D6.1 - Evaluation methodology for TeamMate Car demonstrators (Revised)

Project Number:	690705
Classification **Deliverable No.:**	Public R_D6.1
Work Package(s): **Document Version:** **Issue Date:** **Document Timescale:** Start of the Document: Final version due:	WP6 Vs. 1.0 Month 15 Project Start Date: September 1, 2016 Month 12 (Revision: Month 24) Month 16 (Revision: Month 25)
Deliverable Overview:	**Main document:** D6.1
Compiled by:	Ebru Dogan, VED
Authors:	Ebru Dogan, VED Lynda Halit, VED Mohamed Cherif Rahal, VED Elisa Landini, REL Andrea Castellano, REL Jurgen Pirchen, ULM Maximilan Graff, ULM Fabio Tango, CRF
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
11/09/2018	Modifications following mid-term review	VED
22/10/2018	Modifications following consortium meeting	VED
23/10/2018	Contributions by REL	REL
26/10/2018	Contributions by ULM	ULM
26/10/2018	Contributions by VED	VED
29/10/2018	Compilation and final revisions	VED

[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
CAN	Controller area network
H2A	Human to Automation
HMI	Human-Machine interface
KPI	Key performance indicator
SAE	Society of Automotive Engineers
SUS	System Usability Scale
TET	Time exposed time to collision
TOR	Takeover request
TTC	Time to collision
UX	User experience
V2I	Vehicle to infrastructure

List of Figures

List of Tables

Executive Summary

The aim of AutoMate project is to develop a novel concept of interaction and cooperation based on trust to regulate driver-vehicle interactions in automated vehicle in a safe and acceptable manner. Seven technical enablers have been identified to develop TeamMate system. These enablers will be integrated and evaluated through a series of evaluation studies during the 2nd cycle of the project. One of the challenges of evaluation of an automated vehicle is the lack of metrics and measures consensually accepted by the stakeholders.

In the evaluation studies in 2nd cycle, a range of key performance indicators that are used in diverse research on automated driving has been identified. In first part of chapter 2, these indicators have been grouped in three parties: safety aspects of automated driving, user aspects of automated driving, and the efficiency obtained by automated driving. The indicators are defined and hypotheses for each indicator have been specified. The second part of chapter 2 presents the experimental methodology for evaluation studies, namely, the use cases to be tested and the protocol to be applied.

The indicators that are defined in the current deliverable will be revised before the 3rd cycle of the project based on the feedback from the evaluation studies.

1 Introduction

The research and development activities in the AutoMate project have been organized in 3 cycles to make sure that the development of enabling technologies carried out in the first cycle takes into account user needs, as they are dimensioned during evaluation experiments in the second cycle, in an iterative manner for modification of enabling technologies in the third cycle. The goal of this iterative approach is to progressively improve the TeamMate concept throughout the project, in line with the needs and objectives of each demonstrator at the end of the project.

The current deliverable describes the methodology for the evaluation of TeamMate concept at the end of the second cycle. Metrics and measures may be modified in the third cycle evaluation experiments depending on the outcomes of the evaluation experiments. Next sections will focus on the evaluation methodology and the definition of key performance indicators.

2 Methodology for comparative evaluation

The suitable evaluation methodologies and procedures for automated vehicle are currently not clear, neither are the key performance indicators (KPIs). While a cooperative work for the standardization of metrics and measures are carried out at the international level, a consensual document was not available at the time of the writing of this deliverable. Thus, at the second cycle of the project, a comparative approach for the evaluation of TeamMate concept has been adopted. More precisely, the improvement in human-automation interactions brought about by the enablers of the TeamMate concept in terms of safety and acceptance are evaluated against a baseline vehicle which is not equipped with these enablers.

The first step was to define adequate KPIs (D1.3) for the evaluation of the TeamMate concept (D4.4). The second step was to reduce the number of use cases defined in D1.1 to few, representative use cases which would still allow evaluation of different enablers on selected indicators. A due attention was paid to have the types of use cases that demonstrate situations when the automation supports human (A2H support), as well as the situations when the human supports automation (H2A support), for both are essential for the cooperation between the vehicle and the driver. The third step was to adjust and further detail the KPIs.

Certain indicators are individual values that can be quantified based on previous research. For instance, minimum time to collision, which indicates time left to a collision if the vehicle's current speed is kept constant, can be evaluated with reference to the values recommended in metrics documents by Society of Automotive Engineers (SAE J2944, 2015). Some other indicators, however, cannot be quantified that easily either because there are no reference values (e.g. willingness to pay for an automated vehicle) or because they are specific to the system that is in questions (e.g. trust in a specific technology). This is particularly the case with indicators used for user-related aspects. Comparative approach in evaluation experiments allow us to surmount these concerns, for a significant difference between the TeamMate car and the baseline car indicates improvement on the selected indicators.

3 Definition of KPIs

As described in the previous section, different aspects of the TeamMate car concept will be evaluated by different demonstrators according to the scenario and use cases selected for the empirical experiments. As a consequence, some of the KPIs will be different among the experiments in order to reflect the characteristics of the use cases, while some of them will be common. The objective is to adopt the same objective and subjective measurement tool for the common KPIs as much as possible.

The KPIs are grouped under three headings: safety aspects, user aspects, and efficiency. The next section provides a detailed description of the categories as well as the specific KPIs used by the demonstrators.

3.1.1 Safety aspects

Time to Collision

Time to collision (TTC) refers to the time interval required for a vehicle to collide with an object, which can be another vehicle, infrastructure or a vulnerable road user. It is usually measured in seconds and commonly used as an indicator of safety. Larger values of TTC indicate higher degree of safety as it leaves longer time for driver to react to avoid a collision. Two variants of TTC were used in the evaluation studies: minimum TTC and time exposed TTC.

Minimum TTC refers to the *minimum* time interval required for a vehicle to collide with an object, measured generally in seconds. Figure 1 shows the local minimum value for TTC.

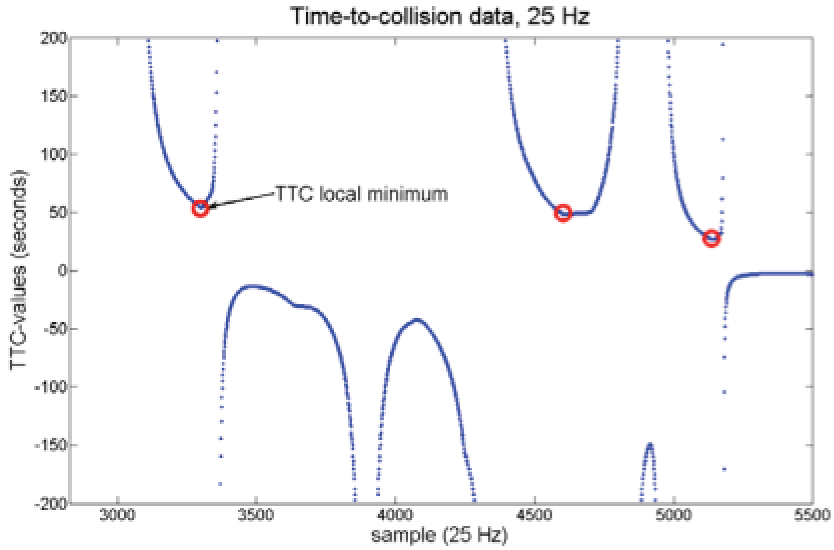

Figure 1. Minimum time to collision (Source: Ostlund, Nilsson, Tomros, and Forsman, 2006, cf. SAE J2944, 2015)

The threshold for the minimum TTC is decided as 3.5 seconds in the evaluation studies following the Hogema and Janssen (1996, cf SAE J2944, 2015).

Time exposed TTC (TET) refers to the time interval during which the TTC is below a certain exposure threshold that is considered to be safety-critical (Minderhoud & Bovy, 2001; cf. SAE J2944, 2015). It is considered to be a safety-relevant measure for it takes into account exposure time. Figure 2 depicts TET as it was originally conceptualized by Minderhoud and Bovy (2001, cf. SAE J2944, 2015). Accordingly, "TET is a summation of all moments (over the time period H) that a driver approaches a front vehicle with a TTC value below the threshold value TTC, the latter is considered to be the boundary between safe and safety-critical approaches."

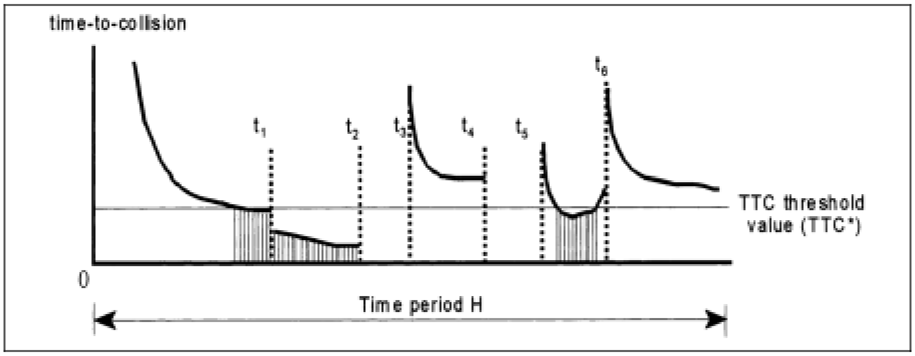

Figure 2. Number of conflicts given arbitrary fluctuations of TTC (Source: Minderhoud & Bovy, 2001, cf. SAE, 2015)

The threshold to distinguish safe and unsafe situations is decided to be 4 seconds, following Van der Horst (1991) and Farber (1991) (cf. SAE J2944, 2015) and the percentage of time above/below this threshold will be calculated.

Number of accidents

Road safety is often assessed by accident records. One of the indicators is the number of accidents involving injury per million kilometres driven. This indicator will be adapted in the evaluation studies and the number of accidents that the vehicle got involved in during the experiment will be used. The success criterion will be zero accidents while driving with the TeamMate car.

Number of lane boundary exceedance

Exceeding the lane boundaries reflect poor vehicle control and jeopardizes safe driving. Number of lane boundary exceedance is defined as the number of time a wheel of the car exceeds the lane line. Zero lane boundary exceedance will be considered as the success criterion in the evaluation studies.

Number of second thoughts

In a two-lane two-way rural road, an attempt of overtaking while another car is approaching on the opposite direction can be considered as a risky behavior. This might reflect an error or a hesitation in decision making process. This indicator will be measured by calculating the number of times the driver attempts an overtake manoeuvre (i.e. left wheels are on the opposite lane) and then he/she drives the car back to the original lane, which will be retrieved from CAN bus data. Completely eliminating error-prone lane change manoeuvres will be considered as the success criterion in the evaluation studies.

KPI	Hypotheses
Minimum TTC	The number of participants below the threshold of 3.5 sec would be higher with TeamMate car.
Time exposed TTC	Time exposed TTC below threshold of 4 sec would be lower with TeamMate car.
Number of accidents	Number of accident would be zero.
Number of lane boundary exceedance	Number of lane boundary exceedance would be lower with the TeamMate car.
Number of second thoughts	Number of second thoughts to engage in a risky lane change manoeuvre would be lower with the TeamMate car.

Table 1. Summary of safety-related indicators and hypotheses for evaluation

3.1.2 User aspects

Trust

Trust refers to user's willingness to depend on a particular technology because of the characteristics of that technology (adapted from McKnight et al., 2011).

We try to answer the question "What about the TeamMate concept makes is trustworthy, irrespective of the people and the human structures that surround it?".

We plan to use a questionnaire described in Körber (2018), tailored to study trust related to automated driving systems to compare baseline and TeamMate vehicles. In the questionnaire, participants evaluates their level of agreement with the following statements on a Likert scale from 1 to 5. Table 2 presents the items of the trust questionnaire.

	1	2	3	4	5
The system is capable of interpreting situations correctly					
The system state was always clear					
I already know similar systems					
The developers are trustworthy					
One should be careful with unfamiliar automated systems					
The system works reliably					
The system reacts unpredictably					
The developers take my well-being seriously					
I trust the system					
A system malfunction is likely					
I was able to understand why things happened					
I rather trust a system than I mistrust it					
The system is capable of taking over complicated tasks					
I can rely on the system					
The system might make sporadic errors					
It's difficult to identify what the system will do next					
I have already used similar systems					
Automated systems generally work well					

Table 2. Trust scale (Körper, 2018)

Acceptance

Acceptance (of a new technology) reflects one’s attitudes towards this technology. In other words, it refers to favorable or unfavorable evaluations of a technology and its usage. Various acceptance scales exist in the literature, mostly based on intention to use.

In the evaluation studies, acceptability and acceptance were measured by a standard scale developed by Van der Laan and colleagues (1997) (Table 3). The scale consists of nine items measuring two factors: *usability* (useful/ not useful, good/ bad, effective/ superfluous, assisting/ worthless, and raising alertness/ sleep-inducing) and *satisfaction* (pleasant/ unpleasant, nice/ annoying, likeable/ irritating, and desirable/ undesirable). Participants evaluated each item on a 5-point Likert scale ranging from -2 to 2. The scale was administered twice. The first administration was meant to measure acceptability and took place before participants tried the system. The second time was meant to measure acceptance of the system; hence, was carried out at the end of the study. Higher scores indicate more favourable responses.

	-2	-1	0	1	2	
Useful						Not useful
Pleasant						Unpleasant
Bad						Good
Nice						Annoying
Effective						Superfluous
Irritating						Likeable
Assisting						Worthless
Undesirable						Desirable
Raising alertness						Sleep inducing

Table 3. Acceptability/acceptance scale developed by Van der Laan and colleagues (1997)

Another acceptance scale that will be used in the evaluation studies is based on Technology Acceptance Model (Davis, 1986), which is tailored to the context of information systems. The questionnaire aims to disentangle two dimensions: *perceived usefulness* is the degree to which a person believes that using a certain technology would improve performance, *perceived ease of use* is the degree to which a person believes that using a certain technology would be easy and free of effort. Higher scores indicate higher level of acceptance.

Participants are asked to provide their level of agreement on a 7 point scale (1=likely and 7 = unlikely) for the following 12 sentences (Table 4)

	1	2	3	4	5	6	7
Using [this car | the TeamMate car] would enable me to accomplish tasks more quickly							
Using [this car | the TeamMate car] would improve my performance							
Using [this car | the TeamMate car] would increase my productivity							
Using [this car | the TeamMate car] would enhance my effectiveness							
Using [this car | the TeamMate car] would make it easier to drive							
I would find [this car | the TeamMate car] useful							
Learning to operate [this car | the TeamMate car] would be easy for me							
I would find it easy to get [this car | the TeamMate car] to do what I want it to do							
My interaction with [this car | the TeamMate car] would be clear and understandable							
I would find [this car | the TeamMate car] to be flexible to interact with							
It would be easy for me to become skillful at using [this car | the TeamMate car]							
I would find [this car | the TeamMate car] easy to use							

Table 4. Acceptance scale based on Technology Acceptance Model

Mental workload

For the acceptance we will also measure the workload and frustration of the driver, by using the NASA-TLX questionnaire (Hart & Stavenland, 1988), shown in Figure 3.

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

Figure 3. Questionnaire for workload and frustration (NASA-TLX)

User experience

User experience (UX) refers to user's emotional experience and attitudes while using a particular product, in this case AV. The main indicator of the UX in the evaluation studies will be system usability, which will be measured by a commonly used scale of System Usability Scale (SUS) shown in Table 5. This scale is composed of 10 items. For each item, participant has to evaluate his level of agreement on a scale ranging from 1 (strongly disagree) to 5 (strongly agree). 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.

	1	2	3	4	5
I think that I would like to use this system frequently					
I found the system unnecessarily complex					
I thought the system was easy to use					
I think that I would need the support of a technical person to be able to use this system					
I found the various functions in this system were well integrated					
I thought there was too much inconsistency in this system					
I would imagine that most people would learn to use this system very quickly					
I found the system very cumbersome to use.					
I felt very confident using the system					
I needed to learn a lot of things before I could get going with this system					

Table 5. System Usability Scale (Brooke, 1996)

Willingness to buy

In order to evaluate potential users’ willingness to, the methodology followed by Kyriakidis and colleagues (2015) will be adopted. Accordingly, participants' responses to the question whether they would buy a vehicle equipped with TeamMate system versus a baseline system will be collected on a scale 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€ will be used to evaluate how much money participants would be willing to spend to purchase the system in addition to the price of the vehicle.

KPI	Hypotheses
Trust	Trust in automation would be higher while using TeamMate car.
Acceptance	Indicators of acceptance, namely, usefulness and satisfaction would be higher after using TeamMate car.
Mental workload	Dimensions of mental workload would be lower while using TeamMate car.
Usability	Usability of TeamMate car would be evaluated higher than baseline car.
Willingness to buy	Participants would be willing to buy and to pay more for TeamMate car than baseline car.

Table 6. Summary of user-related indicators and hypotheses for evaluation

3.1.3 Efficiency

Time to drive a roundabout

This KPI is aimed at measuring how long the vehicle takes to enter a roundabout (with and without the support of the driver).

In details, it will be calculated as T1 – T0, where

- T0: arrival in the roundabout (50 m before the roundabout)
- T1: exit from the roundabout

It will be calculated by gathering CAN bus data. In driving simulator's logs, the T0 is practically collected when the information is shown in the HMI, while the T1 is collected when the car arrives in a predetermined area of the scenario, placed at the end of the roundabout and marked with a virtual marker.

The hypothesis here is that the TeamMate car, and in particular the use of the H2A support in perception, is able to reduce the time needed to enter the roundabout. Since it is a comparative evaluation, no state-of-the-art measures can be used. However, a reduction of "Time to enter a roundabout" of 20% will be considered as acceptable.

Table 7 summarizes all the indicators will be used in the evaluation studies and the use cases in which they will be used.

Indicator	Evaluation aspect	Use case
Minimum TTC	Safety	Martha, Eva
Time exposed TTC		Martha, Eva
Number of lane boundaries exceedance		Martha
Number of accidents		Peter
Number of second thoughts		Peter
Takeover time		Peter
Trust	User	All
Acceptance		All
Mental workload		All
Usability		All
Willingness to buy		Martha and Eva
Time to enter a roundabout	Efficiency	Eva

Table 7. Summary of indicators used for evaluation studies

4 Demonstrator simulators and vehicles: Use cases and experimental protocol

4.1 REL Simulator

4.1.1 Scenario and uses cases

The REL demonstrator will be evaluated by considering the EVA scenario, which takes place in urban and peri-urban traffic context.

The EVA scenario has been identified because it is representative of a limit of the automation: entering a roundabout is a well-known issue for the autonomous driving [[2]] in terms of efficiency and comfort. The scenario implemented in the driving simulator occurs in an urban and peri-urban scenario: the ego-vehicle passes through a simulated city center, while the roundabouts are located at the boundaries of the city center.

The high traffic flows near the roundabout and the different directions of the other vehicles can dramatically affect the time to enter the roundabout and then it can create frustration and reduce the acceptance of the driver.

In order to improve the efficiency of the maneuver, the automation can ask for support to the driver (either in perception or in action).

On the other hand, this scenario is not particularly relevant for the safety of the driver, because the low speed and driving conditions in the roundabout reduce the effect of collisions and the risk of fatalities (compared to other scenarios, such as driving in the highway or in two-lane two-way rural roads).

[2] http://theconversation.com/budget-2017-uks-driverless-cars-stuck-on-testing-roundabout-87805

Therefore, the use cases for the support of the driver to the automation (H2A in perception and in action) have been selected to evaluate the added value of the TeamMate approach (i.e. the cooperation) for the CRF / REL demonstrator.

In the "H2A in perception" use case, the automation is in charge of the vehicle control, and it needs a support in perception from the driver.

The support in perception means that the automation needs a help to continue the driving task without a transition of control to the driver.

This is relevant because the transition of control from the automation to the driver (i.e. the so called "disengagement") represents a highly critical condition for the interaction between the driver and the automation, because it requires the driver to be promptly brought back in the loop while he/she is likely to be performing non-driving tasks.

This case is particularly relevant for the evaluation since this can be considered the most effective example of cooperation (because it reduces the number of disengagements).

The following text provides a simple story (adapted from the EVA use cases) to intuitively describe the scenario for the evaluation of the CRF/REL demonstrator.

H2A support in perception

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 maneuver, the TeamMate car asks Eva a cooperation in perception, asking her to check the available space and to provide a trigger to start the maneuver. 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.*

4.1.2 Baseline

The selection of the H2A use case as the most relevant for the evaluation (to demonstrate the added value of the cooperation in the EVA scenario) also affects the definition of the baseline for the REL demonstrator. Since the demonstrator is aimed to show the value of the driver to support the automation, the baseline is represented by a condition where the driver has no role in the cooperation (i.e. the so called "driverless" approach): therefore, the baseline is the autonomous driving without any support of the driver.

4.1.3 Experimental protocol

As reported in Table 7, since Eva scenario is not safety-critical, mostly comfort- and acceptability-related KPIs will be measured. To measure them in the simulated environment, a within-subject design will be considered: the same subject will be asked to perform the scenario with the baseline and with the TeamMate system, randomized in order to make the dataset more consistent and avoid bias.

Moreover, in order to make the obtained results reliable, a number of at least 20 subjects is expected to be involved in the experiment. The requirements to participate at the experiment will be to have a valid driving license. No previous experience with automated or semi-automated vehicles will be requested. In Eva scenario will be used both objective (e.g. Time to enter the roundabout to measure efficiency) and subjective (e.g. trust and acceptance) indicators. The users will be asked to drive in the urban and peri-urban scenario, and to perform three roundabouts in each configuration (i.e. baseline and

TeamMate). More details about the implementation of the scenarios and experimental setup will be provided in D6.2.

4.2 ULM simulator

4.2.1 Scenario and uses cases

The ULM simulator demonstrator will be evaluated by considering the PETER scenario, which takes place in rural traffic context.

The PETER scenario has been identified since it is representative of a limit of the automation: in a rural road, the automation may not be able to efficiently overtake a tractor, because

1) Its sensors cannot acquire a complete view of the environment (due to the tractor)
2) The automated vehicle is not able to detect the right moment to start overtaking

As a consequence, the automated vehicle may take a lot of time before overtaking, or even do not overtake at all (and just follow the tractor along the rural road).

In order to improve the efficiency of the maneuver, the automation can ask for support to the driver (H2A, either in perception or in action).

H2A in perception aims at demonstrating how human can support automation by providing necessary input at perception level. In this use case, the demonstrator needs Peter's input to fill in missing information beyond its perceptual horizon, which is obstructed by a tractor, in order to carry out the overtaking manoeuvre in a safe manner.

In this use case, the automation is in charge of the vehicle control, and it needs a support in perception from the driver. The support in perception means that the automation needs a help to continue the driving task without a transition of control to the driver.

As already described for the CRF/REL demonstrator, this use case is relevant because the disengagements represent a highly critical condition for the interaction between the driver and the automation. And, with this type of support (H2A in perception) we can reduce the number of disengagements.

The PETER scenario is also particularly relevant for the safety of the driver, because driving in two-lane two-way rural roads with low visibility (due to the tractor) can negatively affect the perception of risk of the driver, and lead to risky driving behaviors.

Therefore, for the ULM simulator demonstrator, both the use case for the support of the driver to the automation (H2A in perception) and the use case for the support of the automation to the driver (A2H in action) have been selected to evaluate the added value of the TeamMate approach (i.e. the cooperation).

The following text provides 2 simple stories (adapted from the PETER use cases) to intuitively describe the scenario for the evaluation of the ULM simulator demonstrator.

H2A support in perception

Peter is driving in a narrow rural road in Automated Mode. The car, arriving behind a tractor, detects that it obstructs the view. Therefore, the vehicle is not confident of the available space sideways to overtake the tractor, due to a limit in perception. Since the vehicle is not sure about the possibility to perform a safe overtake, it would follow the tractor either until the road is wider or the *tractor changes direction. The TeamMate car asks Peter to check. When Peter confirms there is enough space, the TeamMate car performs the overtake in Automated Mode.*

A2H support in action

Peter is driving in a narrow rural road in Manual Mode. He approaches a tractor, that causes limited visibility or the road, but he is in a hurry, so he decides to perform the overtake. The TeamMate car detects a car approaching from the opposite lane. A collision is likely to occur. In order to avoid it, the TeamMate car takes the control of the vehicle and safely plans and execute a safe manoeuvre to drive the vehicle back to the original lane. When the situation is safe, the automation hands over the control to the driver (back to Manual Mode).

4.2.2 Baseline

The selection of both H2A support and A2H support (as well as the corresponding different use cases), requires the definition of 2 different baselines for the evaluation:

- For the H2A use case, the evaluation is aimed at demonstrating the added value of the driver, thus the baseline is the driverless approach (i.e. the autonomous driving without any intervention of the driver)
- For the A2H use case, 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)

4.2.3 Experimental protocol

The TeamMate car concept will be tested in a within-design experimental setup against the baseline car. The use cases (H2A in perception and A2H in action) will be tested in different experiments. This experimental design will allow to compare same driver's evaluation of the two different car concepts and reduce individual differences. The aim is to collect data from 20-30 people for each use-case in simulation studies based on the KPIs.
In order to simulate the use case of the A2H support in action (where Peter is in a hurry and so attempts a risky overtake), the participants will be asked to complete the experiment in a limited time slot, in order to avoid them to just follow the tractor.

4.3 ULM vehicle

4.3.1 Scenario and uses cases

The ULM vehicle demonstrator will be evaluated by considering the PETER scenario.

As for the ULM vehicle, the PETER scenario has been identified because it is representative of a limit of the automation: in a rural road, the automation may not be able to efficiently overtake a tractor.

As a consequence, the automated vehicle may take a lot of time before overtaking, or even do not overtake at all (and just follow the tractor along the rural road for several kms).

In order to improve the efficiency of the maneuver, the automation can ask for support to the driver (either in perception or in action).

In the "H2A in perception" use case, the automation is in charge of the vehicle control, and it needs a support in perception from the driver. The support in perception means that the automation needs a help to continue the driving task without a transition of control to the driver.

As already described for the CRF/REL demonstrator, this use case is relevant because the disengagements represent a highly critical condition for the interaction between the driver and the automation. And, with this type of support (i.e. H2A in perception) we can reduce the number of disengagements.

The PETER scenario is also particularly relevant for the safety of the driver, because driving in two-lane two-way rural roads with low visibility (due to the tractor) can negatively affect the perception of risk of the driver, and lead to risky driving behaviors. However, since a real vehicle will be used on the road for the evaluation, these aspects will not be taken into consideration (for safety reasons).

Therefore, for the ULM vehicle demonstrator, only the support of the driver to the automation (H2A in perception) has been selected to evaluate the added value of the TeamMate approach (i.e. the cooperation).

The following text provides a simple story (adapted from the PETER use cases) to intuitively describe the scenario for the evaluation of the ULM vehicle demonstrator.

H2A support in perception

Peter is driving in a narrow rural road in Automated Mode. The car, arriving behind a tractor, detects that it obstructs the view. Therefore, the vehicle is not confident of the available space sideways to overtake the tractor, due to a limited visibility of its sensors. Since the vehicle is not sure about the possibility to perform a safe overtake, it would follow the tractor either until the road is *wider or the tractor changes direction. To overcome this limit, the TeamMate car asks Peter to check if there is enough space to overtake. When Peter confirms, the TeamMate car performs the overtake in Automated Mode.*

4.3.2 Baseline

The selection of the H2A support implies that the evaluation is aimed at demonstrating the added value of the driver, thus the baseline is the driverless approach (i.e. the autonomous driving without any intervention of the driver).

4.3.3 Experimental protocol

The TeamMate car concept will be tested in a within-design experimental setup against the baseline car. The use case (H2A) will be tested in different experiments. This experimental design will allow to compare same driver's evaluation of the two different car concepts and reduce individual differences. The aim is to collect data from 15-20 people for each use-case in vehicle studies based on the KPIs.

4.4 VED simulator

4.4.1 Scenario and uses cases

The VED simulator demonstrator will be evaluated by considering the MARTHA scenario which takes place in highway traffic context.

The MARTHA scenario has been identified since it is representative of a limit of the automation: in case of roadworks, the automation may not be able to detect the lanes to safely drive in Automated Mode. As a consequence, the automated vehicle may unexpectedly handover the control to the driver (the so called “disengagement”) and this situation could represent a safety critical condition for the driver (as already explained in the previous sections).

In order to improve the efficiency of the maneuver, and avoid the disengagement, the automation can ask for support to the driver (H2A in action). H2A in action was selected in order to demonstrate how human can support the automation when the automation reaches its functional limits. The support in action implies that one of the team member needs direct intervention by the other for a safe driving.

While the H2A use cases selected so far (for EVA and PETER) describe a support in perception, and thus are linked to efficiency, trust and acceptance issues, the H2A in action is also particularly relevant for the safety of the driver, because without his/her intervention, the TeamMate car is not able to continue driving in Automated Mode and it has to perform either a disengagement or a safe maneuver to stop the vehicle.

The MARTHA scenario is also relevant for the safety because it considers a use case where Martha is distracted, and she needs the support of the automation to guarantee her safety. Therefore, for the VED simulator demonstrator, both the use case for the support of the driver to the automation (H2A in action) and the use case for the support of the automation to the driver (A2H in perception and in action) have been selected to evaluate the added value of the TeamMate approach (i.e. the cooperation).

The following text provides 2 simple stories (adapted from the MARTHA use cases) to intuitively describe the scenario for the evaluation of the VED simulator demonstrator.

H2A support in action

The TeamMate car is driving in a highway 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 her to handle the control of the vehicle during the roadwork zone. Martha takes over the control until the end of the roadwork, and is able to shift back to Automated Mode afterward.

A2H support in perception and in action

Martha is driving in a highway 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.

Martha is driving in a highway 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.

4.4.2 Baseline

The selection of both H2A support and A2H support (as well as the corresponding different use cases), requires the definition of 2 different baselines for the evaluation:

- For the H2A use case, the evaluation is aimed at demonstrating the added value of the driver, thus the baseline is the driverless approach (i.e. the autonomous driving without any intervention of the driver)

- For the A2H use case, 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)

4.4.3 Experimental protocol

A 2 x 2 mixed design will be employed in order to evaluate the TeamMate car concept. The use cases (A2H and H2A) will be between-subjects factor, while the car concept (baseline car and TeamMate car) will be a within-subjects factor. This experimental design will allow us to compare same driver's evaluation of the two different car concepts for a given use case and reduce individual differences. We would like to test 40 participants in the evaluation experiment. Participants will be randomly assigned to each experimental group and the groups will be matched in terms of gender, age, and experience. No prior experience with automated vehicle is a requirement.

Two types of data will be gathered:

- Objective measurements: they concern driver performance and trajectory control. This data will be calculated with recorded simulator log files.
- Subjective measurement: questionnaires to evaluate acceptance, mental workload, usability, and trust. In total, three sets of short questionnaires will be used:
    - (1) an introductory questionnaire before the experiment to collect demographic data, driver behavior – such as driver style, habits of secondary tasks, etc..

(2) intermediate questionnaires to make comparison between baseline car and TeamMate car about the KPIs to be measured

(3) final questionnaire more focus on the feeling of the driver about TeamMate vehicle and the intention to buy.

4.5 VED vehicle

4.5.1 Scenario and uses cases

The VED vehicle demonstrator will be evaluated by considering the MARTHA scenario.

The MARTHA scenario has been identified since it is representative of a limit of the automation: in case of roadworks, the automation may not be able to handle the lane change in Automated Mode.

As a consequence, the automated vehicle may unexpectedly handover the control to the driver (the so called "disengagement") and this situation could represent a safety critical condition for the driver (as already explained in the previous sections).

In order to improve the efficiency of the maneuver, and avoid the disengagement, the automation can ask for support to the driver (H2A in action).

H2A in action was selected in order to demonstrate how human can support the automation when the automation reaches its functional limits. The support in action implies that one of the team member needs direct intervention by the other for a safe driving.

While the H2A use cases selected so far (for EVA and PETER) describe a support in perception, and thus are linked to efficiency, trust and acceptance issues, the H2A in action is also particularly relevant for the safety of the driver, because without his/her intervention, the TeamMate car is not able to continue driving in Automated Mode and it has to perform either a disengagement or a safe maneuver to stop the vehicle.

Moreover, the ability to inform the driver 1 km before the roadworks (enabled by the V2I communications) is a sort of A2H support within this use case.

The MARTHA scenario is also relevant for the safety because it considers a use case where Martha is distracted, and she needs the support of the automation to guarantee her safety.

The following text provides a simple story (adapted from the MARTHA use cases) to intuitively describe the scenario for the evaluation of the VED vehicle demonstrator.

H2A support in action

The TeamMate car is driving in an extra-urban road in Automated Mode. Through the V2I communication, it detects that there are roadworks in 1 kilometer. Since the TeamMate car knows that it will not be able to deal with this situation autonomously, it requests Martha, 1 km before the roadworks, a cooperation in action: in particular, it asks Martha to handle the control of the vehicle. Martha is attentive, and she takes over the control until the end of the roadworks, when the TeamMate car can shift back to Automated Mode.

4.5.2 Baseline

For the H2A use case, the evaluation is aimed at demonstrating the added value of the driver, thus the baseline is the driverless approach (i.e. the autonomous driving without any intervention of the driver)

4.5.3 Experimental protocol

An experimentation will be carried out in VED real autonomous vehicle. The objective is to have a representative sample of naïve drivers in order to evaluate the benefits of both TeamMate approaches: when the driver supports the automation, and when the automation supports the driver.

A familiarization phase is necessary at the beginning of the experiment in order for the drivers to get used to vehicles dynamics and rehearse all transitions (Auto <-> Manual).

Each driver has two main driving sessions (one corresponding to the baseline and the other with the TeamMate car). After each driving session, the driver is asked to answer a set of questions to complete the subjective evaluation that will be also used to justify and elaborate the objective results (driver's behaviour and performance).

Two types of data will be gathered:

- Objective measurements: they concern driver performance and trajectory control. In fact, some metrics will be calculated with recorded vehicle CAN data.
- Subjective measurement: questionnaires to evaluate acceptance and trust. In total, three sets of short questionnaires will be used:
  - (1) an introductory questionnaire before the experiment to collect demographic data, driver behavior – such as driver style, habits of secondary tasks, etc..
  - (2) an intermediate questionnaire to make comparison between baseline car and TeamMate car

(3) final questionnaire more focus on the feeling of the driver about TeamMate vehicle and the intention to buy.

4.6 CRF vehicle

4.6.1 Scenario and uses cases

Exactly as for the REL case, the CRF vehicle demonstrator will be evaluated by considering the EVA scenario, which takes place in peri-urban traffic context. The reason is that the CRF vehicle is the implementation of the same scenario in real world: the ego-vehicle passes through a pre-defined test-site (peri-urban roads), in which several roundabouts are present. The traffic flow near the roundabout may vary depending on some external factors, such as time of the day and weather conditions. In order to improve the efficiency of the trip, the automation can ask for support to the driver (either in perception or in action), explaining why and what is required. In addition, such behavior of the TeamMate car can be also relevant for the safety of the driver, because s/he is more ready to support the vehicle by shared control compared to the usual take-over request (TOR), which necessitates recovering full manual control.

Therefore, the use cases for the support of the driver to the automation (H2A in perception and in action) have been selected to evaluate the added value of the approach (i.e. the cooperation) for the CRF demonstrator. In particular, we have two steps in the H2A support:

- The first one is represented by the "cooperation in perception", where the automation needs the support in perception from the driver, that is, the driver has to decide when it is possible to enter the roundabout.

- The second step (consequent to the previous) is represented by the "cooperation in action", where the driving task is shared between the human-agent and the machine-agent, that is, driver is in charge of lateral control, while longitudinal control is under system responsibility.

These types of support mean that the automation needs a help to continue the driving task without a complete transition of control to the driver. This is relevant because the transition of control from the automation to the driver (i.e. the so called "disengagement") represents a highly critical condition for the interaction between the driver and the automation, for it requires the driver to be promptly brought back in the loop while he/she is likely to be performing non-driving tasks, as it is likely to be the case in the current SAE-L3 of automation. Moreover, this is particularly relevant for the evaluation, since it can be considered the most effective example of cooperation by reducing the number of disengagements.

The following text provides a simple story (adapted from the EVA use cases) to intuitively describe the scenario for the evaluation of the CRF demonstrator:

H2A support in perception and action.

The TeamMate car is driving in Automated Mode. When it approaches a roundabout, the car knows that it cannot deal with, due to the lack of lanes on the road [this situation is very common for the roundabout in Italy, in the Orbassano surrounding]. In order to avoid a disengagement of the automated driving function every time and a consequent TOR, the TeamMate car asks Eva for a cooperation in perception first (to check the available space and to provide a trigger to start the maneuver) and then a cooperation in action (she takes care of the lateral control, while the system is in charge for the longitudinal control). Eva checks the traffic and gives the confirmation to enter *the roundabout. The TeamMate car understands the feedback and enters the roundabout in Shared Mode in collaboration with Eva.*

4.6.2 Baseline

The selection of the H2A use case as the most relevant for the evaluation (to demonstrate the added value of the cooperation in the EVA scenario) also affects the definition of the baseline for the CRF demonstrator. Since the demonstrator is aimed to show the value of the driver to support the automation, the baseline is represented by a condition where the driver has no role in the cooperation (i.e. the so called “driverless” approach): therefore, the baseline is the automated driving without any support of the driver, for which, whenever the automation reaches its limits, a TOR is issued to the driver and a disengagement occurs. Furthermore, during the evaluation phase, we will consider also the situation of “normal driving”, to get a more complete picture of the possible benefits that the TM concept can bring.

4.6.3 Experimental protocol

As reported in Table 7, mostly comfort- and acceptability-related KPIs will be measured; for safety-critical aspects, we will consider the number of TOR to the driver (assuming that higher the number, higher the risk and minor the acceptance). To measure them in the real-world environment, a within-subject design will be considered: the same subject will be asked to perform the scenario with the baseline and with the TeamMate system, randomized in order to make the dataset more consistent and to avoid bias. In addition, we to have a third test condition in manual mode might be considered to increase the robustness of the experimental approach.

Around 10-12 drivers are expected to participate in the the experiment. It is worth noting that a special driving license is obligatory to drive the CRF prototype vehicle on public roads. In addition, no previous experience with automated or semi-automated vehicles will be requested. In Eva scenario both objective (e.g. TTC) and subjective (e.g. trust and acceptance) indicators will be used. The users will be asked to drive in the peri-urban scenario, performing all the roundabouts found on the test-site (around 21 in the current configurations). More details about the implementation of the scenarios and experimental setup will be provided in D6.2.

5 Conclusions

The evaluation plan has been defined for all demonstrators by starting from the concept of the project in order 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.

According to the concept, two approaches have been considered for the evaluation:

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.

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

Then, for each demonstrator, specific KPIs have been defined to measure the performance of the TeamMate car against its baseline, and a preliminary experiment design has been defined. The results of the 2nd cycle evaluation will provide input and feedback to the KPIs that will be adopted in 3rd cycle evaluation.

References

Brooke, J. (1996). SUS - A quick and dirty usability scale. *Usability Evaluation in Industry*, *189* (194), 4–7.

Davis, F.D., (1986), *A technology acceptance model for empirically testing new end-user information systems: Theory and results*, PhD thesis, MIT

AutoMate (2017). "*D4.4 TeamMate HMI design, implementation and V&V results from 2nd cycle*". Submitted 31/12/2017

AutoMate (2017). "*Definition of framework, scenarios and requirements incl. KPIs & Baseline for 2nd cycle*". Submitted 31/08/2017.

AutoMate (2017). "*D1.1. Definition of framework, scenarios and requirements*". Submitted 11/01/2017.

Hart, S.G. & Stavenland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Human Mental Workload* (Eds. P.A. Hancock & N. Meshkati), pp. 139-183.

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

Kyriakidis, M., Happee, R., & de Winter, J.C.F. (2015). Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. *Transportation Research Part F, 32*, 127-140.

Mcknight, D. Harrison, et al. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems,* 2.2 (2011): 12.

SAE J2944 (2015). *Operational Definitions of Driving Performance Measures and Statistics, J2944_201506*. SAE International.

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*, 1-10.