D2.5 – Metrics and Experiments for V & V of the driver, vehicle and situation models in the 3rd cycle

Project Number:	690705
Classification **Deliverable No.:**	Public D2.5
Work Package(s): **Document Version:** **Issue Date:** **Document Timescale:** Start of the Document: Final version due:	WP2 Vs. 1.0 Month 26 Project Start Date: September 1, 2016 Month 24 Month 26
Compiled by:	Fei Yan, ULM
Authors:	Fei Yan, ULM Adam Knapp, BIT David Kaethner, DLR Mark Eilers, HMT Elham Fathiazar, HMT Alain Giralt, CAF Paulin Pekezou Fouopi, DLR
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
21.08.2018	Deliverable structure	F. Yan (ULM)
24.09.2018	Update the deliverable structure	F. Yan (ULM)
02.10.2018	Introduction and update the deliverable structure	F. Yan (ULM)
10.10.2018	Inputs for section 3.2.1 from BIT	A. Knapp (BIT)
17.10.2018	Inputs for section 3.1.3 from DLR	D. Kaethner (DLR)
19.10.2018	Addition of input for sections 3.1.2 and 3.2.3, collected from ULM, OFF, and HMT	M. Eilers (HMT), F. Yan (ULM)
22.10.2018	Inputs for section 3.1.1 from CAF	A. Giralt (CAF)
25.10.2018	Executive summary and conclusion	F. Yan (ULM)
26.10.2018	Input for Semantic Enrichment Sub-Module	P.P. Fouopi (DLR)
31.10.2018	Modification based on the internal review feedback	F. Yan (ULM)

[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 Figures

List of Tables

1 Executive Summary

This deliverable D2.5 is an update of the previous D2.1 and D2.3, focusing on the definitions of metrics and experiments for Validation and Verification (V & V) in the 3rd cycle. In project cycle 1, WP2 has defined a process of the model validation as well as also the metrics and experiments of the components for verification, which has been first agreed on and documented in D2.1. The first results of experiment for V & V of the driver, vehicle and situation models has been reported in D2.2. Following the defined process of the model validation in D2.1, the definition of metrics and experiments of the components for verification as well as the results of experiment for V & V in project cycle 2 has been updated in D2.3 and D2.4 separately.

In this deliverable, the metrics of the driver, vehicle and situation models are defined for the 3rd project cycle. Moreover, the planned experiments for verification and validation of these models are presented to measure the metrics. Thus, this document delivers an enhanced version of the components with regard to the driver, vehicle and situation models for Milestone 5, and further contributes to the successful V&V on an integration level in simulators and in real vehicles (Milestone 6) in project cycle 3.

2 Introduction

WP2 delivers the components with regard to the driver models and vehicle & situation models for the TeamMate Car, to monitor, understand, assess and anticipate the driver, vehicle and traffic situation. Following the four-steps process of the model validation, all involved partners have defined the first version of the metrics for verification and validation and also performed first experiments with their components for V & V in project cycle 1. In project cycle 2, partners have updated the defined metrics and experiments for V & V of their models and began to integrate them in the demonstrators.

To reach the milestones M5 and M6 in project cycle 3, the successful V&V both on a component level and also on an integration level in simulators and in real vehicles is needed in this cycle. Accordingly, WP2 is planned to deliver the enhanced and improved versions of these components for M5. Also, the implementation and validation of the components need to be implemented and integrated in a real vehicle in project cycle 3. For this, the metrics and the planned experiments of the driver models, as well as vehicle and situation models in the last period, need to be defined.

This document composes of the chapters of introduction, conclusion, as well as the metrics and experiments of driver, vehicle and situation models which are divided into two parts: driver models and vehicle and situation models. In the section with regard to the driver models, the defined metrics and the planned experiments to measure these metrics (driver state model, probabilistic driver models for intention recognition, and task model for driving) have been updated based on D2.1 and D2.3. In particular, in the section of the oncoming experiments in project cycle 3, we also address the link between the planned experiments of the models and the corresponding demonstrators in AutoMate project. Additionally, concerning vehicle and situation models (V2X communication, Semantic Enrichment Sub-Module (SESM), future situation evolution), the metrics and oncoming tests in cycle 3 have been updated. It provides the ground work for the following integration of the components in demonstrators in project cycle 3.

3 Metric and Experiments of Driver, Vehicle and Situation Models

Following the defined four-step process of the model validation in D2.1, this chapter focuses on the updates of the detailed metrics for driver, vehicle and situation models in project cycle 3. For each model, first a description of the model will be given, followed by the defined metrics and the planned experiments to measure these metrics.

3.1 Driver Models

In this section, the description, the defined metrics and the planned experiments of driver models (driver state model, probabilistic driver models for intention recognition, and task model for driving) are presented.

3.1.1 Driver State Model

3.1.1.1 Model Description

The driver's state model includes two main state models: Drowsiness, and Visual attention and Visual distraction (On-road, Off-road). CAF offers a driver state system and a driver state software component.

- The driver state system includes a mono camera and a software module which compute the models.
- The software component computes the models output from the raw data provided by the SmartEye face tracker.

Besides the model outputs, the driver state system output intermediate data such as eye opening, head gaze, eye gaze, ID of the area the driver observes (cluster, mirrors, etc.), etc. These data are of interest for the validation of the overall system but also for further research on understanding visual strategies and physiological models.

The Driver State System is integrated in the Automate demonstration vehicles and the Relab simulator while the Driver-State SW component is integrated in the ULM and Vedecom simulators. The left image of **Figure 1** shows the position of the camera (blue rectangle) integrated in the Vedecom demonstration car. The right image of Figure 1 shows the camera image and the tracking markers in green overlay.

Figure 1: Driver state camera integrated in the Vedecom demo car.

Figure 2: Camera integrated in the CRF demo car.

3.1.1.2 Metrics

The metrics previously described in the deliverables D2.3 "Metrics and Experiments for V&V of the models in the 2nd cycle" will be used for the 3rd cycle validation.

These metrics can be summarized in the following main ones:

- Output availability:
  - Ratio of the time the model quality is high enough.
- Detection rate of events.

    - Ratio of the numbers of true events over the numbers of ground truth events.
  - Number of false events per hour.

An event is defined as a state model detection as for example the driver is detected drowsy or the driver is detected distracted.

For the third cycle in addition to these generic metrics we will add timing metrics and scenario specific metrics:

  - System response time between the event output and the event ground truth.
  - Detection rate of Automate scenarios.

3.1.1.3 Upcoming Experiment/Test (3rd Cycle)

For the third cycle CAF will perform a set experiments with the following objectives:

  - Validation of generic scenarios defined in previous cycles in demonstration vehicles.
  - Validation of specific Automate scenarios in demonstration vehicles (e.g. The driver is distracted because is reading a message on his/her smart phone placed on the knees).

Within the two cycles, the driver state system was validated in laboratory and in car condition considering an optimal position of the camera placed behind the steering looking at the driver face through the steering wheel. This optimal camera position couldn't be achieved in the Vedecom demonstration car because it would have been too intrusive by occluding some mandatory information displayed on vehicle screens. It is then necessary to perform validation tests of the generic scenarios and specific to the demonstration vehicle in the demonstration vehicle.

As one can see on the Vedecom vehicle the camera was placed in between the two frontal displays about 15 cm to the right and above of about 12 centimetres compared to the nominal position (cf. Figure 2). Consequently, it is expected some performances degradation when the driver turns at the opposite of the camera position, that is left or downward.

Tests will be performed at Vedecom place at Satory. Drivers will be asked to perform the generic protocol defined in the previous cycles and the specific Martha scenario protocol. This data set will be completed by a dataset of drivers collected at the CAF vehicle using a camera placed at about the same position that the Vedecom one.

Because the position of the camera in the CRF demonstration vehicle (see Figure 2) is the same that the position of the camera in the CAF test vehicle, the Eva scenario will be mostly validated in the CAF test vehicle.

3.1.2 Probabilistic Driver Models for Intention Recognition and Behaviour Prediction

3.1.2.1 Model Description

The probabilistic driver model for intention recognition and behaviour prediction is a Dynamic Bayesian Network (DBN) that attempts to model the causal and statistical relationships between a driver's intentions, driving manoeuvres and behaviours, as well as the situational context, as observable by the TeamMate vehicle's sensor platform. For a thorough description of the model, we refer to Deliverable 2.4 "Sensor Platform and Models including V&V results from 2nd cycle".

During the second cycle of AutoMate, the development of the model has been focussing exclusively on the Peter scenario, dealing with overtaking scenarios on rural roads. The goal and purpose of the model in the Peter scenario is to constantly provide the TeamMate vehicle with a probabilistic belief of the current overtaking intentions of the driver of the TeamMate vehicle. The information is intended to be used as follows: if the driver is in control of the TeamMate vehicle (manual driving), the information provided by the model can be used to assess the safety of the intended driving manoeuvre. If the automation is in control (autonomous driving), the information provided by the model can be used to propose appropriate manoeuvres to the automation that reflect the potential intentions of the driver in the given situation. Within the second cycle, the enabler has been integrated in the ULM demonstrator and was evaluated as described in deliverable D6.2 "Results of Evaluation 2nd Cycle". Based on the evaluation results obtained in the second cycle, we will try to improve the performance of the driver intention recognition in the Peter scenario during the third cycle.

To extend the area of application of intention recognition in the third cycle of AutoMate, the driver model will be extended to address intention recognition in the Eva scenario. In this scenario, "the Automated mode hesitates to enter the roundabout with high traffic flows and ask for support of Eva to check the available space and to provide a trigger to enter the roundabout. 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."

For the Eva scenario, the driver model shall attempt to recognize the intention of the driver in entering the roundabout and subsequently provide the proper trigger input in Automated Mode to efficiently enter to the roundabout. Specifically, the model will attempt to recognize the following intentions of the driver: whether the driver intend to wait at the entrance of the roundabout (Pause) or intends to enter the roundabout (Go). At each time point (t), the model will be used to infer a probability distribution over entrance intention and driving behavior, given all available sensory inputs from state of the Teammate vehicle and from the surrounding vehicles. Especially the vehicles inside the roundabout which are in the view field of the human driver will be considered to estimate the traffic flow in the roundabout. In real world scenarios this information could be obtained from stereo cameras (Maximilian et. al., 2012).

Furthermore, we will integrate the enabler to the VED real vehicle demonstrator, which requires the adaptation of the enabler to motorway scenarios.

3.1.2.2 Metrics

Concerning validation, we will rely on the same methods as previously described in the deliverables D2.3 “Metrics and Experiments for V&V of the driver, vehicle, and situation models in the 2nd cycle” and D2.4 “Sensor Platform and Models including V&V results from 2nd cycle”, in that the models for intention recognition will be validated using test sets D_{Test}, consisting of manually annotated time-series representing the ground truth. Utilizing the models on the test sets, we will construct binary confusion matrices as shown in the following table, where we interpret the existence of an intention as positive and the absence as negative.

		Ground Truth	
		Positive	Negative
Predicted	Positive	TP	FP
	Negative	FN	TN

Table 1: Binary confusion matrix. In the case of driver models for intention and behavior recognition in AutoMate, the ground truth is based on a manual annotation of test data.

From the confusion matrix, we then derive the following set of metrics, summarizing different aspects of the performance of the model:

- The *accuracy*, representing the fraction of correctly recognized samples among all samples, defined as

$$ACC = \frac{TP + TN}{TP + FP + FN + TN}.$$

- The *precision*, representing the fraction of correctly recognized intentions among all predicted intentions, defined as

$$Precision = \frac{TP}{TP + FP}.$$

  A high precision indicates that the model only recognizes intentions if there actually exists an intention.
- The *recall* (also known as sensitivity or true positive rate (TPR)), representing the fraction of correctly recognized intentions over the total amount of true intentions, defined as

$$Recall = \frac{TP}{TP + FN}.$$

  A high recall indicates that most of the intentions are recognized as such.
- The harmonic mean of precision and recall, the traditional F-measure or balanced *F-score*, defined as

$$F\text{-}score = 2\frac{Precision \cdot Recall}{Precision + Recall}.$$

- And, for the sake of completeness, the *False Positive Rate* (FPR), defined as

$$FPR = \frac{FP}{FP + TN}.$$

For the third cycle, we will additionally extend the aforementioned metrics for validation by performance metrics that assess the efficiency of the driver intention recognition in terms of computational complexity. For this, we will rely on analytical and empirical approaches. For the analytical approach, we will analyse and report the relationship between the size of the input data and abstract execution time, by estimating the algorithmic complexity in the asymptotic sense, using the Big O notation (e.g., linear or exponential), when varying the complexity of the underlying model. For the empirical approach, we will use the validation data to measure and report the real running time of intention recognition on exemplary systems. If possible, we will repeat this analysis for the intention recognition when integrated in the target demonstrators.

3.1.2.3 Upcoming Experiment/Test (3nd Cycle)

[redacted]Peter scenario

To collect data for training and validation of the driver model for intention and behaviour recognition in the 3rd cycle, the AutoMate partners ULM, OFF, and HMT conducted an experiment for the Peter scenario in the OFF driving simulator. The goal of this experiment was to obtain datasets necessary to improve the driver models for intention recognition and behaviour prediction with a better distinction between the formation of intention to overtake and the execution of an overtaking manoeuvre.

3.1.2.3.1.1 Participants

23 participants with valid German driver license were recruited from the university of Oldenburg. As one of them experienced motion sickness in the very beginning, the left data from 22 participants were valid. Participants (11 males and 11 females) had an average age of 25 years old (SD=6.1) and were licensed on average for 8 years (SD=5.7). They received a compensation of 12 Euro for their 1.5-hour participation.

3.1.2.3.1.2 Apparatus and Materials

The experiment was conducted in the research driving simulator at OFF. The driving simulator is a fixed-based simulator platform, visualizing a maximum field of view of 150 degrees via three beamers (see Figure 3). Two displays with a resolution of 1024*768 pixels are used to simulate the left and right exterior mirror. To apply adjustable force feedback and vibration signals on the steering wheel as well as accelerator and brake pedals, three Lexium Schneider CAN bus servo drives are used. For creating the road geometry, landscape, and traffic scenario, the simulator software SILAB 6.0 is used. To collect input of participants, two black buttons are installed on the left and right side of the steering wheel (see Figure 3).

Figure 3: The participant drove in the OFFIS driving simulator.

3.1.2.3.1.3 Scenario and Traffic Situations

A two-lane (one for each driving direction) German rural road track with a general speed limit of 100 km/h, initially designed by the project partner ULM, was simulated with SILAB for this experiment.

In order to study driver intention to overtake, various traffic situations were triggered by varying relevant factors that potentially influence driver intention:

(1) The type of lead vehicle: The probability of the appearance of a truck is 0.2, while the probability of the appearance of a passenger car is 0.8.
(2) The speed of the lead vehicle: The speed of the passenger car varies from between 65-100 km/h and the speed of the truck is between 70-80 km/h.
(3) The numbers of the oncoming vehicle: no oncoming vehicle; one oncoming vehicle; platoons of (four) oncoming vehicles.
(4) The difference of the time to collision (TTC) among four oncoming vehicles is varied between 6-12 s.

In total, a set of 70 traffic situations were defined to investigate the formation of intention to overtake and the execution of an overtaking manoeuvre.

3.1.2.3.1.4 Procedure

After reading the handout of the instruction and filling the consent form, participants were brought to the driving simulator. Participants drove on rural roads with a general speed limit of 100 km/h. At the beginning of the scenario, a lead vehicle would appear in front of the ego vehicle, when the speed of participants reached 100km/h. Participants were asked to use the left or right button installed on the steering wheel to show their overtaking decisions: Once participants want to overtake the lead vehicle independent of the possibility, they should press the left button once on the steering wheel; once participants don't want to overtake the lead vehicle, they should press the right button on the steering wheel. After pressing the left button, participants overtook the lead vehicle when it was possible for overtaking and then came back to the right lane, which was followed by a new trial. After pressing the right button, all the relevant vehicle would disappear, and the next trial began.

In order to let them get used to the vehicle control in the driving simulator, a training session was begun with 11 trails from the original experiment scenario. These 11 trails were selected to cover the possible overtaking situations, aiming to avoid learning effect for the actual experiment. After the training session, the first session with 35 overtaking situations were presented. After a 5-minute break, the second session with also 35 situations followed. After these two sessions, participants were asked to fill in an online questionnaire regarding experiment and also their demographic data. In the end, they were paid 12 euros for their participation. For each subject the study took around 90 minutes.

3.1.2.3.1.5 Resulting datasets

The test set D_{Test} will be obtained from the experimental data conducted for training and evaluation of the probabilistic driver models in the first cycle. More specifically, the experimental data will be split into a training set D_{Train}, including approx. 70% of the experimental data, and a test set D_{Test}, including the remaining experimental data. The driver model will be learned exclusively using the training data D_{Train} and subsequently validated on the test set D_{Test}. For this, both data sets need to be manually inspected by experts to annotate each data sample with the currently assumed intention and shown driving manoeuvre resp. behaviour of the driver.

[redacted]Martha scenario

No dedicated experiments are planned for the adaptation of the driver model for intention recognition in motorway scenarios (Martha) in the third cycle. Instead, we will rely on exemplary real world datasets provided by VED to manually adapt the model to the new scenario.

[redacted]Eva scenario

For the development and evaluation of the driver model for intention recognition in roundabout scenarios (Eva) in the third cycle, we will consider datasets collected during evaluation experiments in WP6 (c.f. D6.2 "Results of Evaluation 2nd cycle", Section 3.1.2). Additionally, further experiments will be performed individually by REL and OFF, using similar scenarios as described in D6.2, but covering a more variant traffic flow inside the roundabout to insure a better generality of the model. These datasets will be manually annotated by experts according to the behavior of the driver, e.g., the behavior "Enter" will be manually selected as the point in time where the human driver accelerates to enter the roundabout, and the intention "Enter" will subsequently automatically annotated to be existing at some constant time before the acceleration.

3.1.3 Task Model for Driving

3.1.3.1 Model Description

The task model for driving is a framework to model specific scenarios during the execution of the driving task, or specific interactions between a driver/user and a driver assistance system. It is not generative (i.e., does not produce a behaviour like a software might do), but allows for relatively quick and cheap modelling of scenarios of interest. The resulting models can be used to broaden the understanding of the scenario of interest by the stakeholders, inform HMI design decisions, and also serve to communicate the assumptions one might have about any such scenario and the involved drivers' information processing.

3.1.3.1 Metrics

Deriving actual metrics from/for task models is usually not done, since its purpose is to break down the task in chunks in order to gain a better understanding of the task itself (e.g. Ormerod & Shepherd, 2003; Fastenmeier & Gstalter, 2007). However, there are two interesting metrics we will try to produce: The intra- and intersubject similarity between different overtaking maneuvers (i.e., within and between subjects), and the similarity of models as produced by different modellers (analogue to an interrater reliability).

3.1.3.2 Upcoming Experiment/Test (3nd Cycle)

To collect data for validation of the driving task model for goals, and cognitive, perceptual and motor operators, the AutoMate partner DLR will conduct a study for the Peter scenario in a DLR driving simulator (see Figure 4, showing an impression of the road and scenario). A similar (but smaller) study had been conducted already by the AutoMate partner ULM, specifically to gather data for the model construction. These data had been very helpful to reach a first impression and model the rough layout of the task.

Figure 4: The participant drove in the DLR driving simulator.

However, more validation regarding the approach and the resulting models has been deemed necessary by the project reviewer at the Midterm Review

Meeting, voicing concerns that the resulting models and their underlying constructs should receive better validation. Therefore, the goal of the study is to collect data to a) construct models of the overtaking and b) thereby achieve a validation both of the postulated goals and operators.

Since the construction of the model is rather time consuming, the study is conceptualised as a case study with ca. 10 subjects who will be fully informed of the study's aim. They will drive a scenario which is very similar to the ones at the simulators of ULM and OFF. Data will be collected from an eye tracker, a recording of the subject's *thinking aloud* during the drive (voicing their current plans and goals according to an instruction), and driving simulator data such as position, velocity or the commandeering of the overtaking maneuver using the indicator lever.

3.2 Vehicle and Situation Models

In this section, the description, the defined metrics and the planned experiments or tests of vehicle and situation models (V2X communication, Semantic Enrichment Sub-Module (SESM), future situation evolution) are presented.

3.2.1 V2X Communication

This subsection presents the test scenario of the V2X communication components (ITS-Stations) including the metrics used for the validation.

3.2.1.1 Model Description

One ITS-S (on-board unit) is integrated in Vedecom's vehicle and it is responsible for the proper reception and processing of Decentralized Environmental Notification Message (DENM). The other ITS-S (road side unit – RSU) assembles and broadcasts the DENM message containing the road works warning (RWW) in the relevant geographical area.

3.2.1.2 Metrics

The geolocations of vehicle and the incoming V2X messages are logged with timestamps during the field tests.
The following metrics are defined.

[redacted]Communication distance

The first reception of DENM can be easily detected based on the logs, and from that position the distance can be calculated. The communication distance depends on the environment. The following table contains the threshold values that should be reached during the experiments.

Type of area	Threshold [m]
Dense urban	200m
Suburban	300m
Rural	400m

Table 2: The threshold values used in the experiments.

Related requirements from D1.5: R_EN1_tool2.1, R_EN1_tool2.2, R_EN1_tool2.3.

[redacted]Reliability of communication

Since the content of the DENM messages are constant in the given scenario, i.e. the same messages are broadcasted by the RSU over and over, the content have to be the same at the receiver side as well. It is important to verify this, because otherwise it can confuse the vehicle's other systems, i.e. the confidence level will decrease. Related requirements from D1.5: R_EN1_tool2.1, R_EN1_tool2.2, R_EN1_tool2.3.

3.2.1.3 Upcoming Experiment/Test (3nd Cycle)

Beside the experiments of the Vedecom vehicle, field tests will be carried out in Budapest using two traditional cars equipped with ITS stations. The initial distance of the cars is around 600 metres. One car behaves as a normal car approaching a roadwork area. The other vehicle does not move and behave as an RSU broadcasting the RWW message. The message is broadcasted in a 500m area around the RSU carrying car. The message contains 30 km/h speed limit for the roadwork area and in its 100m proximity (i.e. around the vehicle). It also informs the approaching car about the closure of the most right lane that is 50m around the RSU carrying car. The following figure (see Figure 5) illustrates the test setup.

Security features for the communication are turned off. The positions and the messages are logged with timestamps as mentioned, and the logs are processed offline after the tests.

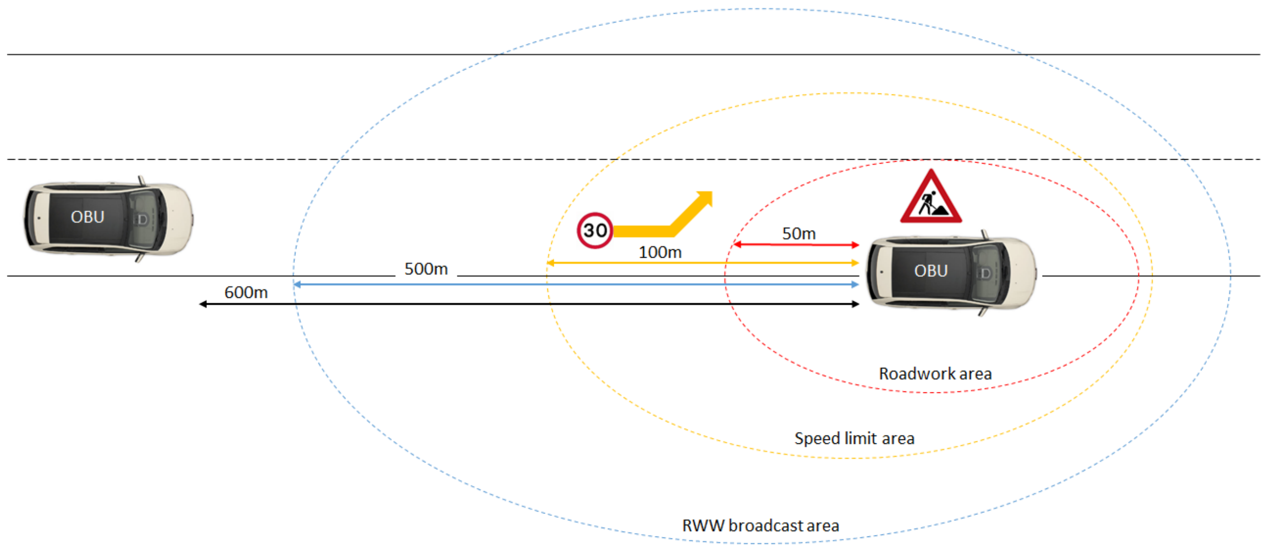

Figure 5: Test scenario for V2X communication

3.2.2 Semantic Enrichment Sub-Module (SESM)

3.2.2.1 Model Description

To interpret the situation around the ego-vehicle, automated driving systems need to extend the objects provided by the perception layer with new information. For this purpose, semantic models can be used to model further attributes of the objects from the perception layer, the relations between those objects as well as logical rules, can be useful for understanding driving situation. Furthermore, inference algorithm can be used to infer new information about the objects given the semantic model and some evidence. The SESM therefore consists of a semantic situation model and an inference engine. The semantic situation model contains following elements:

1. Concepts: scene elements, e.g. traffic participants (vehicle, pedestrian, cyclist, etc.), road markings and traffic infrastructures (traffic light and traffic sign).
2. Relations between the concepts, e.g. "vehicle is on lane", "traffic light has state red"
3. Traffic rules modelled as logical rules. These logical rules represent the permissible maneuvers of traffic participants given the concepts and they relations.

The inference engine reasons about the permissible traffic participants maneuvers given the modelled situation and the evidence from a scene. The traffic participants permissible maneuvers can be combine with the dynamic of the traffic participants to predict the future evolution of the situation. The main advantage of the SESM is the formal and explicit modelling of the knowledge which allows easy verifying and validating the system. More information on the SESM can be found in the deliverable D2.2 "Sensor Platform and Models including V&V results from 1st cycle".

3.2.2.2 Metrics

For evaluating the SESM-sub-module ground truth data need to be generated. The metrics then measure the difference between the output of the SESM and the ground truth data to estimate the accuracy of the proposed sub-module. The metric for measuring the accuracy in the second cycle was the F_1-score $F_1 = \frac{2}{\frac{1}{recall}+\frac{1}{precision}}$. This metric was introduced in the deliverable D2.3 "Metrics and Experiments for V & V of the driver, vehicle and situation models in the 2nd cycle" and described in the deliverable D2.4 "Sensor Platform and Models including V&V results from 2nd cycle". This metric will be used also for the third project cycle. An acceptable F_1-score on the component level should be in the range of over 90% as specified in the requirement "R_EN3_model1.3" in the deliverable D1.5 "Definition of framework, scenarios and requirements incl. KPIs & Baseline for 3rd cycle". Furthermore, we will evaluate the module based on average time and memory the module needs to process a frame.

3.2.2.3 Upcoming Experiment/Test (3nd Cycle)

As mentioned in the section above, ground truth data will be generated in the third cycle to evaluate the sub-module. The relevant scenarios presented in the deliverable D2.4 "Sensor Platform and Models including V&V results from 2nd cycle" will be extended with complex situations including road with two lanes. These scenarios will be generated in the simulation. Based on the simulation the proposed-sub module will be evaluated. Since the simulation provides scenarios, where all traffic participants follow the traffic rules and the proposed sub-module is a formal model without uncertainties, a F_1-score of 100% will be expected. This F_1-score was reached with some simple scenarios as ground truth. An evaluation with real traffic data will provide a F_1-score less than 100% because of the uncertainties of the perception. Moreover, some traffic participants could violate the traffic rules. Since these sub-module is a prior for predicting the future evolution of the situation, therefore the accuracy of these sub-module in real traffic can be used as weight for the prior.

3.2.3 Future Situation Evolution

3.2.3.1 Model Description

The goal and purpose of the prediction of the future evolution of the traffic situation is to provide the TeamMate vehicle with a temporal-spatial prediction of the location and orientation of other traffic participants. Such a prediction is required by the online risk assessment to derive a safety corridor in which the TeamMate vehicle can maneuver safely and to assess the safety of potential maneuvers.

A detailed description of the underlying models and algorithms for the prediction of the future evolution of the traffic situation has already been provided in the previous deliverables D2.2 "Sensor Platform and Models including V&V results from 1st cycle" and D2.4 "Sensor Platform and Models including V&V results from 2nd cycle" and any repetition will thus be omitted.

3.2.3.2 Metrics

Concerning the validation of the prediction of the evolution of the traffic situation, it is most important that the predicted regions encompass the true (future) location of any predicted vehicle and that such predictions can be provided fast enough to be usable in the context of the TeamMate vehicle. In the following, we will define a set of metrics to measure the performance of the prediction in respect to these goals. We note that any validation must be performed using a set of independent test data D_{Test}, representing ground truth time-series of traffic situations.

In the deliverable D2.3 "Metrics and Experiments for V&V of the driver, vehicle, and situation models in the 2nd cycle" we proposed the concept of the "correct classification rate" as the ratio of correct predictions and the number of total predictions (for a more detailed description and utilization of this metric, we refer to D2.4 "Sensor Platform and Models including V&V results from 2nd cycle"):

$$CR_\delta^i = \frac{\#_s}{\#_s + \#_f},$$

with $\#_s$ representing the sum of successes, $\#_f$ representing the sum of failures, i representing a temporal prediction horizon, and $\delta, 0 < \delta < 1$ representing the coverage probability for the prediction.

The proposed metric has the potential drawback that the correct classification rate can be arbitrarily increased by increasing the volume of the underlying prediction area, e.g., by using imprecise models for prediction or artificially inflating the variance. To account for this drawback during the third cycle, we will therefore include the volume of the prediction area as an additional metric, with smaller volumes, which can be interpreted as more certain predictions, being preferred over bigger volumes.

For the third cycle, we will extend the metrics for validation by performance metrics that assess the efficiency of the prediction of the future evolution of the traffic situation in terms of computational complexity. For this, we will rely on analytical and empirical approaches. For the analytical approach, we will analyse and report the relationship between the size of the input data and abstract execution time, by estimating the algorithmic complexity in the asymptotic sense, using the Big O notation (e.g., linear or exponential), when varying the number of objects and the prediction horizon. For the empirical approach, we will use the validation data to measure and report the real running time of the prediction on exemplary systems. If possible, we will repeat this analysis for the prediction of the future evolution of the traffic situation when integrated in the target demonstrators.

3.2.3.3 Upcoming Experiment/Test (3nd Cycle)

No dedicated experiments will be performed for the prediction of the evolution of the traffic situation during the third cycle. Instead, we will rely on the experimental data obtained from the experiments for the driver model for intention recognition (Section 3.1.2.3) and, where available, real data sets provided by the demonstrator owners for evaluation.

4 Conclusion

In this deliverable, the metrics of the driver, vehicle and situation models are defined for the 3rd project cycle. Besides, the planned experiments for verification and validation of these models are presented to measure the metrics. With these, this document delivers an enhanced version of the components with regard to the driver, vehicle and situation models for Milestone 5, and further contributes to the successful V&V on an integration level in simulators and in real vehicles (Milestone 6) in project cycle 3.

5 Reference

Muffert, M., T. Milbich, D. Pfeifer, and U. Franke (2012). „May I enter the roundabout? A time-to-contact computation based on stereo-vision". In the proceedings of the IEEE Intelligent Vehicles Symposium (IV) 2012, pp. 565-570.