, AT = After test drive, LS = Likert scale, FT = Force timing characteristics, D = Distance characteristics)

Criterion	**Explanation**	**Measured**
Take – Over characteristics	Criterion describes system intervention when approaching and recentering from lane limits	AM, LS + FT
Perception of Warnings	Describes how well the warnings were perceived, or if they were perceived at all	AM, LS + D

Steering take-over force	Describes steering torque required to overdrive lane markings with LKA active without indicators	AM, LS
Deactivation perceptions	System capability of reaching over control when leaving its ODD so driver can safely control vehicle	AM, LS
Warning sequence while deactivating	Describes how warnings were perceived optically, acoustically, and haptically with respect to the situationally adapted salience	AM, LS
Perceived reliability	Criterion describes consistent and precise ability to hold in lane limits without wrong warnings in all conditions	AT, LS

3.4 Hypothesis and Evaluation

As the aim of the study is comparing the vehicles as a benchmark, all criteria's values are compared in non-directional hypothesis. Accordingly, the following set of hypotheses is defined, where Criterion C_i denotes all Criteria listed in tables one, two, and three, leading to **29 pairwise criterion comparisons** (8 criteria (6) for HMI ACC (LKA), 5 criteria (6) for functions ACC (LKA)).

$H_{0,i}$: The VW Taos' average subjective score of Criterion C_i doesn't differ from the corresponding score of the benchmarking vehicle.

The (AB/BA, two-period, two-sequence) cross-over design is evaluated using the approach described in [3]. Accordingly, tests are performed to determine whether a carryover effect exists by comparing within-subject-sums. A carryover effect is assumed below a significance level of $\alpha = 0.10$, according to literature [3]. In case no carryover effect could be verified, a test for treatment effects is conducted by comparing within-subject differences using a significance level of $\alpha = 0.05$. All evaluation is performed using the R programming language to ensure a reliable, reproducible, and transparent evaluation procedure. Interviews are summarized using a LLM approach of synthesis.

4 Results

4.1 Limitations

As the procedure was implemented for the first time, some effects might have influenced the results. Mainly, the family and friends recruiting approach might have led to a distortion in which participants favorized the VW Taos over the benchmarking vehicle (Demand characteristics [12]). Similarly, the experiment leads' expectations being company internal might have further influenced the result favorizing the VW Taos.

The benchmarking vehicle performed worse or equal in nearly all (28/29) categories. Such a strong result can indicate the benchmarking process might have not considered the best system. Therefore, while results obtained from the study might reflect accurate comparisons, they don't help in understanding areas of improvement comparing with the best-in-category systems.

Furthermore, because the experiment was performed under real-life conditions, certain influences can't be controlled. In some instances, an experiment was postponed or split due to bad weather conditions. This led some participants to drive unpaired, causing the overall traffic conditions in both groups not to be symmetric. While influencing the results, issues like these seem to be the prize for ensuring real-life conditions.

4.2 Demographics and conditions

Participants' average age was 41.2 (SD = 12.9) years old; the gender ratio of participants was 48:52 female-to-male. Seven initially scheduled participants dropped out. Regarding the MDSI self-described themselves as average with a tendency to Calm – Defensive.

4.3 ADAS HMI

From criteria used to assess the HMI, the three subjectively most important categories, on average for the ACC were the perceived safety, the display, and the activation usability. For the LKA, the most important categories were the perceived safety, the display, and the supervision workload.

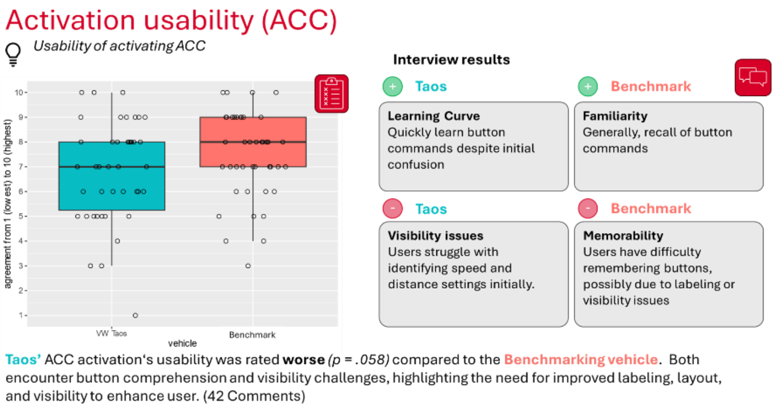

Fig. 5. Dashboard created for HMI – evaluation type criteria including an analysis on strengths and weaknesses.

To gain an overview on the results, an overview dashboard (cf. Fig. 5) is created were results of all significance testing of all categories are summarized. For each category, a detailed result dashboard is created. In it, a boxplot with all the responses overlayed is portrayed together with a Strength – and – Weaknesses synthesis of all comments to the respective field. A summary below helps management take decisions backed with hypothesis – tested results. This procedure allows quickly identifying fields of improvement and giving feedback on the spot. It was noted that some responses lack insight because interviewing questions were open.

4.4 ADAS functions

For the ACC criteria, participants preferred by the customers in order: Braking at object, Cut-In Braking, Distance while cruising, whereas for the case of the LKA the criteria preferred in order are reliability, steering intervention, and effort against intervention.

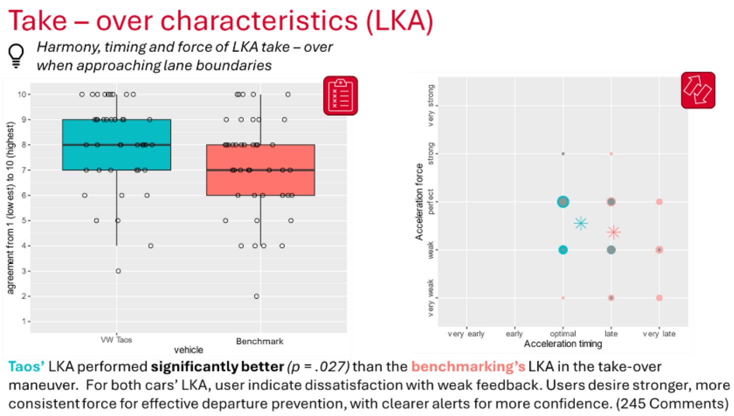

Fig. 6. Dashboard created for Function – evaluation type criteria including trade off charts. Bubbles represent frequency of characteristics combined, while stars represent a weighted average of responses.

The procedure for evaluation and improvement is the same in the case of functions like for the HMI. Like for the HMI, the in – detail dashboard (cf. Fig. 6) contains a boxplot as well as an actionable summary. The difference in this case is that for criteria in which characteristics are used (e.g. very early/very late), a trade off chart is displayed. They provide meaningful feedback to calibrate function behavior and allow tailoring functions to local market needs. An average deviation from the center (optimum) directly indicates a potential for improvement.

5 Discussion

The methodology developed in this work introduces a procedure to a part of development where none existed before. The procedure allows introducing a test procedure to ADAS (LKA and ACC) that raises both to its intentions of holistic development and being tailored to understanding the needs of customers in the region of SAM. This evaluation approach is vital because customers experience the vehicle as an entire system and not a part.

So, how did the method perform? According to us, the granularity in which criteria show potential for improvement is fine enough to identify concrete measures while not falling prone to losing the bigger picture. However, the largely one-sided test results revealed the need to implement a proper benchmarking approach to help select an optimal comparison vehicle to gain more feedback on potential for improvement.

The one-sided results also raised the question whether the friends-and family approach skewed participants opinions towards preferring the VW. For the recruiting, a conflict between the goals *safety* (avoidance of theft), *representativeness* of customer ship, and *unbiased* feedback (e.g. no internals as participants) was revealed. While the procedure as it is was sufficient to implement and test the new approach, for the future, the creation of a database of secure participants that represent our customer is being discussed.

The need to both including safety considerations and finding representative customer ship stem from the unique conditions in Brazil: high financial inequality (Gini coefficient 49.7 considered "intermediate-high" [13]) and a high degree of criminality. Whether and how cultural differences influence test results can only be answered if a test like this one was performed in two different countries, yet under similar circumstances. This would be a scenario where driving simulators would be of use. We also learned that it is important including local, mother tongue speaking experimenters for conducting tests to avoid communication problems and better predict participants' behavior.

For future projects, implementing a similar procedure in the format of an expert evaluation is planned. The expert evaluation would differ from the test person study by letting experts rate categories in more detail and objectively measuring KPIs. Especially for aspects like the HMI, this study's result results lacks indicating concrete ways of improvement. In this way, both designs can be used complimentary: One to find potential for improvement, and the other to understand what to change. In case it remains difficult to understand how these changes can be implemented, a design thinking phase could be introduced to identify solutions for remaining weak points. As an additional tool, certain safety-critical maneuvers can be tested in a driving simulator to collect information in an environment with less pressure. Once finished, such a holistic evaluation approach can be applied also for other and coming ADAS developments (e.g. parking assist) and other testing in which humans interact with the vehicle.

References

1. We Are Social. Special Report Digital 2023. Link: https://wearesocial.com/cn/blog/2023/01/digital-2023/ (Accessed 06.05.2024)
2. Data Reportal. Digital 2023 April Global Statshot Report. https://datareportal.com/reports/digital-2023-april-global-statshot (Accessed 06.05.2024)
3. Wellek S., Blettner M.: On the proper use of the crossover design in clinical trials: part 18 of a series on evaluation of scientific publications. Dtsch Arztebl Int. 2012 Apr;109(15):276-81. doi: 10.3238/arztebl.2012.0276. (2012)
4. Aydogdu S., Seidler C., and Schick B. Trust Is Good, Control Is Better? – The Influence of Head-Up Display on Customer Experience of Automated Lateral Vehicle Control. In HCI in Mobility, Transport, and Automotive Systems: First International Conference, MobiTAS 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26-31, 2019, Proceedings. Springer-Verlag, Berlin, Heidelberg, 190–207. https://doi.org/10.1007/978-3-030-22666-4_14 (2019)
5. Aydogdu S., Schick B., and Wolf M., "Claim and Reality? Lane keeping assistant - The Conflict between Expectation and Customer Experience," in Fahrzeug- und Motorentechnik, Aachen. (2018)
6. Nesensohn, J., Lefèvre, S., Allgeier, D., Schick, B., Fuhr, F.. An Efficient Evaluation Method for Longitudinal Driver Assistance Systems within a Consistent KPI based Development Process. In: Pfeffer, P.E. (eds) 11th International Munich Chassis Symposium 2020. Proceedings. Springer Vieweg, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-63193-5_7. (2021)
7. Bengler K., Rettenmaier M., Fritz N., and Feierle A. From HMI to HMIs: Towards an HMI Framework for Automated Driving. Information 11, no. 2: 61. https://doi.org/10.3390/info11020061. (2020)
8. Sauro J. and Dumas J. Comparison of three one-question, post-task usability questionnaires. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09). Association for Computing Machinery, New York, NY, USA, 1599–1608. https://doi.org/10.1145/1518701.1518946 (2009)
9. Brooke, J. SUS: A Quick and Dirty Usability Scale. In: P.W. Jordan, B. Thomas, B.A. Weerdmeester & I.L. McClelland (Eds.), Usability Evaluation in Industry. London: Taylor & Francis, 189-194. (1996)
10. International Organization for Standardization. Intelligent transport systems: Adaptive cruise control systems – Performance requirements and test procedures. ISO Standard No. 15622:2018. (2018)
11. International Organization for Standardization. Intelligent transport systems: Lane keeping assistance systems (LKAS) – Performance requirements and test procedures. ISO Standard No. 11270:2014. (2014)
12. Orne, Martin T. Demand Characteristics and the concept of Quasi-Controls. in Artifacts in Behavioral Research: Robert Rosenthal and Ralph L. Rosnow's classic books : a re-issue of Artifact in behavioral research, Experimenter effects in behavioral research and The volunteer subject, pages 110 pp. (2009)
13. UNU WIDER. World Inequality Database – WIID. Link: https://www.wider.unu.edu/database/world-income-inequality-database-wiid#WIIDcomp (Accessed 06.05.2024)