D6.2 – Results of Evaluation 2nd Cycle

Project Number:	690705
Classification	Public
Deliverable No.:	D6.2
Work Package(s):	WP6
Document Version:	Vs. 1.0
Issue Date:	Month 23
Document Timescale:	Project Start Date: September 1, 2016
Start of the Document:	Month 19
Final version due:	Month 23
Deliverable Overview:	**Main document:** D6.2
Compiled by:	Stefan Suck, OFF
Authors:	Sebastian Feuerstack, OFF Mark Eilers, HMT Elsa Yousfi, VED Elisa Landini, REL Andrea Castellano, REL Jürgen Pichen, ULM
Technical Approval:	Fabio Tango, CRF
Issue Authorisation:	Andreas Lüdtke, OFF

DISTRIBUTION LIST		
Copy type[1]	Company and Location	Recipient
T	AutoMate Consortium	all AutoMate Partners

RECORD OF REVISION		
Date	Status Description	Author
18.06.2018	Document Structure	OFF
12.07.2018	Experiment description and results	VED
20.09.2018	2nd experiment results	VED
20.09.2018	Experiment description	REL
24.09.2018	Experiment description and results	ULM
26.09.2018	Experiment Results	REL
28.09.2018	Including Partner Contributions after Reviewers Feedback	OFF

[1] Copy types: E=Email, C=Controlled copy (paper), D=electronic copy on Disk or other medium, T=Team site (Sharepoint)

Table of Contents

List of Abbreviations

A2H .... *Automation to Human*
H2A .... *Human to Automation*
KPI .... *Key Performance Indicator*
WP .... *Work Package*

List of Figures

List of Tables

1 Executive Summary

This document describes the comparative evaluation experiments of the second project cycle to demonstrate the benefits brought by the bidirectional cooperation between the human and the automation. For each simulator demonstrator a TeamMate system setup was instantiated, meaning several enablers were integrated into a driving simulator, and compared against a simulated baseline vehicle by driving use-cases of the three AutoMate scenarios (Eva, Peter, Martha). Each subsection of section 3 is dedicated to one of the three simulator demonstrators.

Section 3.1 is about the experiments for Eva roundabout scenario demonstrated in the REL simulator. The evaluation was carried out in the "H2A support in perception" use case and the "H2A support in action" use cases concerning mostly comfort and acceptance and trust features. The baseline, an autonomous vehicle which follows the driverless approach, was compared against a TeamMate car with integrated TeamMate HMI and Driver Monitoring System. The TeamMate car can ask for support if it hesitates to enter the roundabout or share control with the driver, while the baseline has to wait or the automation must be deactivated. In the measured KPIs the TeamMate system showed significant improvements in the performance against the baseline.

The demonstration of the Peter rural road scenario in the ULM simulator is described in Section 3.2. The evaluation was carried out in the "H2A support in perception" use case and the "A2H support in action" use cases. The baseline car was a state-of-the-art automated car. For the TeamMate car the Augmented Reality HMI, multiple Interaction Modalities, the Driver Intention

Recognition, and the Online Risk Assessment were integrated. To overtake a slowly driving car in the scenario the baseline had to deactivate the automation otherwise it had to follow. While the TeamMate car could overtake the car by assessing the risk and the usual driver intention (A2H case) automatically or the driver could assess the situation and command the automation to initiate the automatic overtaking manoeuvre (H2A case). The evaluation results show a benefit of the TeamMate car compared to the baseline car in all measured KPIs.

In Section 3.3 the experiments for the Martha extra-urban road scenario in the VED simulator are described. The "H2A support in action" use case and the "A2H" use cases are considered. The H2A baseline followed the driverless approach while the TeamMate car was equipped with the TeamMate HMI and V2I communication. To pass the roadwork side in this scenario the control was handed to the driver. The take-over request was issued earlier in the TeamMate car due to V2I communication. In all measured KPIs a benefit of the TeamMate approach can be observed. The baseline for the A2H case was manual driving, while the Driver Monitoring System was integrated into the TeamMate car. The resulting distraction due to a secondary task in this scenario could be detected by the TeamMate system. Thus, the automation could offer to take control of the car to avoid critical situations. Increased safety with the TeamMate system in comparison to the baseline is shown in the results.

2 Introduction

This document describes the evaluation results of the project cycle 2. During this cycle the enablers were integrated into the demonstrators in WP5, and the performances of the 1st version of the demonstrators is evaluated against their corresponding baseline in WP6.

The aim of the evaluation is to demonstrate the benefits brought by the bidirectional cooperation between the human and the automation in terms of safety, comfort, trust and acceptance.

The deliverable D6.1 has presented the methodology that is now being applied for the evaluation of the demonstrators. For the evaluation two approaches are considered:

1. When the automation supports the driver (A2H support)
2. When the driver supports the automation (H2A support)

Both approaches provided benefits: A2H mainly in terms of safety, H2A mainly for efficiency. In previous deliverables, especially D1.3 and D6.1, scenarios and use cases have been identified and assigned to the demonstrators to highlight the benefit of each approach against its baseline. In fact, different baselines have been identified, to measure the performance of the TeamMate car in the different use cases and approaches:

- A "manual driving" baseline for the A2H support, to quantify the impact of the support of the automation
- An "autonomous driving" baseline for the H2A support, to quantify the impact of the support of the driver to the automation

For each demonstrator, specific KPIs have been defined to measure the performance of the TeamMate car against its baseline to measure how the cooperation between the driver and the TeamMate car can provide a benefit in terms of safety, efficiency and, as a consequence, in terms of trust in the automation and acceptance of the new technology.

3 Comparative Evaluations in Simulators

This chapter describes the comparative evaluation of a TeamMate car against a baseline car in three driving simulators with three different scenarios and several use cases. Despite the common aim to demonstrate the benefits of the cooperation of driver and automation partially different features and KPIs are evaluated due to the different scenarios.

3.1 REL Simulator

3.1.1 Specific aim

One of the experiments to evaluate the benefits introduced by AutoMate has been performed in REL driving simulator. In this phase, the aim of this experiment was to evaluate the added value of the features developed in the project after the first integration on the driving simulator demonstrator.

In order to evaluate the AutoMate’s added value, the EVA use case, including the roundabout, has been implemented in a driving simulation scenario, using SCANeR Studio 1.7, a driving simulation software engine able to reconstruct driving situations and collect data on the driving behaviour.

The scenario used in the simulation included three roundabouts, in order to measure different elements and, indeed, different KPIs and features; the detailed evaluation scenario will be described in the following paragraph.

In EVA scenario will be evaluated mostly comfort- and acceptability-related features. The scenario will implement two types of support, i.e. the H2A support in perception and the H2A support in action.

Since, as stated in D6.1, the baseline is the autonomous driving, in the first case the aim is to evaluate the benefits given by the support in perception in terms of efficiency and acceptability. In the second case, the aim is to compare the typical vehicle-driver interaction (i.e. the takeover request, implemented in the baseline), with a more complex form of interaction that can be described as cooperation in action, i.e. the shared control.

3.1.2 Description of Evaluation Scenario

This chapter will describe in detail the scenario in which the test took place. The following use case belonging to EVA scenario was implemented:

> *"The TeamMate car is driving in Automated Mode. When it approaches a roundabout, it detects high traffic flows that can affect the efficiency (i.e. the TeamMate car evaluates that it may take some time to enter the roundabout in Automated Mode). To speed up the manoeuvre, the TeamMate car asks Eva for cooperation in perception, asking her to check the available space and to provide a trigger to start the manoeuvre. Eva checks the traffic and gives the confirmation to enter the roundabout. The TeamMate car understands the feedback and enters the roundabout in Automated Mode"*

The scenario was designed to be realistic and complex enough to allow the users to have a representative and meaningful driving behaviour. Moreover, in order to collect enough driving data, the scenario was not limited to a roundabout, but included a series of driving tasks in a mixed (urban and extra-urban) environment.

The scenario development can be logically divided into two parts:

- The **TERRAIN** development, i.e. the physical shape of the environment in which the simulation takes place (that is common for baseline and TeamMate scenario).
- The **SCENARIO** and **SIMULATION** development, in which the traffic rules, the car behaviour and interaction logics are implemented (these are different for baseline and TeamMate driving scenarios).

The terrain in which the simulation occurred was designed with a focus on three roundabouts, since the roundabout is the focus of the use case. However, in the driving simulator a complex scenario with high traffic flows was created in order to allow 8 minutes of driving simulation for each condition (i.e. baseline and TeamMate).

Three roundabouts were designed in order to present to the user three different interaction modalities in that situation:

- At the first and the second roundabout ("Small") the car (both in baseline and in TeamMate scenarios) is able to perform the roundabout, but interaction could be needed to increase the efficiency.
- At the third roundabout ("Big"), since there are no lanes on the road, cooperation in action is needed.

Figure 1 shows the terrain itinerary in the simulation environment, while Figure 2 shows a detail of the simulation from the driver's point of view before the small roundabout.

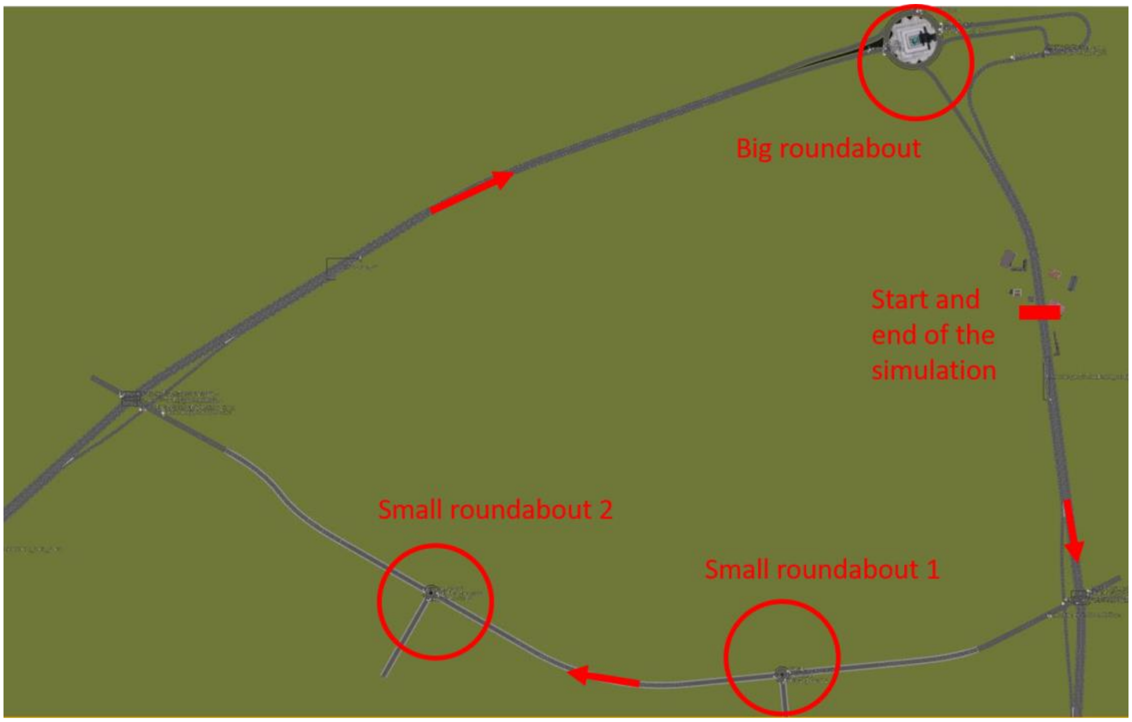

Figure 1: Baseline and TeamMate itinerary in EVA scenario

Figure 2: Representation of EVA scenario in REL driving simulator

The aim of EVA scenario is to show the value of the driver to support the automation: for this reason, the baseline is represented by a condition where the driver has no role in the cooperation (i.e. the “driverless” approach): therefore, the baseline is the autonomous driving.

In the baseline scenario, the simulation started in Automated Mode. When arriving at small roundabouts the car hesitates, since there are traffic flows and the car waits until it deems the situation safe enough to enter the roundabout.

After the first two roundabouts, the baseline car continues its driving path in autonomous mode. When it approaches the third (big) roundabout, it asks for support in action from the driver (i.e. a take-over request) since it is not able to approach it without the reference of the lines. Thus, the driver has to take full control.

In this scenario, the take-over request is given to the driver through a very simple video + audio message: to do this, a basic message was given, in order to ask for the transition of control, using the same screen of the TeamMate HMI, communicating with the driving simulator according to the same protocol.

A simplified representation based on a static text message and a “hands-on the steering wheel” icon for the take-over was given. In the baseline no distinction is done according to the attention level, since the baseline doesn’t include the enablers developed in AutoMate, i.e. the Driver Monitoring System.

Figure 3 shows the schematic representation of the baseline driving scenario.

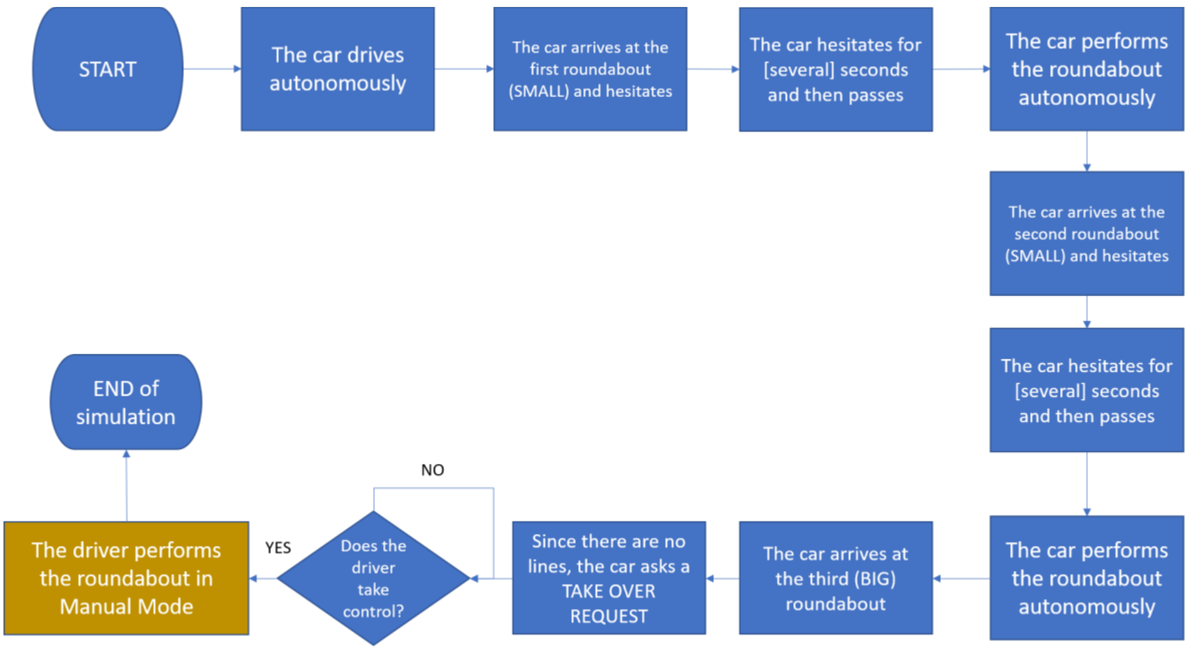

Figure 3: EVA driving scenario (baseline)

Also in the TeamMate scenario, the simulation starts in Automated Mode. When it arrives at the first two roundabouts (small), since there are high traffic flows, the TeamMate car is not confident in entering it and hesitates.

The hesitation and the subsequent "over-safe" manoeuvre can be frustrating for the driver. So, in this case, if the Driver Monitoring System detects that the driver is attentive when the car is stopped at the roundabout, triggers the suggestion, given through the TeamMate HMI, for the support in perception. If the driver is distracted the assumption is that no support is needed, and the car behaves like in baseline scenario (i.e. it takes more time to perform the roundabout). If the driver gives the support and confirms that there is enough room to enter the roundabout, the TeamMate car performs it autonomously.

When approaching the third roundabout (big) a support in action is needed: through the HMI the car asks the activation of the Shared Control, since there are no lines on the road. The car waits until the support in action is given: the expected feedback, asked through the HMI is to put the hand on the steering wheel and take the lateral control. When the car recognizes a deviation in the steering torque in the proximity of the roundabout (that is not given by the automation but is somehow "forced" by the driver) it recognizes that the support in action is accepted and the vehicle cedes the later control (while it maintains the longitudinal).

After the support in action, the simulation ends. Figure 4 shows a schematic representation of the TeamMate scenario.

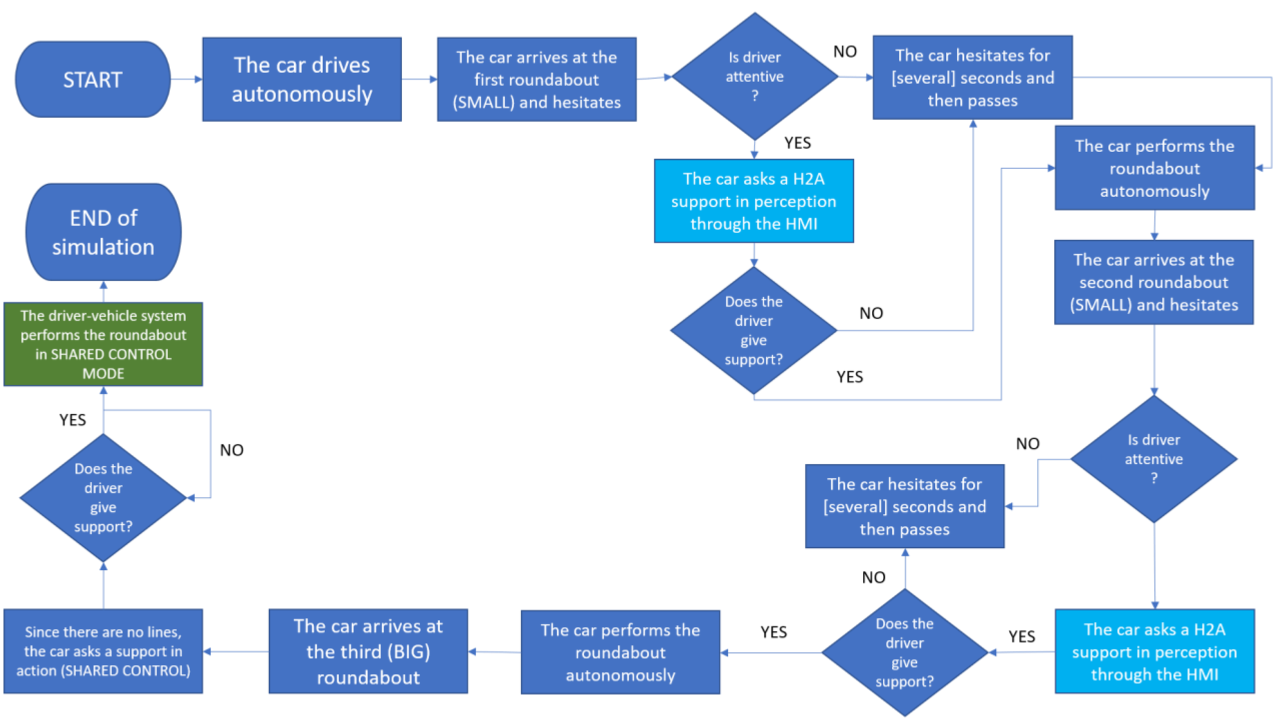

Figure 4: EVA driving scenario (TeamMate)

3.1.2.1 Main differences between the baseline and the TeamMate scenarios

As emerged from the previous paragraphs, the main differences between the baseline and the TeamMate scenarios are in the interaction modality at the critical event, i.e. the roundabout.

While in the baseline less interaction is needed, in the TeamMate car the support in perception allows the driver to give small specific supports to the driver. One scope of the EVA scenario is, in fact to, demonstrate that if the driver is maintained in the control loop (by giving him/her small tasks related to the driving) and information about the reasons of the car behaviour are communicated, the trust in the automated system can increase.

3.1.3 KPIs selected for EVA scenario

As stated in D6.1, the KPIs selected for this scenario are mainly linked to the efficiency, the comfort, the acceptance and the trust. In this phase, other KPIs, partially introduced in the same document, have been considered in the evaluation. For each KPI a threshold has been defined as success criteria: the threshold is represented as percentage increase of each indicator, measured through the comparison between the results collected in the baseline and the TeamMate configurations.

KPI category	KPI	KPI ID
Efficiency	Time to enter the roundabout	KPI1
Trust	Trust in automation	KPI2
Acceptance	Acceptance	KPI3
Driver's performance	Workload and frustration	KPI4

Willingness to buy	Willingness to buy	KPI5

Table 1: KPIs measured in EVA scenario

As stated before, the KPIs measured in EVA scenario are mostly related to comfort and acceptability issues: for this reason, consistently with the concept (see D4.4), the use case used to quantify AutoMate's added value concerned the H2A support (both in perception and in action).

The time to enter the roundabout has been considered an indicator of efficiency, both for the effect on the driver (as direct measure of satisfaction) and from a "traffic perspective" (i.e. the other cars in the traffic scenario that could be affected by the wait for the decision of the preceding driver-vehicle system).

The trust is widely considered as a crucial indicator for automated vehicles, being also a key factor for their adoption (Kaur & Rampersad, 2018). Strictly related to trust, also the user acceptance of the technology was measured. Both these indicators used self-reported subjective measures.

For this reason, the baseline is represented by a full autonomous car, in which the only possible interaction modality is represented by the Take Over Request, in case of disengagement.

In order to evaluate the cognitive workload and other characteristics related to the performance (i.e. the frustration), the NASA-TLX questionnaire was administered.

Another important KPI used to measure the TeamMate system against the baseline is the willingness to buy. To do that, two questions have been used to compare the system developed in AutoMate with the full autonomous driving (and also in comparison with full manual cars).

KPI ID	KPI	Recording tool	Type of data
1	Time to enter the roundabout	Simulator's logs	Objective data
2	Trust in automation	Questionnaire "*Trust in automation*" by Koerber (2018)	Self-reported
3	Acceptance	Technology Acceptance Model	
4	Workload and frustration	NASA-TLX	
5	Willingness to buy	Custom questionnaire	

Table 2: Recording tools to measure the KPIs

Finally, one of the research questions of the experimentation will be to measure the relation between the attention in relation to a vehicle inefficiency in Automated Mode and the frustration, measured trough the specific item of the NASA-TLX.

The hypothesis is that unnatural and uncomfortable car behaviour can affect the driver's level of frustration: the data about attention at roundabouts and frustration will be crossed in order to evaluate a possible correlation. This issue cannot be considered as a KPI, but it has been considered as an important feature to evaluate the effectiveness of the TeamMate approach. It will be measured by counting the number of users that will be attentive at the small roundabouts; moreover, the "frustration" collected as indicator of the NASA-TLX questionnaire, will be considered an indicator to measure this item.

3.1.4 Experimental Protocol

The test has been performed in REL facilities. The participants were welcomed and an explanation of the experiment was given, including the steps of the procedure and the scope of the experiment. The project's concept was introduced to the subjects in order to explain the meaning of the cooperation as intended in the project. Risks and constrains were explicitly explained to the participants who were informed that the experiment could be stopped at any moment.

The participants were asked to read and sign an informed consent form: since the DMS filmed frontally the participants, a note was added in the consent form to authorize the collection of this data. The recordings were then delated, and only the data related to distraction were collected and stored.

The participants were asked to have a 10 minutes dry-run, in order to become familiar with the driving simulator and the different types of supports that the car can ask or suggest to the driver. In this phase, the users were asked to drive both in manual and in automated driving.

A mobile phone, running a video, was located at the right of the driver, in order to be a source of potential distraction, simulating a real driving scenario in which the user, being in automated mode, can also have different tasks without having the hands on the steering wheel and the eyes on the road.

After the dry-run, the first scenario was administered, followed by the survey. Then, after the first part of the questionnaire, the users drove the second scenario and completed the questionnaire. The two scenarios were randomized in order to avoid biases.

3.1.5 Participants

20 participants have been recruited for the experiment. Only participants with valid driving license have been recruited for the test. The sample was balanced for gender (9 males and 11 females), in order to avoid biases.

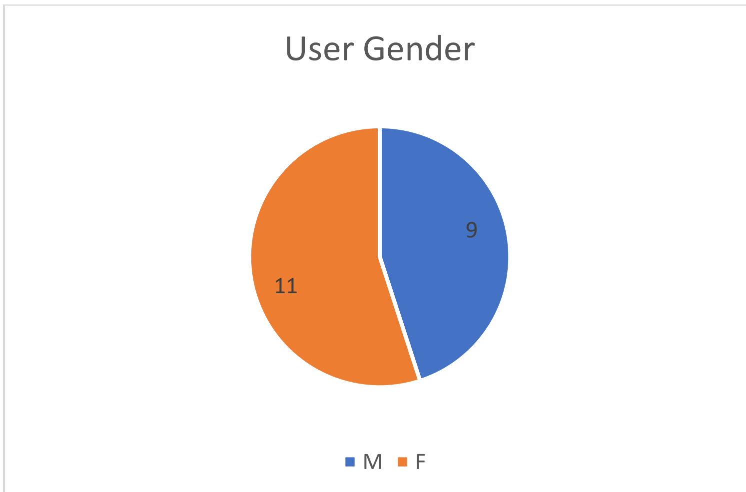

Figure 5: Gender of participants at REL evaluation

The average age was 28,35 years. The mean years of driving experience was 9,3 years; the participants travel about 17.000 km/year in average.

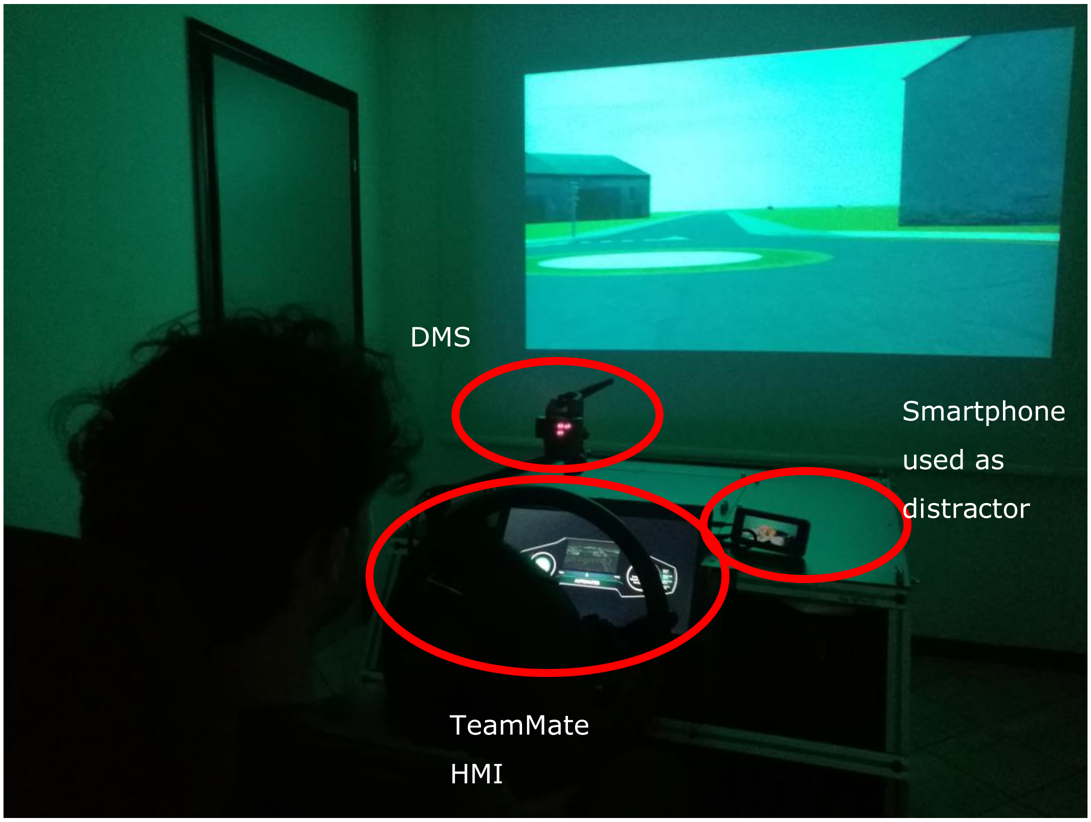

Figure 6: TeamMate scenario during the support in perception

A mobile phone was located on a support next to the instrument cluster. The TeamMate HMI was visualized on a 13.3" screen Full HD display, connected to the driving simulator.

Figure 6 shows the experimental setup of the evaluation in REL driving simulator including the Driver Monitoring System, TeamMate HMI, and a smartphone as possible distraction.

3.1.6 Results

In this chapter the results of the evaluation are described. Each item has been scored on a 7-point Likert scale and then normalized (max. -3 for negative responses, max. +3 for positive responses) to use the same measuring scales.

KPI ID	KPI	baseline results	TeamMate results
1	Time to enter the roundabout	27,12 sec	17,83 sec
2	Trust in automation	+0,25	+0,47
3	Acceptance	+0,41	+0,97
4	Workload and frustration	6,51	4,82
5	Willingness to buy	-0,1	+1,15

Table 3: KPI results in EVA scenario

Table 3 shows the results of the evaluation. The TeamMate configuration increases the performance of all the indicators against the baseline.

Figure 7 shows the comparison in the Time to enter the roundabout, collected from the simulator logs. The data have been collected from the moment in which the car approaches the roundabout autonomously (i.e. when it starts to slow down) to the moment in which, after the roundabout, the steering wheel resumes the 0° position. In the baseline scenario, this time is constant, and is 27,12 seconds. In the TeamMate scenario, this time depends from the moment in which the user gives the support in perception (if he/she is detected as attentive). The average time to perform the roundabout, on a sample of 40 roundabouts, has been measured in 17,83 seconds. The difference between the two configurations is 9,25 seconds.

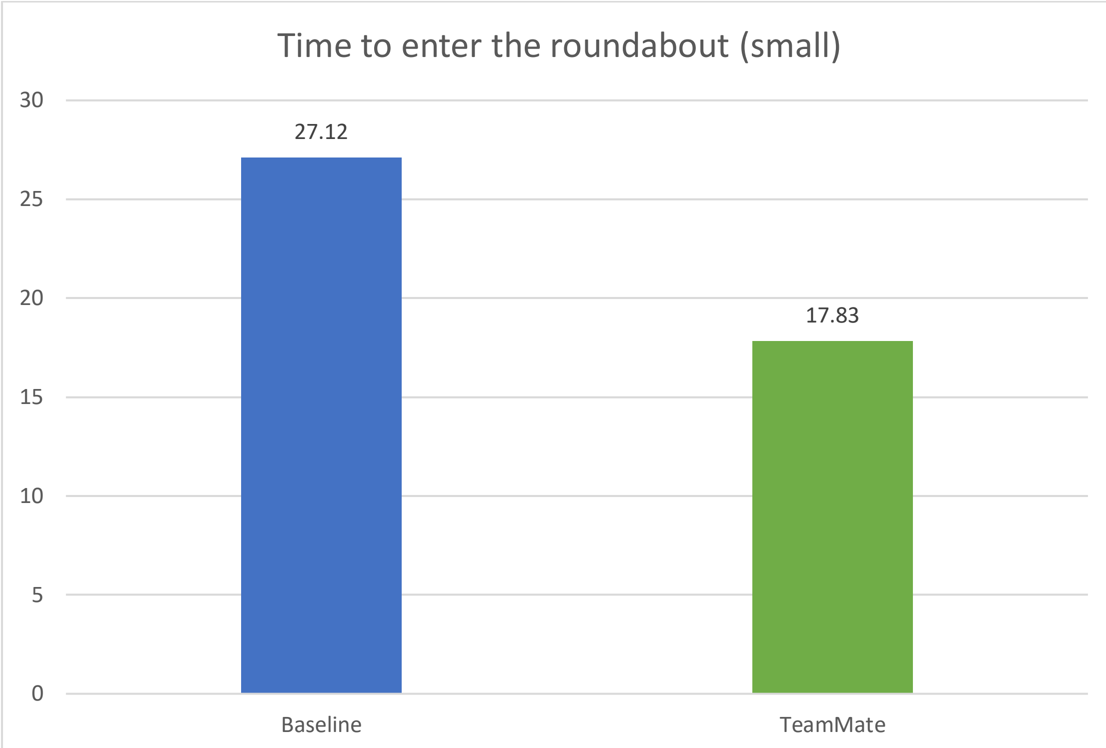

Figure 7: Time to enter the roundabout (small) in baseline and TeamMate

Figure 8[2] shows the compared results of trust between baseline and TeamMate scenarios. The baseline score is +0,25, while the TeamMate score is +0,47 (with a Δ of +0,25). In particular, the TeamMate has been considered reliable (+0,60 than the baseline), and the state of the system was considered clear (+0,55 than the baseline). However, the user found that the TeamMate system not very predictable, also in comparison with the baseline (TeamMate predictability = -0,25 than the baseline).

[2] Please note that the charts are represented on a resized scale (in this case between -1 and 1) even if the real scale is between -3 and 3 for legibility reasons.

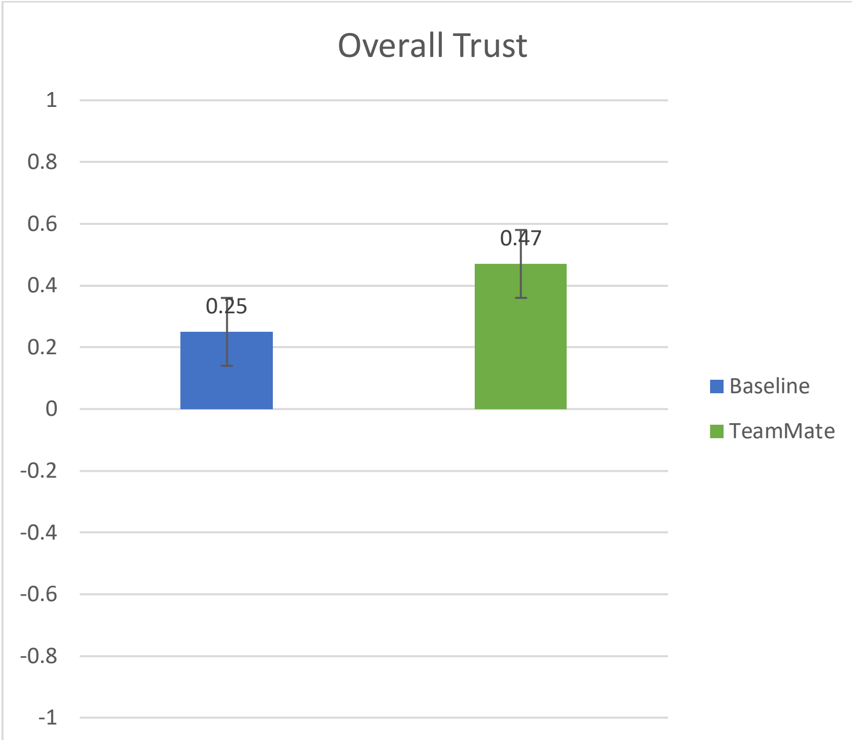

Figure 8: Comparison of Trust between baseline and TeamMate in EVA scenario

Another relevant KPI, i.e. the KPI3 ("Acceptance") showed encouraging results, since the overall acceptance of the TeamMate system was found to be higher than the baseline acceptance in the EVA scenario (TeamMate acceptance = +0,97; baseline acceptance = +0,41; Δ of the increase = +0,55).

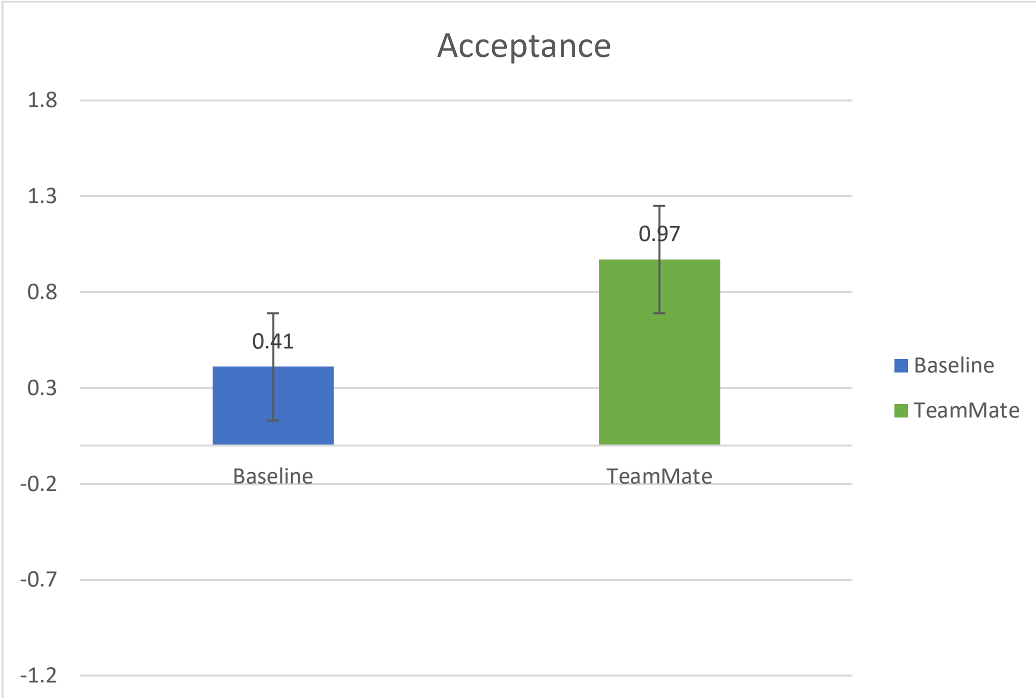

Figure 9: Comparison of Acceptance between baseline and TeamMate in EVA scenario

In particular, the users found that the TeamMate car can allow them to do what they want (+0,75 against the -0,15 of the baseline; Δ=+0,90). From qualitative data, collected through the users' comments, it was clearly expressed by the subjects that they appreciated the feature of being partially in control of the perception and decision, rather than completely rely to the vehicle. Moreover, the users appreciated the possibility ensured by the TeamMate system of increase the effectiveness in terms of time needed to complete the tasks, since they found the baseline, in full automation, too cautious and restrictive in the driving behaviour. However, from the comments collected, some users pointed out that they would prefer to drive manually in complex situations, since the automated car (both in baseline and in TeamMate configurations) would take longer then manual driving.

In general, the users pointed out that the TeamMate car could be able to increase the perceived performance (+1,00 of TeamMate against -0,15 of the baseline, Δ=+1,15).

Figure 10: Comparison of Perceived Performance (as part of Acceptance) between baseline and TeamMate in EVA scenario

The KPI4, i.e. the workload, measured through the NASA-TLX, showed an increase of the performance with TeamMate car. Figure 11 shows the compared results of the baseline and the TeamMate workload, normalized in order to put the activities in the same scale (between 0 and 20). The overall workload score for baseline system is 6,81, the score for TeamMate system in 4,82. The results, even if not particularly prominent, can be considered as an important tendency.

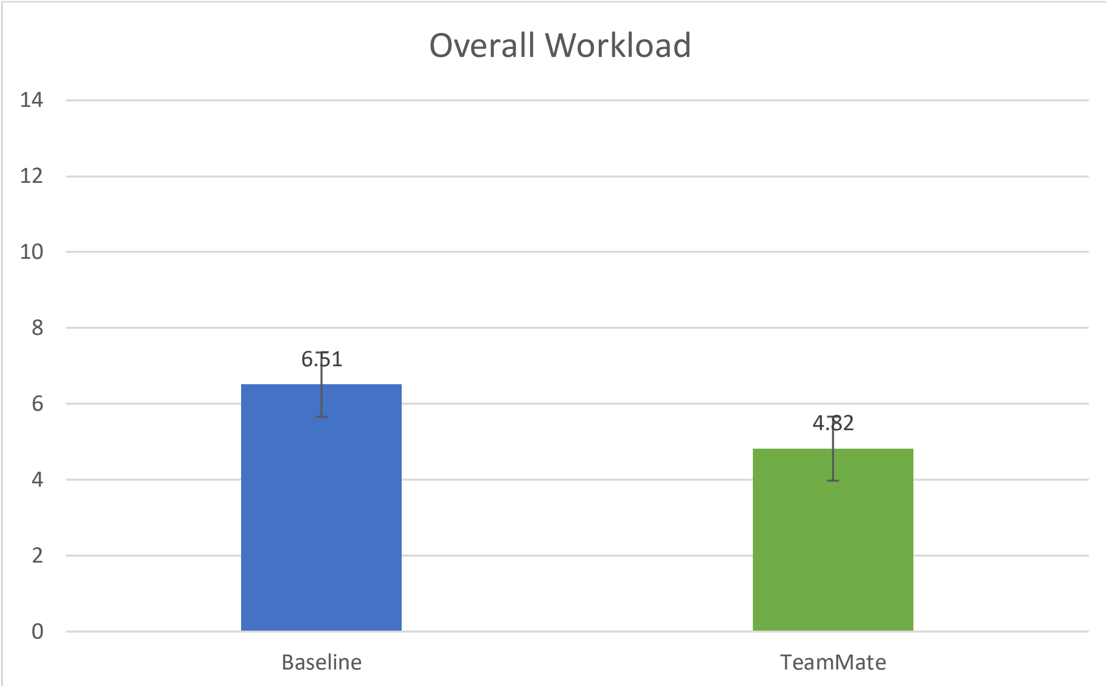

Figure 11: Comparison of Overall Workload between baseline and TeamMate in EVA scenario

By analysing the single items that make up the NASA-TLX questionnaire, the TeamMate system showed to be able to reduce the frustration of the driver; this can be considered as an interesting observation, since one of the hypotheses was that the TeamMate, and in particular the "Support in Perception", can be able to reduce the frustration. The score of baseline system was 8,4 on a 0-20 self-reported scale; the score of TeamMate system was 5,85 (on the same scale) as shown in Figure 12.

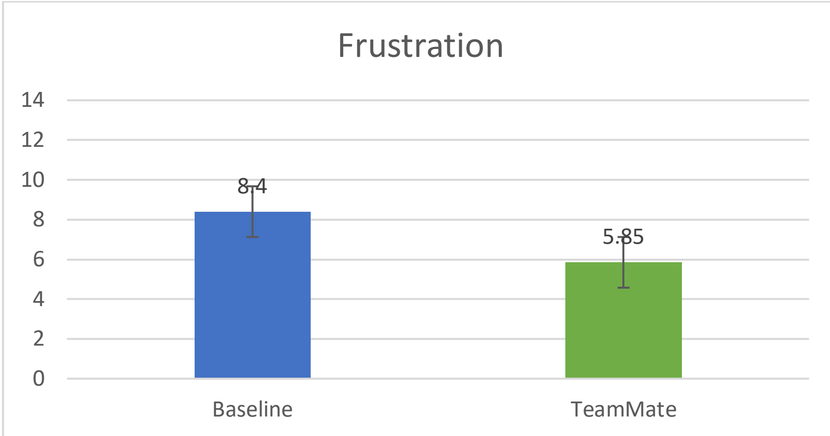

Figure 12: Comparison of Frustration (as part of Workload) between baseline and TeamMate in EVA scenario

Figure 13 shows the results of the KPI5 ("Willingness to buy"). The TeamMate car obtained a score of +1,15 against a score of -0,10 obtained by the baseline. This is particularly relevant due to the amplitude of the result and the relevance of this market-related issue in the adoption of highly automated vehicles.

Figure 13: Comparison of Willingness to buy between baseline and TeamMate in EVA scenario

These data are corroborated by the results of the item about the "Willingness to pay", i.e. the amount of money the users are willing to spend to have the automation features. This item was collected by considering the car behaviour and the features as an "upgrade" of an existing manual car (the exact question was: *"How much you'll be willing to spend more than a manual car for a vehicle that can behave also like the one you've used in the experiment?"*) The TeamMate car showed an important raise in the willingness to pay (i.e. 4872,50 € against 3547,50 of the baseline, +37,4%.)

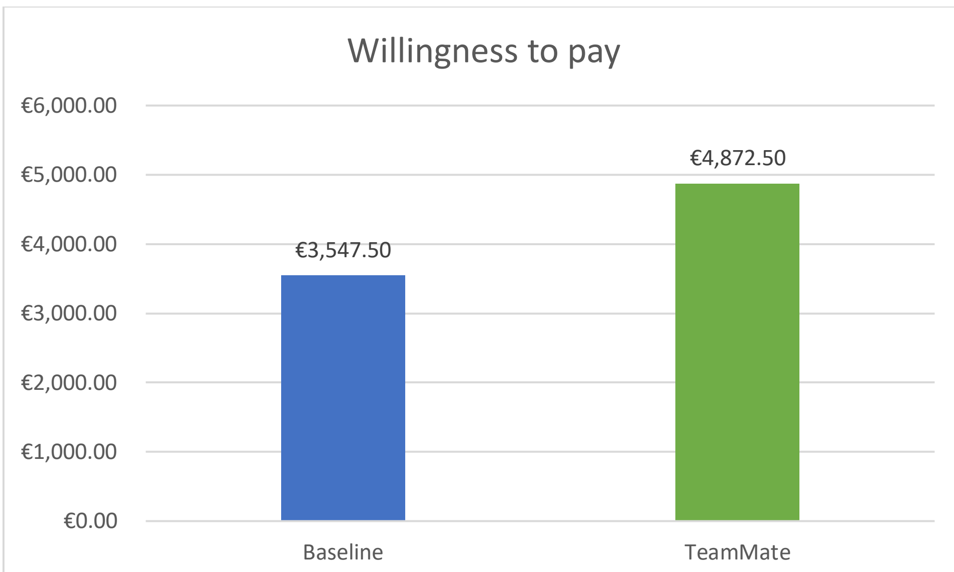

Figure 14: Comparison of Willingness to pay between baseline and TeamMate in EVA scenario

Finally, other indicators, not considered as KPIs were collected during the experiments. One accident occurred during the baseline scenario, right after the take-over request at the third (big) roundabout. In general, the Take Over was considered by users as a critical situation, and a source of potential danger. The users well accepted the shared control, as possible partial mitigation of transitions, and considered as good the approach of keeping the driver in the control loop with small tasks (such as the support in perception) in order to minimize the risks in transitions of control. Finally, the data about distracted drivers at the small roundabouts were collected; only three subjects out of 20 were distracted before a small roundabout and did not receive the support in perception. Moreover, this happened only at one roundabout for each user (in total, 3 out of 40 roundabouts).

3.1.7 Discussion and next steps

All in all, in EVA scenario, the TeamMate system showed significant improvements in the performance against the baseline.

The TeamMate system showed to significantly improve the acceptance and reduce the frustration. Also the trust slightly increased in comparison with the baseline. The temporal efficiency increased, and this is a relevant indicator both from vehicle and traffic perspectives. The time reduction, in fact, can be considered a relevant indicator that affects several parameters, from traffic fluidity, to the acceptance and all comfort-related factors. The EVA scenario was actually designed to measure mostly comfort-related parameters, and the evaluation shows that some encouraging considerations reinforce the approach followed in AutoMate.

Some safety-related parameters have been considered as having a not prominent part in this assessment phase: however, some indicators (e.g. the number of accidents and the comments collected during the experiment) show that also the safety-related parameters are satisfactory, even if only indicative.

Moreover, some interesting behaviours were observed during the experiments. In general, the users are more prone in looking at the road (and being attentive) when the car is approaching a turn.

In general, the possibility of deciding if give a support to increase the efficiency of the manoeuvre was well perceived by the users. Several observations collected during the experiments can be considered as inputs for the design in the 3rd cycle. In particular, some observations on how to improve the HMI will be used in WP4 in the last project's cycle.

Due to the comments received in the test phase, and by the fact that the possibility of integrating other enablers during the last cycle is under consideration, another evaluation test will be done after the third project's cycle, in parallel with the evaluation on real vehicles.

3.2 ULM simulator

3.2.1 Description of the Evaluation Scenario

H2A support in perception

Peter is driving in a narrow rural road in Automated Mode. Arriving behind a slowly driving car, the TeamMate car detects that it obstructs the view. Therefore, the vehicle is not confident of the oncoming traffic due to a limit in perception. Since the vehicle is not sure about the possibility to perform a safe overtaking manoeuvre (no confidence in perception), it would follow the slowly driving car (car-following mode) until the sensors provide enough information to safely overtake it or until the slowly driving car changes the lane. The TeamMate car asks Peter to check by himself, in order to support the automation. When Peter confirms there is enough space, the TeamMate car performs the overtaking manoeuvre in Automated Mode (even in Sharing Mode would be possible).

A2H support in action

Peter is driving in a narrow rural road in automated mode. There is a slowly driving car driving in front of the TeamMate car. The TeamMate car decides for Peter if he wants to overtake based on Peters normal behaviour. The TeamMate car supports Peter in understanding the intention of the automation by showing him the crucial information, like the trajectory for the following manoeuvre or, in case Peter would overtake in a situation in which is not possible, by informing him that it is too dangerous to overtake now.

3.2.2 Baseline

The baseline car is a state-of-the-art automated car that can overtake both, longitudinal and lateral control of the car. Such as the TeamMate car, the baseline car can drive through the rural road fully autonomously. If there is a slowly driving car in front, the baseline car would follow it without suggesting or be able to conduct an overtaking manoeuvre. Hence, in such a situation Peter has to shut off the automation and overtake the slowly driving car in manual mode.

3.2.3 Participants

In total 26 participants took part in the experiment (13 women, 13 men) with an average age of 27.35 years (SD = 11.31). All of them had a valid driving license for at least one year (Mean = 9.85; SD = 12.01), had normal or corrected to normal vision. Each driver received a 12 euros allowance.

3.2.4 Material

3.2.4.1 Simulator

The experiment was carried out in the static ULM driving simulator, equipped with three high definition projectors offering a 160° horizontal field of view. Rear view was displayed on three digital mirrors. A 12'' screen set right to the steering wheel was used as a dashboard. A 17'' touch screen, mounted in the central console, was used as the central control unit. The scenarios were created and simulated with the SILAB simulation software.

3.2.4.2 Automated system and HMI

The automated mode could be activated only while on ones lane by pressing a big button ("Automation einschalten" - "turn automation on", see Figure 15). If the driver pressed the brake pedal or turned the steering wheel to a certain degree (10 degrees more than it currently should be turned by the automation), the automation was turned off. This method is used by current driving assistance systems such as ACC.

Figure 15: Touch screen with automation and overtake button

When the driver turned on the automation, the button switched the colour to a lighter blue and the label switched to "Automation ausschalten" – "Turn automation off”. By pressing this button again, the automation was turned off and the driver had to drive in manual mode again.

In TeamMate mode the automation was equipped with an Augmented Reality (AR) system provided by OFF. While planning a manoeuvre, the TeamMate car showed different AR displays to the driver based on the situation.

Figure 16: Baseline condition with no AR system

Figure 17: TeamMate car follows a slowly driving car because driver intention recognition predicted the driver wants to follow

Figure 18: TeamMate car overtakes because driver intention recognition and risk assessment both indicate to overtake

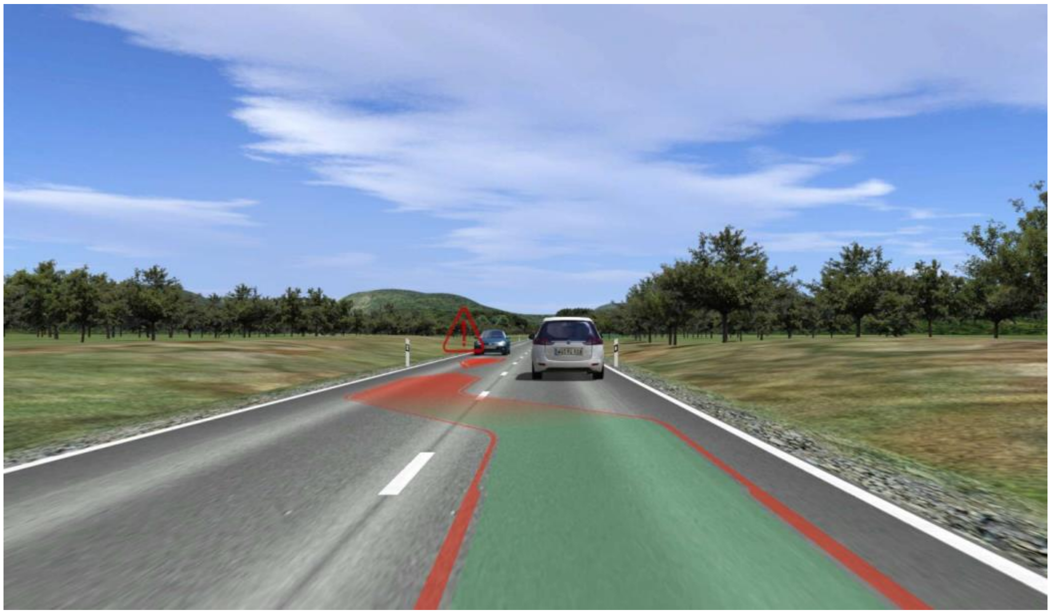

Figure 19: TeamMate car predicts that driver wants to overtake but there is oncoming traffic

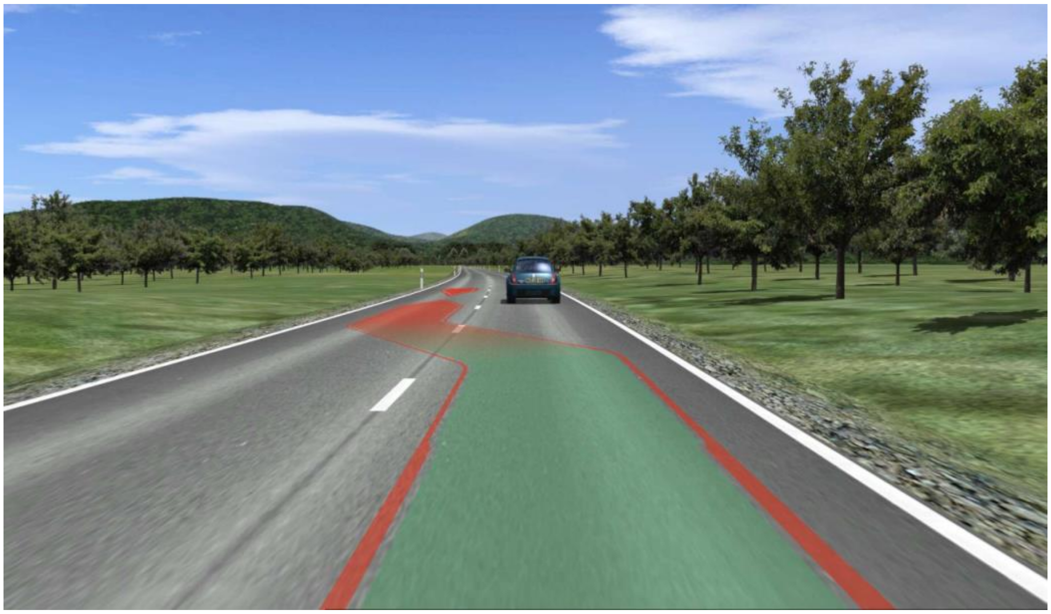

Figure 20: TeamMate car cannot overtake on its own and needs driver's help

In the H2A scenario, the driver had the possibility to overtake the other car by either turning the steering wheel slightly to the left or pressing a button (Fig.15) on the control unit.

The driving track was a one lane rural road with an opposing lane. There was oncoming traffic throughout the track and the road structure (curves and sight) changed in each overtaking scenario. The speed limit was 100 km/h and there were slowly driving cars with a speed of 60 km/h on Peter's lane.

3.2.4.3 Scenarios

A training scenario was designed to familiarize the participants with the driving simulator and the automated system. It was a 4.5-km long rural road with curves, with no other traffic or crossing lanes.

There were two experimental scenarios, one for the A2H support and one for the H2A support.

The A2H scenario was a 20 km long rural road with a speed limit of 100 km/h. The car drove through the rural road in autonomous mode and there were six cars with a speed of 60 km/h on the ego cars lane. Depending on the driver intention recognition and the risk assessment, in the TeamMate scenarios, the slower driving car was either overtook or followed. The baseline car always followed the car. After around one kilometre, the slower driving car changed the lane on an intersection. After each intersection, a one kilometre long straight road with no traffic followed, where the experimenter had the chance to ask the participant questions about the overtaking manoeuvre.

In the H2A scenario, the rural road was 14 km long and had a speed limit of 100 km/h. Four slower driving cars were blocking the sensor view, hence the overtaking manoeuvre had to be initiated by the driver. In the baseline scenario, the driver also had to overtake manually before switching on the automation again. There were also intersections where the slower driving cars changed their direction and a straight road to question the participant about the manoeuvre. The overtaking situations had different environmental conditions: straight road with no oncoming traffic, foggy road with no oncoming traffic, straight road with oncoming traffic and a curvy road with no oncoming traffic.

3.2.4.4 Questionnaires

Trust, acceptance and criticality questionnaire

Trust was measured after each overtaking manoeuvre. The participants were asked 'how much they trusted the system' on a scale from 0 to 10 (Brown & Galster, 2004). The participants were also asked if they agreed with the systems behaviour. The participants were asked to rate the system behaviour from 0 (don´t accept it at all) to 10 (perfect behaviour). As a third question, the participants were asked how critical the situation was in their opinion using the situation criticality scale (Neukum et al., 2008).

NASA-TLX

Driver workload after each scenario was measured with a German translation of the NASA-TLX (Hart & Staveland, 1988). This questionnaire is composed of six Likert scales ranging from 0 to 100. Each scale aims at evaluating a dimension of workload: mental demand, physical demand, temporal demand, performance, effort and frustration linked to the completion of a specific task.

System Usability Scale

Systems usability was assessed using a German translation of Brooke's (1996) System Usability Scale (SUS) composed of 10 items. For each item, participants had to evaluate their level of agreement on a Likert scale ranging from 1 to 5. According to the calculation procedure provided in Brooke (1996), this questionnaire provides a usability score ranging from 1 to 100. Higher score means higher rated usability.

Experimental design

The experimental design is a within subject design. There are two factors:

- System with two possible versions - baseline or TeamMate
- Interaction modality with two different modalities – turning the steering wheel or pressing a button. This factor was only tested in the cooperative use cases where the driver had to initiate the overtaking manoeuvre

3.2.5 Procedure

The participants were welcomed, given the informed consent and the data protection agreement. Then the eye tracker was calibrated individually. Participants who had never been in the driving simulator received a short introduction to the functions and control of the driving simulator. Each participant then got an introduction to the function of the automation and how to turn it on and off. Afterwards they had to drive through an about 5 – minute long training course, where the participants were instructed to turn the automation on and off with the different modalities.

Then the aim of the experiment and the two different car concepts were explained to the participants. The order in which the participants drove with the car concept or the scenario (A2H and H2A) was counterbalanced on all participants.

Each scenario started with a 2 km long rural road before entering the experimental conditions, so that the participants can turn on the automation and adapt to the environment.

The functions and limitations of the baseline car were explained to the participants before both baseline scenarios (A2H and H2A) together with the length of the course in minutes. Participants were introduced that they could turn off the automation and overtake manually if they wanted to.

The autonomous scenario (A2H) started with an explanation of the TeamMate car. It was explained that the car will overtake autonomously when the car thinks that the driver would want to overtake and that it is safe to overtake. They were introduced, that they can turn off the automation at any time but should use the automation if they are in accordance to the behaviour of the system.

The cooperative scenario (H2A) was driven through two times. Once with each interaction modality (steering wheel and touch button). The interaction modality was explained to the participants and they were introduced, that if they want to overtake and think it is safe they should initiate the manoeuvre using the scenarios interaction modality. It was also explained that they can turn off the automation if they feel unsecure by the behaviour of the TeamMate car.

After each overtake (autonomously or manually), the participants were asked if they trust the system, how critical the situation was and if they were in accordance to the behaviour of the system. These questions were asked right after switching back to the initial lane and, in case of a manual overtake, when the automation was turned on again. The questions were asked using a microphone while the simulation was still running.

After each scenario, the participants had to fill out the Nasa-TLX and the SUS on a Microsoft surface tablet, sitting in the driving simulator to keep the immersion.

After driving through all scenarios, the participants had to fill out the demographic questionnaire. At the end, the participants were interviewed shortly about the TeamMate car and paid.

3.2.6 Results and Discussion

3.2.6.1 Results A2H

There was a significant difference in the rating of the behaviour between the TeamMate car (M=8.19, SD=1.20) and the baseline car (M=7.68, SD=1.12) conditions; $t(25) = 2.28$, $p = 0.032$.

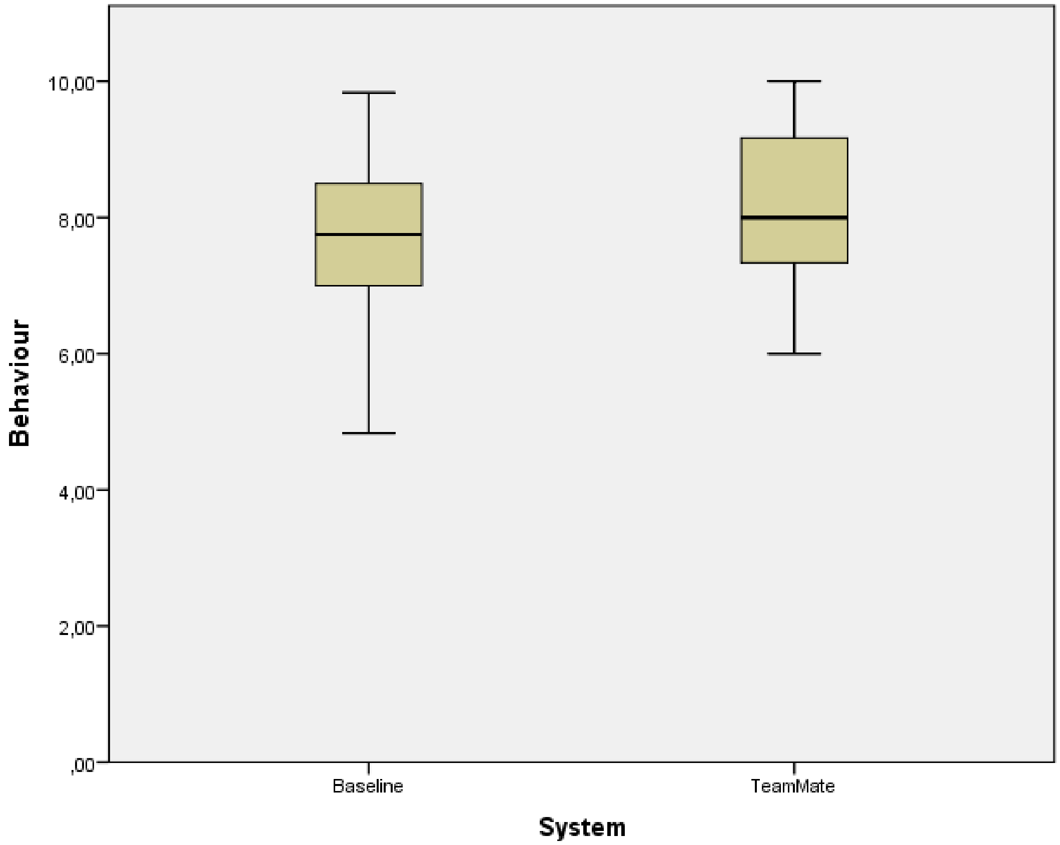

Figure 21: System behaviour rating of the baseline and the TeamMate car (A2H)

The criticality was not significantly different between the TeamMate (M=2.00, SD=1.14) and the baseline (M=2.09, SD=1.21) car; t (25) = -0.42, p = 0.68. Neither was the trust significantly different between the TeamMate (M=74.33, SD=21.16) and the baseline (M=76.37, SD=17.68) car; t (25) = -0.71, p = 0.48. Although the overall workload was lower after using the TeamMate car (M=27.01, SD=9.90) over the baseline car (M=31.55, SD=14.53), it was not significantly different; t (25) = 1.64, p=0.113. The usability was also not significantly different between the TeamMate (M=76.635, SD=11.725) and the baseline (M=75.577, SD=13.366) car; t (25) = -0.348, p = 0.730.

There were no second thoughts from the participants while driving in the TeamMate mode.

The participants rated the system behaviour as very acceptable (M= 7.67, SD = 2.083).

3.2.6.2 Results H2A

Each participant overtook the slower driving car in each condition (different interaction modalities and baseline).

The workload in the H2A scenarios was analysed using a repeated measurement ANOVA. Mauchly´s test indicated that the assumption of sphericity had been violated ($X^2(2) = 10.113$, $p = .006$), therefore the degrees of freedom were estimated using the Greenhouse-Geisser correction ($\varepsilon = 0.98$). There was a significant effect $F(1.488) = 4.962$, $p = 0.02$, where the workload of the baseline (M = 32.820, SD = 13.888) was significantly higher than of the touch interaction (M = 27.596, SD = 9.990).

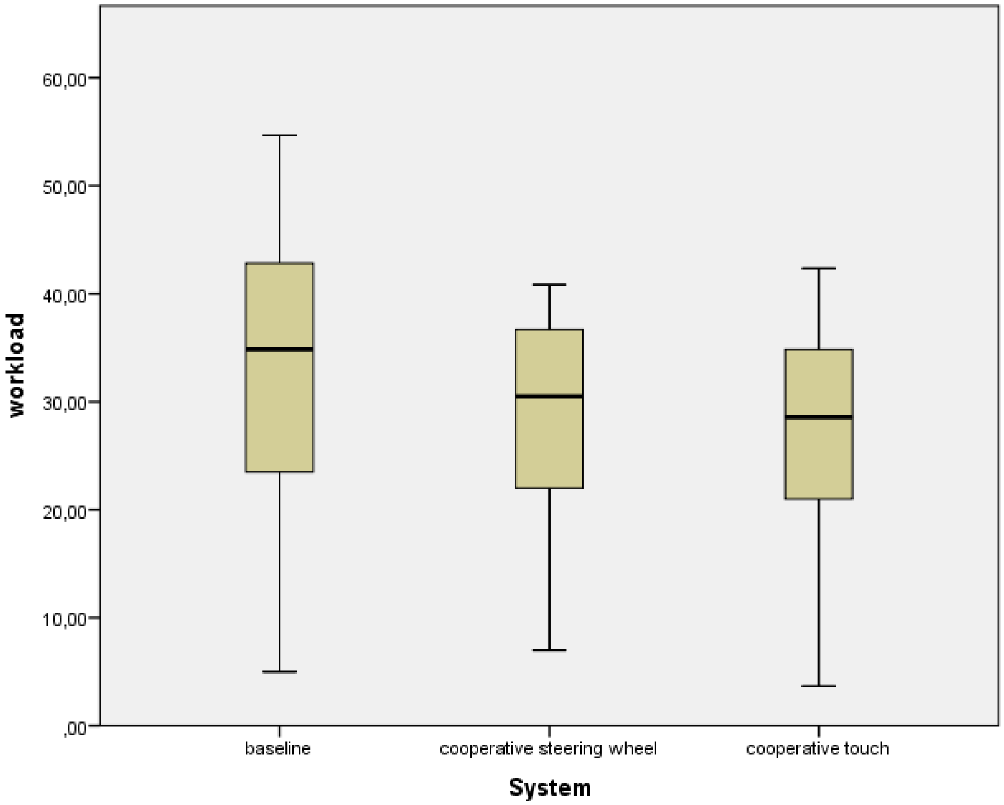

Figure 22: Workload with different interaction modalities and baseline

Pairwise Comparisons

(I) Faktor1	(J) Faktor1	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Lower Bound	95% Confidence Upper Bound
baseline	Steering wheel	4,417	2,172	,158	-1,156	9,990
	Touch	5,224*	1,832	,026	,525	9,924
Steering wheel	baseline	-4,417	2,172	,158	-9,990	1,156
	Touch	,808	1,222	1,000	-2,327	3,942
Touch	baseline	-5,224*	1,832	,026	-9,924	-,525
	Steering wheel	-,808	1,222	1,000	-3,942	2,327

Table 4: Pairwise comparison of the workload

Mauchly´s test indicated that the assumption of sphericity is given for the Usability ($X^2(2) = 1.365$, $p = .505$). There was a significant effect $F(2) = 4.705$, $p = 0.02$, where the usability of the baseline (M = 32.820, SD =

13.888) was significantly higher than of the touch interaction (M = 27.596, SD = 9.990).

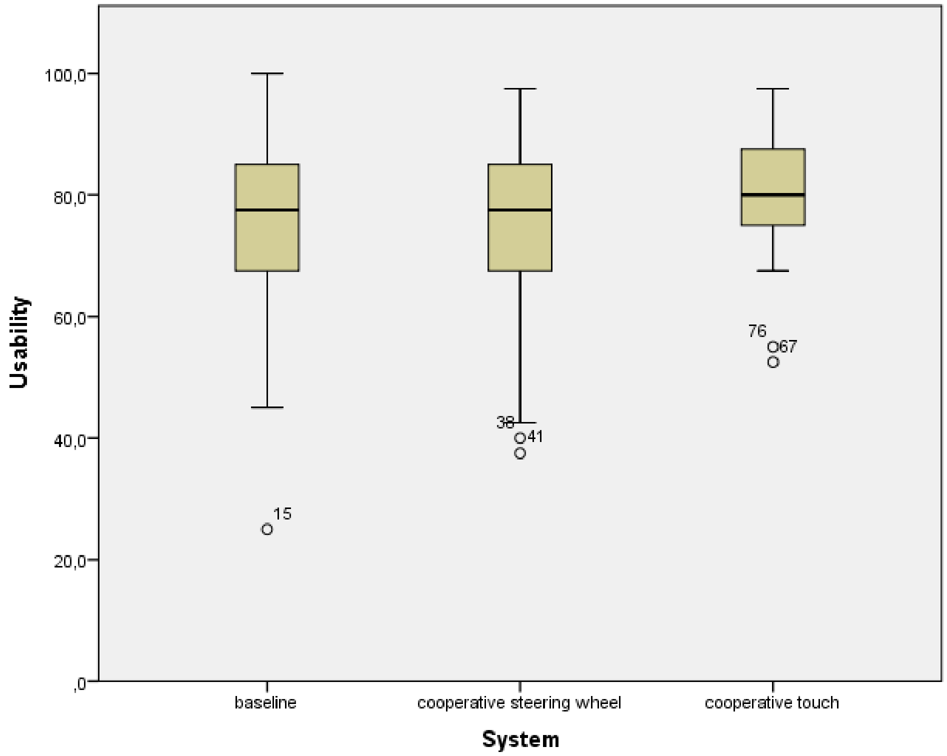

Figure 23: Usability comparison between different systems

Pairwise Comparisons

(I) Mode	(J) Mode	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference Lower Bound	95% Confidence Interval for Difference Upper Bound
baseline	Steering wheel	-,577	2,903	,844	-6,556	5,402
	Touch	-5,481*	2,527	,040	-10,685	-,277
Steering wheel	baseline	,577	2,903	,844	-5,402	6,556
	Touch	-4,904	2,400	,052	-9,846	,039
Touch	baseline	5,481*	2,527	,040	,277	10,685
	Steering wheel	4,904	2,400	,052	-,039	9,846

Table 5: Pairwise comparison of the usability

Comments throughout the experiment

Most of the participants found that the TeamMate system was too conservative. They wanted the system to overtake earlier and were okay with the system overtaking inside a curve. The parameters of the automation like the safety distance of the ACC or the trajectory was criticised by some participants.

Debriefing

Some participants said that they preferred the baseline system because the behaviour of the TeamMate system did not quite fit their own behaviours. They said that if they have to do any driving related task, they prefer driving manually for a while because they can drive as they want to. The participants defined the driving related task as for example checking the situation and initiating the manoeuvre. Some people preferred the steering wheel as an interaction modality because it was more intuitive and they could better control when to start the manoeuvre. In addition, some of the participants criticized the lack of a haptic feedback while interacting with the steering wheel.

3.2.6.3 Discussion

The *trust in automation* (KPI1) was slightly lower with the TeamMate system. Participants knew that the baseline system will always follow the slower car and could start the manoeuvre whenever they liked and perform the manoeuvre in their preferred way. The *usability* (KPI6) of the TeamMate system was rated higher but not significantly. The participants accepted

(KPI2 Acceptance) the TeamMate system while the *workload* (KPI3) was significantly lower compared to the baseline. There were no *accidents* (KPI4) or *second thoughts* (KPI5) with the TeamMate system.

Overall the results showed that users accept the TeamMate car. A benefit can be seen in the descriptive data, although it is not significant for the workload and Usability. Several open points for improvement were found, such as the too conservative driver intention recognition or the missing feedback of the steering wheel interaction. The workload is lower in the TeamMate condition but not significantly. The reason for that could be that the TeamMate car should be equipped with the TeamMate cluster HMI and the individual components (like the interaction modality) should be improved within the work packages two and four. The behavior of the automation by means of the driving behavior was not done by a trajectory planer, but was set manually by the developers. Thus, these parameters should be refined in the next cycle to improve the driving experience with the TeamMate car. It is planned that the in work package three developed trajectory planner will be included in the simulator to address this problem. There is a fundamental difference in the driving experience while using a driving simulator compared to a real car. None of the participants wanted to intervene the automation of the TeamMate car and highly accepted it. That suggests that they agree with the TeamMate concept overall but there is room for improvement in the next cycle. The usability of the system was not significantly higher than the baseline, although it was rated over 70, which means the system is usable (Bangor, Kortum, & Miller, 2009).

It was expected that on the curvy road no user would overtake because the driver intention recognition algorithms predicted that the user did not wanted to. Other than expected, the participants overtook even in the curvy scenarios. This means that the driver intention recognition should not be too conservative when it comes to curves in the road structure. The analysis of the workload and the usability suggest that the Touch lowered the workload of the participants and improved the usability. The steering wheel interaction improved the values but not significantly. The comments of the participants like the missing feedback can be improved in the next cycle.

3.3 VED simulator

H2A and A2H use cases have been implemented on VEDECOM's diving simulator and tested with naive drivers on two distinct experiments. Method and results for each experiment are presented and discussed in two distinct sections.

3.3.1 H2A support

3.3.1.1 Description of the Evaluation Scenario

The TeamMate car is driving in an extra-urban road in Automated Mode. Through the V2I communication, it is informed that there is a roadwork zone in 1 kilometre and that the lanes might be no longer visible. Since the TeamMate car knows that it will not be able to deal with this situation autonomously, it shares the information with Martha and asks het to handle the control of the vehicle during the roadwork zone. Martha takes over the control until the end of the roadwork, and can choose to shift back to Automated Mode afterward.

For this scenario, the evaluation is aimed at demonstrating the added value of the driver. Thus the baseline is the driverless approach without V2I communication when manual takeover is issued when roadwork zone is detected later by car sensors.

3.3.1.2 Participants

19 participants took part in the experiment. Among them, four were excluded from data analysis due to simulator scenarios dysfunctions. The remaining 15 drivers (8 women, 7 men) were in average 37 years old (SD =

9.75), held a valid driving license for at least 5 years (Mean = 16.78; SD = 9.67), had normal or corrected to normal vision, and drove on average 19368.42 km per year (SD = 13969.05). Drivers who already drove VEDECOM's simulator were excluded from the experiment. None of them complained about motion sickness. Each driver received a 50 euros allowance.

3.3.1.3 Material

3.3.1.3.1 *Simulator*

The experiment was carried out in VEDECOM static driving simulator, equipped with four 32'' 16/9 screens offering a 120° horizontal field of view. Rear view was displayed on three digital mirrors. A 10'' screen set right to the steering wheel was used as a dashboard. A 10'' Microsoft Surface tablet, mounted in the central console, was used to perform non driving related task during automated mode activation. Scenarios were generated by SCANER® studio software 1.6 developed by Oktal.

Figure 15: The dashboard in manual mode.

3.3.1.3.2 Automated system

The automated mode could be activated only if the vehicle was travelling on the right lane and below the speed limit. When those conditions were not met, the vehicle was in manual mode. As depicted in figure 15, in manual mode the driving mode, speed, tour by minute and destination were displayed on the dashboard.

Once the conditions of activation were met, a specific short acoustic signal was issued, and the message "automated mode available" ("mode autonome disponible") was displayed on the center on the dashboard (cf. figure 16).

Figure 16: The dashboard when automated mode is available.

Automated mode activation was possible by pressing a dedicated button positioned on the right of the steering wheel. A different short tone was issued to validate the automated mode activation. The automated system maintained the vehicle in the center of the lane and set speed according to speed limit. The system kept a minimum time interval with followed vehicles of 1.8 seconds, and did not performed lane change autonomously. The system could be deactivated at any moment by pressing the brake or the accelerator. In automated mode, the system status on the dashboard changed from "manual" ("manuel") to "automated" ("autonome") (cf. Figure 17).

Figure 17: The dashboard in automated mode.

In case of a takeover request, a different acoustic signal was repeated as long as the system was activated, and a specific pictogram was displayed on the dashboard as depicted in figure 18.

Figure 18 : The dashboard when a manual takeover is requested.

In case the driver could not takeover manual driving on time, a safe manoeuver was performed in order to stop the vehicle. Once the safe manoeuver initiated, the message "safe maneuver" ("mise en sécurité") was displayed in the center of the dashboard, as depicted in figure 19.

Figure 19: The dashboard in case of a safe manoeuver.

3.3.1.3.3 Scenarios

Driving track consisted on a highway with a speed limit set at 130 km/h, composed of straight lines and slight curvatures. Most of the track is composed of 2 lanes in each directions, with a minor part of the track composed of 3 lanes in each direction. A roadwork zone was situated on a road portion with 2 lanes in each direction. The roadwork zone covered entirely the right lane on a distance of 67 meters. A small truck was positioned on the right lane to symbolize the beginning of the roadwork zone (cf. figure 20).

Figure 20: Screenshot of the roadwork zone causing the takeover request.

A Training scenario was created. This scenario aimed at getting participants used to drive the fixed based simulator in manual mode, and to learn how to activate and deactivate the automated system; thus experiencing a takeover request. This training scenario began with 10 kilometers of manual driving, followed by an activation of the automated mode for 2km after a takeover request was issued by the car. After manual takeover, participants were instructed by the experimenter to take the next exit, then scenario ended.

There was two experimental scenario: one with the baseline system, the other one with the TeamMate system. The two scenarios began with 1 km of manual driving, then automated mode was activated. A takeover request was issued after about 8 minutes of automated system activation. The takeover request was caused by a roadwork zone blocking the right lane. A specific pictogram was displayed in the center of the dashboard in order to symbolize the roadwork zone (cf. figure 21).

Figure 21: Pictogram symbolizing the roadwork zone displayed in the centre of the dashboard during the takeover request.

In the scenario with the baseline system, the takeover request was issued 200 meters before the roadwork zone. This distance was chosen in order to simulate the capacities of obstacle detection by car sensor. In the scenario with the TeamMate system, the takeover request was issued 1 km before the roadwork zone, in order to simulate the capacity of the V2I, allowing an anticipated takeover request.

3.3.1.3.4 Questionnaires

Trust questionnaire

Trust in the baseline system and in the TeamMate system was assessed with a French translation of a questionnaire composed of 19 items described in Körber (2018). Level of agreement with each item is assessed on a scale ranging from 1 to 5. Six dimensions of trust are assessed with this questionnaire: reliability of the system, predictability of the system, familiarity with the system, intentions of developers, propensity to trust and trust in automation (toward the system).

Acceptance questionnaire

A French translation of the questionnaire described by Van Der Laan, Heino and De Waard (1997) was used to compare baseline to TeamMate acceptance. This questionnaire aims at assessing two dimensions of automotive technologies acceptance: perceived usefulness and satisfaction, with nine items. Each item is composed of one scale ranging from -2 to 2 with a pair of opposed adjectives (ie. "*useful*" versus "*useless*", or "*assisting*" versus "*worthless*"). According to the calculation procedure described in Van Der Laan, Heino and De Waard (1997), this questionnaire provides a perceived usefulness score and a satisfaction score ranging from -2 to 2. Higher score indicates higher perceived usefulness and higher satisfaction.

NASA-TLX

Driver workload during manual takeover was assessed using a French translation of the NASA-TLX (Hart & Staveland, 1988). This questionnaire is composed of six scales ranging from 0 to 100. Each scale aims at evaluating a dimension of workload: mental demand, physical demand, temporal demand, performance, effort and frustration linked to the completion of a specific task.

System Usability Scale

Systems usability was assessed using a French translation of Brooke's (1996) System Usability Scale (SUS) composed of 10 items. For each item, participant has to evaluate his level of agreement on a scale ranging from 1 to 5. According to the calculation procedure provided in Brooke (1996), this questionnaire provides a usability score ranging from 1 to 100. Higher score means higher rated usability.

Willingness to buy

Participants' willingness to buy a vehicle equipped with the baseline system and the TeamMate system was assessed by mean of a scale ranging from 1 to 5. Participants were asked if they would buy the vehicle equipped with the TeamMate system and with the baseline system. Responses were collected by mean of two scales ranging from 1 to 5, with 1 corresponding to a lower willingness to buy and 5 corresponding to a higher willingness to buy. Additionally, a scale ranging from 0 € to 50 000€ was used to evaluate how much money participants were willing to spend to purchase the system in addition to the price of the vehicle.

3.3.1.4 Procedure

The participants were welcomed, given the financial allowance and explained the aim of the experiment and the general steps of the procedure. Risks and constrains were explicitly explained to the participants who were informed that they could stop the experiment at any moment. Participants were asked to read and sign an informed consent, and to answer a questionnaire to collect sociodemographic data.

Then, baseline car and TeamMate car description were given to participant along with a description of the different HMI states. Drivers were told that they would drive two automated car, which could be driven in two modes: manual and automated. The two cars were described as having the same functioning and the same dashboard, but that one of the two was equipped with V2I communication system. This feature was described as allowing the vehicle to receive and to send messages from the infrastructure which could be useful to control the vehicle and could allow to be informed in advance of events on the road

It was said that in manual mode, the driver was responsible for the entire vehicle control. Automated mode was described as working thanks to sensors that allowed the vehicle to perceive the environment, and that road markings and GPS data were used in order to locate and to move in the environment. It was said that it was possible to activate automated mode under the following conditions: to be on a dual carriageways, to comply with speed limit, and to be riding in the right lane. Participants were told that they would be informed if activations conditions were validated by mean of an acoustic signal and a message on the dashboard. Then participants were shown the automated mode activation button, and explained that once activated, another acoustic signal would be released, and that they would have to release pedals and steering wheel.

Participants were explained that in automated mode, the vehicle would maintain itself on the center of his lane and would adapt speed according to speed limit and to others vehicles. Experimenter said that participants did not have to monitor the road environment or the system functioning during automated mode activation, and that they were free to engage in any task on the tablet.

Participants were informed that the automated system wasn't able to deal with all the driving situations and that if, for example, the infrastructure was damaged or, if the driving situation was too complex, then the system would issue a takeover request to give back vehicle control to the driver. Participants were told that if the system couldn't be deactivated on time, it would complete a safe manoeuver.

The participant carried on with the training scenario. In manual mode, participants were instructed to complete lane changes and to use the brake pedal as many time as they needed to get used to simulator's command. After 10 kilometers of manual driving automated mode became available and participants were instructed to activate it. During automated mode activation experimenter commented on the system functioning and instructed the participant to deactivate the automated system with each pedal. Driver could test the activation/deactivation process as many time as they needed.

Afterwards, the two driving scenarios were completed in a counterbalanced order. Before each scenario, participants were instructed to imagine that they were going to be on a daily commute, to comply with speed limit, and to drive in the right lane as often as they could. Participant were told which system they would use and were instructed to activate the automated system as soon as it would be available, and to engage use the tablet if they wanted to.

After each scenario, participants responded to questionnaires to evaluate trust, acceptance, mental workload and usability. At the end of the last scenario, participants answered to the questionnaire evaluating their

willingness to buy both system. The experiment ended with a semi-structured interview. The whole protocol lasted about 2 hours.
The experimental design is a within subject design, with the factor system with two modalities: baseline or TeamMate.

3.3.1.5 Results

6 KPIs were analysed:

- Minimum time to collision. This indicator, expressed in seconds, is calculated from the takeover request to the lane change. The threshold selected for this KPI is 3.5 seconds, as defined in SAE (2013) as a reference value.
- Mental workload. The six dimension of the NASA-TLX were analysed separately.
- Trust. The five dimensions of the trust questionnaire were analyzed separately.
- Acceptance. The two dimensions evaluated with this questionnaire were analyzed separately.
- Usability.
- Willingness to buy.

The KPI were analysed by mean of an ANCOVA with system (baseline or TeamMate) as an independent variable and scenario order as a covariate (1 or 2). In addition to that, the number of collision between the truck at the beginning of the roadworks zone and the ego vehicle was also recorded. The proportion of manual takeover with a collision was compared between the two conditions with a Fisher test. The level of statistical significance is set at .05.

The only **collision** that occurred after manual takeover was with the baseline system. However, the proportion of manual takeover with a collision is not significantly different between the two condition (Fisher: p = 1).

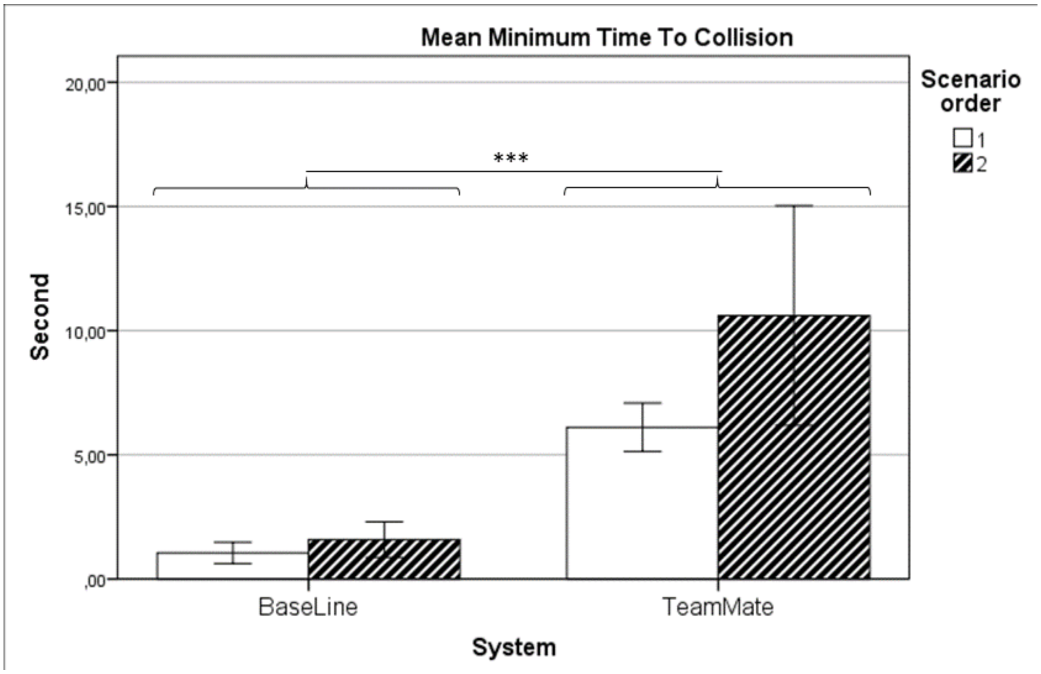

Figure 22: Mean minimum time to collision according to System (BeseLine or TeamMate) and scenario order (1 or 2). Errors bars represent standard errors.

After controlling for the effect of scenario order, a significant effect of system on **minimum time to collision** is observed (F (1, 29) = 32.38, p < .001, η_p^2 = .55). Mean minimum time to collision is higher in TeamMate condition (MEAN = 8.51; SD = 5.06) compared to baseline condition (MEAN = 1.34; SD = 0.85) (cf. Figure 22). In baseline condition, every participant had a minimum time to collision inferior to the threshold set at 3.5 seconds. In the TeamMate condition, 1 participant had a minimum time to collision inferior to 3.5 seconds. The proportion of manual takeover with a minimum time to collision inferior to 3.5 seconds with the baseline system is significantly higher than with the TeamMate system (χ^2 (1) = 22.63, p < .001).

Our results point out that manual takeover with baseline system endangers the driver by inducing a risk of collision and an unsafe time to collision. The TeamMate feature tested on this experiment showed a benefic effect on manual takeover safety by allowing the driver to keep a safe time to collision. Moreover, a feature of the HMI was identified as potentially unclear. Indeed, during manual takeover, the vehicle speed is displayed on a different place which misled a driver into thinking that speed was the distance with the roadwork zone.

After controlling for scenario order, a significant effect of system is found on **mental demand** (F (1,28) = 6.83, p = .01, η_p^2 = .2) and **performance** : (F (1,28) = 4.97, p = .03, η_p^2 = .16). Mental demand is rated as higher in baseline condition compared to TeamMate condition (Mean = 45.73; SD = 31.55) while performance in TeamMate condition (Mean = 79; SD = 16.41) is rated as higher than in baseline condition (Mean = 64.07; SD = 19.69) (cf. Figure 23).

Effort during manual takeover with the baseline system (Mean = 52.67; SD = 31.51) is evaluated as higher than with the TeamMate system (Mean = 33.80; SD = 25.23), but this effect is only marginal (F (1,28) = 3.17, p = .09, η_p^2 = .1). **Frustration** with the baseline system (Mean = 44.67; SD = 33.21) is also evaluated as higher than with the TeamMate system (Mean = 27.40; SD = 24.56), but the effect of type of system is not significant (F (1,28) = 2.54, p = .12, η_p^2 = .09)(cf. Figure 23).

After controlling for order, a marginal effect of type of system is found on **physical demand** (F (1,28) = 3.45, p = .07, η_p^2 = .11) and on **temporal demand** (F (1,28) = 3.16, p = .09, η_p^2 = .1), with higher physical and temporal demand with the baseline system (physical demand : Mean = 49.27; SD = 33.8; temporal demand : Mean = 47.93; SD = 36.34) than with the TeamMate system (physical demand : Mean = 28.93; SD = 24.38; temporal demand : Mean = 28; SD = 22.4) (cf. Figure 23).

Error! Reference source not found.). **Satisfaction** is also rated as higher with the TeamMate system (Mean = 1.52; SD = 0.43) compared to the baseline system (Mean = 0.75; SD = 1.22) (cf. figure 24).

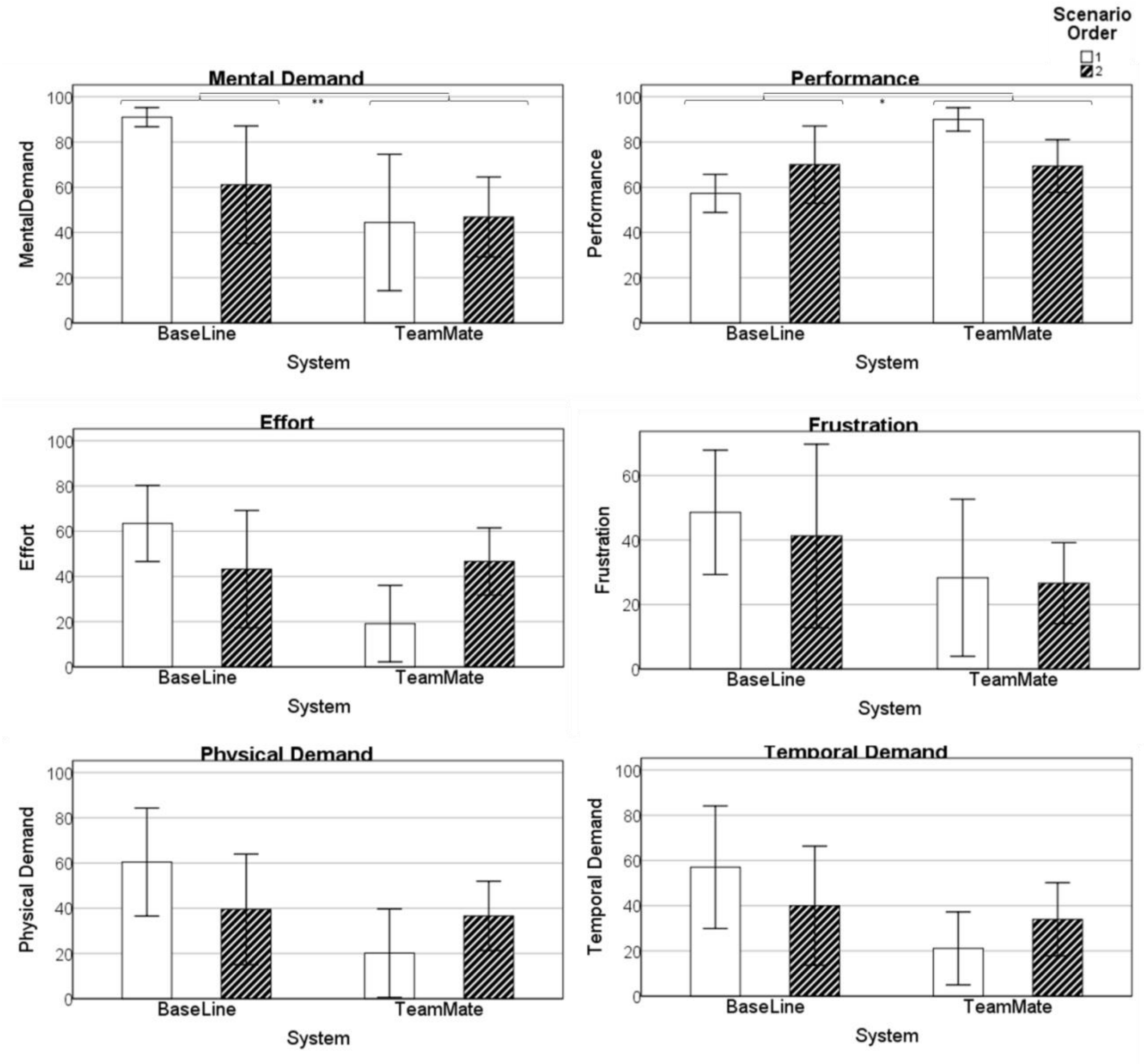

Figure 23: Mean mental demand, physical demand, temporal demand, performance, effort and frustration evaluated with the NASA-TLX, as a function of system (baseline or TeamMate) and scenario order (1 or 2). Errors bars represent standard errors.

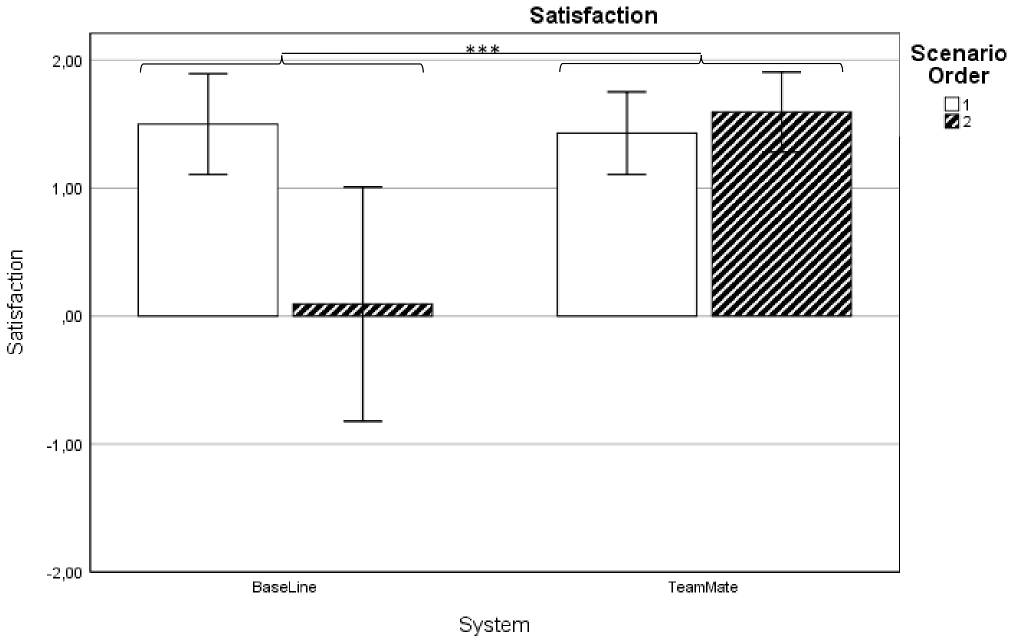

Figure 24: Mean satisfaction as a function of system (baseline or TeamMate) and scenario order (1 or 2). Errors bars represent standard error.

System **usability** is rated as higher in the TeamMate condition (Mean = 81.17; SD = 15.05) than in the baseline condition (Mean = 73.83; SD = 19.77). However, after controlling for order, the effect of type of system on usability is not significant (F (1, 28) = 1.27, p = .27, η_p^2 = .04) (cf. figure 25).

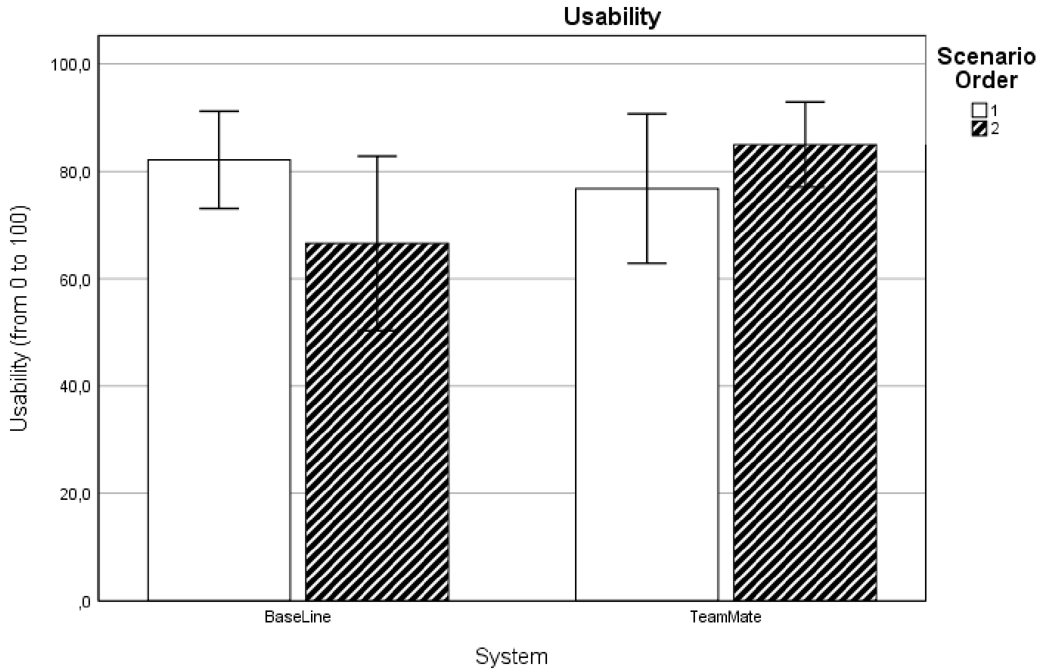

Figure 25: Mean usability as a function of system (baseline or TeamMate) and scenario order (1 or 2). Errors bars represent standard errors.

After controlling for order of presentation, **willingness to buy** for the TeamMate system (Mean = 3.40; SD = 1.07) is significantly higher (F (1, 28) = 5.1, p = .03, η_p^2 = .15) than for the baseline system (Mean = 2.40; SD = 1.3) (cf. Figure 26).

Figure 26: Willingness to buy as a function of system (baseline or TeamMate) and scenario order (1 or 2). Errors bars represent standard errors.

The amount of money (in euro) participants were willing to spent to purchase the system in addition to the price of their new car is higher for the TeamMate system (Mean = 13857.14 €; SD = 11179.43) compared to the baseline system (Mean = 9071.43 €; SD = 10457.15). However, after controlling for scenario order, this effect is not significant (F (1, 28) = 1.41, p = .24, η_p^2 = .05) (cf. Figure 27).

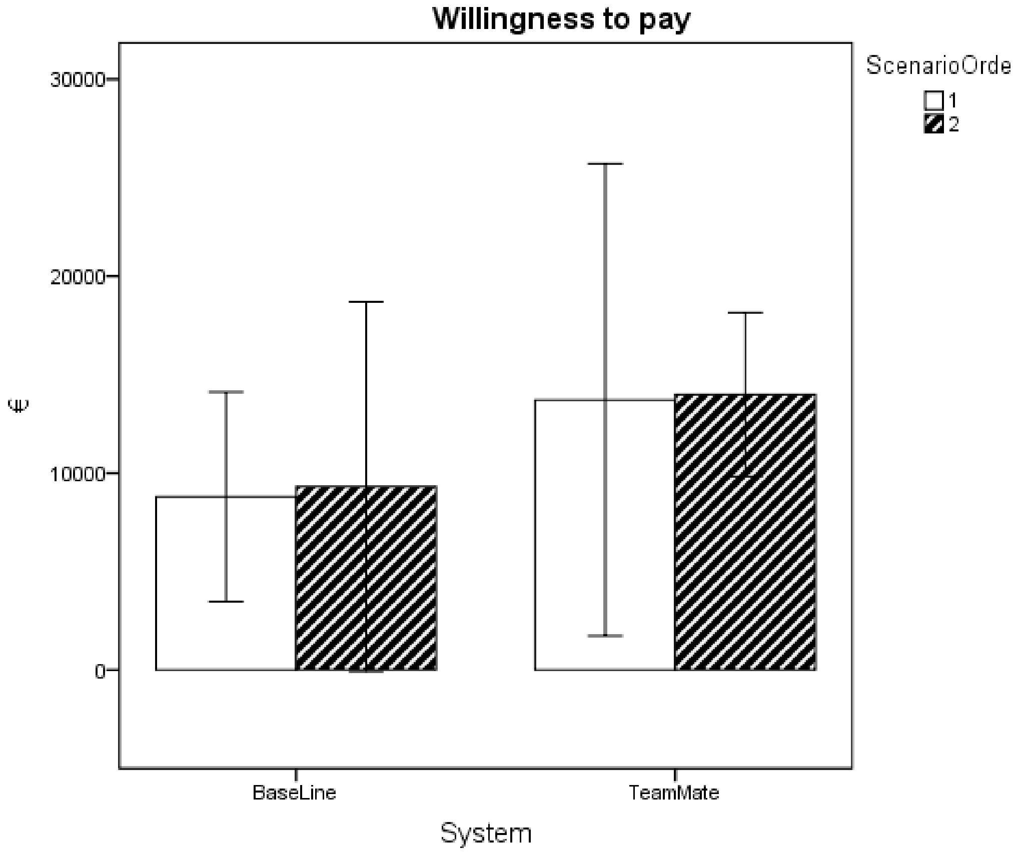

Figure 27: Willingness to pay as a function of system (baseline or TeamMate) and scenario order (1 or 2). Errors bars represent standard errors

Trust with the TeamMate system (Mean = 3.58; SD = 0.67) is rated as slightly higher than with the baseline system (Mean = 3.22; SD = 0.86), but after controlling for scenario order, this difference is not significant (F (1, 28) = 1, p = .73, η_p^2 = .16) (cf. figure 28).

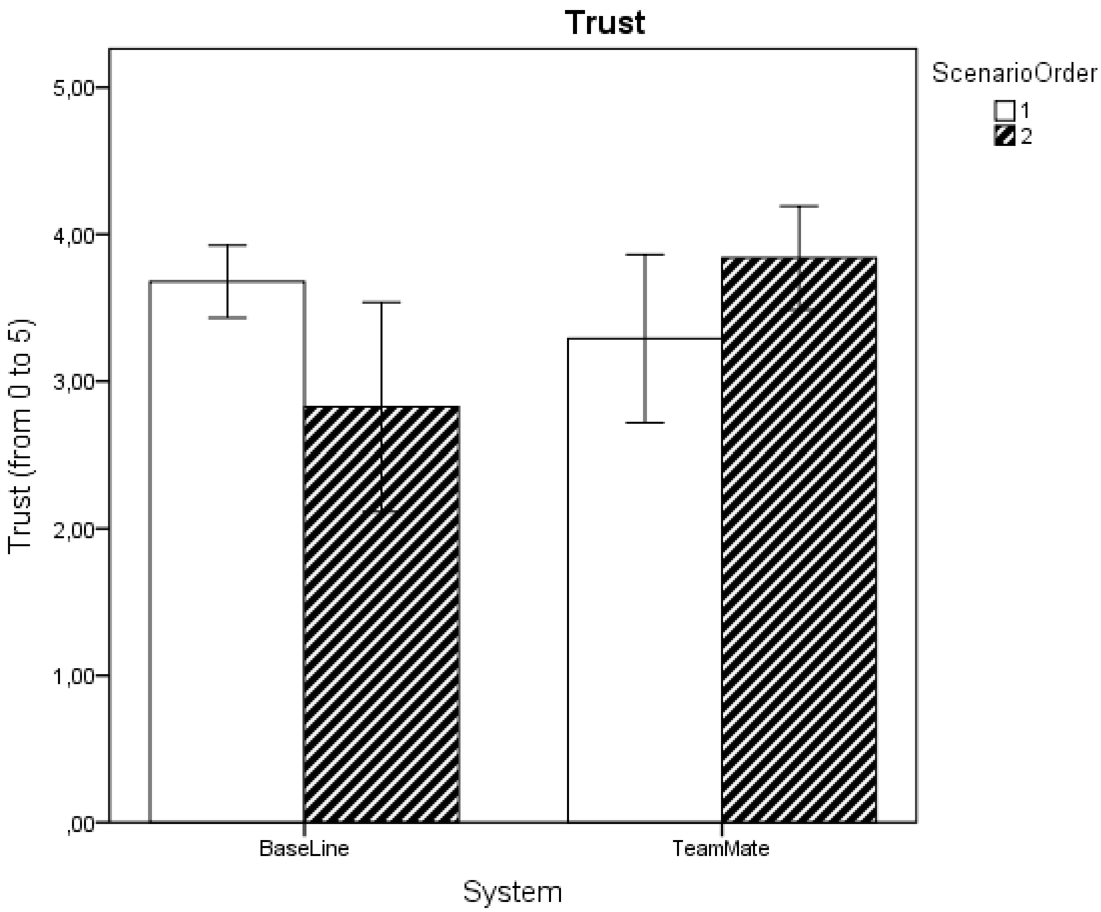

Figure 28: Mean trust as a function of system (baseline or TeamMate) and scenario order (1 or 2). Errors bars represent standard errors.

Our results points out the positive effect of V2I communication system on driver's mental workload, especially on mental demand and on auto-evaluated performance, during manual takeover. The subjective evaluation of the system is also improved by the V2I system by allowing a higher acceptance and a higher willingness to buy. However willingness to pay is not significantly higher for the TeamMate system, and trust and usability are not influenced by the TeamMate feature.

3.3.2 A2H support

3.3.2.1 Description of Evaluation Scenarios

Scenario 1

Martha is driving in an extra-urban road in Manual Mode. She receives an email and begins to read it. The TeamMate car detects that she is distracted, so it informs her that it will take the control of the vehicle, and then it automatically shifts to automated mode.

Scenario 2

Martha is driving in an extra-urban road in Manual Mode. She receives an email and begins to read it. The TeamMate car detects that she is distracted, so it informs her and proposes her to activate automated mode. Martha can choose to activate automated mode to finish reading her email, or to keep driving in manual mode.

The evaluation is aimed at demonstrating the role of the automation to promptly and efficiently address safety-critical conditions, thus the baseline is the manual driving (i.e. when there is no support of the automation).

3.3.2.2 Participants

20 participants took part in the experiment (10 women). Among them, 4 were excluded from data analysis due to simulator scenarios dysfunctions and 2 did not completed the experiment due to motion sickness. The remaining 14 drivers (7 women, 7 men) were in average 33.93 years old (SD = 7.77), held a valid driving license for at least 5 years (Mean = 14.86; SD = 8.35), had normal or corrected to normal vision, and drove on average 18285 km per year (SD = 10865). Drivers who already

drove VEDECOM's simulator were excluded from the experiment. Each driver received a 40 euros allowance.

3.3.2.3 Material

The experiment was conducted on VEDECOM static driving simulator described in section **Error! Reference source not found.**. The participant drove the same track described in section **Error! Reference source not found.** in a segment without the roadwork zone.

3.3.2.3.1 *System*

The driver monitoring system developed by CAF and described in deliverable 2.1 was used in this experiment to detect driver's distraction.

In case the driver is classified as distracted by the driver monitoring system, the TeamMate - Proposed system issued a vocal message to inform the driver on his/her attentional state and proposed to activate the automated mode ("*Vous avez l'air distrait, voulez-vous activer le mode autonome?"; "you seem distracted, do you want to activate automated mode?"*). The automated mode could be activated by pressing the dedicated button. If activated, a vocal message was issued to confirm it (*"mode autonome activé" - "automated mode activated").* In automated mode, the status of the system was updated on the dashboard as depicted in Figure 17. The automated mode could be deactivated at any moment by pressing the brake or the accelerator. The automated maintained the vehicle in the center of his lane, set speed according to speed limit, and maintained a minimum inter vehicular distance of 1.8 seconds with the vehicle followed.

In case the driver is classified as distracted by the driver monitoring system, the TeamMate - Forced system activated automated mode and issued a vocal message (*"vous êtes distrait, j'active le mode autonome" - "you are* *distracted, I activate automated mode").*Once in automated mode, the status of the system is updated on the dashboard as depicted in Figure 17. The automated mode could be deactivated by pressing the brake or the accelerator only if the driver was classified as "*attentive*" by the river monitoring system. The automated mode maintained vehicle in the center of his lane, set speed according to speed limit, and maintained a minimum inter vehicular distance of 1.8 seconds with the vehicle followed.

3.3.2.3.2 *Scenario*

A training scenario aimed at getting participants used to drive the simulator in manual mode and to observe the HMI. In this scenario participants drove 10 kilometers in manual mode.

The three experimental scenarios began with manual driving. After 6 kilometers travelled in manual mode, a vehicle overtook the participant and stayed in the right lane and maintained an inter-vehicular time of 2 seconds. After 10 seconds, a sound was issued, and an email was received on the infotainment tablet. The email needed to be answered by "*yes*" or "*no*". In the three experimental scenarios, after this sound was issued, in case detection of distraction triggered the brake of the followed vehicle until is speed was set to 80 km/h, and accelerate up to 130 km/h.

In the baseline scenario, the participant did not had the opportunity to activate automated mode. In the TeamMate - Proposed scenario, in case distraction is detected by the driver monitoring system, the vocal message used to propose automated mode activation was issued and participant could activate it. In the TeamMate - Forced scenario, in case distraction is detected, activation of the automated mode was triggered.

Two scenarios aimed at demonstrating the operating mode of the TeamMate systems. For the TeamMate - Propose scenario, the automated mode activation was proposed every 20 seconds. For the TeamMate - Forced scenario, the automated mode activated by itself every 20 seconds. The scenarios lasted until participants felt at ease with the system.

3.3.2.4 Procedure

The participants were welcomed, given the financial allowance and explained the aim of the experiment and the general steps of the procedure. Risks and constrains were explicitly explained to the participants who were informed that they could stop the experiment at any moment. Then participants were asked to read and sign an informed consent, and to answer a questionnaire to collect sociodemographic data.

Participants were told that they would drive four scenarios including a training scenario. Participants took the training scenario during which they were instructed to complete lane changes and to use the brake pedal as many time as they needed to get used to simulator's command.

Afterward, participants took the three experimental scenarios in a counterbalanced order. Before each scenario, the operating mode of the vehicle that had to be driven was explained to the participant.
Before the scenario in manual condition, it was explained to the participants that they would drive a classical vehicle in manual mode, and that they would be responsible of the entire control of the vehicle.
Before TeamMate scenarios, participants were told that the vehicle was equipped with a distraction detection system based on cameras, and an automated mode. Automated mode was described as working thanks to sensors, cameras that allowed the vehicle to perceive the environment, and

that road markings and GPS data were used in order to locate and to move in the environment. Participants were told that in automated mode, the vehicle would maintain itself on the centre of his lane and adapt his speed according to speed limits and to others vehicles. Participants were shown a screenshot of the dashboard in automated mode.

Before AutoMate - Proposed scenarios, participants were told that if the system classified them as "*distracted*", a vocal message would be issued to propose them to activate an automated mode. Then participants were instructed that automated mode could be activated by pressing a button, and that once activated, another vocal message would be issued, and that they would have to release pedals and steering wheel. Participants were explained that the automated mode could be deactivated at any moment by pressing brake or accelerator. Experimenter played the two vocal messages. Participants then drove the demonstration scenario.

Before TeamMate - Forced scenario, participants were told that if the system classified them as "*distracted*", the automated mode would be activated, and that a vocal message would be issued to inform them. The experimenter played the vocal message. The experimenter explained that the automate mode could be deactivated by pressing the brake or the accelerator only if they were attentive. Participants then drove the demonstration scenario.

Before each scenario, participants were instructed to comply with speed limit, to drive in the right lane as often as they could, and to be attentive to the road, unless they received an email. In this case, they were instructed to answer it. The experiment ended with a semi-structured interview. The whole protocol lasted about 1 hour and a half.

The experimental design is a within subject design, with the factor system with three modalities: baseline or TeamMate - Proposed, or TeamMate - Forced.

3.3.2.5 Results

2 KPIs were analysed: time exposed to time to collision inferior to 4 seconds and number of lane exceedance. Time exposed to time to collision inferior to 4 seconds was recorded from the beginning of the completion of the email to the end of the reacceleration of the leading vehicle. The number of participant exposed to a time to collision inferior to 4 seconds was computed for each condition. Number of lane exceedance during the completion of the email was computed for each condition.

The proportion of participant who experienced time to collision inferior to 4 seconds and lane exceedance was compared between conditions by mean of a Fisher test. The level of statistical significance is set at .05.

In the manual condition, 1 participant choose to change lane before starting the completion of the email, 6 changed lane after the braking of the followed vehicle, and 7 stayed in the right lane even after the braking of the leading vehicle. In the TeamMate - Proposed condition, 1 participant changed lane before beginning to complete the email task, and another participant chose not to activate automated mode. In the TeamMate - Forced condition 1 participant changed lane before beginning to complete the email.

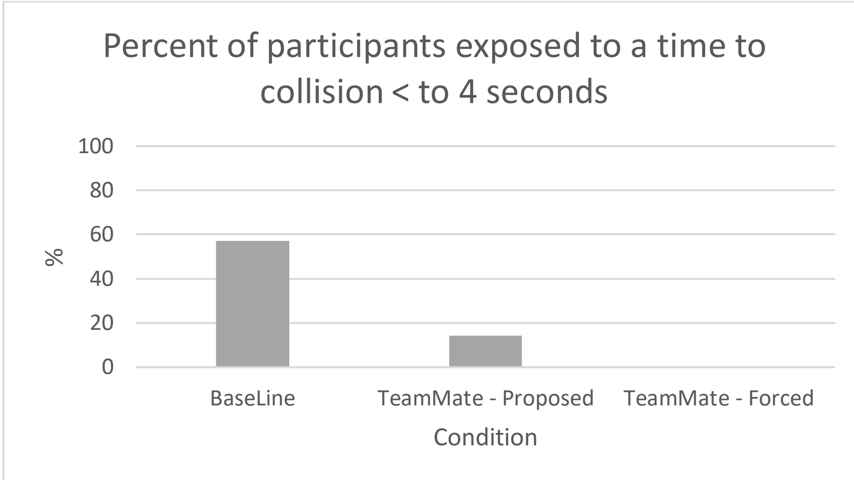

Figure 29: Percent of scenario exposed to a time to time to collision inferior to 4 seconds according to condition (baseline or TeamMate - Proposed or TeamMate - Forced.

The number of participants exposed to a time to collision inferior to 4 seconds was computed. They were 8 in baseline condition, 2 in TeamMate - Proposed condition and 0 in TeamMate - Forced condition. In TeamMate proposed condition, one of the participant that was exposed to a time to collision inferior to 4 seconds is the one who chose not to activate the automated mode. Compared to baseline condition, the proportion of participant who experienced a time to collision inferior to 4 seconds is significantly lower in TeamMate -Forced condition ($p < 0.01$), and in TeamMate - Proposed condition ($p < 0.05$). The difference between the two TeamMate conditions is not significant.

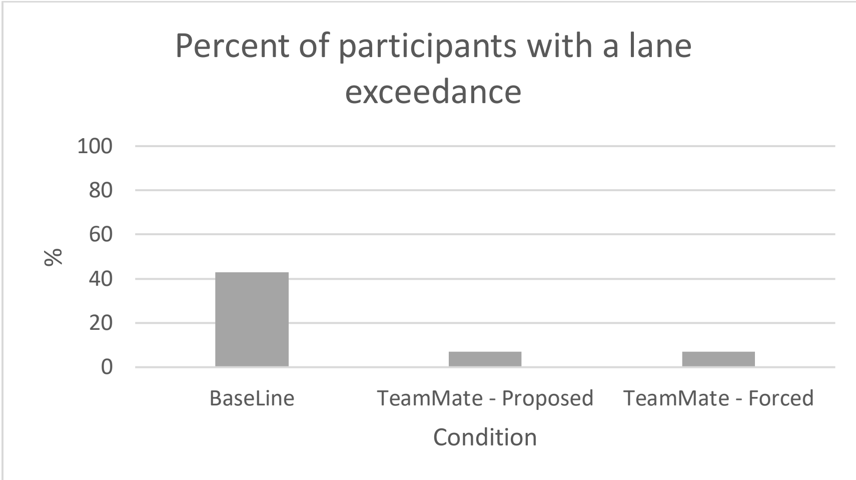

Figure 30: Percent of participants with a lane exceedance according to condition (baseline or TeamMate - Proposed or TeamMate - Forced.

In baseline condition, 6 participants experienced lane exceedance. The proportion of participant who experienced lane exceedance is marginally higher than in TeamMate - Proposed and in TeamMate - Forced conditions ($p = 0.8$). Indeed, 1 participant for each TeamMate condition experienced lane exceedance Figure 30.

After controlling for order, no significant effect of system is found neither on perceived usefulness (F (1, 27) = 0.23, p = .64, η_p^2 = .3) nor on satisfaction ((F (1,27) = 0.002, p = .96, η_p^2 = .00). **Perceived usefulness** is rated as slightly higher with the TeamMate - Forced system (Mean = 1.22; SD = 0.74) compared to the TeamMate - Proposed system (Mean = 0.20; SD = 0.57) (cf. Figure 31). Mean **Satisfaction** is rated equal with the TeamMate - Forced system (Mean = 1.42; SD = 0.55) compared to the TeamMate - Proposed system (Mean = 1.42; SD = 0.47) (cf. figure 24).

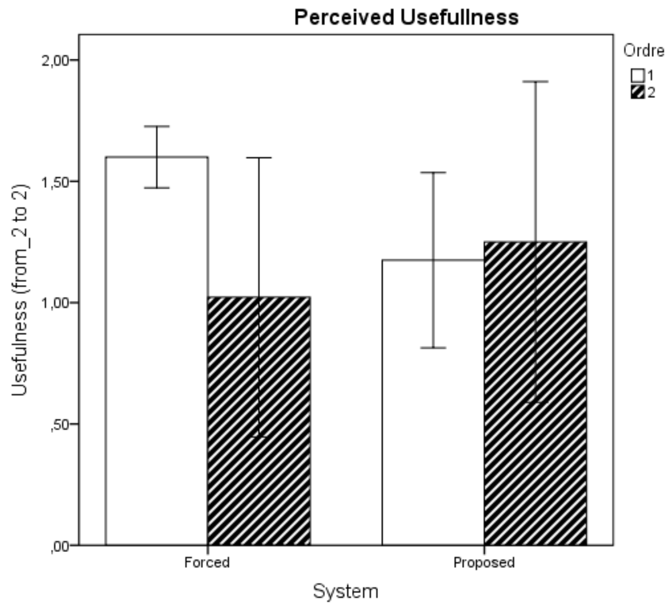

Figure 31: Mean perceived usefulness as a function of system (Proposed or- Forced) and scenario order (1 or 2). Errors bars represent standard error.

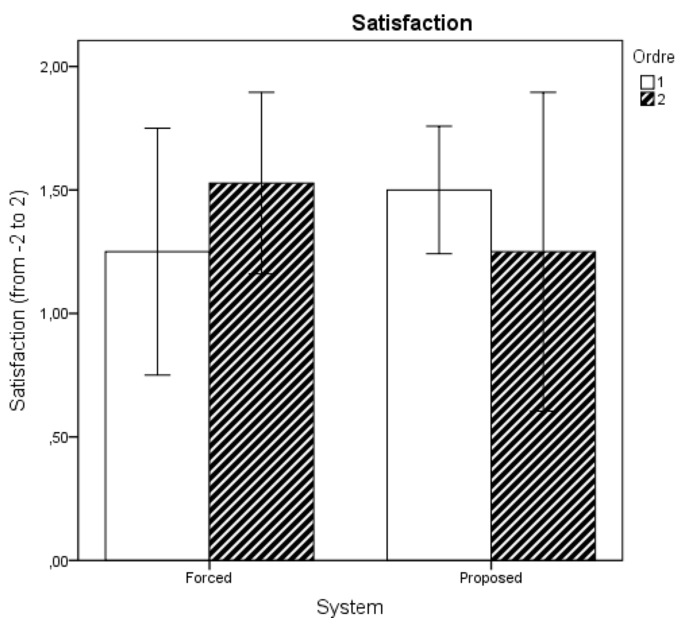

Figure 32: Mean satisfaction as a function of system (Proposed or Forced) and scenario order (1 or 2). Errors bars represent standard error.

Safety is increased with TeamMate system by reducing the exposition to a time to collision inferior to 4 seconds and lane exceedances. The TeamMate - Forced system allows a greater improvement of safety. System acceptance was not found to favour a system over another, even if participants acknowledged the higher safety benefit of the Forced system. However, participants evoked concerns about possible unwanted activation of the system, and some were reluctant to delegate the decision to activate automated mode to the system.

4 Conclusions

The previous chapter described the comparative simulator evaluation of TeamMate car against a baseline car in several use cases of the Eva, Peter, and Martha scenario. The aim of the evaluation was show the benefits of the bidirectional cooperation between the human and the automation in terms of safety, comfort, trust and acceptance.

In all scenarios and uses cases the TeamMate system brought improvements compared to the baseline system. This is reflected by the measured KPIs for both systems, where the TeamMate system in most categories reaches higher ratings than the baseline or is at least on par.

A critic point mentioned by the participant was the too conservative Driver Intention Recognition in the Peter scenario, which might be caused by the data this model was trained on. This issue could be handled with more data or the integration of the Online Learning enabler. While one has to consider that the manoeuvres performed by the participants where not too risky.

For the Martha scenario some participants had concerns about possible unwanted activation of the automation. This reaction is not surprising for the case that a system one is unfamiliar with tries to override actions.

Overall the cooperation introduce via the TeamMate concept brings benefits in several situations. For the 3rd cycle the objective is not only to create the AutoMate system with real vehicles but also to integrate more enablers for all three scenarios.

5 References

Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies*, *4*(3), 114-123.

Brown, R. D., & Galster, S. M. (2004). Effects of Reliable and Unreliable Automation on Subjective Measures of Mental Workload, Situation Awareness, Trust and Confidence in a Dynamic Flight Task. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 48(1), 147–151. doi: 10.1177/154193120404800132

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp. 139-183). North-Holland.

Kaur, K., Rampersad, G., *Trust in driverless cars: Investigating key factors influencing the adoption of driverless cars*, Journal of Engineering and Technology Management, Volume 48, April-June 2018, Pages 87-96

Körber, M. (2018). Theoretical considerations and development of a questionnaire to measure trust in automation. In Proceedings 20th Triennial Congress of the IEA. Springer.

Neukum, A., Lübbeke, T., Krüger, H. P., Mayser, C., & Steinle, J. (2008). ACC-Stop&Go: Fahrerverhalten an funktionalen Systemgrenzen. In Maurer, M. & Stiller, C. (eds.) *5. Workshop Fahrerassistenzsysteme – FAS 2008* (pp. 141-150).

SAE (2013). Operational definitions of driving performances measures and statistics.

Van Der Laan, J. D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies*, *5*(1), 1–10. http://doi.org/10.1016/S0968-090X(96)00025-3

Van Der Laan, J. D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies*, *5*(1), 1–10. http://doi.org/10.1016/S0968-090X(96)00025-3