Information Society
Technologies

PReVENT

Preventive and Active Safety Applications
Integrated Project
Contract number FP6-507075
eSafety for road and air transport

Subproject

PReVAL
a PReVENT Project

Deliverable type	SP Deliverable
Deliverable name	D16.3 Proposal of procedures for assessment of preventive and active safety functions

D16.3 Proposal of procedures for assessment of preventive and active safety functions

Version number	1.10
Lead contractor	VTT
Confidentiality status	Public
Due date of deliverable	31.8.2007
Actual submission date	21.12.2007
File name	PR-16100-SPD-070831-v10-D16_3_ProcedureProposal

SIXTH FRAMEWORK PROGRAMME

Authors

Johan Scholliers, VTT

Karsten Heinig, University of Hannover

Jean-Marc Blosseville, LCPC

Mariana Netto, LCPC

Virpi Anttila, VTT

Susanna Leanderson, Volvo Technology

Johan Engström, Volvo Technology

Mikael Ljung Aust, Volvo Car Corporation

Falke Hendriks, TNO

Jeroen Ploeg, TNO

Jian Chen, ika RWTH Aachen

Subproject Leader

Johan Scholliers

VTT

P.O.Box 1300

33101 Tampere, Finland

tel +358 20 722 3642

email Johan.Scholliers@vtt.fi

IP Coordinator

Matthias Schulze

Daimler AG

HPC 050 – G021

71059 Sindelfingen / Germany

Phone +49 7031 4389 603

Mobile +49 160 86 33308

Fax +49 7031 4389 210

E-mail matthias.m.schulze@daimler.com

Revision chart and history log

Version	Date	Reason
0.1	3.4.2007	table of contents, generated
1.0	12.10.2007	version for peer reviewers
1.01	12.11.2007	Comments of peer review integrated.
1.02	12.11.2007	Integration of function description
1.04	12.11.2007	Comments from Susanna Leanderson Chapter 2.2 Chapter 3.3
1.05	11.12.2007	Reworking of section 3.2 (V-shape definition) and section 4
1.06	18.12.2007	JMB and MN : Changes in table 9 (scenarios) and conclusions as well as other corrections - note : changes in ch. 4 completed.
1.10	19.12.2007	Final version

Table of contents

Authors ii

Subproject Leader ii

IP Coordinator ii

Revision chart and history log iii

Table of contents iv

Executive summary 1

1 Introduction 3

2 Methodology 5

- 2.1 Technical evaluation procedure 5
- 2.2 Procedure for human factors evaluation 5

3 Framework for the evaluation of safety functions 7

- 3.1 Situational Control concept 7
- 3.2 Adapted V-shape design cycle 11
- 3.3 Relation between technical, human factors and safety potential evaluation 13

4 Proposal for technical evaluation procedure 18

- 4.1 Assessment framework proposal 18
- 4.2 Evaluation specification 19
  - 4.2.1 System and functions description 20
  - 4.2.2 Assessment Objectives 24
  - 4.2.3 Expected system impacts 25
  - 4.2.4 Key indicators 25
- 4.3 Test Definition 29
  - 4.3.1 Synthetic Scenario 29
  - 4.3.2 Tools and methods 31
  - 4.3.3 Data collection and measurement plan 33
  - 4.3.4 Reference case 34
- 4.4 Execution and reporting 35
  - 4.4.1 Verification 35
  - 4.4.2 Validation 36
- 4.5 Best practices 36
- 4.6 Conclusions and Final Remarks 37

5 Human Factors Evaluation procedure 39

- 5.1 General approach 39
  - 5.1.1 A framework for ADAS human factors evaluation 39
- 5.2 Procedure 42
  - 5.2.1 Overview 42

5.2.2 Step 1: System and functions description 44
5.2.3 Step 2: Hypothesis generation 45
5.2.4 Step 3: Test scenario selection 50
5.2.5 Step 4: General selection of methods 52
5.2.6 Step 5: Detailed study design 60
5.2.7 Detailed study design of methods 60
5.2.8 Step 6: Reporting and interpretation 64
6 Safety Potential Assessment Procedure 66
6.1 The behavioural effect approach 66
6.2 The procedure of the methodology 68
6.2.1 Description of the functions 68
6.2.2 Literature survey on the effects of the functions 68
6.2.3 Relevant safety impact mechanisms 69
6.2.4 Quantitative assessment of the impacts 70
7 Conclusions and next steps 73
References 75
Annex 1 Keywords 77
Annex 2 Glossary 78
Annex 3 Dummy targets 82

Executive summary

PReVAL addresses the possible safety impacts of functions developed and demonstrated in the PReVENT integrated project. One of the major aims of the PReVAL project is the development of a harmonized framework for the assessment of preventive safety applications and advanced driver assistance functions.

The work described in this deliverable is a major step towards this objective. In deliverable D16.1 [1] the evaluation methodologies used by the PReVENT subprojects and related projects, such as eIMPACT, APROSYS, AIDE and CONVERGE were reviewed. Based on this work, evaluation procedures are proposed for technical performance and human factors, which are described in this deliverable.

The safety potential of an application is determined by the technical performance, the reliability of the system, the human factors (the interaction between the driver and the vehicle) and the traffic safety level. For the safety impact analysis, the behavioural effect approach, which has been developed and used by the eIMPACT project, is considered.

As a general concept to link the different evaluation procedures, the concept of situational control is introduced. Situational control refers to the level of control jointly achieved by the driver and the vehicle in a specific driving situation. In general, the purpose of preventive safety systems and advanced driver assistance systems could then be understood as the increase of situational control. Consequently, the general goal of the system evaluation is to assess the extent, to which this is achieved.

PReVAL proposes to extend the V-shape design cycle, which is common in the automotive industry, with steps focused on evaluation. These steps include the evaluation specifications and the test definition. In most cases, the system or function specifications are not detailed enough to design the evaluation tests. The evaluation specifications contain all the information needed for design of the tests for the different evaluations. A key part of the evaluation specifications is the function description, which is common for the three procedures.

Chapter 1 of this deliverable situates the deliverable within the PReVAL work. Chapter 2 gives an overview of the methodology used to develop the procedures.

Chapter 3 introduces the concepts common to the different procedures: the situational control concept and the adapted V-shape design cycle.

Chapter 4 describes the proposed technical evaluation procedure. The technical evaluation consists of two phases: the technical verification at component or subsystem level, and the validation or functional assessment at system level, with as main aim to test whether the goals and specifications of the complete system are met. The different phases of the adapted V-shape design and evaluation cycle are elaborated: the functional and technical specifications; the evaluation specifications; test definition for both phases (including test scenario definition, method selection,

measurement plan and reference case), and the actual verification and validation phases

Chapter 5 describes the proposed human factors evaluation procedure. The human factors part of the evaluation focuses on the driver-related impacts on situational control. A human factors evaluation potentially involves all aspects related to the driver's response to the system, including acceptance, usability and (intended and non-intended) behavioural effects that can be assessed empirically. The human factors evaluation consist of 6 steps: (1) the system and function description; (2) hypothesis generation; (3) scenario definition; (4) general selection of methods; (5) detailed study design and (6) reporting and interpretation.

Chapter 6 gives an overview of the safety potential assessment method. This method has been described already in D16.1, but is repeated here, so that the report gives a complete and stand-alone overview of all the proposed procedures. PReVAL uses the behavioural effect approach, which has been developed and is used by the eIMPACT project. The underlying rationale of the selected approach is to assess the impact mechanisms evaluating first how the functions affect driver behaviour and travel behaviour. The relevant impact mechanisms are identified. Starting from the impact mechanisms and accident data, and penetration rate estimations, a quantitative analysis is performed.

Chapter 7 describes the steps to come to a consolidated framework, which will be described in D16.4. "Project final report and recommendations for future assessment"

1 Introduction

PReVAL evaluates the safety impact of PReVENT systems and develops a harmonized evaluation framework for the assessment of preventive safety systems and advanced driver assistance functions

Figure 1 shows the methodology of the PReVAL work:

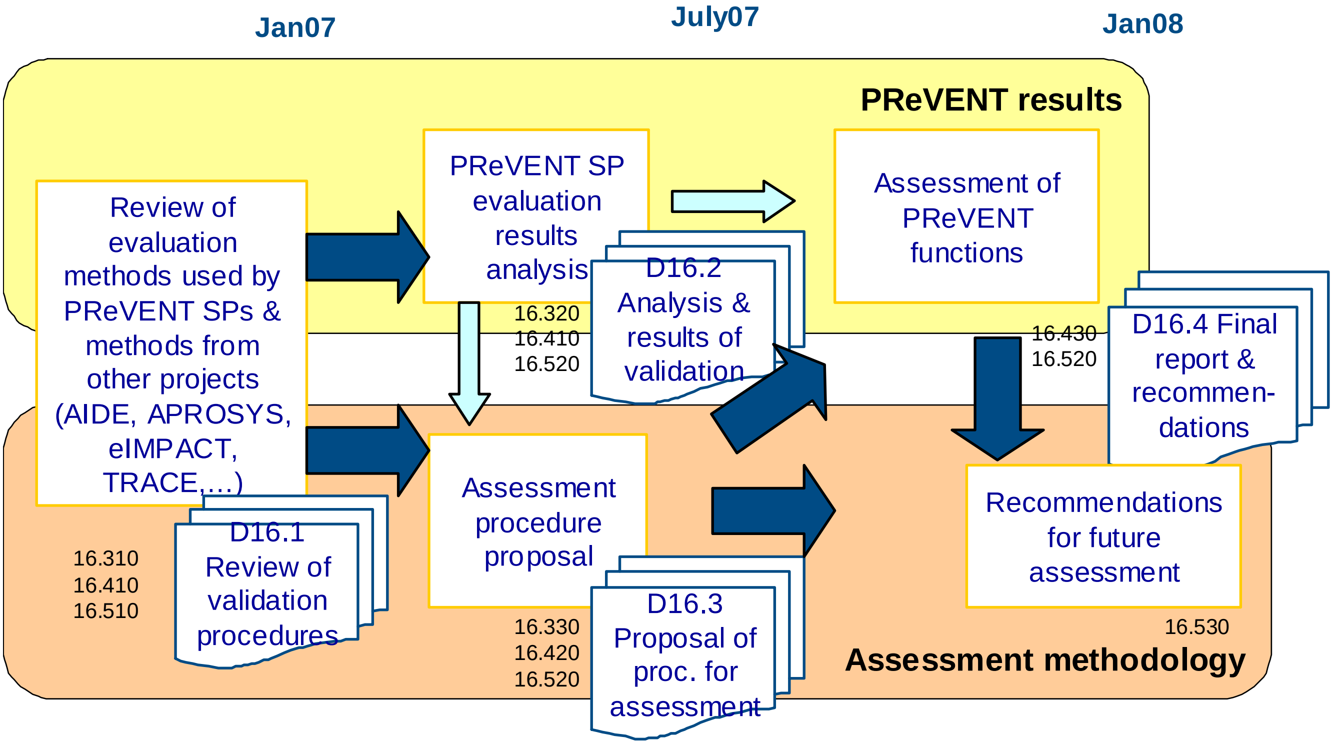


Figure 1: PReVAL project phases

The work in PReVAL is concentrated on two lines:

- the analysis of the PReVENT evaluation results
- the development of a harmonized framework for the assessment of preventive safety functions and advanced driver assistance systems. The work reported in this deliverable is related to the development of the assessment framework.

The development of the assessment framework goes through the following steps:

- review of the evaluation methodologies used by the PReVENT subprojects. This work has been described in deliverable D16.1. All of the PReVENT subprojects have performed evaluation activities, designed to the particular needs of the respective systems. Additionally, the evaluation procedures proposed by other related projects, such as APROSYS, AIDE and eIMPACT have been investigated. In this way, on the one hand, best practices in evaluation could be identified, and, on the other hand, ways to generalise the very application specific evaluation activities from the PReVENT subprojects or related projects, so that they produce comparable results, could be found out.

- development of procedures for the assessment of active and preventive safety functions. This work is reported in this deliverable. The work is performed in the following PReVAL tasks:
    - WP16.330: Technical performance – Proposal of an agreed and consistent assessment framework procedure for technical evaluation
    - WP16.420 Determination of an assessment procedure for use in evaluating integrated HMI with ADAS functions
    - The methodology, used for the safety impact analysis, is the methodology which has been developed and used by the eIMPACT project; it is based on the behavioural effect approach. This method is described in deliverable D16.1 [1] and copied here in order to provide a total view of the framework.
- the proposed framework is then applied to selected PReVENT functions, and feedback on the procedures is received from experts.
- The different procedures (technical, human factors, safety potential) are integrated to obtain one holistic framework. The final framework will be reported in PReVAL's final report, D16.4.

The deliverable starts with an overview of the methodology for the development of the different procedures. Chapter 3 gives an overview of the common concepts and the interfaces between the different procedures. Chapter 4 describes the procedure for technical evaluation, and Chapter 5 the procedure for the human factors evaluation. Chapter 6 gives an overview of the safety assessment methodology. Chapter 7 describes the steps to come to a consolidated framework, which will be described in the final report D16.4.

2 Methodology

This deliverable will propose a framework for the assessment of preventive safety systems. The proposed framework will be refined in the last phase of the project based on the feedback of the subprojects and other experts. The final framework will be presented in the final deliverable D16.4.

Work in the PReVAL project is organised according to the different aspects of the evaluation: technical performance, human factors and safety impact assessment. The procedures for technical and human factors evaluation have been developed in parallel, with discussions in order to harmonise the framework. For safety impact assessment, the methodology developed and used by the eIMPACT project, is adopted.

2.1 Technical evaluation procedure

The main objective of this task is to develop a procedure, which can lead to standardisation and harmonisation. It is based on a selection of best evaluation practices, that allow: 1) comparison between versions of the same function (accuracy and reproducibility); 2) sound evaluations of the technical performances (completeness of the tests, etc.). The procedures for technical evaluation from the different PReVENT subprojects and the APROSYS project have been investigated in D16.1, and are used as basis for the procedure development. The evaluation results of the PReVENT subprojects have been analysed in D16.2, and lessons learned from these evaluations, are used as input for the procedure development.

The work is performed in the following steps:

- A new technical evaluation workflow is proposed, which is based on CONVERGE, the evaluation methodology used in the PReVENT subprojects, and the experience of APROSYS. This workflow is reviewed by the various partners, and is also sent to the INSAFES project as basis for their evaluation plan. This workflow is also the basis for the human factors evaluation.
- All partners add precise consideration of PRevENT like functions e.g. illustration by examples that facilitate the writing of a validation plan of preventive safety functions

2.2 Procedure for human factors evaluation

The procedure for human factors evaluation aims at retrieving information on the function from a human perspective. This includes assessing how the function interacts with the driver from a subjective as well as an objective perspective. Subjective assessment here means assessing to what extent the driver finds the system useful and acceptable. This must be evaluated because it influences the driver's willingness to make use of the function when driving. Objective assessment here means assessing the function's influence on driving performance. This must be evaluated because it directly influences the functions preventive efficiency.

The results from a human factors evaluation of the subjective perspective shall include to what extent the function is accepted by the driver, how satisfactory the function is to the driver and how the driver apprehends the usability and usefulness of the function. The results from the objective evaluation shall include the degree to which warnings are timely comprehended and appropriately responded to, as well and the degree to which other driver behaviours are altered as a consequence of the function's presence in the vehicle.

The development of a generic procedure for human factors evaluation of preventive safety systems and ADASs is based on existing methods, knowledge and experience on human factor and HMI test methodologies both from within and outside PReVENT. These methods, knowledge and experience are presented in deliverable D16.1.

Examples on what founds the basis for deriving the procedure is

- Guidelines and checklists (e.g. RESPONSE, AIDE questionnaire etc.)
- Acceptance rating scales (e.g. van der Laan etc.)
- Verbal reports
- Focus groups
- Methods for assessment of timely comprehension and appropriate response (ISO N512)
- Driving performance measurement (e.g. INVENT-Traffic Safety Assessment)

The procedure development work is performed in the following steps:

- **Define the main dimensions of evaluation of functional performance with respect to human factors.** Identify the scope and basic aspects necessary to consider when performing human factors evaluation of an ADAS.
- **Define the general procedure.** Derive the general main steps to be followed when performing human factors evaluation. This should be aligned with the workflow for technical evaluation.
- **“Fill” the procedure with content**, that is, define specific HMI evaluation methods. The purpose is not to define new methods, but rather use what is already available (i.e. selection of the methods which have been reviewed in deliverable D16.1, and experience gained from the evaluations performed in the different PReVENT projects).

3 Framework for the evaluation of safety functions

Two major concepts are introduced, which are used by the different assessment procedures:

- "situational control", referring to the level of control jointly applied by the driver and the vehicle in a specific driving situation
- adapted V-shape design cycle: including evaluation specific phases in the V-shape design cycle, which is used to describe the design and assessment cycle.

The deliverable starts with a description of the common concepts, and then describes the procedures for technical and human factors evaluation in more detail.

3.1 Situational Control concept

Background

The term situational control is intended as a general concept linking technical, human factors and general safety impact assessment of preventive safety systems and advanced driver assistance systems within a common framework.. Situational control is based on the established term **Controllability** by RESPONSE 3, as the likelihood that the driver can cope with driving situations including ADAS-assisted driving, system limits and system failures [23]. According to legal requirements, an ADAS is considered safe, as long as the driver is able to control the vehicle.

Controllability is also associated with liability issues, ensuring that the function is safe in all circumstances. Situational control has another perspective, searching for the safety impact of the functions on traffic safety as a whole. Introducing the concept of Situational Control does not mean that PReVAL wishes to define the driver control down to its finest mechanisms. Rather we are seeking a common terminology in which we can discuss the safety impact induced by a function in certain situations.

Definition

In modern ADAS-supported driving the vehicle control is shared between the driver and the vehicle. For present purposes, it is thus useful to consider the driver-plus-vehicle as a single system - the *Joint Driver-Vehicle System* (JDVS [15]). In general terms, *control* can be understood as the ability to direct and manage the development of events [15]. Controlling a process involves selecting actions in order to achieve and/or maintain a consistent goal state. In driving, goals and the corresponding control processes can be described at different levels. A widely adopted scheme has been proposed by Michon [18] who proposed a description of the driving task on three levels:

(1) the strategic level. Tasks on the strategic level involve e.g. route planning and navigation.

(2) the tactical level. Tactical level tasks include the selection of speed and following distance, decision to overtake etc.).

(3) the operational level. The operational level refers to the moment-to-moment vehicle handling.

Finally, in the context of driving, a *situation* can be understood as a set of relevant driver-, vehicle and environment elements within a volume of time and space. Based on these concepts, the *situational control* can be defined as *the degree of control that a JDVS exerts over a specific situation.* It should be stressed that the term, as defined here, applies to all three levels of the driving tasks in Michon's model.

Every driving situation can be characterised through a number of parameters which describe different aspects of the situation. A parameter is a measurable quantity that represents properties of driver, vehicle and/or environment respectively. The list of possible parameters is logically infinite, but in traffic safety normally two subsets of parameters are used. One subset could be called **situational descriptors.** This is the range of parameters which describe aspects of the situation that are used to classify the situation as belonging to a certain group of traffic situations. Situational descriptors usually include parameters like road type, weather conditions, driver age, etc.

The other subset of parameters includes those for which staying below or above certain values is necessary for maintaining control of the driving situation. Slightly rephrased, for every driving situation, the driver's situational control can be described through a series of distributions (normal or other) of parameters which are judged essential for control of the situation. To distinguish this subset of parameters from the situational descriptors used for situation classification, they will be called **indicators,** i.e. indicators of control over a certain situation.

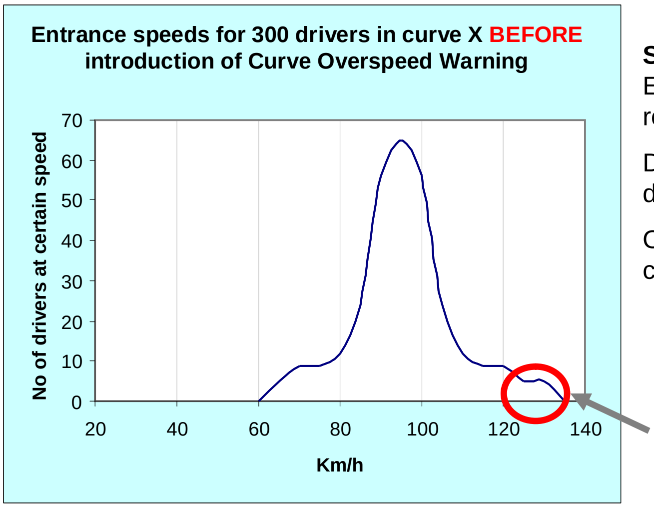
The situational descriptors could be seen as the independent variables, the variables that are manipulated and set by the experiment designer. The indicators can be seen as the dependent variables; the response that is used as a measure of driving performance.

The (control) indicators are measurable quantities that are used for evaluating the functional performance. The indicators should be very specific on what shall be measured, but can be based on general characteristics of driving performance, for example lane keeping, which could be assessed in several ways through different metrics (for example, standard deviation of lateral position, nr of lane exceedances, etc). Other indicators could be headway, reaction time, entrance speed into curves, etc.

For a driving situation or set of situations, the distributions in the chosen indicators define what can be called the normal variability of driver and vehicle behaviour. From the safety perspective, this normal variability will include all normal driving as well as all incidents and accidents associated with the situation(s). For example, the driving situation of entering a curve can be assigned a number of (control) indicators which are judged important for successful negotiation, such as entrance speed, road friction, curve radius, etc. Each of them will have a certain distribution if it is measured or calculated. Within the values for the indicators are all driver behaviours for which entering curves results in successful negotiation (the absolute majority), and values, which accident investigations show to be related to an incident/accident outcome, and which can be coupled to a relevant harm metric, such as number of injuries above a certain severity. In the further process of defining the experiments, all or a subset of these important indicators will be selected for measurement or calculated from measurements.

Current state of
situational control

(normal) distribution of curve
entrance speed for situations X_1-X_n



Specification of situations X_1-X_n
Environment characteristics (rural road, lighting, time of day, etc...)

Driver charachteristics (alert, tired, distracted, etc...)

Other descriptors used for situation classification...

Values of control indicator Y typically associated with harm metric Z (number of injuries, accident frequency, air pollution, etc)

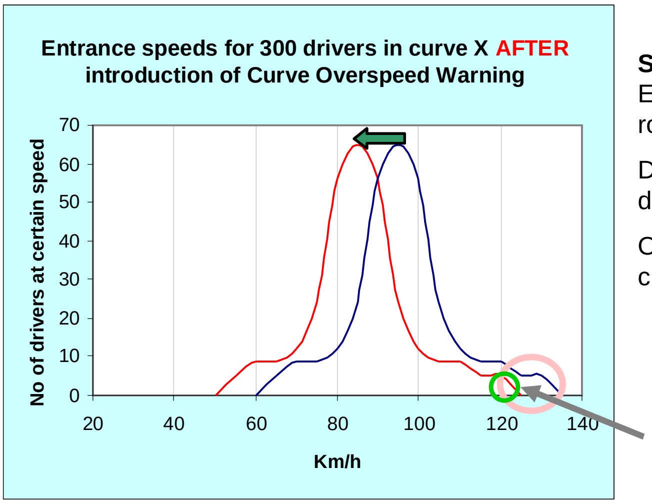
Figure 2: Illustration of the distribution of an indicator, with highlighting of the values associated with a chosen harm metric.

When an ADAS is introduced, it is generally expected that one or more of the indicators which influence situational control will change. For example, the main intended effect of a curve speed warning would be a reduction of speed when entering the curve. However, the system may also have unintended effects on the chosen indicators. For example, drivers may come to rely too much on the system so that curve entrance speed is only adapted when a warning is given, which could result in non appropriate speeds in curves or situations which are not covered by the system.

Effects on situational control can be of two types. It can either alter one or more of the indicator distributions by “pushing” them to the right or left (as in Figure 3 for entrance speed in curves, in the case of a curve speed warning), or they can change the shape of the distribution curve, by for example narrowing, skewing or clipping it (i.e. certain values are not there anymore). The task for evaluators of active safety systems then is to establish the baseline distributions for the control indicators (current state of situational control, above), measure the changes which are due to system introduction (new state of situational control, below), and then look at how these changes in the indicators affects the values associated with the chosen harm metric. This is further illustrated in Figure 3.

System introduction creates new state of situational control

Changes in (normal) distribution of curve entrance speed for situations X_1-X_n



Specification of situations X_1-X_n
Environment characteristics (rural road, lighting, time of day, etc...)

Driver charachteristics (alert, tired, distracted, etc...)

Other descriptors used for situation classification...

CHANGE in values of control indicator Y typically associated with harm metric Z (number of injuries, accident frequency, air pollution, etc)

Figure 3 Illustration of the effect of an ADAS on a situational control parameter, and the associated change in the chosen harm metric

Application

When using the concept of Situational Control, the challenge is to define the indicators which describe and quantify a Joint Driver-Vehicle System's control over various situations (as classified by the situational descriptors) and to select a subset of relevant indicators for experimental measurement.

The situational control concept can be viewed as a general term applicable not only for driver-vehicle systems and road safety, thus in principle indicators can be anything from number of injuries or accidents to the degree of air pollution or fuel consumption, but always strongly correlated with the objectives of the experiment. For present purposes, the indicators relate to road safety.

In modern vehicles, the ability to maintain situational control may also be supported by one or more ADAS. In this case, different functional responsibilities may be assigned to the driver and the ADAS. For example, in the case of ACC, the longitudinal control within safety margins is fully automated by the system. Situational control should therefore be viewed as a property of the entire Joint Driver-Vehicle System. In this way, the situational control concept is also applicable in cases when the system takes over control entirely (and the driver is not in the loop).

When a preventive safety system or ADAS is introduced, the situational control can be affected in a number of different ways (intended and unintended). The task for the evaluators of the ADAS is to formulate more precisely what the hypothesised effects are, i.e. how the system is going to alter the distributions of the indicators associated with the situation. In other words, for every preventive safety system we need to formulate a number of hypotheses which say something about how a JDVS normally maintains control of the situation(s) in which the system is intended to work, how that control may be lost, and the way(s) in which the system is hypothesised to enhance the control process in order to prevent risks and/or accidents.

When selecting indicators and the metrics to measure them with, the following aspects need to be taken into account:

1. Measurability. The indicators must be measurable: metrics should be defined in the test definition phase how to retrieve parameter values.
2. Relevance. The choice of indicators and situational descriptors must correlate well with the objectives of the experiment, in terms of providing representative and relevant information on the situation that shall be evaluated, for assessment of the functional performance.
3. Validity. The indicators, situational descriptors and metrics shall be selected with a high confidence that you actually measure what is intended to measure.
4. Reliability. The measurement process must be repeatable. A specific indicator shall be possible to measure several times in the same or very similar situations, given the same or very similar constraints, and provide the same result.
5. Sensitivity. Relevant changes in the experimental situation must be reflected in changes in the values of the indicator, that is, the indicators have to be sensitive to the situation they are expected to reflect.

3.2 Adapted V-shape design cycle

The main purpose of evaluation is to assess whether the system works as required, i.e. if it achieves the desired improvement of situational control. CONVERGE provides an evaluation framework which is valid for any research and development project, thus it is very generic and needs to be adapted for the case at hand related to preventive advanced driver assistance systems (ADAS) as those proposed in PReVENT.

The adaptation proposed by PREVAL consists in adapting CONVERGE thanks to the introduction of the V-shaped cycle (well known in the automotive industry) in order to build an adapted evaluation plan for PReVENT type function. First of all, let’s consider the simplified V-shaped scheme :

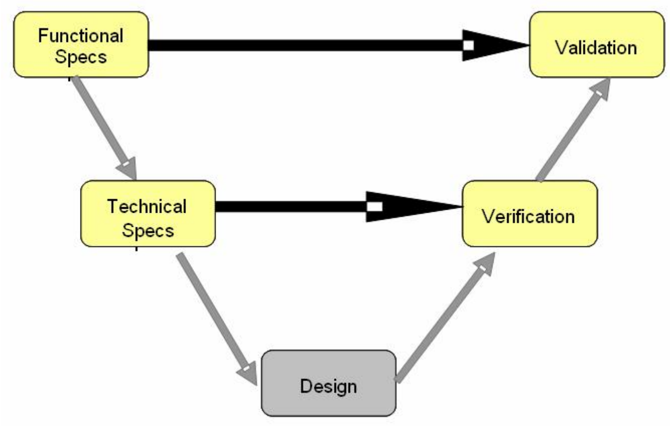


Figure 4: Simplified V-shaped design cycle

This figure explicits the evaluation phase in two levels:

- The function level: by relating the functional specifications with the technical validation of the entire system (upper arrow)

- The component level: by relating the technical specifications of the subsystems with an evaluation at a component level phase: verification (lower arrow).

In order to carry out these two levels of evaluation, a thorough test definition for each level is necessary. Our first figure is then enriched as follows:

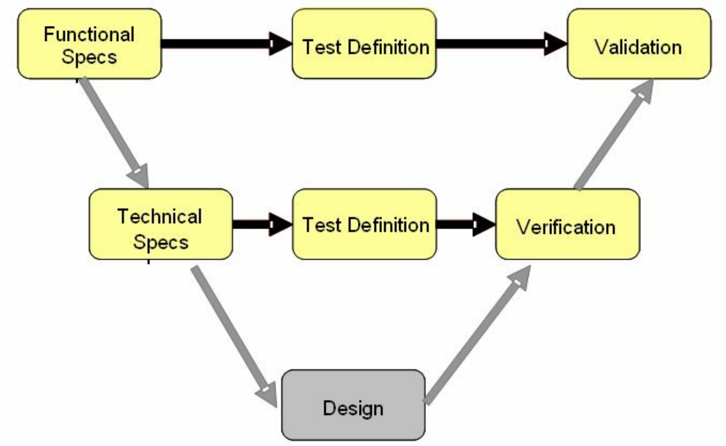


Figure 5: V-shaped design cycle including test definition phases

We could stop our evaluation diagram at this stage. However, we choose to enrich it with an additional step. The reasons are briefly described the following ones:

The initial car manufacturer specifications (functional and technical) may not be described enough for evaluation purposes; they also have evolved during the function design phase: some practical modifications could have been introduced. The specifications must be enriched by considering high level aspects (cooperation level with the network or other vehicles for example). The design step allows to collect and make explicit some elements related to the function description (like self protection or partner protection, type of target, etc.); all those elements could imply to adapt the system description for evaluation purposes.

Our figure is then modified with the introduction of an additional step that we will call *evaluation specifications* where all specifications that are necessary to design tests should be included.

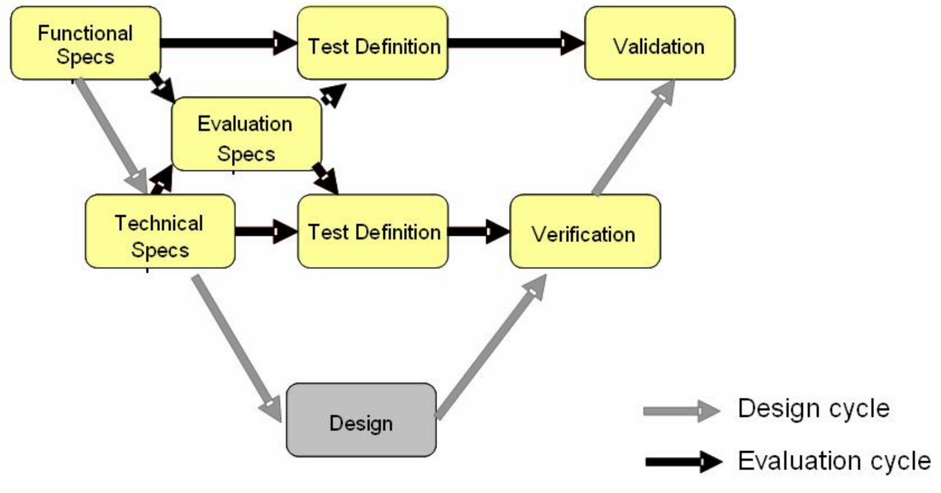


Figure 6: Adapted V-shaped design and evaluation cycle

The evaluation specifications include the system and functions description, which is common for all evaluations (technical, human factors, safety potential), and the assessment objectives and key indicators. The framework has been developed for the technical evaluation, but will also be applied to the human factors evaluation.

3.3 Relation between technical, human factors and safety potential evaluation

The safety potential of preventive safety systems and advanced driver assistance systems is affected by different factors, such as technical reliability of the system, HMI and changes in driving behaviour. The safety implications can be classified into three aspects: technical performance, driver behaviour and driving performance (interaction between the system and the user); and the traffic safety level (safe operation of the traffic system, interaction between users and non-users) [3]. All these different aspects need to be taken into account when considering the safety potential of a preventive safety function.

Preventive safety functions consist of 3 layers: Perception, Decision and Action. Preventive safety functions can be classified according to the type of the action:

- warning functions: the driver is informed about a dangerous situation, and is responsible for taking an appropriate action.
- assisting/intervening functions: the system applies the vehicle control actuators, but the driver has the ability to override. (e.g. speed alert system with override possibility)

- controlling functions: the system applies automatically vehicle control actuators, and the driver has no possibility to override (pre-crash systems and reversible protections systems).

The goal of the technical evaluation is to validate whether the system meets the specifications. It is therefore mainly related to the Perception and Decision layers, and for controlling and intervening functions also on the Action layer. The human factors evaluation is concentrated on the Action layer for warning and intervening functions.

Figure 7 illustrates the main relations between three types of system evaluation. The technical performance (effect or output) of the system is the major contributor to the control of the driving situation. However, high usability and acceptance of the system is of key importance in order for the system to at all have a changing effect on the driving situation. The overall changes in driving performance, caused by the technical effects of the system, the user's experienced usability and acceptance is together summarized by situational control.

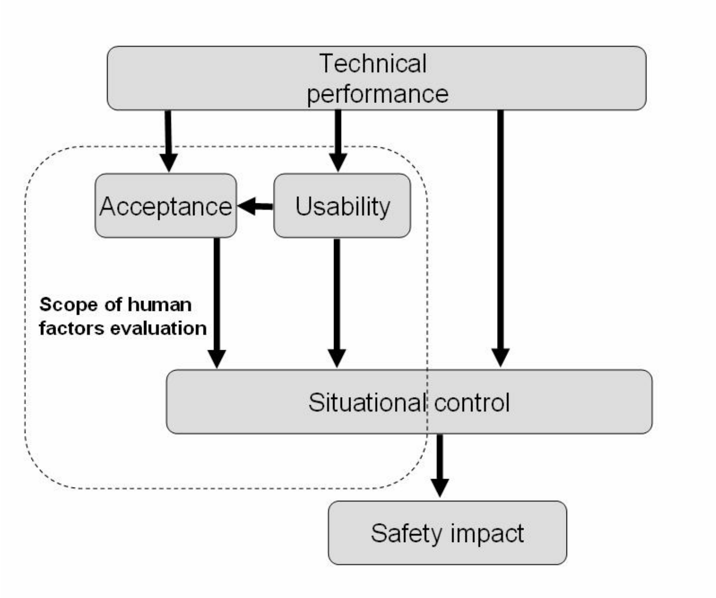


Figure 7 The main relations between technical performance-, HF- and safety impact evaluation.

The technical evaluation focuses on the technical performance of the system and its ability to affect situational control independently of the driver. For controlling systems, acceptance and usability are not of relevance to consider, since there is no interaction with the driver. In this case, the situational control concept is also applicable, but it is now only influenced by the technical performance, and how the system technically can improve the control, without interacting with the driver. An example is ESC (electronic stability control).

The goal of human factors evaluation is to assess the ability of the function to affect situational control *via the driver*, e.g. by means of providing changing behaviour. Naturally, technical and human factors evaluation is closely linked. For example, the acceptance of the function (assessed in the human factors evaluation) could be expected to depend strongly on the false alarm rate (assessed in the technical evaluation).

While indicators in human factors evaluation address driving performance related indicators such as reaction time, the indicators used in technical evaluation are for example indicators used for assessing how the system fulfils the technical specifications e.g. timing of warnings, brake pressure etc. The indicators selected for the technical evaluation are directly influencing or interacting with the human factor related indicators reflecting the situational control. The degree of interaction between technical evaluation and human factors evaluation, with respect to indicators, depend on the type of function; whereas it is a warning function, an assisting/intervening function or a controlling function,

As an example, the active steering function, as the one developed in the SAFELANE project, not only warns the driver, but also intervenes by means of applying a torque to the steering wheel, in the event of an unintended lane drift.

An indicator reflecting technical performance can be for instance the timing of the warning issued. This indicator is not necessarily a situational control indicator itself; however it is directly influencing the situational control, according to Figure 7. Situational control is then reflected by indicators representing the following actions, taken by either the system or the driver.

An indicator reflecting human factor performance can be for instance an indicator related to the driver's response to the warning such as the reaction time. This may be defined as the overall time between the instant when the warning is issued and the time when the driver takes an action; e.g. starts to steer back into the correct lane. This indicator is then directly reflecting situational control; it is desired to get a fast response from the driver. An indicator reflecting both technical *and* human factor performance can be for instance the overall response of the JDVS represented by the overall time duration of the unintended lane drift, starting at the time where the vehicle cross the border line of the lane and ending at the time when the vehicle is back in position in the target lane.

This indicator would then be composed by both the system response as well as the driver's response, both reflected in torque on the steering wheel.

In this example the timing of the warning can be seen as an independent variable as input for evaluating situational control. The timing of the warning might indicate good technical performance according to the functional specification, because it meets the specified requirement. However, it is no guarantee that the situational control indicators evaluated from the consecutive actions show satisfactory results.

The example above presents how the technical performance not only influence the situational control by the warning issued, but is also directly constituting a part of it through the actions taken by the system.

An example of a warning system where the technical performance only influences the situational control is a warning system that issues a warning at a certain TTC, but does not take any actions itself. The situational control indicators are then only connected to the driver's response.

An example of a controlling system is the ESP (or in some applications ESC), which does not interact with the driver. In a situation where ESP is activated, the overall response of the JDVS is only connected to the technical performance. The situational control concept is applicable here as well, but now the situational control is represented only by indicators reflecting the technical performance, such as time for activating the brakes automatically or time for decreasing the throttle lever.

The goal of safety impact analysis is to make an aggregate-level assessment of the effects on harm metrics (such as number of accidents or fatalities) based on general assumptions of technical performance, behavioural effects and accident statistics. Such an assessment could be informed by previous technical and HF assessment studies of the system under evaluation but this is not a necessary condition (if empirical data is lacking, expert opinions is another option (see D16.1).

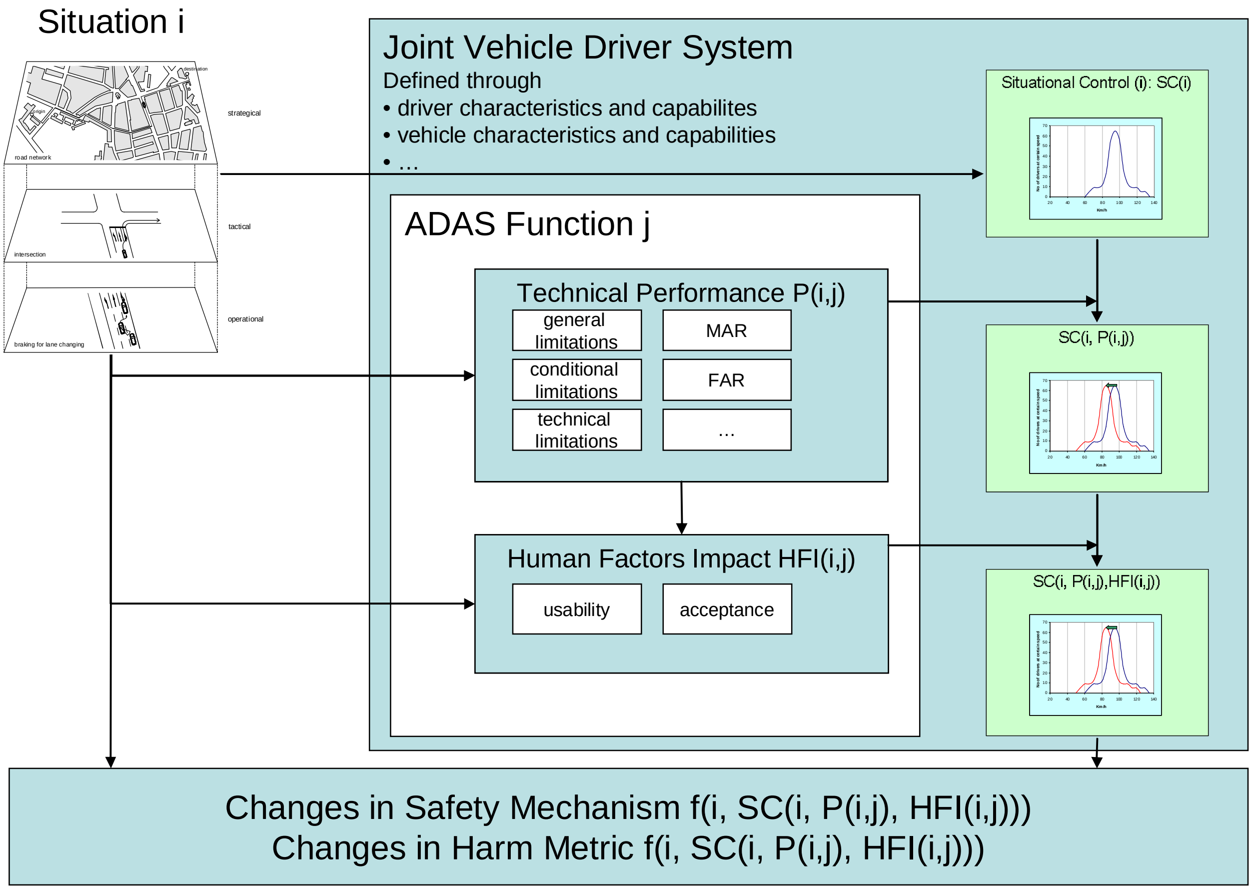


Figure 8: Main relations between the three domains of evaluation with respect to the Situational Control concept.

Figure 8 shows the overall relations between Michon's Levels of the driving task, the Joint Driver Vehicle System and the changes in the safety mechanisms. While Situational Control of the JDVS without the support of a system like an ADAS can be seen as the baseline condition, any introduction of an ADAS will change the control of the JDVS over the actual driving situation. The three sketches on the right side of Figure 8 show these changes in distributions of any relevant indicator which are induced by an ADAS. The upper one shows the distribution of the indicator with no ADAS at all (baseline condition). The second shows the respective distribution changed by the introduction of an ADAS and is only depending on the technical performance of this system. This can either be the case when the ADAS is a fully automated system and the driver is not in the loop at all, or in case the driver acts as the “perfect driver” with a “perfect HMI” with no drawbacks from any imperfection at all. Since this is a quite theoretical situation, the “real” changes in Situational Control with a driver in the loop depend not only on the technical performance, but also on the impact of human factors, who are in return depending of the technical performance (cp. the TTC-triggered warning system example above). The analysis of changes in Situational Control together with the consideration of the actual driving situation is then taken into account for the estimation of changes in the respective safety mechanisms or related harm metrics. A mapping of the test results on technical and human factor performance is first needed, in order to translate these test results into the corresponding safety mechanism, for making them applicable for the safety impact estimation.

4 Proposal for technical evaluation procedure

4.1 Assessment framework proposal

The goal of the technical evaluation is to validate whether a system meets its specifications. CONVERGE provides an evaluation framework which is valid for any research and development project, thus it is very generic and needs to be adapted for the case at hand related to preventive advanced driver assistance systems (ADAS) as those proposed in PReVENT.

The adaptation proposed by PREVAL consists in adapting CONVERGE thanks to the introduction of the V-shaped cycle (well known in the automotive industry) in order to build an adapted evaluation plan for PReVENT type function. The adapted V-shape design cycle has been introduced in Section 3.2.

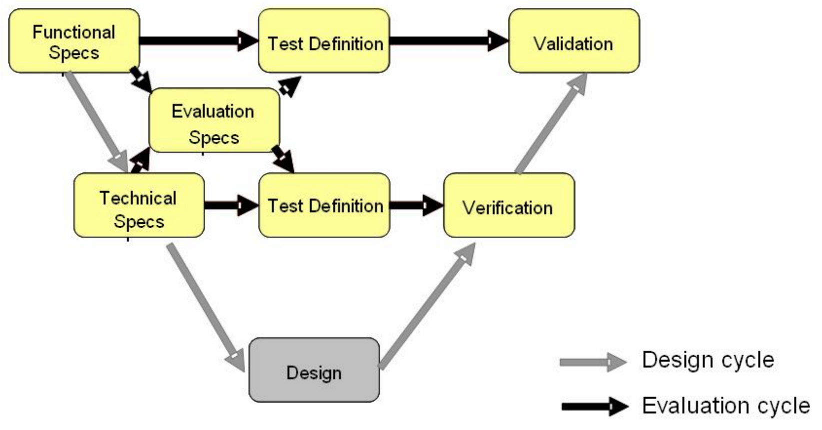


Figure 9: Adapted V-shaped design cycle

What about the relation of evaluation cycle with the CONVERGE methodology?

Figure 10 shows the relationship between the adapted V-shaped design cycle and the CONVERGE evaluation methodology.

a) All specifications that are necessary to design the tests are encircled in dark grey (system description; applications targeted; system objectives). They correspond to the high level specifications in CONVERGE.

b) The definition of the tests is encircled in mid grey. This relates to the lower level CONVERGE steps (study design: indicators, reference case, data collection, measurement conditions, statistical considerations, measurement plans).

c) The execution of the tests carried out and the reporting are figured out in the boxes “validation” and “verification”

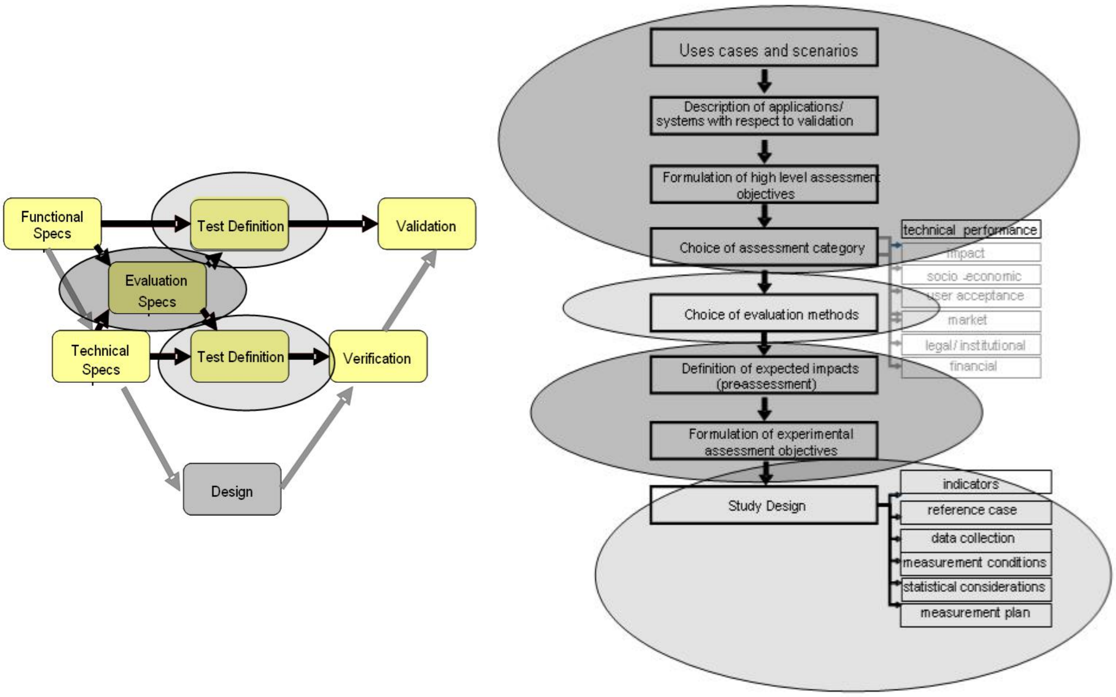


Figure 10: Comparison of the PReVAL workflow with CONVERGE

4.2 Evaluation specification

This section details the elements of the methodology outlined above. It includes the following items

- System and functions description,
- Assessment objectives
- Expected impacts

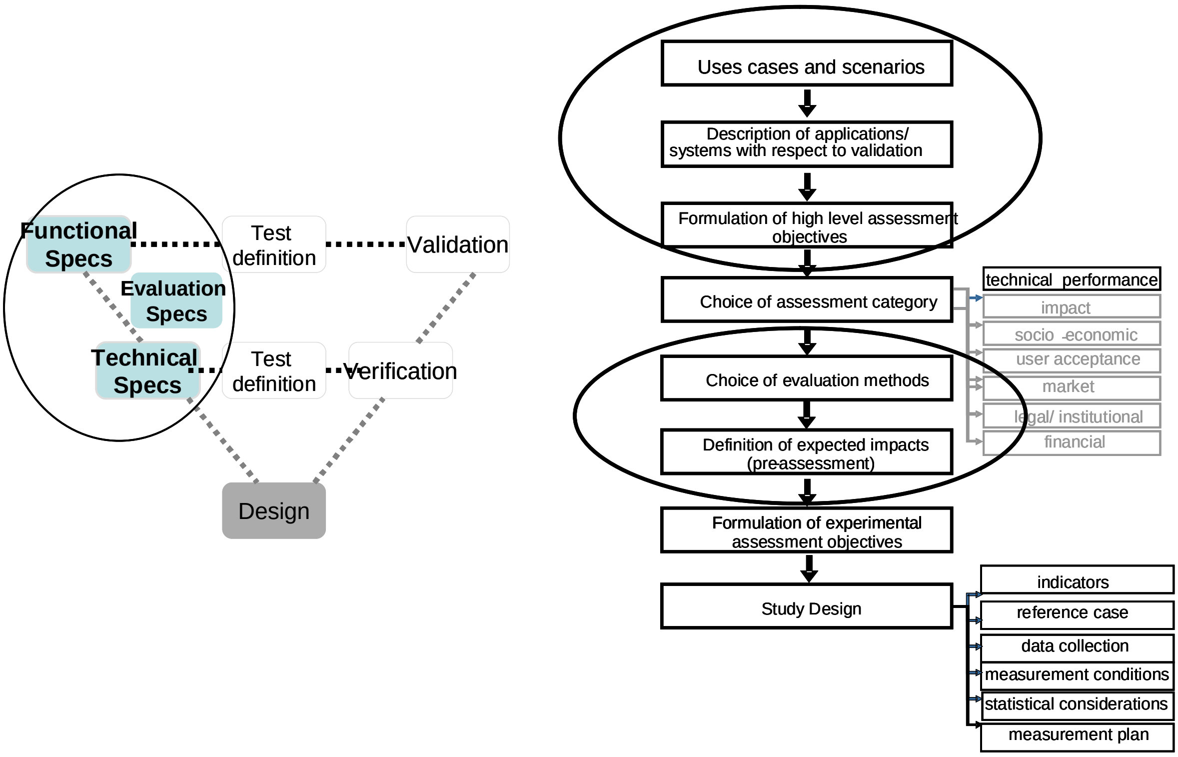


Figure 11: Elements of the method addressed by the Evaluation Specifications

4.2.1 System and functions description

The **System and functions description allows** the evaluator to understand what the system is supposed to do and how it works. The items described below aims at classifying and facilitating the reporting of the information:

- General information (commented below in 4.2.1.1)
- Functionality and uses cases (commented below in 4.2.1.2)
- Classification or what type of accidents are we dealing with (commented below in 4.2.1.3)
- Description of subsystems (commented below in 4.2.1.5)
- Minimal technical specification(commented below in 4.2.1.6)
- Limitations in operational conditions (commented in 4.2.1.4)

In the ideal case, all the information related to the function description is included in the functional specification documents. If this information is available for the evaluators, the tables can consist of accurate references to the data in the specification documents.

4.2.1.1 General information

The system and its functions can be described according to the tables below. The text in italics provides examples, based on a lane departure warning system

Table 1: General information- identification of system

<table>
<tr><th>Type of data to include</th><th>Example</th><th>Comment</th></tr>
<tr><td>Name of the system</td><td><i>Lane departure warning system</i></td><td>A one line description for identification purposes.</td></tr>
<tr><td>General function</td><td><i>Providing driving feedback to the driver in lane drift situations.</i></td><td>A brief “headline” description of the function performed by the system.</td></tr>
<tr><td>Manufacturer</td><td></td><td>Name and contact points to get further information if necessary.</td></tr>
<tr><td>Build status</td><td><i>Research</i></td><td>A brief description on the status of the function, whether it is in production or under development</td></tr>
<tr><td>Type of vehicles</td><td><i>Trucks</i></td><td>A brief description of the type of vehicles.</td></tr>
<tr><td>Available documentation</td><td><i>Function and System requirement specification</i></td><td>Information on specifications that can be used- e.g system requirement specifications and design specifications</td></tr>
<tr><td>Cooperation level</td><td><table>
<tr><td><i>Level of Cooperation</i></td><td><i>System</i></td></tr>
<tr><td><i>None, stand alone system</i></td><td><i>X</i></td></tr>
<tr><td><i>Cooperative V2I</i></td><td></td></tr>
<tr><td><i>Cooperative V2V</i></td><td></td></tr>
</table></td><td>Information on whether it is a stand alone system, cooperative V2I system or cooperative V2V system.</td></tr>
<tr><td>Major technologies</td><td>Perception subsystem based on image processing mainly
Decision subsystem : estimate the level of reliability of the situation and decide the cooperation mode
Action subsystem : sounds and haptic feedback</td><td>subsystems and components that will be detailed in the subsystems description such as sensor systems, actuators and decision algorithms)</td></tr>
</table>

4.2.1.2 Functionality and use cases

The section gives a brief overview of the functionality of the system. The intended goal and the benefit of the system are described, as well as the intended behaviour of the driver. The description should also provide accurate information on the timing schedule.

The concept rationale should be related to a sound accident analysis. The definition of relevant use cases/scenarios is important, since they should be representative for the most important situations in the real world. The description of the use cases is used e.g. for the derivation of hypotheses related to driver behaviour in the human factors evaluation.

The description can also include other requirements, related to dependability and maintenance of the system.

Table 2: Functional specifications

Type of data to include	Example	Comment
Benefits	*To support drivers on tactical level in situations of unintended lane changes for heading back in correct position in the current lane. Unintended lane changes are avoided*	A short summary of the intended goals and benefits with the system. Where appropriate references should be made to the three levels of the driving task strategic, tactic, operational (Michon)
Intended driver behaviour	*Drivers react to warning and/or haptic support, and steer back to the correct lane*	How does the function change the driver behaviour or drivers ability to cope with the traffic situations compared to driving with vehicle without the function?
Interaction of the function with the driver	*an auditory warning and/or haptic support*	How does the function interact with the driver?
Time schedule	*e.g; the system acts in few seconds before an actual departure, a TLC of 3s is a maximum, 2s is a minimum*	The time schedule for providing warning or action. The schedule should be detailed such that it provides a basis for time and distance assessment during verification at technical level (for example sensors needed, at what time before the dangerous event the decision has to be made, and when is action required)
Functionality/ Main-use cases	Main use cases: to be precised 1. *The system acts when the turn indicator is not activated* 2. *The system provides a warning in case of an unintended lane change, which is defined as an immediate crossing of the line markings is estimated….*	See below

The description of the use cases can include:

- **Situational descriptors.** These are factors that describe the driving situation addressed in the use case

  - **Type of target objects** (e.g. passenger car / heavy goods vehicle (truck) / bus / motorcycle / pedestrian)
  - **Road type** (motorway, rural, urban)
  - **Environmental Conditions** (weather, light and road conditions).
  - **Traffic situation** (speed…)
- **Use cases and relevance. Safety preventive functions have to be based on** accident analysis and statistics. This is discussed in the next section "classification".
- Finally, some other requirements, related to dependability and maintenance, are included:

Table 3 : Functional dependability

Type of data to include	Examples	Comment
Possible dependability requirements	system in fault is detected No missed alarm tolerated Very few number of false alarm tolerated	e.g. fault tolerance, fail safety, reliability etc. Dependability is defined as the trustworthiness of the system such that reliance can justifiable be placed on the service delivered, i.e. that the system gives appropriate alarms and takes correct action at the right moment with a certain confidence level, for a wide set of operating conditions, in spite of the occurrence of disturbances and failure modes.
Maintenance issues	Not relevant for research prototype	In terms of availability and reliability the expected system maintenance might be important, e.g. service intervals and software updates. This holds especially for map-based systems: the frequency and method of content update influences the reliability of the system very much

4.2.1.3 Accidents targeted classification

The system needs to be classified to enable road safety impact assessment and comparison with other systems. The classification can therefore be consistent with the classification of the CARE database, which can be used by the safety impact assessment.

Table 4 : Function classification

<table>
<tr><th>Category</th><th></th><th>Comment</th></tr>
<tr><td>Accident and incident type</td><td>lateral (lane change, crossing, …) / road departure</td><td></td></tr>
<tr><td>Self Protection or partner protection</td><td>self / other car occupants/ other vehicle or vulnerable users…</td><td></td></tr>
<tr><td>Action mode</td><td><table>
<tr><td><i>Action mode</i></td><td><i>System</i></td></tr>
<tr><td><i>Informative/Advisory/Warning</i></td><td><i>X</i></td></tr>
<tr><td><i>Support</i></td><td><i>X</i></td></tr>
<tr><td><i>Intervene</i></td><td></td></tr>
</table></td><td>Information whether it is an Informative/Advisory/Warning system, a support system or an intervening system</td></tr>
<tr><td>Type of target object for detection</td><td>none</td><td>Type of target objects which are detected by the system</td></tr>
<tr><td>Road types</td><td><table>
<tr><td><i>Road types</i></td><td><i>System</i></td></tr>
<tr><td><i>Urban</i></td><td></td></tr>
<tr><td><i>Rural</i></td><td></td></tr>
<tr><td><i>Highway</i></td><td><i>X</i></td></tr>
</table></td><td>A definition of road context in which the system is, and is not intended to operate. This includes general locations of roads and physical requirements of road markings, gradients, curvatures, widths etc.</td></tr>
<tr><td>Road section type</td><td>straight roads or high radius curves</td><td></td></tr>
</table>

4.2.1.4 Limitations on operational conditions

This section deals with the conditions for which the system is designed and the conditions for which the system is not expected to operate efficiently.

Table 5: Function limitations

<table>
<tr><th>Type of data to include</th><th>Example</th><th>Comment</th></tr>
<tr><td>Driver restrictions</td><td></td><td>This information shall be provided where applicable, if there are any restrictions or special driver skill requirements defined by the system designers or manufacturer.</td></tr>
<tr><td>Vehicle requirements</td><td></td><td>Which requirements does the function set to the vehicle?</td></tr>
<tr><td>Traffic environment</td><td><i>Except urban roads</i></td><td>A description of the traffic context within which the ADAS is, and is not, intended to operate.</td></tr>
<tr><td>Infrastructure requirements</td><td></td><td>Which requirements does the function set to infrastructure?</td></tr>
<tr><td>Environmental conditions</td><td><table><tr><td><i>Weather</i></td><td><i>System</i></td></tr><tr><td><i>Adverse (Rain)</i></td><td><i>X</i></td></tr><tr><td><i>Adverse (Snow)</i></td><td></td></tr><tr><td><i>Normal</i></td><td><i>X</i></td></tr></table>
<table><tr><td><i>Lighting specifications</i></td><td><i>System</i></td></tr><tr><td><i>Dark</i></td><td><i>X</i></td></tr><tr><td><i>Light</i></td><td><i>X</i></td></tr></table></td><td>Additional requirements or restrictions resulting from environmental conditions like weather (adverse/normal) and lighting specifications (light/dark)</td></tr>
<tr><td>Other system limitations</td><td></td><td>System limitations in terms of the conditions under which the technology is not expected to function effectively (e.g. high speed, short or long distance to target object etc.)
Performance risks, defined as situations in which the system works correctly, but with less effect (e.g. due to a slippery road, or late detection by sensors) have to be described here as well</td></tr>
<tr><td>Interaction of the function with other applications</td><td></td><td>How does the function interact with other applications?</td></tr>
</table>

4.2.1.5 Description of subsystems

The description of the subsystems addresses the **technical specifications** that contain the minimal information to enable the verification of subsystems at component level. Table 6 gives the information which should be available from the description of the subsystems

Table 6: Description of the subsystems

Technical specifications		
Type of data to include	**Example**	**Comment**
Perception/ Sensors	Camera + image processing	Specify the sensors of the system, e.g. type, range, necessary/optional etc. Describe additional sources which are necessary for the system, but are not provided by conventional sensors, e.g.

		map data
Decision level	Decision algorithm that decides the choice of the support	Provide relevant information about the algorithms, as far as they are necessary for understanding the system functionality. Try to provide easy understandable flow charts when possible
Action level	none	Provide information about the actuators of the system, specifications like intervening modes, torque etc.
HMI	Loudspeaker Vibrating steering wheel	Provide information about the HMI of the system, if there is any. Include pictures of the HMI or other descriptions to show the main aspects of the HMI.

4.2.1.6 Minimal Technical Specifications

Necessary technical specifications that are not provided by the manufacturer can be derived from functional specifications and describe minimal requirements to guarantee operation of a subsystem. They can be determined by simulations or standards and must be defined such that they can be verified.

Examples are:

- Range, field of view, drop out rate and drop out duration of sensors
- Position accuracy, the accuracy of the position of the target objects; distance to the previous vehicle
- Speed accuracy, the accuracy of the relative speed of objects;
- Detection rate, is the number of detected objects related to the number of existing objects. The value can be given in percent (%).
- Time to detect, time collision (TTC) at which the system should give a warning.
- Time to detect line, time to line crossing (TLC) at which the system should give a warning.

4.2.2 Assessment Objectives

Assessment of the system can be carried out in different ways, for different purposes and at different levels of depth. There are different types or categories of assessment: technical assessment, human factors assessment and impact assessment. This section deals with technical assessment objectives.

A technical assessment aims to determine how far a system meets technical requirements and expected objectives. As announced in section 4.1, we propose to perform the technical assessment at two levels:

- subsystems are evaluated at component level (Verification). The main purpose of this assessment is to verify that the subsystems are working according to the technical specifications.

- the system is evaluated at functional level (Validation). The main purpose of this assessment is to test if the system works according to the specifications and to assess the improvement of the system on situational control.

4.2.3 Expected system impacts

The technical objectives of the system should be described in such a way that it is possible to evaluate the performance of the system, e.g. reducing speed, preparation for impact or collision mitigation by warnings. The definition leads to the selection of the key indicators that we will done in the next step.

In this phase, hypotheses are made about the possible effects of the function on situational control through the driver or independently if the system is autonomous (collision mitigation for example).

4.2.4 Key indicators

The expected impacts have defined how the performances should be characterized. The definition of **key indicators** (also called “criteria of response”, “measures of effectiveness”, “variables”, assessment criteria, etc.) allows assessing if and how far the objectives will be achieved.

The **key indicators** are defined such that they enable measuring the effectiveness of the system. To achieve this, the key indicators should address the **nominal behaviour** of the system (they must be able to reflect clearly the related performance, e.g. what is the amount of speed decreased) and the **dependability** (they must be capable of reliable assessment using the experimental methods and measurement, e.g. what is the reproducibility of this speed decrease).

Indicators can be classified in different types:

A. Reliability related indicators

The reliability can be defined as the probability of a component, subsystem of complete system to function correctly over a given period of time under a given set of operating conditions

However, the reliability that the driver experiences in practice, is associated the capability of the system to provide reliable warnings. Thus system reliability is associated with a high number of true positives (TP, correct action when needed) and true negatives (TN, no action if none necessary) and a low number of false positives (FP, or false alarm) and false negatives (FN, or missed alarms).

TP, TN, FP and FN can be defined per km, per hour or per event. The relation between the actual situation and the prediction is shown in Table 7.

Table 7: Relation between actual and predicted data

		Actual data	
		Negative (safe)	Positive (threat)
Prediction	Negative (safe)	a	c
	Positive (threat)	b	d

Table 8: Definition of rates [10].

Rate		Definition
P (Precision)		$d/(b+d)$
TP (True Positive rate)		$d(c+d)$
FN (False Negative rate)		$c/(c+d)$
TN (True negative rate)		$a/(a+b)$
FP (False positive rate)		$b/(a+b)$
Reliability index	if positive and negative datasets have similar size	$(a+d)/(a+b+c+d)$
	otherwise	$\surd(d^2/(b+d)(c+d))$

In some tests, the decision if an event results in a true and false positive is judged by the operator of the vehicle. It is therefore important to distinguish between false positives and nuisance alarms. An example: a false positive for FCW would be the alarm going off while the vehicle is completely alone on the road. A nuisance alarm is when you get a FCW when not wishing for one, i.e. the function works fine but the driver does not think the information/ warning/ intervention is appropriate at that point in time. Which of these is being tested for needs is to be specified in the description.

The safety of the system as perceived by the driver can be defined in terms of the ability of the system to adequately respond to a hazardous situation. This is expressed by the **Correct Alarm Rate** (CAR), **False Alarm Rate** (FAR) and **Missed Alarm Rate** (MAR). FAR is related to FP, MAR to FN. CAR, MAR and FAR can be expressed with respect to all times an alarm should be given, or with respect to all situations, and with respect to driven kms, time, scenarios,..

Example:

- for collision mitigation system, the missed alarm rate is calculated as the number of missed critical scenarios over the total amount of critical scenarios (FN), FAR is the number of wrongly detected critical situations related to a time or distance of normal driving.

Please note that the sum of CAR, MAR and FAR can be more than 100%, depending on the definition of CAR, MAR and FAR.

B. Indicators related to the decision criteria.

Indicators are related to the parameters used inside the decision system for the function activation. Measurement of these indicators serves two purposes: firstly - to know if the system activates well when it should (this can be checked by verifying the correct calculation of these variables) and secondly – to verify if these activation criteria have been chosen “correctly”.

A reference measurement is needed in order to calculate and evaluate the accuracy of the different indicators (see below).

Examples of commonly used indicators:

- Time to collision (TTC).
- Extended definition of TTC. TTC is normally used in vehicle longitudinal guidance (E.g. application in ACC). It can also be extended to the lateral situation (lane change) and 2D situation (intersection)
- Minimum TTC during a scenario. An increase in the minimum TTC indicates an improvement in situational control.
- Time to line crossing (TLC) for lane keeping systems
- Time-to-Object, for map based warning systems, the following indicators can be used: analogous to the TTC or TLC warnings are provided a certain time before the vehicle reaches a certain object, e. g., a curve or a black spot. The Time-to-Object is defined by the distance divided by the vehicle speed. The correct Time-to-Object is depending from positioning and map matching as well as from correct position coding of the object in the map data base.
- Headway, or distance to preceding vehicle
- Lateral offset

Indicators are also needed to measure the quality of the different subsystems. Examples:

- Detection rate, the number of detected objects related to the number of existing objects.
- Map data errors: Map object data are the basis for a warning. Data errors can occur due to coding errors (wrong coding or missed data) and the change of reality in comparison to stored data, e. g., by road construction.

As an example, in SAFELANE, the danger is estimated on the basis of the distance between the vehicle edge and lane markers (lateral offset) and the TLC (estimation of a lane crossing in the predicted path). Assessment if the parameters are selected correctly, can be performed in two ways: either by subjective opinion of the operator, or by defining a more common restrictive variable that one should not allow the vehicle to overpass and use it as indicator to see if the chosen criteria is “correct”.

C. Indicators related to the performance related to the situational control

For intervening systems, indicators are needed for measuring the improvement of situational control, i.e. to analyse the quality of the intervention. These indicators are especially related to intervening systems.

Examples of indicators are:

- Lateral acceleration for lane change assistant; In this case the lateral acceleration that is not a technical indicator in case of a warning system (it is more an indicator for impact) becomes a technical indicator for an intervening trajectory correction system. It can also be used as a comfort indicator.
- Duration of lateral excursion, i.e. the time spent by drivers outside the safety envelope round the centre of the lane.
- Longitudinal deceleration in brake assistant.
- Impact speed reduction for collision mitigation systems.

The indicators selected for the technical evaluation are directly influencing or interacting with the human factor related indicators reflecting the situational control. The degree of interaction between technical evaluation and human factors evaluation, with respect to indicators, depend on the type of function; whereas it is a warning function, an assisting/intervening function or a controlling function,

An indicator reflecting technical performance can be for instance the timing of the warning issued.

An indicator reflecting human factor performance can be for instance an indicator related to the driver's response to the warning such as the reaction time.

An indicator reflecting both technical *and* human factor performance can be for instance the overall response (e.g. starting at the time where the vehicle cross the border line of the lane and ending at the time when the vehicle is back in position in the target lane).

This indicator would then be composed by both the system response as well as the driver's response, both reflected in torque on the steering wheel.

An example of a warning system where the technical performance only influences the situational control is a warning system that issues a warning at a certain TTC, but does not take any actions itself. The situational control indicators are then only connected to the driver's response.

An example of a controlling system is the ESP (or in some applications ESC), which does not interact with the driver.

4.3 Test Definition

Once the different categories of assessment objectives have been identified, the next stage is to design the assessment task in more detail: tests need to be described. The result is a complete technical evaluating test plan for the two levels: system and components. Both test plans contain the scenario, the key indicators, the test method, a definition of the reference measurements, a description of the data collection and the measurement plan.

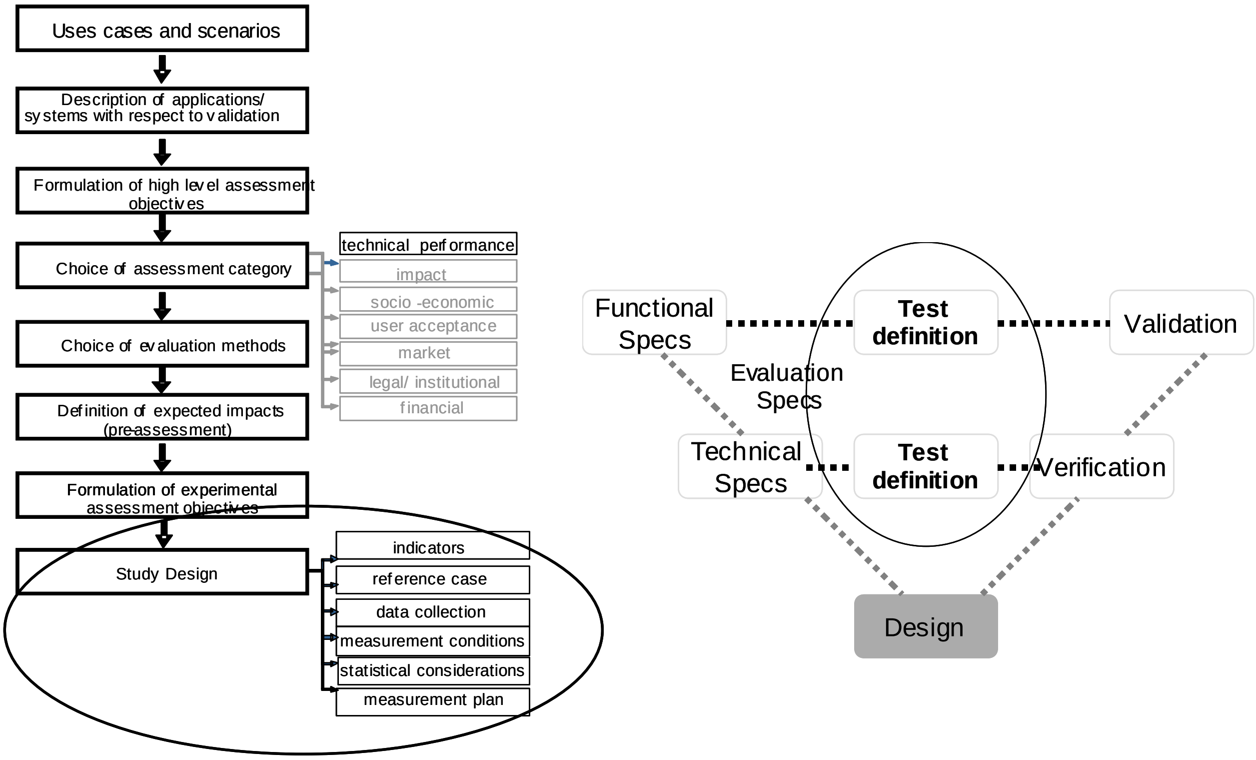


Figure 12: Part of the methodology addressed by the "Test Definition" section

4.3.1 Synthetic Scenario

This item addresses the global situation in which the function can be evaluated. It is naturally a traffic situation (validation level). It could be also a more artificial situation (verification level), which we call a "synthetic scenario".

The scenarios for the verification and the validation phase can be very similar. However, evaluating subsystems first is natural and conduct to a "more comfortable evaluation".

Example: the lane tracker subsystem can be tested in many different perturbed situations. In the validation phase, the whole system is tested, and the subsystems are considered to perform well. Scenarios are designed to verify if the system makes the correct decisions and takes correct actions in case of various conditions of road, speed, drift angles…

The test scenarios should be consistent with the use cases.

The scenarios are adapted to the level of interactions targeted. The table below gives examples that refer to PReVENT like functions classified in 3 groups: G1 short interactions, G2 medium interactions, G3 loose interactions.

Table 9 : Examples of scenarios for PReVENT like functions

	Scenario characteristics for assessing <u>subsystems</u>: verification phase	**Scenario characteristics for assessing <u>systems</u>: Validation phase**
G1 : ADASE dealing with short interactions: **Prevention of Collisions**	Test scenarios mainly related to the perception layers : Qualification of sensors through • Relative trajectories related to positive (targets should be detected) and negative tests (should no be detected when passing by) • Nominal and adverse environmental conditions.	Autonomous and semi-autonomous intervention : <u>Frontal collision</u> : stationary or moving targets; frontal or lateral collisions, single or multiple objects, various exposition durations, non-crash or pre-crash … <u>Collision in intersection</u> : on-coming traffic while turning left, crossing traffic while turning left or straight crossing
G2 : ADASE dealing with medium interactions : safe speed, safe distance, safe lateral interactions, safe lane changes…	Test scenarios related to the 3 layers : perception, decision, action. Emphasis on decision layer. The tests are made in simulation, test benches or test tracks Various vehicle interactions, road characteristics, lane markers… Nominal and adverse environmental conditions…	Tests are made in simulations, on tracks and open roads : The tests include various traffic conditions, driving parameters, operating conditions, sensor parameters, driver characteristics. Lateral interactions : various characteristics of relative trajectories in case of lane changes, drifts toward another vehicle or other objects…
G3 : ADASE dealing with loose interactions: **Awareness of hazards, hot spots, speed limits …**	Various technical simulated conditions to test modules : the localisation module, the communication module, the hazard position relevance check, the access to map attributes…	Various scenarios of reduced visibility, reduced friction, warning dissemination, arriving in hot spot, speed limit changes…

Comments :

- It’s important that the scenarios are easy to reproduce in the same conditions for more repetitions in order to have a significant and coherent data statistics.
- <u>Manoeuvres and trajectories</u>: manoeuvres have to be described in a very detailed way, containing all parameters that have to be varied (e.g. longitudinal and lateral velocities, positions and accelerations for host and target).
- <u>Dummy targets:</u> a clear specification of target objects in terms of target shapes and parameters relevant for the sensors technology (e.g. reflection coefficient, temperature of the object, radio signals the object can transmit,…) are needed.

- <u>Situational descriptors</u>: the conditions influencing the data collection should be as far as possible controlled and homogeneous.

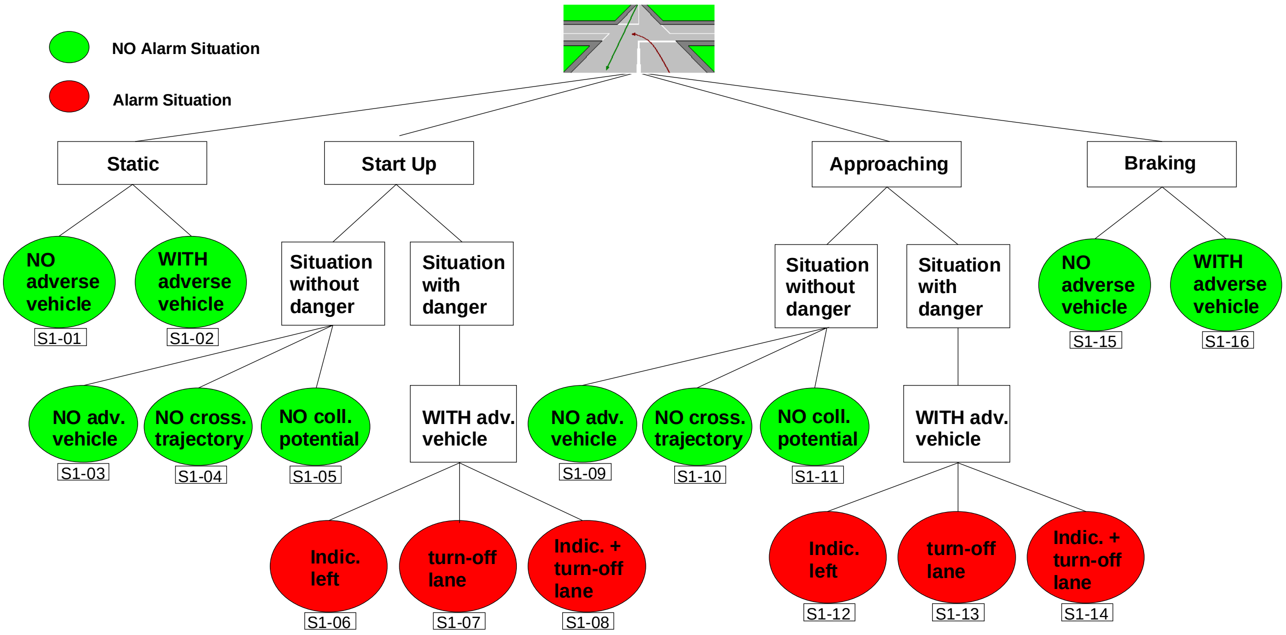


Figure 13: Example of scenario followed in INTERSAFE (PReVENT)

4.3.2 Tools and methods

4.3.2.1 Tools and methods: verification phase

Several tools can be used to verify the system components. Although there is a very large range of tools available, these can be grouped into three classes only, based on the application of either simulation models or real hardware:

- **Simulation,**

  During the design of the application, extensive use is made of simulation tools such as Matlab/Simulink. These can obviously also be used to evaluate the decision and control algorithms

- **Component Hardware-in-the-Loop (HIL) tests on dedicated test rigs**, to evaluate sensors (e.g. radar and reflector), algorithms and actuators

  Especially for sensors, the use of dedicated hardware-in-the-loop set-ups is customary, as is the case for control systems. It is very hard to provide a general description of such a HIL-set-up due to the application specific nature. Probably the most general set-up for this purpose is provided by PRESCAN (refer to section 4.6.2). Note that target representativeness is an important notion here in order to achieve sufficiently reliable test results. This can be achieved through *target standardization*.

- **Real World** (such as test track or road tests), e.g. to evaluate sensors, algorithms and actuators

  Although road tests primarily serve validation purposes (system level, par. 4.6.2), it might still be relevant to perform road tests for component testing, e.g. verification of sensor reliability. To this end, we can distinguish between testing on a test track with representative, standardized targets (see above), i.e. a relatively controlled environment, and real-world tests. For component testing, both types might be relevant.

It is in the nature of the verification process, primarily dealing with components rather than the whole (prototype) system, that emphasis is put on Hardware-in-the-loop tests. However, numerical simulation is sometimes used in the verification process as well. Several questions rise immediately: what can we assure from a simulation? How far is it from the reality? In sum, what are the advantages and the limits of the simulation in a validation process of preventive safety functions? Should we consider the cost/benefit? If a very accurate and complex simulation has to be established to test a sensor, can it be beneficial?

The validation through simulation has its place within subsystems and system complete evaluation as well (this case mostly for traffic taking into account and complex systems where road/track tests are not possible). An important fact is that simulation can be considered as a first step in a subsystems evaluation framework. As an example, in SAFELANE, the vision system has been validated through LCPC's simulator video sequences and after that with real recorded road data. These simulation test results were very performing. But if the subsystem had not passed the simulator test, it would have indicated that improvements would have been needed before any other tests were to be done.

4.3.2.2 Tools and methods: validation phase

The methods here have to be defined according to the validation procedure chosen as a function of sensor module characteristics, decision module characteristics and actuator characteristics. Then, based on the requests for the function reliability tests, road or track tests should be necessary, besides simulation:

- **Simulations:**

  During the design of the application, extensive use is made of simulation tools such as Matlab/Simulink. These can obviously also be used for evaluation. For this purpose, a promising possibility would be to build a virtual test environment, independent of the design test environments, within which the software module will be evaluated for various traffic scenarios, environmental conditions etc. This type of simulation is e.g. supported by VEHIL [24].

  Results of the simulations indicate which scenarios need to be tested further in Hardware-in-the-Loop tests and real world tests.

- **Hardware-in-the-Loop (HIL) tests:** the entire system, implemented in the car, can be tested in a dedicated test facility, allowing testing a vehicle in laboratory conditions, while realistic road conditions are being simulated.

  An example of such a facility is the TNO VEHIL test facility [9] where the complete vehicle with ADAS is placed on a chassis dynamometer. The vehicle is able to ride and brake as if on the road. The dynamometer simulates road behaviour based on a simulation model of the test vehicle. Other road users are represented by so-called moving bases, highly dynamic automatic guided vehicles. A central controlling computer creates a virtual relative world for the test vehicle and co-ordinates the interaction between chassis dynamometer and moving bases. The computer also provides a real-time visualisation showing the corresponding road behaviour of the test vehicle. This can be used as input for a vision system of the ADAS.

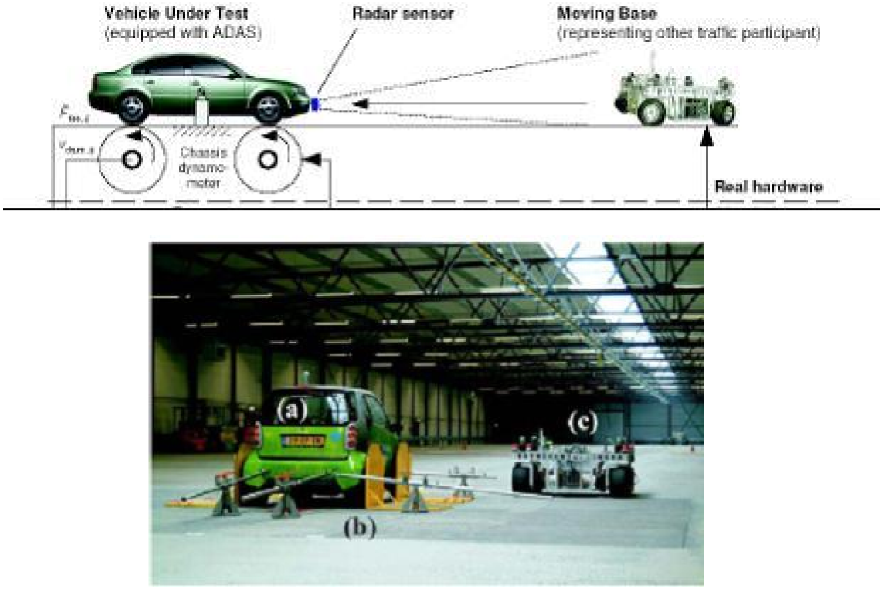


Figure 14 : VEHIL installation in TNO

- **Trials: Real World (such as test track or road tests).**

  When the system has been successfully tested in laboratory conditions, and has been assured that the system is robust to faulty operation by the users and is capable to recover from system failures, the system is ready for tests on test tracks and the road to evaluate the system in real world conditions. Trials can also be used to gather information for the human factors evaluation (Section 5.2.5.2) Trials can be performed with professional test drivers (Table 17), with end users (Table 18). For gathering information on long-term behaviour, FOTs can be used (Table 19).

4.3.3 Data collection and measurement plan

The measurement plans should include the following considerations:

- **Blocking**: the conditions influencing the data collection should be as far as possible controlled and homogeneous. A series of tests done in the same conditions is called a “block”. A block of tests refers to more or less the same conditions (of weather conditions for example).
- **Statistical relevance**: the number of tests performed should be related to the expected level of statistical confidence. For example, if N is the number of tests, f the observed

percentage, p the probability of the phenomenon, the following statements can be taken into account:

  - N=10; f=0,9; 0,71<p<1,
  - N=100; f=0,9; 0,84<p<0,96 with 95% significance i.e. 5% probability to be out of the interval(under hypothesis regarding p distribution)

The measurement plan should also guarantee:

- **Completeness:** concentrating the resources on most important aspects is better than spreading efforts with the consequence of a low statistical significance.
- **Insularity:** all the influence factors are considered
- **No disturbance** of the validation process: no bias except accidental ones introduced in the measurement plan

Example: in a thorough measurement plan, the lane tracker elaborated in the PReVENT/SAFELANE project, has to take into account different aspects linked to subsystems/functions tests on adverse/normal and on-line/off-line tests. The sensor was tested first showing that it was able to track a lane (with respect to several parameters like curvature, lane width, etc). Those tests were carried out partly on a simulator partly on the road. Nominal and different adverse conditions including many disturbing factors have been considered. This can be done off-line by recording first many sequences on road and then treating them in the lab. The same can be done for the decision system. At the end, the whole function was tested on test and real roads.

4.3.4 Reference case

The notion "Reference case" is used in two meanings, which are relevant for the evaluation. It refers both to:

- the "**reference situation**" that is, when evaluating an ADAS, the same tests are performed with the ADAS activated and deactivated. This allows to evaluate the improvement of situational control through the use of the system.

- the **"reference measurement",** which allows to evaluate the correct working of a component or system. This measurement, which is related to the evaluation indicators, should be accurate, so that it allows a reliable computation of the performance of the component or system. As an example, in order to compute if a lane departure warning system has activated correctly, e.g. when the vehicle front wheels touch the lane markers, another measure, in addition to lane detection, is required. This measure can be e.g. the differential GPS, if sufficient accuracy can be guaranteed. This is then the "reference measurement".

  It is however not always possible to provide an accurate reference measurement, and the assessment of correct working of the system can only be made on operator judgement. In this case, the computation of the performance of the system is more prone to errors.

4.3.4.1 Reference Case for the validation

- For the validation, the reference measurement system should allow to verify that all elements and parameters of the decision system are measured correctly.

The picture below illustrates the measurement scheme used in the French ARCOS project.

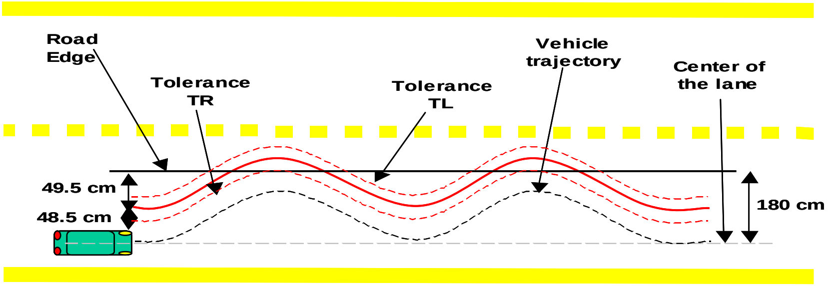


Figure 15: measurement scheme used in the French ARCOS project.

4.4 Execution and reporting

4.4.1 Verification

In this part the actual verification is performed according to the verification plan as described in the previous sections. The results of the verification of all sub-functions and subsystems of the total system need to be analyzed and reported. Conclusions about the performance with respect to the defined specifications and limitations that are revealed during the evaluation can be drawn here.

If the technical specifications are not met, either the specifications should be redefined, or the components should be improved such that the specifications are met.

Detailed verification of the subsystems also provides a kind of sensitivity analysis of the system performances. When the system is a quite early prototype (as it is the case for PReVENT functions) the results obtained by subsystems verification give a more precise knowledge of the topics to be addressed in further research.

4.4.2 Validation

In this phase of the evaluation, the real validation is executed according to the validation plan. The results of the validation of the system must be analyzed and reported. The performance of the system is evaluated with respect to the specifications and possible limitations are reported in this phase.

If the functional specifications are not met, either the specifications should be redefined, or recommendations should be given to redesign the system such that the specifications are met.

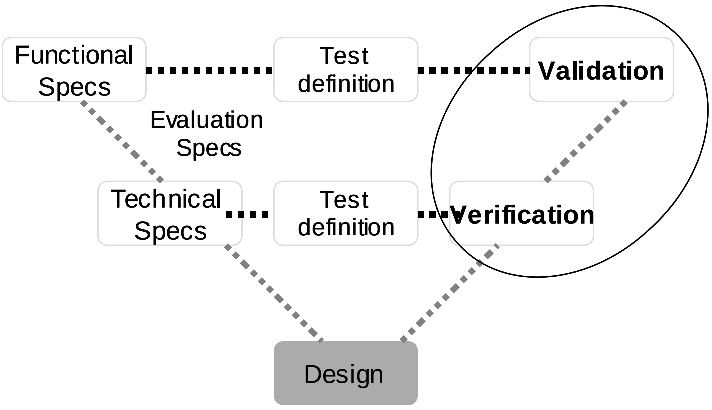


Figure 16 : Part of the methodology addressed by the validation and verification step.

4.5 Best practices

In this section, we extract some important features that can be illustrated through the actual validations performed in PReVENT subprojects.

Assessing subsystems prior to assessing the complete function: although this point can be judged as of no use in the evaluation of overall function performances, its application in several subprojects (WILLWARN, SAFELANE…) show that it provides sound assessment elements like the knowledge of some limitations and a kind of a sensitivity analysis to operational conditions…

A central point in the methodology: ways to measure the indicators of success: the “reference measurement”: in an evaluation procedure, two kinds of reference measurements are needed, spatial and temporal ones. These two kinds of measurements can be of absolute or relative nature. Two good examples of those practices can be found especially in APALACI and INTERSAFE.

Combination of simulation environment with hardware-in-the-loop tests: the validation phase combines a high number of scenarios that can be performed in a first stage through a simulation study with identification of most critical cases for subsequent test in a hardware-in-the-loop environment: SASPENSE follows this process.

Representativeness of the tests, lacks of standards: this topic relates to the realism of the interactions considered in the tests carried out (either in reality on test tracks, open roads… or in simulation) to assess the technical performances. A central point regards the definition of the dummy targets against which the perception systems are confronted. Due to the absence of standards, the diversity of target parameters (shape, colours) is very high in PReVENT SPs (the same in all projects of the same nature). A methodology begins to exist to decide the characteristics of the targets in the case of radar and lidar (cf. ISO15622); there is no such standard element for deciding how to define targets used to validate sensors based on cameras and image processing. Annex 3 gives a first contribution to this issue.

4.6 Conclusions and Final Remarks

Based on the experiences gained in the PReVENT subprojects, a framework has been defined and commented to guide future projects in the technical assessment phase.

The assessment process starts from the general CONVERGE scheme in which the V-shape design cycle is introduced.

The framework starts with the definition of the targeted functionalities. This phase makes a distinction between functional specifications (at system level) and technical specifications (at subsystem level. The result of this first phase identifies as clearly as possible the technical elements (at system and subsystem levels) that should be evaluated

- General information
- Functionality and uses cases
- Accidents targeted (what type of accidents are we dealing with)
- Description of subsystems including some possible intrinsic limitations (e.g. sensor FOV) or limitations with respect to operational conditions.

The second step of the PREVAL assessment framework allows to progressively formulate the test plan that includes the definition of

- Assessment objectives (technical translation of the function objectives)
- Key indicators (way to measure the success),
- Reference measurement (how to measure the indicators),
- Test scenarios (in relation to the functional use cases)
- Measurement plans (sound statistical framework)

The last step relates to the testing itself (validation at system level and verification at component level) and the description of the results.

The framework description is illustrated by examples and best practices drawn from the PReVENT subprojects.

As standards do not yet exist, PREVAL points out some key elements that allow to conceive a sound assessment process:

Particular attention has been given to the identification of a compromise between addressing the full complexity of the driving context and the limited resources allowed in the evaluation process (identification of use cases, test scenarios –positive and negative, dummy targets- are of prime importance in this process)

<u>A second point of importance is related to the statistical aspects</u> : blocking tests of the same nature, avoiding bias, considering disturbing identified factors…were identified

<u>A third aspect is related to tools and methods used</u>. To this regard, the automotive sector is presently facing a major difficulty as the aeronautic one was some decades ago: how to assess a complex technical process that should make very limited faults whose probability is very low (less than 10^{-8}). New tools and methods should be introduced: dedicated HIL test benches or pure digital simulations for example.

5 Human Factors Evaluation procedure

5.1 General approach

5.1.1 A framework for ADAS human factors evaluation

The present chapter describes the general framework proposed for human factors evaluation of Advanced Driver Assistance Systems (ADAS) developed in PReVAL. A human factors (HF) evaluation potentially involves all aspects related to the driver's response to the system, including acceptance, usability and (intended and non-intended) behavioural effects that can be assessed empirically[1].

On the most general level, the goal of an ADAS HF evaluation is to assess the extent to which the system succeeds in generating the intended behavioural responses from the driver. This depends on a number of factors including acceptance, usability and the technical performance of the system. The HF evaluation should also address potential unintended (side-) effects of the system.

In order to obtain a common view of the overall aim of an ADAS evaluation, incorporating also the technical aspects, the concept *situational control* was introduced (see Section 3.1). Briefly, situational control refers to the level of control jointly exerted by the driver and the vehicle (including ADAS) in a specific driving situation. The general purpose of an ADAS could then be understood as the enhancement of situational control. Consequently, the general goal of ADAS evaluation is to assess the extent to which this is achieved. The human factors part of the evaluation then focuses on the driver-related[2] impacts on situational control.

In the following section, the role of human factors evaluation within the general ADAS assessment procedure is described. The later sections include a more detailed description of the proposed HF evaluation framework.

5.1.1.1 Role of human factors evaluation in the general design and evaluation process

Human factors evaluation is often conducted as part of a general ADAS assessment procedure which may also include the evaluation of technical performance as well as safety impact analysis on the aggregate level (e.g. the reduction of accidents of a certain type). The framework for the technical evaluation of ADAS functions, proposed by the PReVAL project, has been described in Section 3.2, and is depicted in Figure 4. In terms of this framework, *validation* deals with functional-level evaluation of performance while *verification* deals with subsystem and component-level evaluation. The technical evaluation (addressed in Section 4) includes both validation and verification, while the

---

[1] The term "human factors evaluation" was chosen rather than "HMI evaluation", since the latter could be interpreted as only dealing with the operation of the human machine interface.

[2] Potentially, this could also involve passengers and other road users. However, the present methodology focuses on driver-related factors.

human factors evaluation only deals with validation (functional-level).

5.1.1.2 Human factors ADAS evaluation

This section describes the framework adopted for ADAS human factors evaluation in PReVAL, based on the situational control concept described in section 3.1.

Key distinctions

The proposed framework is based on a number of key distinctions. The first can be made between *intended* and *unintended* effects. Intended effects are those intended by the designers and normally implied by the functional specification (e.g. reduced response time as a result of Forward Collision Warning). By contrast, unintended effects refer to side effects that were not anticipated by the designers. It is important to stress that unintended effects may be both positive and negative for safety (An example of a positive effect would be the increased use of the turn signal as a result of lane departure warning. An example of a negative unintended effect would be reduced headway and increased speed when using Antilock Braking Systems, ABS).

A second distinction could be made between *long-term* and *short-term testing.* For present purposes, “short-term” refers to tests within the time frame of hours, i.e. tests that can be realistically performed in a controlled setting, e.g. a simulator. By contrast, “long-term” refers to tests with exposure times at the time frame of days, weeks, months and even years, investigating the process whereby the driver integrates the system into his/her general driving activity. A typical example of this is field-operational tests (FOTs). In a HF evaluation procedure like the one developed in PReVAL, empirical measurement is only feasible for short term effects. However, it is important to also consider potential long term effects in the HF evaluation, even if they cannot be addressed empirically. One way to do this is to review results from existing long-term studies.

A third distinction concerns the *specificity of the target situations*. The target situations for an ADAS function are the situations in which the function has its intended effects. Some ADAS functions, such as e.g. Intelligent Speed Adaptation or Adaptive Cruise Control are targeted for rather general situations (e.g. non-urban driving above 60 km/h), while others are designed to take effect only in very specific situations. For instance, a Curve Speed Warning is expected to reduce speed when entering curves and Forward Collision Warning is only activated if the distance to an obstacle in front is critical. These situations are often, but not necessarily, time- and/or safety-critical. For some functions, such as e.g. Lane Departure Warning, one could look both on the effects in general scenarios (“normal driving”) as well as the effects in specific lane-departure situations (in SAFELANE, the latter option was chosen, see PReVAL D16.1 [1]). The specificity of the target situation is naturally of great importance for the design of the evaluation study, especially the scenario definition.

Finally, behavioural effects of ADAS functions could be classified with respect to Michon's [18] three levels of the driving task (operational, tactical and strategic). For example, an (intended) effect of an ABS on the operational level is enhanced braking control (e.g. in terms of reduced skidding). However, this may be compensated on the tactical level by increased speed or reduced headway (e.g. [19]). An example of an effect on the strategical level would be the decision to stop and take a break after receiving a warning from a drowsiness detection system.

Dimensions of ADAS HF evaluation

The main goal of most ADAS functions (i.e. all functions that are not entirely automated) is to change driver behaviour in some way (e.g. slowing down or reacting more quickly to a sudden hazard). Thus, according to the present framework, the overall goal of HF evaluation is to assess the ability of the ADAS function to do this, that is, to affect behaviour-related situational control parameters in the intended way. This ability can be viewed as dependent on two main factors: (1) *acceptance* and (2) *usability*.

Usability refers to how easy the system is to use for the intended purpose, as determined mainly by the human-machine interface design. In the case of ADAS functions, a key issue is naturally the extent to which the warning is comprehended in the intended way by the driver. Usability could be conceptualised in terms of the match between the user's mental model of how a system works and the actual operation of the system [12].

Acceptance relates to the extent to which the driver views the system as useful and satisfactory. This is determined e.g. by the technical performance (e.g. if warnings are not precise enough, the driver may be annoyed and loose confidence in the system) as well as usability (if the system is difficult to use due to bad design of the human-machine interface, the acceptance will be affected).

A HF evaluation study may involve both the direct assessment of behavioural situational control parameters (e.g. whether the intended change in speed occurs) and the specific assessment of acceptance and/or usability. The general relations between these dimensions are illustrated in Figure 17.

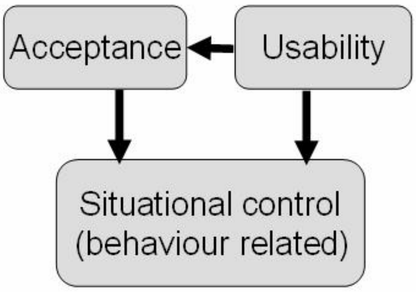


Figure 17 Relation between the main dimensions in ADAS HF evaluation

Scope of PReVAL ADAS HF evaluation methodology

Summarising the framework outlined above, the main scope of PReVAL HF evaluation methodology can be defined as:

- Both intended and unintended effects (with the emphasis on the former)
- Short-term testing (long-term effects should be considered but not addressed empirically)
- Target situations, depending on the goals of the study
- Functions that support the operational and tactical level of the driving task

The scope of the HF evaluation in PReVAL is presented in Figure 18.

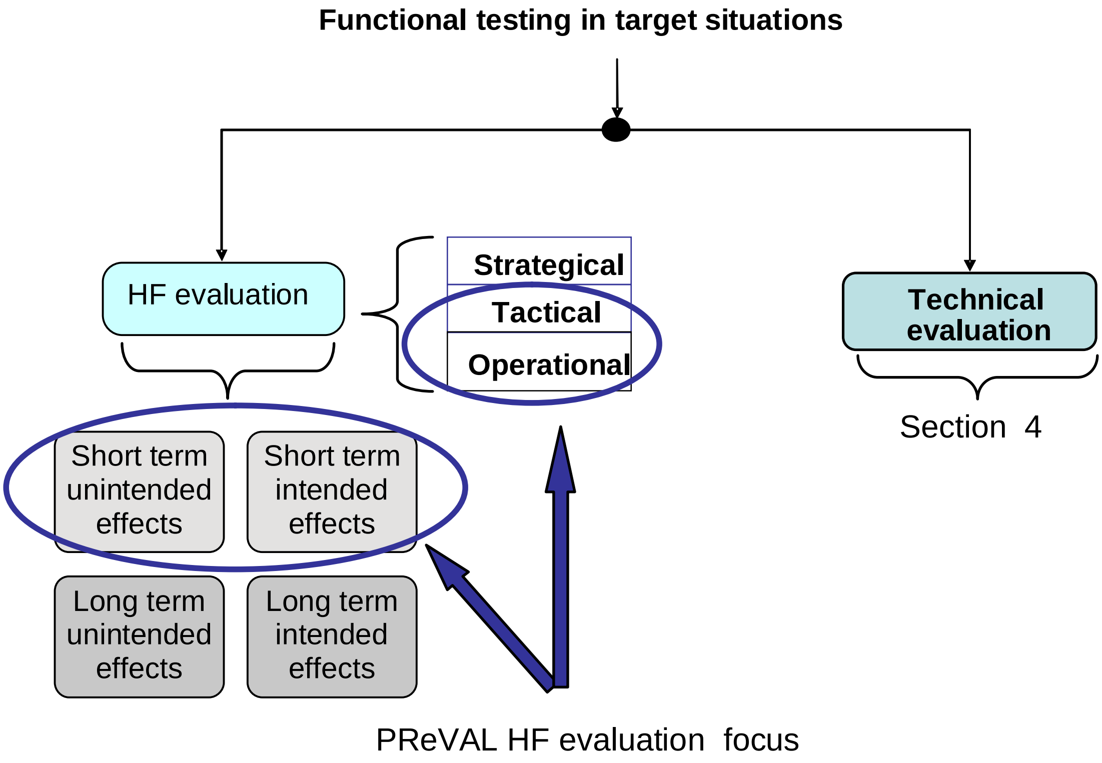


Figure 18 Scope of PReVAL Human Factors evaluation

5.2 Procedure

5.2.1 Overview

This chapter provides an overview of the proposed procedure for ADAS human factors evaluation and its relation to the general PReVAL evaluation framework. Figure 19 presents a general overview of the proposed procedure

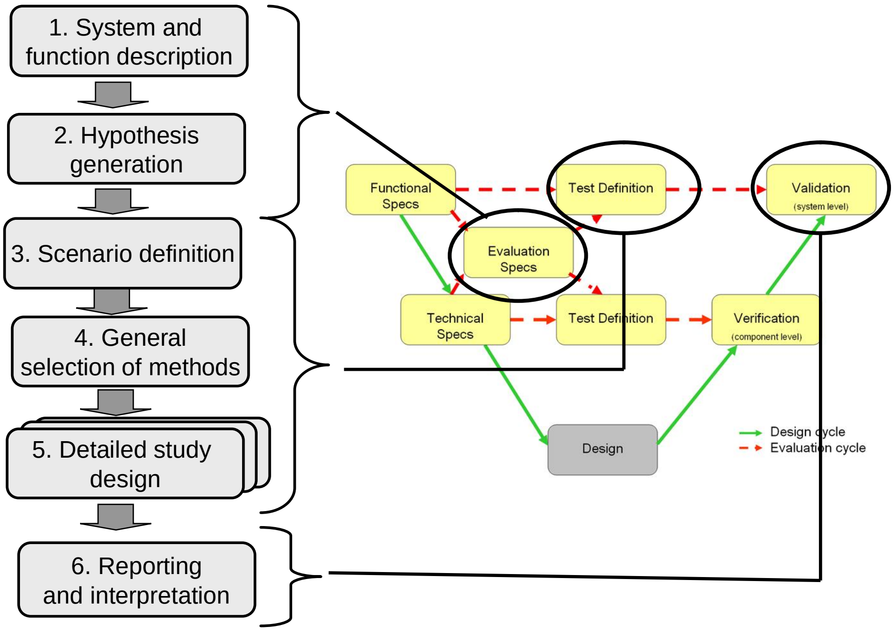


Figure 19 General procedure for human factors evaluation and its mapping onto the general evaluation framework adopted in PReVAL.

Step 1: System and function description

In this step, the overall system and its individual functions are described. The input needed is included in the "System and functions description", described in Section 4.2.1. Compared to the technical evaluation, technical details are of less relevance, but the description of the use cases and the HMI description are the most important.

Step 2: Hypothesis generation

In this step, the main hypotheses to be addressed by the evaluation should be clearly stated. The hypotheses should generally be about the potential effects of the functions on situational control, or the factors affecting it, i.e. usability and acceptance (see Figure 17). Hypotheses about the intended effects should normally follow directly from the functional specification. In this step potential characteristics for evaluating the hypotheses should be derived by identifying corresponding indicators. The corresponding parameters and metrics are then specified in Step 3 when defining the detailed test scenarios. The characteristics could be described quite generally, for example *improved lane keeping*. A corresponding indicator could then be for example the variation in lateral position.

Hypotheses about unintended effects are also of importance to consider. However they need to be based on existing empirical or theoretical knowledge and will generally not follow directly from the system specifications. They should be included when motivated based on existing research.

Step 3: Test scenario selection

In order to address the general hypotheses defined in the previous step, general test scenarios are selected. The key issue here is to define the level of specificity of the scenarios and select the driver, vehicle and environment parameters that are most relevant for testing the hypotheses defined in the previous step. The parameters to be measured should also be identified in this step. The information collected in this step should be integrated in the evaluation specs and are similar to the test scenarios defined for the technical evaluation. The specific test scenarios, defining the detailed evolution of events (e.g. exactly when a lead car should brake) are defined in step 5.

Step 4: General selection of methods

In this step, the general methods to be used for addressing the stated hypotheses (Step 2) in the selected test scenarios (Step 3) are defined. In general, the selection of an evaluation method depends on the quality of results needed, on the availability of resources, the stage of development and the effects that the experimenter is interested in, i.e. those defined in the hypotheses. The choice of indicators also affects the selection of methods.

A simple example could be the choice between a test track and a driving simulator study. In some cases, several types of methods may be chosen.

Step 5: Detailed study design

This step involves the detailed design of the study/studies to be performed, including the number and characteristics of subjects and the experimental design (e.g. within/between groups), the dependent and independent parameters, definition of metrics (including operational definitions), methods for data analysis and general study logistics. This also includes the detailed test scenarios (e.g. exactly when a lead car should brake).

Step 6: Interpretation and reporting of results:

This involves the actual execution of the study, the subsequent data analysis and the final interpretation and documentation of the results.

5.2.2 Step 1: System and functions description

5.2.2.1 Objective

To identify the functionality based on the functional specifications in order to be able to derive hypotheses on human factor related effects of the function.

5.2.2.2 Approach

The description of the functions and the system is described in Section 4.2.1.

5.2.3 Step 2: Hypothesis generation

5.2.3.1 Objective

The intention with this step is to define clearly the prerequisites and goal with the evaluation. This is done by identifying the hypotheses, or theories that shall be tested, on what kind of effects the ADAS has on the driving performance and driver behaviour in a particular target situation.

The evaluation itself aims to provide results in terms of scientific facts that can be used for conclusion on whether the theory/hypothesis is correct or not, thus the hypotheses can either be verified or disproved when assessing the evaluation results.

The hypotheses shall be defined primarily in terms of *intended effects* but also in terms of *unintended effects*.

5.2.3.2 Approach

Intended and unintended effects

A main distinction is needed between *intended* and *unintended* effects. As mentioned in the introduction chapter, intended effects are effects that are intended by the designers of the system. These are normally included in the system descriptions, thus the hypotheses of intended effects should to large extent be possible to derive directly from the system description. For instance, an intended effect of a lane departure warning system is improved lane keeping performance (e.g. reduced number of lane exceedances).

In addition to designed or intended effects on driver behaviour, the ADAS functions may have unintended effects on driver behaviour. ADAS functions may change user behaviour indirectly in many, largely unknown ways, where the driver adapt to the changed situation.

As Saad et al. [5] indicate; “these systems will mediate drivers' interactions with their driving environment (vehicle, road infrastructure and other road users) by creating new sources of information and/or offering new modes of action regulation. They will thus alter the conditions in which the driving task is currently performed and, as a result, changes in drivers’ activities can be expected.”

Unintended effects refer to side effects that are not anticipated by the designers of the systems. These could be either positive or negative. A potential positive indirect effect of a lane departure system is the increased use of the turn indicator (due to the fact that, in most systems, the warning is typically turned off upon activation of the turn signal). An example of a potential negative effect is the misuse of the system as an alarm clock.

In contrast to the intended effects, the unintended effects are rarely covered in the system description, and in general these are very difficult to define and predict in advance. To be able to derive theories on unintended effects in advance you need to investigate the existing empirical or theoretical knowledge. They should be included when motivated, based on existing research. In this document, the unintended effects are not the focus; however it is of importance to take them into account.

A common concept in this context is *behavioural adaptation*, which has been defined by an often cited OECD report [20] as “those behaviours which may occur following the introduction of changes to the road-vehicle-user system and which were not intended by the initiators of the change” (p. 23). Thus, according to this definition, behavioural adaptation only includes unintended effects. However, other authors have proposed more general definitions, e.g. Summala [21] who proposes that “a general definition of behavioural adaptation says that the driver is inclined to adapt to changes in the traffic system, whether they be in the vehicle, in the road environment, in road and weather conditions, or in his/her own skills or states, and that this reaction occurs in accordance with his/her motives” [21]. For present purposes, the latter definition will be adopted.

As briefly mentioned in the introduction behavioural effects of ADAS systems can be classified with respect to the levels of Michon. Michon [18] defines a three-level model of the driving task, where a distinction is made between

(1) the strategic level (e.g. general trip planning)

(2) the tactical level (e.g. deciding to overtake) and

(3) the operational level (online vehicle handling).

Based on this distinction the different effects, both intended an unintended of an ADAS system can be characterised in terms of the matrix as in Table 10. The hypotheses should provide the goals of the tests performed with mapping closely to Michon’s levels. Statements on what effects to test and on what level according to Michon should be made.

Table 10 Michon’s levels for characterising behavioural effects of ADAS

Driving task level	**Type of effect**	
	Intended	**Unintended**
Strategic		
Tactical		
Operational		

The hypotheses generation in this procedure cover mainly how to generate hypotheses on intended effects on driver behaviour and driving performance, which are evaluated by defining indicators and parameters for quantitative measurements. In addition subjective data on acceptance and usability should be collected in the evaluation process and when defining the hypotheses, since these are important factors to assess.

Driver performance/behaviour, acceptance and usability contribute together to the overall effects on situational control as presented in Figure 17.

Hypotheses generation- intended effects

The hypothesis generation of intended effects can be divided into the following parts:

1. **Analysis of use cases from system description**

   In this step the functional description of the system and its use cases shall be analysed from the driver interaction perspective, hence what are the intended effects on driver behaviour and driving performance in different target situations.

2. **Categorisation of use cases.**

   In this step the intention is to break down the uses-cases into appropriate events connected to a particular target situation where a specific driver behaviour is expected. The intention is to categorise the use cases so that they clearly describe a specific driving situation with specific expected driver behaviour. A use case might consist of a chain of events, each with a specific driving manoeuvre and expected driver behaviour, and it might be of importance to separate these sub-events since there might be different indicators to be tested and measured for each event.

   For example, if a lane departure warning system issues an auditory warning when the driver's vehicle has crossed a lane marking, with intention to make the driver steer the vehicle back in the lane, and the system also provides an active steering support once the driver has started to steer back, this chain of events might be seen as two separate, but connected events and two separate hypotheses:

   - Hypothesis 1: In case of an unintended lane departure, the system issues an auditory warning for making the driver alert to the lane departure. The driver understands the warning, accepts it and responds to the warning by re-allocating attention to the road and making a corrective steering manoeuvre to avoid a full lane departure.
   - Hypothesis 2: In case the driver starts to steer back to the correct lane at an unintended lane departure, the system provides an active steering support in terms of a torque on the steering wheel. This active steering support helps the driver to reach the target lane faster than he/she would have done, without the support.

   The example above shows that the use cases from the system description might be necessary to break down in order to define distinct hypotheses.

3. **Definition of driving performance characteristics.**

   In this step the characteristics for evaluating the effects on driving performance by should be defined together with potential indicators. The indicators and the corresponding parameters for evaluating are then further defined in step 3 Examples of indicators for the above stated hypotheses could be reaction time (hypothesis 1) and time delay for steering back into the correct lane (hypothesis 2) and driver's subjective opinion on usability and acceptance of system feedback (hypothesis 1 and 2).

An example of hypotheses and indicators for driving performance on a lane departure warning system is presented in Table 11.

The hypotheses are named H1...H4 and are mapped to the Michon's tactical and operational level as well as divided in to short term and long term effects.

Table 11 Hypothesis generation for intended effects

Hypothesis generation for intended effects of lane departure warning system				
	Short term		**Long term**	
Michon level addressed	**Target situation**	**Driving performance indicator**	**Target situation**	**Driving performance indicator**
Tactical level	**H1**:System will reduce number of lane ecxeedances.	Number of lane exceedances		
	H2: In case of an unintended lane departure the system alerts the driver on the lane departure. The driver responds to the warning by re-allocating attention to the road and making a corrective steering manoeuvre to avoid a full lane departure.	Steering response time Time/area spend outside lane boundary		
Operational level	**H3**: In case of an unintended lane departure the system supports the driver to steer back in to the target lane, by supplying a torque on the steering wheel directed towards the target lane.	Time elapsed between start of corrective steering manoeuvre and time for correct positioning in the target lane. Time/area spend outside lane boundary	**H4:** System will decrease the lateral drift in the lane.	Variation in lateral position.

Hypothesis generation – unintended effects

This section provides an overview of the mechanisms of unintended effects. It should be emphasized that since it is difficult to define hypotheses on unintended effects, the procedure presented in this document focus on evaluation of intended effects. However the unintended effects are important to consider to greatest possible extent.

One way of deriving potential unintended effects from literature could be to review what is available regarding observed unintended effects of similar systems, earlier launched or if there are knowledge on acceptance and usability from assessment of similar HMI's. Further knowledge on unintended effects might be achieved by performing FOT's.

Many researchers (e.g., Draskóczy et al. [4], Saad et al. [5]) have listed various unintended effects of ADAS functions that may be present. The following list is based on the analysis of Draskóczy et al. [4] (see also section 6.1):

- Modification of user behaviour

  Behavioural adaptation will often not appear immediately after a change but may show up later and even after years of usage. This delay is one factor that makes it very hard to predict and measure unintended effects. Behavioural adaptation may appear in many different ways (e.g. by change of usage of the car, by change of headway in a car following situation, by change of expectation of the behaviour of other road users). For more discussion of behavioural adaptation (concept and findings), see Saad et al. [5]. In addition, the distraction resulted from the use of function can be included in this category. For example, if the function provides warnings, there is always a question whether this additional information has harmful effects beside of potentially beneficial ones. Furthermore, there have been lots of discussions dealing with delegation of responsibility to ADAS functions which can lead unintended modification of driver behaviour.

- Modification of interaction between users and non-users

  ADAS functions will change the communication between equipped road users. This change of communication may influence the traditional communication with non-equipped road users. To a large extent this problem may appear in the interaction between drivers and unprotected road users. In addition, ADAS functions can have effects on traffic flow (e.g. ACC) and thereby on interaction between users and non-users.

- Modification of road user exposure

  This is certainly an area where ADAS will have a large impact for example by changing travel pattern, modal choice, route choice etc. For example, well-accepted and attractive function (and HMI) is likely to increase mileage. Furthermore, it is likely that there are many functions that affect route choice, especially if the use of the function is affected by the road environment (road type, intersection type, etc).

In hypothesis generation of potential unintended effects we can assume a priori that indirect modification of user behaviour typically occurs at the tactical and operational levels, modification of interaction between users and non-users at the tactical level and modification of road user exposure at the strategic level. Examples of hypotheses on unintended effects connected to Michon's levels are provided in Table 12.

Table 12 Example for lane departure warning function

Driving task level	Type of effect	
	Intended	**Unintended**
Strategic		Drive longer hours without break
Tactical		Increased use of turn indicator Increased speed
Operational		

The hypotheses generated on unintended effects can be quite general but they should be based on relevant literature. It is acknowledged that empirical results presented in the literature may be insufficient, too general or missing. However, all relevant findings (supporting or not) the hypotheses should be presented.

This phase is related to the analysis of the safety mechanisms in the safety impact analysis (see Section 6.2.3). The results of the hypothesis generation provides input to the safety impact analysis.

The final step is to generate specific research questions based on the hypotheses. These research questions specify which aspects of human behaviour will be measured, scenarios, metrics, etc. An individual study usually covers only some hypotheses.

5.2.4 Step 3: Test scenario selection

5.2.4.1 Objective

Identify and describe on a general level the test scenarios to be included in the evaluation of the hypotheses defined in step 2.

5.2.4.2 Approach

The test scenario is formally defined in terms of a set of *parameters* (indicators) and their temporal evolution. Thus, a test scenario could be viewed as an instance of a *situation* (see p. 8). The key issue in defining the test scenario is thus to identify the parameters that are *relevant* given the hypotheses defined in step 2.

A general view of a scenario, adopted from ADVISORS [3], is illustrated in Figure 20. As illustrated in Figure 20, a scenario specification includes parameters describing the driver, the vehicle, the road infrastructure, traffic conditions and environmental conditions. For simplicity, the three latter categories will henceforth be merged into a single Environment category.

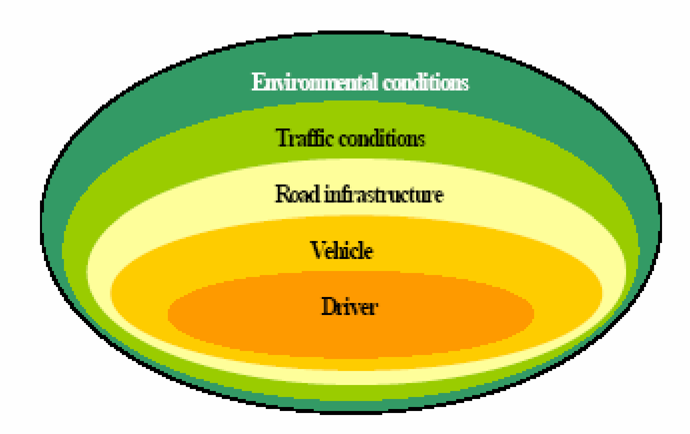


Figure 20 Illustration of scenario, adopted from ADVISORS [3]

The test scenarios should, at this stage, be defined in terms of: (1) the general Driver-Vehicle-Environment conditions; (2) a narrative describing the temporal development of events, including the intended effects of the function; (3) a list of *potential indicators* to quantify these effects. This should be directly based on the results from the two previous steps. The narrative may also be complemented with illustrations, e.g. pictograms of the general flow of events. A template for this information is given in Table 13. One such table should be completed for each defined evaluation scenario. The intended use of the template is illustrated with an example of lane departure warning evaluation.

Table 13: Template for test scenario specification, with a hypothetical example from evaluation of a lane departure warning function

Test scenario X: Lane departure resulting from visual distraction/occlusion	
General Driver-Vehicle-Environment conditions	
Driver (of host vehicle)	
Age	25-55
Gender	Male and female
Driving experience (total mileage)	>10000 km
Vehicle	
Vehicle type {passenger car, light truck, heavy truck, bus, N/A}	Passenger car
Environment	
Time of day	Day (daylight conditions)
Weather conditions	No precipitation or fog (good visibility)
Road type	Rural road, two lanes, speed limit of 70-90 km/h, lane width of approximately 3.5 meters
Road surface condition	Dry surface
Road segment type	Straight driving
Surrounding vehicles	Free driving (no surrounding vehicles within a time headway of 5 seconds)

Specific flow of events

Narrative

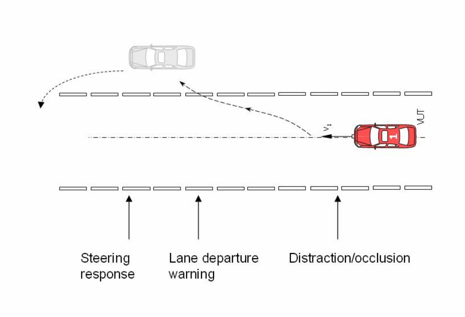
Prior to ADAS activation

The driver drives normally in the right lane at a speed of 70-90 km/h. At a certain point in time, the driver gets visually distracted or occluded which induces a heading error which, in turn, leads to a lane departure. The lane departure warning system is then activated (given the warning trigging criteria).

Intended effect

The driver responds to the warning by re-allocating attention to the road and making a corrective steering manoeuvre to avoid a full lane departure.

Illustration



Main potential indicators:

Steering response time

Time/area spend outside lane boundary

…

5.2.5 Step 4: General selection of methods

5.2.5.1 Objective

Identify general methods (e.g. simulator, test track, expert evaluation etc.) suitable for addressing the stated hypotheses/research questions.

5.2.5.2 Categorisation of methods

The evaluation methods can be grouped into three domains [7]

- Inspection methods,
- Inquiry methods
- Testing methods

Inspection Methods are used to inspect usability related aspects of a system. Inspection can be used to assess a generic system at any time during the development phase. This is particularly useful in order to identify potential problems at an early stage of the development, and not overload developers with wrong or inconsistent elements. Example: Expert Panels.

Inquiry Methods rely on users to collect qualitative data about the system. They can involve a wide range of people, providing sensible data to examine. They are a subjective source of data, that is they might be influenced by personal likes, needs, background. Because of this, people who participate in inquiry methods must be carefully chosen in order to reflect an accurate statistical sample. This way, it is possible to give consistency and validity to observed data. Example: HMI Concept Simulations, Simulator studies based on with subjective evaluations. The results from the subjective evaluation can consist of numbers/grades within a certain range that can be compiled and quantitatively assessed, or it can be verbal statements that are only used for qualitative assessment.

Trial Methods work on a representative set of users, whose task is the utilisation of a system, or prototype, in order to allow for an assessment of the human factors related changes in Situational Control induced by the system. In this context, it is important to understand how much the system impacts driving performance, that is whether the driver can safely interact with the system while maintaining full control of the car, how the system influences the driver's performance, and providing no harm to him/herself, the car, pedestrians, other vehicles, etc. In the following we will concentrate on testing methods applicable for mature prototypes or systems close to market, since the Human Factors evaluation procedure concentrates on the validation of mature prototypes, whose components have already been verified. We assume that critical situations due to system failures which cannot be remedied do not occur. Example: Test drives with professional test drivers or subjects providing data for objective quantitative assessment on test tracks or on public roads, Field Operational Tests (FOT). The outputs from the tests are quantitative data that are quantitatively assessed.

Trial methods are also used in the technical evaluation to gather information on technical performance and reliability. (Section 4.3.2).

Often all of the evaluation methods presented above are used, in order to get a complete base for the evaluation based on both qualitative and quantitative data.

5.2.5.3 Method selection

The main input variables for the selection of one or more methods are:

- **Effects:** The effects of the system under evaluation are a main input parameter for the selection of methods. Effects can be distinguished according to the time scale in which they might occur (long term/ short term), and represented as intended/ unintended effects.
    - Time scale
        - **Short term effects** can be observed during **simulator studies** and all kinds of **test drives** with professional test drivers or subjects.
        - In case **long term effects** are an important aspect of the evaluation, simulator studies or test drives where

    an experimenter is necessary all the time are usually not applicable. In that case it is necessary to carry out large scale tests (**Field Operational Tests**), to provide subjects with test vehicles for a longer time period (days, weeks, months) and to record data automatically during the every day use. In case budget does not allow the necessary apparatus, one might consider **expert panels** as a valuable source of data for long term effects. Such a panel should consist of experts from different disciplines like behavioural scientist, traffic psychologists, development and traffic engineers.

  - Intended/ unintended effects
    - **Intended effects** are indicated by the hypotheses, which give direct advice what to look for while the identification of **unintended effects** is a difficult task since usually the experimenter does not know what he should look for. The support from experts, again in form of an **expert panel**, accompanied by literature studies on similar systems, can give valuable hints what kind of unintended effects might occur. The detailed experimental design should then be modified or amended accordingly.

- **Scenarios**: The distinction of target situations also affects the selection of methods:
  - If mainly **target situations** with a clear definition of circumstances with respect to independent variables like the driving situation and environmental conditions are addressed in the study, tests either with **professional test drivers** or tests with **naives**, both on closed **test tracks** should be conducted. Professional test drivers can give very exact evaluations of the system and can also easily handle critical driving situations. In that case they act as a “gold standard”. Naives can act in the relative security of the test track and can fully concentrate on the vehicle system without having to deal with high demands resulting from real traffic situations.
  - In case behaviour in other driving situations but those reflected by the target situations is of interest, **test drives** with **naives on public roads** accompanied by the experimenter, or **Field Operational Tests**, are inevitable. One has to consider bias effects in case the subjects are observed by a human observer. FOT can give an extremely valuable set of data about all kinds of effects in all kinds of situations and they are the only way to analyse the behaviour of drivers in both, normal driving and target situations in their natural driving environment with the natural occurrences of all kinds of circumstances. On the other hand they are the most expensive kind of study, both in terms of time and financial resources. The analysis of huge amounts of data is an additional effort which has to be considered.

- **Reliability of Results**: The intended **reliability of the results** in terms of statistical significance and power is one of the major boundaries when selecting an appropriate method and is clearly linked with **resources**. Considerations on sample size and sample composition are included in Section 5.2.7.
- **Build Status**: The **build status** of the system under evaluation sets natural boundaries for evaluation activities. If tests with users in prototypes or with close-to-market systems are conducted, the system must be robust to faulty operation by the users and the systems must be capable to recover from system failures.

5.2.5.4 Overview of methods for evaluation

The following non-exhaustive tables comprise of the main aspects, pros and cons and general information of the different methods, built on Response-3 'Code of Practice work' [23].

Table 14 Expert panels

Inquiry methods	
Method	**Expert Panels**
Description	Expert Panels can be used in any stage of development. The panels can make their assessment on the basis of professional interpretations of verbal and written descriptions or prototypes of systems in all stages of development. Usually they are used when quantitative data is lacking or the availability of resources precludes quantitative assessment based on experiments. Depending on the system, different expertise is needed. An expert panel may include research and development engineers, human factors specialists, accident research specialists etc.
Pros	Beneficial cross-checks will result from the different points of view gathered, as experts can come from different technical horizons Rapid detection of critical problems and their causes can be achieved Less time consuming and well-rounded feedback, as expert evaluators may operate on sketches, drafts or mere descriptions of systems and functions. Can be used in any stage of development, quantitative data not necessary. If well composed, valuable substitute for quantitative data in case of lacking resources. High reliability of feedback on a qualitative basis.
Cons	The method relies on previous experience only, even if it is from experts, therefore it has only limited efficiency if there is a fully innovative function Experts' judgements are often divergent, and synthesis may not always be easy to do It is sometimes difficult to define what an expert is and which point of view he represents actually There is no guarantee that the approach is exhaustive, i.e., that issues will not be overlooked or forgotten Due to highly specialised knowledge, average end-user concerns and experiences can be left out which might be important to consider. Based on the transfer of characteristics of other systems to the system under evaluation. Therefore, their reliability might be limited for completely

	innovative systems.
Recommendations	Panel should consist of experts of different disciplines to avoid biased results. To be supplemented with end user participation in evaluation panels.
Expected Output	High level evaluation; inter-disciplinary knowledge sharing, qualitative assessment of the system under evaluation.

Table 15 HMI concept simulation

Inspection methods	
Method	**HMI concept simulation**
Description	A concept simulation uses computer based visualisations of the mere interface.
Pros	Abstract scenario testing possible when fine-tuning HMI concepts. Opens the possibility of end-user feedback to HMI design from beginning stages of conceptualisation, since there is no need of real hardware, maybe only suitable “rapid prototyping” software to show the functions
Cons	Lack of realism, Interaction with driving task is ignored The assessment of such a simulation can only comprise the interface, but not the whole system. Therefore it is of limited use for the assessment of whole systems.
Recommendations	To be supplemented with in-car testing of HMI concepts.
Expected Output	A good starting point for the fine-tuning of HMI systems, allowing for feedback in the conceptual stages of development.

Table 16 Simulator experiment

Inquiry methods	
Method	**Simulator experiments**
Description	Driving simulation uses models of vehicle dynamics and virtual driving scenarios. This allows artificial driving situations and repeatable tests with various subjects. Potentially hazardous traffic scenarios can also be tested because in contrast to real driving the virtual scenario is harmless.
Pros	Applicable at early stages of development given enough details on the system. Allows direct information gathering about user’s interaction with the system including objective and subjective data. Allows testing of scenarios that are difficult to test on the roads (e. g. highly critical situations, control taking by vehicle). Allows evaluation of risky driving behaviour and situations without endangering subjects. Depending of simulator type (mock-up, fixed base, moving base), high realism of the experiments. High repeatability of tests in a completely controlled environment allows collecting good statistic samples with comparatively low expenditure of time. Depending on the simulators realism, highly reliable results can be achieved with direct transferability to the real in-car systems.
Cons	Lack of realism in case of mock-up simulators. User behaviour might be modified due to the presence of the experimenter

	and the perceived influence of simulation on real life consequences. Lack of correspondence between visual and motion information, which might lead to unrealistic driver behaviour. This drawback is minimized to a certain degree with moving-base simulators, but with this type, simulator sickness might occur and influence test results.
Recommendations	Validity tests should be conducted, to be supplemented with track-studies and FOTs
Expected Output	High repeatability of results. Allows for testing of systems that have not been fully developed yet or are in conceptual stages.

Table 17 Professional test drivers at test tracks or real roads

Trial methods	
Method	**Test Drives with prototype vehicles with professional test drivers on closed test tracks and on public roads.**
Description	Professional test drivers are trained to analyse systems in a standardised way and to compare the system under evaluation with existing systems.
Pros	Highly experienced drivers might imply that problems related to human factors when driving with the system might be found quickly e.g since they are used to the vehicle and the driving environment. More realism (relative to simulators) Allows close monitoring of driver behaviour in a wide range of driving conditions by professional drivers Professional test drivers can act as “the standard driver”, making their opinion a valuable source of feedback without having to deal with a representative sample of subjects. Depending on test drivers experience and capabilities, dangerous situations can be tested (system failure for active systems).
Cons	Both professional drivers and closed test track environments preclude average drivers/ driving conditions, hence modifications in behaviour cannot be studied with a high level of realism The use of professional drivers precludes behavioural data from a large section of the driving population sample Hard to standardise, since independent variables cannot be controlled in all cases (e. g. weather, traffic situations, etc), high costs, high effort preparing the tests,
Recommendations	To be supplemented with using samples representatives of the end users and with tests with naives or FOTs
Expected Output	High reliability of results. Advancement in complexity of behaviour studies from simulator tests, Feedback on prototype vehicle in public road environment (with realistic traffic conditions)

Table 18 Test drives with end users

Trial methods	
Method	**Test Drives with prototype vehicles with end users (subjects, naives) on closed test tracks and on public roads.**
Description	Test drives with naives allows testing of driver behaviour and performance while driving the vehicle equipped with the ADAS system in defined situations in a realistic environment. Test on test tracks allow a better repeatability and safety of the tests, while driving on public roads has a very high validity of the

	test results.
Pros	Enables collection of rich data on end user driving behaviour High realism in data collected Both end user driving behaviour and complex traffic flows studied.
Cons	The limitation of a close test track on realism remains Dangerous situations usually cannot be tested (system failure for active systems), hard to standardise, since independent variables cannot be controlled in all cases (e. g. weather, traffic situations, etc), high costs, high effort preparing the tests,
Recommendations	Test on test tracks should be supplemented with tests on public roads if the systems development stage allows it.
Expected Output	High reliability of results. End user driver behaviour analysed. Collection of very realistic data possible.

Table 19 Field operational tests

Trial methods	
Description	For FOTs usually a fleet of test vehicles is equipped with the system under development and handed over to subjects for a certain time for their everyday use, usually weeks to months. FOTs are experiments with very high effort in terms of time and financial resources. In return they have the highest achievable reliability.
Pros	Highly realistic data collected. Behavioural modifications over time (long term intended and unintended effects) can be studied. Since the subjects are not accompanied by an observing experimenter, some unexplainable artefacts might occur in the data.
Cons	Very expensive in terms of time and financial resources Huge amounts of data have to be analysed. Very extensive preparation necessary in case of cooperative ADA Systems or map based systems, since infrastructure must be ready to interact with ADAS or additional electronic map data has to be collected and processed.
Recommendations	Since FOTs are extremely laborious, the benefit has to be analysed critically and compared to the potential benefit.
Expected Output	Highest reliability of results. Most realistic data collection possible.

5.2.5.5 Techniques for data collection and analysis

Different techniques can be used for collecting data for assessment on the effects of an ADAS system on Situational control.

Subjective data collection; Questionnaires

Subjective data collection by questionnaires can be used for evaluating usability, acceptance and also to some extent effects on driving performance and driver behaviour of a system.

Questionnaires are frequently used for evaluating the human factors related impact of a system, in particular for evaluating usability, acceptance and workload of In-Vehicle Information

Systems (IVIS). Standardized questionnaires exist for evaluation of Usability, Acceptance and Workload, but for ADAS in particular there are currently no standardized questionnaires and the experience on how to evaluate ADAS with help of questionnaires is scarce.

Questionnaires can be used both in expert evaluations and in evaluations with other subjects; however the questions will differ depending on who is evaluating the system.

Questionnaires are documents with a number of questions, designed for providing answers to a specific aspect of the system. The questions can be either single or multiple choice questions or answered by selecting a number on a scale. The questions can also be freely answered by own words and sentences.

Freely answered questions will be subjected to qualitative analysis, while single or multiple choice questions might be subjected to quantitative analysis by application of statistical methods to the results from the questionnaires. Predefined scales that are frequently used are proposed by e.g Van Der Laan [16] and Brooke [15].

An extensive analysis of questionnaires and scales are presented in the AIDE project, deliverable D2.1.1.

In this procedure, questionnaires are recommended for subjective evaluation of ADAS, however how to specify the questionnaires to be used are not proposed within this procedure. Questionnaires should preferably be developed based on existing knowledge and standards (e.g. provided in the AIDE project). The questionnaires available should be adapted for usage on ADAS, and for the specific system to be evaluated.

Subjective evaluation can be made both in stationary vehicles and in moving vehicle during test drives. However, it depends on the HMI design to what extent it is possible to evaluate the system in a stationary vehicle.

Objective data collection

For assessment of driving performance and driver behaviour, objective data collection is of great importance for quantitative analysis, possibly complemented with data from questionnaires.

Objective data collection is made in moving vehicles, by collecting data parameters and log files during test drives.

An example on how to collect data on usability, acceptance and driving performance and driver behaviour for evaluation of situational control is presented in Figure 21.

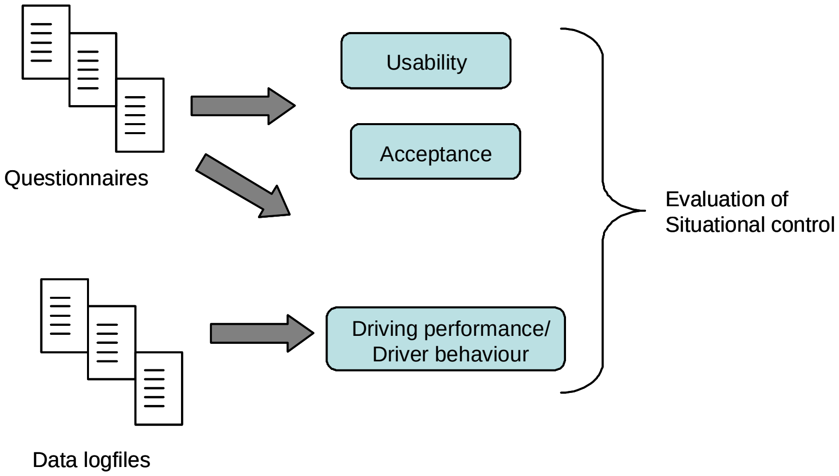


Figure 21 Example on data collection for evaluation of situational control

5.2.6 Step 5: Detailed study design

5.2.6.1 Objective

Design the study/studies in detail, i.e., after it has been decided which method to apply.

5.2.6.2 Approach

The design should be specified in detail prior to conducting the evaluation in practice. Some general issues that needs to be considered are

- Outline of the experiments.
- The detailed setup of the test scenarios. If many test scenarios that needs to be conducted, there might be a need of reducing the amount of test scenarios for time or resource reasons
- Specifications of the details in test setup (e.g. exactly when shall the lead vehicle brake)
- Control of whether the tests are repeatable
- If there are several test sites involved for conducting the tests, control how the consistency between the test site environments can be assured in order not to affect the test setup and results. A reference test for testing the test environment could be one way to assure the quality.
- Control of whether the indicators and parameters are measurable

5.2.7 Detailed study design of methods

5.2.7.1 Expert panel method

The following description of what this method should entail is derived from the Response-3 ‘Code of Practice ‘work [23].

Depending on the specific system to be evaluated expert panels may consist of:

- R & D Engineers
- Human Factors specialists
- Marketing specialists
- Accident research specialists
- Traffic psychologists etc.

Expert panels' assessments are useful to quickly identify:

- What is and is not acceptable as function/system behaviour, when a choice between system features is concerned,
- Which problems could be expected and will need later empirical assessment.; a prime example of this are the possibilities for unintended side effects to occur, which can be brought up relatively easily but which cannot be checked presently except by empirical research
- How some issues may be solved (type of warning signals, maximum forces required by controls, etc).

As experts assessment relies on personal knowledge built by practice and almost never made explicit, by definition expert points of view may differ on what is important or not. To make comparable judgements between experts, checklists may be useful to permit an assessment of critical items.

Expert judgment elicitation procedures have been formalized in several ways, mainly in the process industry (chemical, nuclear), where it is essential that potential risks are identified in a very early stage of the design process. A prime example of this is the HAZOP procedure (e.g. [22]). In the present domain, however, these procedures will probably have only limited application potential, mainly because they are direct towards hazard identification rather than everyday use and its associated HF issues.

Input/Requirements

A clear description of the function and its use cases should be available as result from step 1 in the procedure.

There should be a list of critical issues to stimulate complete assessment (limiting omissions due to different expert focus on problems)

There should be a panel of 6 to 8 experts from different horizons depending on the subject covered by the assessment (number of functions, alternative solutions, etc). At least 2 Human Factors experts are needed in the Expert panel when the focus is on HF.

5.2.7.2 Concept simulations

The description is also derived from the Response-3 Code of Practice [23].

In this form of evaluation, HMI concepts are visualised by tools like Macromedia Flash, VAPS or similar, so that a panel of "users" (which – again – may be experts) is able to watch what the system looks like and how the interaction works. The HMI should be represented in real sizes and colours, sounds (and optional: placement in the cabin). It is used to explore the interaction between user and system. Different HMI concepts might be compared by usability criterions like interaction time, number of errors, etc.

5.2.7.3 Experiments

For the sake of compactness, no distinction will be made here between the design of simulator experiments and on-the-road studies, although it is recognized that some choices to be made do depend on the particular type of experiment.

An empirical form of evaluation is, of course, always to be preferred over the weaker (by nature mostly non-quantitative) methods described before. In that case, the following are the most important issues:

Sample size - Sample size should follow from so-called statistical power considerations. The required minimum sample size is a function of the size of the effect one wants to detect, if it exists; of the probability with which one is willing to find it, if it exists; and of the significance level applied. The source for consultation on these issues, and for the tables that can be used to assist in the selection of these parameters, is Cohen [8].

For commonly accepted levels of these variables (i.e., for a medium-sized effect, with probability of finding it 80%, at a one-sided significance level of 5%) it follows from available power Tables that one should have a sample of at least 18 subjects in a within-subjects condition[3] and of about 26 subjects in a between-subjects condition. In practice, if the experiment has even a single between - Ss variable, as in a mixed design, that means that a minimum of 26 Ss should be used per between-Ss condition. Using less than this “magical number” of subjects leads to less reliable results and to an increased risk that one does not find a result that really exists and that one is interested in.

Sample composition - One may have reasons to define subgroups of Ss, according to age, experience, etc. It should be stressed that the only good reason to make a subdivision is if an interaction is expected between some system property and some user characteristic. For example, if readability of lettering on a display is non-optimal one could expect elderly users to suffer from that in particular, which could then be a reason to introduce age as a subdivision. If there are no such reasons, a workable default sample composition is:

- Same-gender Ss only, so as to achieve a more homogeneous sample.
- Age between 25 and 50.

---

[3] This assumes, furthermore, that the test-retest correlation – i.e., repeatability of measurements for individual drivers -, is high (about .80-.90), which is usually the case for driving performance measures

- Driving experience of at least 10 k/year for at least the three preceding consecutive years. It is also of importance that the subjects in the test are experienced with driving the specific vehicle, so that they are distracted by the vehicle dynamics or forced to focus on the driving task in a sense the decrease their ability to do a reliable assessment of the system. For example, for being able to evaluate a system in a truck you will need to be experienced with driving a truck and its vehicle dynamics

Review of selected subjective and objective parameters - It is useful to distinguish between (a) A minimum set of mandatory parameters and (b) A non-minimum set, i.e., that is not mandatory, but for which there are prescriptions in case one wants to include them. Furthermore, depending in the nature of the target situation a distinction should be made between (a) Parameters that describe 'overall' driving behaviour in a target situation that appears over a time frame of a number of seconds, minutes or more, like average speed, speed variability, lane keeping; and (b) Parameters that describe driver behaviour in a target situation that appears over a very short time interval (number of seconds), like reaction time at a lane departure.

Care should be taken in adhering to agreed technical definitions, prescriptions, and measurement procedures for selected parameters. E.g., if applying steering wheel reversal rate parameters one should be very strict in specifying which one is being considered. Collections of technical prescriptions are available from, e.g., the AIDE project [11].

Definition on metrics- In this step the metrics should be defined on how to measure and derive data on the indicators, earlier defined in step 3.

Experimental design – A central issue is whether to apply within- or between-Ss designs, or a combination of those (mixed design). There are textbooks dealing with this, but in the end the choice must be based on good judgment from the experimenter. Baseline (control) conditions, where the system that is evaluated does not function, should always be included.

Logistics – This moves into the practicalities of getting the study up and running. It includes such issues as:

- Preparing and giving instructions to subjects and experimenter
- Carrying out pilots and revising the experimental procedure on the basis of it
- Taking ethical considerations into account, i.e., guidelines and procedures that deal with the treatment of subjects in behavioural studies, in particular when these will be performed in instrumented vehicles on the open road.

Analyzing the data - Nothing special needs to be done, once the a priori considerations on the experimental design (power!) have been followed. ANOVA is the usual form of quantitative analysis. In some cases, non-parametric forms of analysis must be used. This need not be explained here in detail, since textbooks are available.

One specific recommendation is that results should not only be reported in terms of statistical significance, but also in terms of effect sizes. When an effect is statistically significant this by itself does not tell us how ‘big’ it is. Effect sizes tell us, and – since they are standardized scores – also permit us to compare across variables and even across different studies. [4] With respect to this last aspect they furthermore allow us to do meta-analysis, so that the results of a set of studies can be compared and combined.

5.2.8 Step 6: Reporting and interpretation

5.2.8.1 Objective

Document the results and findings from the evaluation with interpretation and conclusion on to what extent the hypotheses and expectations were met.

5.2.8.2 From expert panels

Since this method does not yield results of a quantitative nature (except if simple scales have been used) the report will not contain a formal statistical analysis, but it will mainly be in a descriptive form.

The following issues should at least be dealt with, in terms of discussion issues and conclusions:

- The identification of potential or actual problems and causes (diagnosis) should be discussed
- Comparative results should be reported, i.e., some attention should be devoted to discussion of the merits of alternative solutions, where conceivable.
- Concrete supports to decision and choice between alternatives should be given, i.e., amongst different solutions or different warning modalities.
- Possible remaining HF problems should be identified, including possible misuse situations.

5.2.8.3 From an HMI conceptual simulation

The following issues should be discussed in the (largely descriptive) report:

- Judgment on usability and comprehensibility of the dialog, bases on the limited quantitative measurements that may have been collected.
- Result on how the system fits into the whole concept of vehicle systems and the influence from the system under consideration to the others.
- A comparison of merits and demerits of alternative conceptualisations.

---

[4] Note that some commonly reported ANOVA outcomes, like ε^2 and ω^2, are *not* measures of effect size. They estimate the percentage of variation in the dependent variable that is explained by the independent variable, and thus are measures of association, like common r^2.

5.2.8.4 From experiments

This is not the place for a treatise on how to write a report on an experimental study, if only because much of it should be left to the author's craftsmanship. However, the following are to be specifically considered when reporting on a HF-oriented empirical study.

Referring to a priori hypotheses - The simplest, but often adequate, way of reporting is to run through the a priori hypotheses once more and state whether they have been confirmed or not, in terms of statistical significance.

Focus on effect sizes - If significance is obtained, however, the discussion should then immediately focus on effect sizes, first of all in terms of how big the experimental manipulations have in fact turned out to be. A secondary issue is how the comparative sensitivity of the dependent measures looks, which is also easy to judge from the effect sizes. This can tell us which is the 'best' dependent variable, in terms of potential to discriminate between experimental treatments. I.e., it helps us identify which is the parameter to be taken most seriously when thinking about the results.

Extrapolation to accident risk effect estimates - Recently, algorithms have become available for extrapolating behavioural results to expected accident risk effects in the aggregate. These are the German I-TSA procedure and the AIDE algorithm, both described in Janssen et al [11]. Unfortunately, it is just too early to definitely have reasons to prefer one above the other. In the meantime, it is recommended to behavioural researchers to look into both and make up one's mind, and at least apply one of them so as to get to a risk effect estimate.

Template for summary Report – Table 20 could be useful as a way of summarizing an experimental study and its results.

Table 20 Template for summary report

Objective with test	**Comment**
System and functions investigated	
Hypotheses addressed	
Scenarios used	
Methods used	
Experimental design; **Parameters, indicators and metrics**	Independent/dependent variables; within/between-Ss factors
Experimental vehicle	Whether simulator, or instrumented vehicle (closed-track vs on-road)
Participants	Description of sample size and composition
Tools	Specifics of measurement procedures
Results and conclusions	Significances, effect sizes, extrapolation to risk effects
Recommendations	

6 Safety Potential Assessment Procedure

PReVAL uses the behavioural effect approach, which has been developed and is used by the eIMPACT project. The approach is described in the deliverable D16.1.

6.1 The behavioural effect approach

The underlying rationale of the selected approach is to assess the impact mechanisms evaluating first how the functions affect driver behaviour and travel behaviour. Based on earlier research results concerning the relationships between driver behaviour and crash risk and/or consequence or desktop estimates based on expert judgments, these behavioural changes are projected into relative and/or absolute changes in fatality numbers. In other words, the change in driver behaviour, for example average speed, is transformed to change in accident risk and consequences.

Draskóczy, Carsten and Kulmala [4] compiled the following list of mechanisms, via which ITS affects safety:

1. Direct in-car modification of the driving task by giving information, advice, and assistance or taking over part of the task. This may influence driver attention, mental load, and decision about action (for example, driver choice of speed)
2. Direct influence by roadside systems mainly by giving information and advice. Consequently the impact of this influence is more limited than of the in-vehicle systems.
3. Indirect modification of user behaviour in many, largely unknown ways. The driver will always adapt to the changing situation. This is often called behavioural adaptation and will often not appear immediately after a change but may show up later and it is very hard to predict. Behavioural adaptation may appear in many different ways (for example, by change of usage of the car, by change of headway in a car following situation, by change of expectation of the behaviour of other road users)
4. Indirect modification of non-user behaviour. This type of behavioural adaptation is even harder to study because it is often secondary. Non-equipped drivers may for example change their behaviour by imitating the behaviour of equipped drivers (for example, driving closer or faster than they should, not having the equipment).
5. Modification of interaction between users and non-users. ITS will change the communication between equipped road users. This change of communication may influence the traditional communication with non-equipped road users. To a large extent this problem may appear in the interaction between drivers and unprotected road users.
6. Modification of road user exposure by for example information, recommendation, restrictions, debiting. This is certainly an area where introduction of ITS will have a large impact for example by changing travel pattern, modal choice, route choice etc.
7. Modification of modal choice by for example demand restrains (area access restriction, road pricing, area parking strategies), supply control by modal interchange and other public transport

management measures, travel information systems. Different travel modes have different accident risks, therefore any measure which influences modal choice, has also impact on traffic safety.

8. Modification of route choice by demand restraints by route diversions, route guidance systems, dynamic route information systems, hazard warning systems monitoring incidents. Different parts of the road network, i.e. different categories of roads, have different accident risks, therefore, any measure which influences route choice by diverting traffic to roads of different category, has also impact on traffic safety.
9. Modification of accident consequences by intelligent injury reducing systems in the vehicle, by quick and accurate crash reporting and call for rescue, by reduced rescue time, etc.

Based on the above analysis, we summarise the mechanisms as follows:

1. direct modification of the driving task (#1 above),
2. indirect modification of user behaviour and non-user behaviour (#3 and 4 above),
3. modification of interaction between users and non-users (#5 above),
4. modifications of road user exposure (e.g. change of modal choice or route choice) (#6, 7 and 8 above) and
5. modification of accident consequences.(#9 above).

Mechanism 2 is not relevant in this context, since PReVENT deals mainly with in-vehicle systems.

This *behavioural effect* approach is somewhat different from the so-called *accident causation approach* that may be more familiar for many ITS experts. Therefore, the rationale of these approaches is discussed in the following.

The accident causation approach first focuses on the information in accident statistics and in-depth accident studies and aims to identify, which crashes, injuries and fatalities could have been prevented by the system. The actual effectiveness estimate can be based in the assessments of the in-depth accident investigation teams or desktop estimates based on expert judgments. However, one can assume that in a number of instances this type of assessment does not consider all possible relevant safety mechanisms. Specifically, the accident causation approach is likely to miss some effects and especially behavioural adaptation effects. The same holds also for systems and functions affecting exposure. As Nilsson [13] illustrates, the number of fatalities (or human loss in general) is a product of three factors, namely crash risk, risk of fatal injuries in a crash and exposure (Figure 22) This suggests that it is important to consider each of these dimensions.

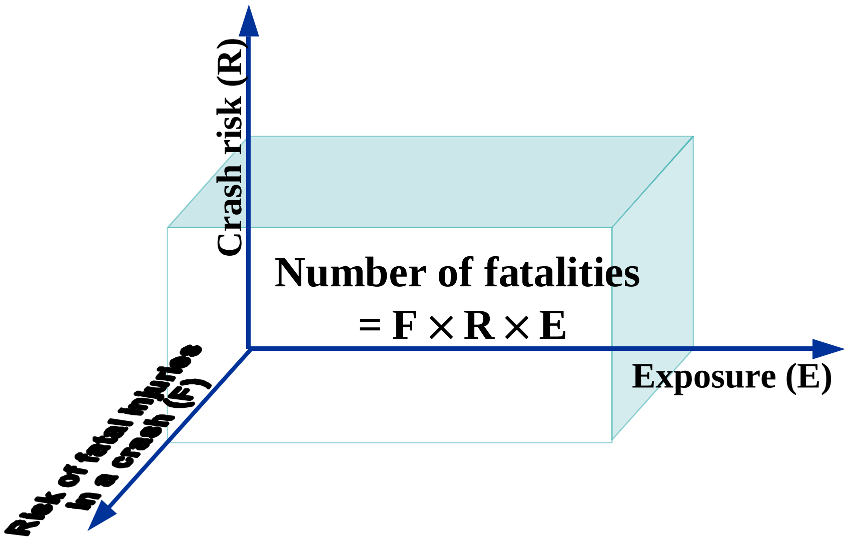


Figure 22. Conceptual framework for traffic safety i.e. the size of the safety problem (the number of human injuries and fatalities) illustrated as a function of the product of the three variables [13]

Furthermore, when systems are affecting crash risk, the two approaches seem different. Although the intended effects related to crash risk are usually covered with both approaches, in many cases other effects are not covered by the accident causation approach. Finally, concerning consequences and systems only affecting the consequences of the accident, either during the crash or after it, it seems that the main effects are estimated quite similarly when applying either the accident causation or behavioural effect approach [14].

It should also be pointed out that both approaches are sensitive to assumptions made in the assessment process. Assumptions have to be made in all cases as all information required is rarely available. When using the behavioural effect approach, researchers tend to have some information of the function’s effect on driving and travelling behaviour on the basis of simulator, modelling or field studies. This can then be transformed to the overall accident/injury/fatality effectiveness of the function. When using the accident causation approach, the effectiveness assumptions used are in many cases just guesses, although frequently based on earlier accident statistic based effectiveness estimates of similar types of functions.

6.2 The procedure of the methodology

6.2.1 Description of the functions

The starting point for the analysis is the “System and functions description” (see Section 4.2.1). It is essential that all safety functions are defined. For safety impact analysis, the user perspective and traffic flow viewpoint are important.

6.2.2 Literature survey on the effects of the functions

A state of the art review concerning the known safety impacts has to be carried out. The results of the literature study have to be classified as follows:

- Empirical evidence on safety impacts (verified results based on valid experimental design, for example)
- Expert evaluations of safety impacts (predicted results)
- Indirect evidence on safety impacts, which means more general assessment of the effects based on knowledge of driver behaviour, traffic flow, and effects of comparable functions, e.g. road side telematics (potential results)

6.2.3 Relevant safety impact mechanisms

The relevant impact mechanisms will be described in detail qualitatively. Parallel, both the positive and negative safety effects will be rechecked. This phase includes:

- Checking the positive and negative safety effects, identification of additional mechanisms
- Qualitative descriptions of the relevant main mechanisms:
    - What is the user reaction to the function?
    - How can user reactions and modification of the situation be ‘translated’ into expected safety effects?
    - How do these reactions depend on the characteristics of the function?
    - How do these reactions depend on penetration degree?
    - What other effects can be identified (comfort, environment, traffic flow, delay etc.)
- Identification of the relevant state of the art knowledge connected to the impacts
- First quantitative estimates of the impacts.

The results of the hypotheses generation step in the human factors evaluation (Section 5.2.3) can be input to this step.

A following example dealing with the night vision function illustrates this phase, although only as a non-exhaustive description.

Night vision

Mechanism 1&5: Direct modification of the driving task, accident consequences

The function is designed to improve driver visual information acquisition when the visibility conditions are limited because of low ambient illumination (i.e. at night). Specifically, the function enables drivers to detect potential hazards or obstacles and to take evasive actions earlier than without the function. Consequently, the use of the function is expected to reduce the crash risk of all night-time crashes and especially those occurring on road with no road lighting. The function could also reduce crash severity (Mechanism 5, accident consequences), because the driver may reduce the speed earlier than with no function and thereby the speed in a potential crash would be lower. Overall, these effects can be called *intended effects* as well.

Mechanism 2: Indirect modification of user behaviour

Night vision might induce drivers to use higher driving speeds at night, which in turn would increase crash risk and also crash severity. This *unintended* effect happens in course of time when the driver learns to rely on the function and may assume that he is able to detect all hazards and obstacles so well that the crash risk is not elevated although he increases the driving speed. Because the effects of speed on crash risk and crash severity are fundamental, this may not be the case and, in reality, the crash risk increases. In addition, it is important to understand that this behavioural adaptation can happen with or without any conscious decision

Mechanism 3: Modification of interaction between users and non-users

The function is likely to have an overall impact on driver behaviour so that the driver does not search night-time pedestrians as effectively with no function. This unintended effect results from the delegation of responsibility to the function. This may result in reduced safety if some pedestrians or other obstacles are not detected by the driver (although there was no malfunction).
The key difference between Mechanism 2 an 3 is the fact that Mechanism 3 focuses on interaction between the user and non-user, while the Mechanism 2 deals with effects that do not have to involve a non-user. Otherwise, there are obvious similarities between the two mechanisms. Furthermore, it is acknowledged that in the analysis of some functions it may be difficult to define whether a given effect includes to Mechanism 2 or 3. However, this leads to no principal problem because the main issue is to cover each identified effect. At the same time, one must be careful not to take the same effect into account more than once.

Mechanism 4: Modification of road user exposure

The function will likely increase exposure by making driving at night and unlit roads less unattractive. This effect can be particularly substantial to such drivers who currently tend to avoid driving in such condition (e.g. elderly drivers). While increasing the mobility of such drivers, this effect will increase the accident risks

Based on the thorough descriptive analyses, a list of the relevant impact mechanisms is provided for each function. Each item on the list will be a target for the quantitative analyses of the impacts.

6.2.4 Quantitative assessment of the impacts

6.2.4.1 Classification

The safety analysis utilises disaggregated and classified accident data. The classification of accidents is based on environmental and situation-specific factors. The specific classification depends on the classification used in the CARE accident database. The following aspects have to be considered:

- Road class: motorway / rural area / urban area
- Vehicle type: heavy vehicle / car
- Accident type: Frontal, rear, single vehicle, etc.
- Road weather: adverse / good
- Lighting conditions: day / night

Furthermore, it is important that the classification of each variable is mutually exclusive. Each case in the crash data appears in the data only once (e.g. for one crash only one crash type is chosen).

6.2.4.2 Preliminary assessment

After going through the items, the safety effects of the function are to be given as percent changes. These percents are based on the existing data gathered in the functions description and literature survey. It is assumed that every cell is not relevant for each specific function, and therefore many cells are filled with a zero effect. The starting point for this phase is the list and the qualitative descriptions provided in “definition of relevant safety mechanisms”. The analyses will be conducted for each effect mechanism listed.

The assessment will be repeated for each function concerning the fatal accidents i.e. fatalities. A tentative procedure for carrying out this phase is described below. For many functions it might be easiest to start with the accident types.

For ESC, for example, there could be two assessments:

1. ESC / Mechanism 1: Direct in-car modification of the driving task
2. ESC / Mechanism 3: Indirect modification of user behaviour

Please tick the relevant accident types for the effect (specified as function + mechanism).

Give the first estimated overall effect of the function on the table for each accident type for the number of the fatalities. Please note that the effects may be positive or negative and that the estimated effects will be summed up later to give the estimated total effect for the safety system.

Please provide for the report the following information for each estimate:

- *verified, give the reference*
- *expert evaluation, give the reference if available*
- *potential effect, give the reference, indirect evidence*

Is the safety impact dependent on road type?

- *yes, please modify the estimates accordingly*
- *no, use same impact % for all road types*

Is the safety impact dependent on vehicle type?

- *yes, please modify the estimates accordingly*
- *no, use same impact % for all vehicle types*

Is the safety impact dependent on weather conditions?

- *yes, please modify the estimates accordingly*
- *no, use same impact % for all weather conditions*

Is the safety impact dependent on lighting conditions?

- *yes, please modify the estimates accordingly*
- *no, use same impact % for all lighting conditions*

6.2.4.3 Modification of assessment based on estimated penetration

Firstly it should be considered whether the safety impacts should be analysed taking into consideration the different target year or more precisely, the penetration level of the studied function. In other words, it should be considered whether the effects for an equipped vehicle are independent of the percentage of vehicles equipped.

For example, if we assume that the behavioural modification and its safety impact percentage due to ESC will change according to how much of the vehicle fleet is equipped with the function, the following analyses are produced (in this example for the years 2010 and 2020):

1. ESC / Mechanism 1: Direct in-car modification of the driving task

2. ESC / Mechanism 3/ 2010 (penetration level of A%): Indirect modification of user behaviour

3. ESC / Mechanism 3 / 2010 (penetration level of B%): Indirect modification of user behaviour

4. ESC / Mechanism 3 / 2020 (penetration level C%): Indirect modification of user behaviour

5. ESC / Mechanism 3 /2020 (penetration level of D%): Indirect modification of user behaviour

6.2.4.4 Overall quantitative assessment

The actual number of fatalities affected by the function will be calculated for each of the four afore-mentioned market penetration scenarios in the following manner:

- The numbers from the accident database, which are used as the basis, are adjusted for the overall traffic and safety changes from the current situation for example to 2010 or 2020 (depending on the year/penetration level that is decided to be studied).
- Following calculations are performed in each data row;
    1. calculating the change in the number of crashes by changing first the number according to the % change in exposure and then to the % change in crash risk
    2. calculating the change in the number of fatalities by first changing the number according to point 1 and then according to the change in % fatalities/crash
    3. calculating the change in the number of severe injuries by first changing the number according to point 1 and then according to the change in % severe injuries/crash
    4. calculating the change in the number of slight injuries by first changing the number according to point 1 and then according to the change in % slight injuries/crash
- The numbers are added up to produce the overall change in the numbers of crashes and their consequences.

7 Conclusions and next steps

This deliverable describes the framework proposed by the PReVAL project for the assessment of preventive and active safety functions.

The safety potential of a function is influenced by the technical performance, the interaction between the driver and the vehicle, as well as the traffic safety level. The work has been performed according to these three aspects. Procedures have been proposed for the technical evaluation and for human factors evaluation. For the safety impact, the behavioural effect approach is applied.

The framework for the technical evaluation is based on a modification of the V-shape design cycle. Steps for evaluation are included in this V-shape cycle. The approach for human factors evaluation can also be mapped on this modified V-cycle.

The concept of situational control has been introduced, referring to the level of control jointly exerted by the driver and the vehicle in a specific driving situation. This concept links the different aspects of the assessment to each other.

The three different procedures (safety impact assessment, technical evaluation, human factors evaluation) have to be further integrated. The following steps are performed to come to a holistic framework:

- the proposed framework for technical evaluation and human factors evaluation is sent to the INSAFES project as input to their evaluation plan. The human factors evaluation is benchmarked for selected PReVENT functions. Two PReVENT functions, one integrated in a passenger car (MAPS&ADAS) and one integrated in a truck will be selected for evaluation. This evaluation includes an expert evaluation by a team of Human Factors expert. The main goal of this evaluation is to evaluate the developed procedure by applying the method in practice. The purpose is to contribute to the already performed evaluations in the subprojects, not to re do the evaluation.

- the interfaces between the different evaluation aspects will be refined through the work done in INSAFES and by the application of the human factors procedure within PReVAL. Experience and lessons learned will be used as input to the final framework.

- The proposed framework is sent to different experts for feedback. A draft version of the procedures have been sent to the INSAFES project, which takes the recommendations from PReVAL into account for their evaluation.

- at the end of the project (on 10.1.2008), a workshop will be held together with experts from the eIMPACT project to come to a harmonized holistic approach.

- Based on the feedback and the experiences, the framework will be updated. The final assessment framework will be included as an Annex to the deliverable D16.4.

References

[1] PReVAL consortium, Review of validation procedures for preventive and active safety functions, PReVAL deliverable D16.1, 11.6.2007

[2] Carsten, O.M.J. and Nilsson, L. (2001). Safety Assessment of Driver Assistance Systems, European Journal of Transport and Infrastructure Research, No. 3 (2001), pp. 225-243.

[3] Brook-Carter, N. et al. (2002) An Integrated Methodology and Pilot Evaluation Results. ADVISORS Deliverable D4.2

[4] Draskóczy, M; Carsten, O.M.J. and Kulmala, R. (1998). Road Safety Guidelines. CODE Project, Telematics Application Programme, Deliverable B5.2.

[5] Saad, F., Hjälmdahl, M, Cañas, J., Alonso, M. Garayo, P., Macchi, L., Nathan, F., Ojeda, L., Papakostopoulos, V., Panou, M. and Bekiaris, E. (2004). Literature review of behavioural effects. AIDE D1_2_1). Retrieved April 20, 2007from http://www.aide-eu.org/res_sp1.html

[6] ZHANG, X., KOMPFNER, P., WHITE, C. & SEXTON, B. (1998) Guidebook for Assessment of Transport Telematics Applications: Updated Version, Deliverable 2.3.1 of the CONVERGE Project. Project Deliverable, CONVERGE Consortium.

[7] CHERRI, C., NODARI, E. & TOFFETTI, A. (2005) Review of existing Tools and Methods, Deliverable 2.1.1 of the Integrated Project AIDE. Brussels, AIDE Consortium.

[8] COHEN, J. (1988) Statistical Power Analysis for the behavhioral sciences, New York, Academic Press.

[9] Gietelink O.J., Ploeg J., de Schutter, B. and Verhaegen, M. 2004. VEHIL: A Test Facility for Validation of Fault Management Systems for Advanced Driver Assistance Systems. in: Proceedings of the 1st IFAC Symposium on Advances in Automotive Control, Salerno, Italy,

[10] Gietelink, OJ. Design and Validation of Advanced Driver Assistance Systems, PhD thesis (draft version) 2007.

[11] W. Janssen, E. Nodari, J. Plaza, J. Östlund, A. Keinath, A. Toffetti, M. Alonso, M. Rimini-Doering, R. Brouwer, V. Portouli (HIT), D. Horst, C. Marberger & H. Vega (2007). Development of AIDE evaluation methodology. AIDE Deliverable 2.1.4, in press.

[12] Norman, D. 1988. The design of everyday things. New York: Currency-Doubleday

[13] Nilsson, G. (2004). Traffic safety dimensions and the power model to describe the effect of speed on safety. Lund: Lund Institute of Technology, Department of Technology and Society.

[14] Kulmala, R. (2006). Observations on the use of the accident causation approach in assessing the safety effects of systems. International Cooperation on Theories and Concepts in Traffic

Safety (http://www.ictct.org/index.htm). The ICTCT-workshop of 2006 in Minsk.

[15] BROOKE, J., SUS. (1996) A “quick and dirty” usability scale http://www.usabilitynet.org/trump/documents/Suschapt.doc

[16] VAN DER LAAN et al. (1997) A Simple Procedure for the assessment of Acceptance of Advanced Transport Telematics.

[17] Hollnagel, Erik and Woods, David D. (eds.) (2005): *Joint Cognitive Systems. Foundations of Cognitive Systems Engineering.* Boca Raton, Florida, USA, CRC Press

[18] Michon, J. A. (1985). A critical view of driver behavior models. What do we know, what should we do? In L. Evans & R. Schwing (Eds.), Human behavior and traffic safety (pp. 485-525). New York: Plenum press.

[19] Fosser, S., Saetermo, I.F. & Sagberg, F. (1997). An investigation of behavioural adaptation to airbags and antilock brakes among taxi drivers. Accident Analysis and Prevention. 29

[20] OECD 1990. *Behavioural adaptation to changes in the road transport systems.* OECD, Paris.

[21] Summala, H. 1997. Hierarchical model of behavioural adaptation and traffic accidents. In Rothengatter, T., and Vaya, E. C. eds. *Traffic and Transport Psychology: Theory and application,* 41-52. Oxford: Pergamon Press.

[22] Kirwan, B. 1994. *A Practical Guide to Human Reliability Assessment*. London: Taylor and Francis.

[23] RESPONSE 3, Deliverable D11.2, Code of Practice for the Design and Evaluation of ADAS, 2006

[24] PreScan, www.PreScan-tno.com.

Annex 1 Keywords

Active Safety System, ADAS, Advanced protective safety, Integrated safety systems, Preventive safety applications, Evaluation, Assessment, Validation, Safety impact analysis, Human-Machine Interface, Technical evaluation

Annex 2 Glossary

Abbreviation	Explanation
ABS	Antilock Braking System
ACC	Adaptive Cruise Control
ADAS	Advanced Driver Assistance System
ANOVA	Analysis of Variance
CAR	Correct Alarm Rate
CPU	Computer Processing Unit
ESC	Electronic Stability Control
FAR	False Alarm Rate
FCW	Forward Collision Warning
FN	False negative rate
FOT	Field Operational Test
FP	False positive rate
HF	Human Factor
HIL	Hardware-in-the-Loop
HMI	Human Machine Interaction Human Machine Interface
ITS	Intelligent Transport Systems
I-TSA	INVENT Traffic Safety Assessment
IVIS	In-Vehicle Information Systems
JDVS	Joint Driver-Vehicle System
MAR	Missed Alarm Rate
OEM	Original Equipment Manufacturer
TEM	Trajectory Estimation Module
TLC	Time to line crossing
TN	True negative rate
TP	True positive rate
TTC	Time to collision
V2I	Vehicle to Interface
V2V	Vehicle to Vehicle

Term	Definition	Comment	Source
Acceptance	The degree to which drivers consider a function to be useful and satisfactory.	Acceptance may be influenced by the technical performance as well as the usability of the system.	Original definition
Application	A program (such as a word processor or a spreadsheet) that performs one of the important tasks for which a computer is used		AIDE Glossary
Assessment	The process of determining the performance and/or impacts of a candidate application, usually in comparison to a reference case (existing situation or alternative applications), and usually including an experimental process based on real-life or other trials, often involving users.		CONVERGE

Assessment objectives	A precise statement of an individual objective which an application should be judged against. It should be associated with a precise definition of the associated indicator(s) and definition of success.		CONVERGE
Control	Achieving and/or maintaining a consistent goal state		Original definition
Controllability:	likelihood that the driver can cope with driving situations including ADAS-assisted driving, system limits and system failures		RESPONSE
Dependability	The trustworthiness of a computer system such that reliance can justifiably be placed on the service it delivers.		EAST-EAA, AIDE Glossary
Driving performance	The degree to which the goals associated with the driving task are attained		AIDE glossary
Driving task	All aspects involved in mastering a vehicle to achieve a certain goal (e.g. reach a destination, including tracking, regulating, monitoring and targeting)		AIDE glossary
Evaluation	The process of determining the value of an application in comparison to alternative applications and/or to a "base case", and deriving recommendations for decision makers based on identifying requirements on and analysing results of related experiments.		CONVERGE
Function	A description of what something does or is used for		RESPONSE (option a)
HMI- Human Machine Interaction	All the possible modes by which interaction (direct or indirect) between the driver and one or more vehicle systems takes place.		Response Glossary
HMI- Human Machine Interface	A set of components that govern the interaction between the user and one or more systems.		AIDE Glossary
Human factors evaluation	Assessment of a function taking into account driving performance, driver behaviour, usability and acceptance.		Original definition
Impacts	Changes or effects brought about by an application resulting from its use in an experimental or real application.		CONVERGE
Indicator	Parameter that is used for estimating the performance (or impacts) of a function.	Indicators form subsets of parameters. They are usually measured or can be derived from measurements and operationally defined in terms of metrics. When simulation is used instead of measurement, indicators will usually be outputs of the simulation. It may be necessary to use more than one indicator for each assessment objective.	Original definition

Intended effects	Effects of a function that are intended by the designer of the function.	Intended effects are normally (explicitly or implicitly) implied by the functional specifications. For instance, an intended effect of a Forward Collision Warning function is a reduction in brake response time to a sudden unexpected front obstacle. This corresponds to the term “direct effect” in the AIDE glossary.	Original definition
Joint Driver-Vehicle System (JDVS)	The system comprised of the driver and the vehicle.		Based on Hollnagel and Woods (2005)
Long term testing	Assessment of effects that appear on a time scale of days or longer.		Original definition
Metric	Operational definition of an indicator.	A metric defines how an indicator is measured or derived from measurements. It gives clear advice about the techniques and technologies to be used during the measurement and, if the indicator is derived from direct measurements, the metric describes the respective mathematical operations; the process how to get values of the indicator.	Original definition
Parameter	An independent variable used to express the coordinates of a variable point and functions of them	Parameters are measurable quantities that, for the purpose of safety system assessment, represent properties of the driver, the vehicle and/or the environment. Parameters can be used to describe situations.	EAST-EAA (Webster)
Short term testing	Assessment of effects that appear on a time scale of hours or less.		Original definition
Situation	A set of relevant driver-, vehicle and environment elements within a volume of time and space.	A situation can be described in terms of parameters and their temporal evolution.	Original definition, based on the existing definition of situational awareness by Endsley.
Situational control	The degree of control that a Joint Driver-Vehicle System (JDVS) exerts over a specific situation.		Original definition
System	A collection of components organized to accomplish a specific function or set of functions.		AIDE (EAST-EAA) + RESPONSE
Target situations	The situations where a function has its intended effects.	In the case of active safety functions, target situations are often, but not necessarily,, critical.	Original definition

Technical evaluation	Assessment of the technical performance of a function or system		Original definition
Technical performance	The degree to which a system meet functional and technical specifications.	Technical performance can, for example, be quantified in terms of false alarm rate, missed alarm rate, etc. It also involves aspects related to system dependability. Technical performance is directly related to situational control for fully automated functions, i.e. functions without the driver in the loop. However, whenever the driver is in the loop both technical and human factors testing are needed to assess the impact on situational control e.g. in terms of reliability, time accuracy, missed alarm rate, false alarm rate etc.	Original definition
Unintended effects	Effects of a function that are not intended by the designer of the function.	Unintended effects are not implied by the functional specification. They can be either positive or negative with respect to safety. This corresponds to "indirect effects" in the AIDE Glossary	Original definition
Usability	The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.	In the case of warnings, comprehensibility is a key aspect of usability.	ISO 9241-11
Use case	An intended or desired flow of events or tasks that occur within the vehicle and are directed to or coming from the driver in order to accomplish a certain system-driver interaction		AIDE Glossary
Validation	The process of evaluating a system or component during or at the end of the development process to determine whether it satisfies the expectations.		RESPONSE
Verification	Assuring, e.g. by testing, that a component, a subsystem, a system or a process is working as required and specified.		RESPONSE

Annex 3 Dummy targets

The validity of the tests carried out to assess the technical performances of objects detection devices depends highly of the characteristics of the dummy targets against which the system are confronted.

In INTERSAFE: silver motorcycle, black compact car, silver estate car, red middle size car, black large size car, pedestrian (dark clothing, wooden)

Figure 23: Examples of dummy objects used to evaluate PReVENT functions

Due to the absence of standard, the diversity is very high. Furthermore, the question of their justification arises. This is a point of method that can be improved in the future.

A methodology exists to decide the characteristics of the targets in the case of two sensor types : radar an lidar.

Detectability specifications for Lidar and Radar (ISO 15622:2002)

The test target are defined as possessing a *CTT* (Coefficient for Test Target) for lidars or a RCS for Radars.

Infrared LIDAR:

The test target is defined by a coefficient of a reflector, which represents the reflectivity of a dirty car without any retro-reflector.

$$CTT = \frac{I_{ref}}{E_t}$$

where

I_{ref} = radiated intensity in a given direction, out of the reflector, measured in front of the receiver surface. [W/sr]

E_t = intensity of irradiation, out of the transmitter $\left[\frac{W}{m^2}\right]$

CTT = Coefficient for Test Target $\left[\frac{m^2}{sr}\right]$

The *CTT* only describes the quality of a reflector (damping). The smallest acceptable test target is a corner reflector with the required CTT. It is permissible to use a test object with a larger surface of reflection, if it meets the same CTT requirement.

The infrared test target is defined by an infrared coefficient for test target (CTT) and the cross section of the test target. The minimum cross section for test targets is 20 cm^2. Test target is a diffuse reflector with a CTT = (1 ± 0,1) m^2/sr.

Millimetric wave RADAR: The radar test target is defined by a Radar Cross Section RCS. For the frequency range between 20 GHz and 95 GHz. The RCS for test target shall be 3 m^2.

Detectability for image processing

A contribution to this specification has been undertaken first in ARCOS in which geometrical pedestrians outlooks have been defined. Several structure types (based on gradients) were explored and used for test. The searched quality was the reproductivity of the test conditions. The dummies obtained are easy to define and easy to reproduce anywhere in Europe. However, the representativeness was not ensured:

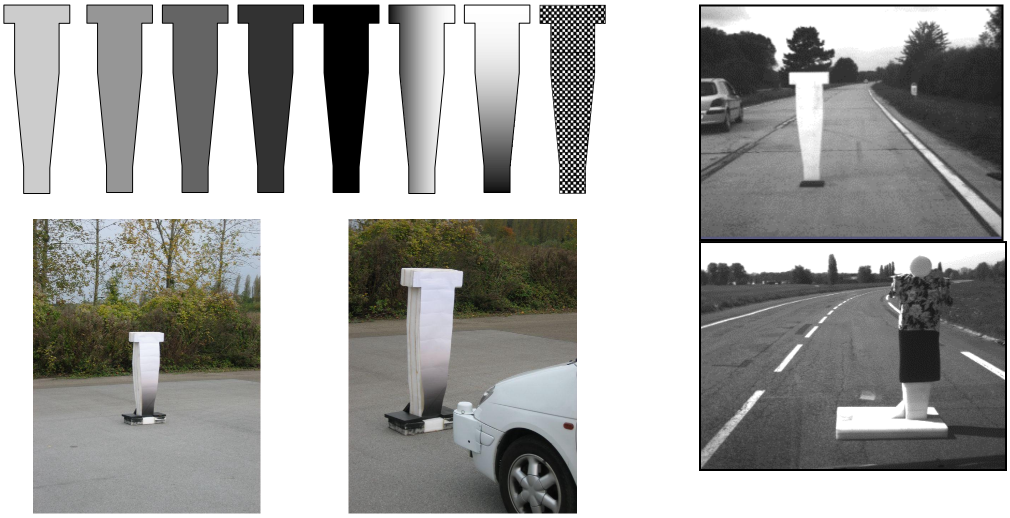


Figure 24: Dummies used in the ARCOS project

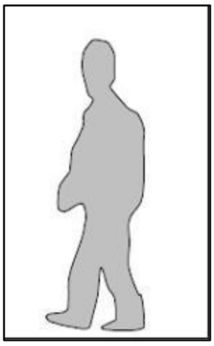
Currently another study allows contributing to the way to define a standard pedestrian dummy.

Classification approach[5]

The outlook of a dummy is composed of an outlook and a structure. The outlook used is based on the mock-up used in street lighting studies (Figure 25)

---

[5] Proposition d'une méthodologie pour la définition d'un piéton standard. Nicolas Hautière, Romain Gallen, Franck Turban, Jean-Marc Blosseville. Rapport technique, Laboratoire Central des Ponts et Chaussées, octobre 2007.



Functional

Figure 25: Mock-up of pedestrian used in street lighting studies

In order to initiate a study on representativeness a database of real pedestrian images has been collected (Figure 26). A search for a representative structure has been approached on different criteria: a 2D histogram of normalized red and blue colours (for colour images), gray level co-occurrence matrices and a histogram of gradients orientation (for black and white images). Several explored calculation methods have been conducted to the following results.

The classical classification methods can be used in order to identify classes of dummies that can be approximated by one representative. All those pedestrian outlines are representative for a great variety of real pedestrians. Another approach allows extracting the “limit cases”, i.e. the outlines that can be difficult to detect. In the collection of pictures below (Figure 26), the red framed outlines represent the most difficult cases based on normalized colour histogram classification, the green ones, the most difficult cases based on gray level co-occurrence matrices, the blue ones are limit cases for both criteria. The two approaches are thus complementary.

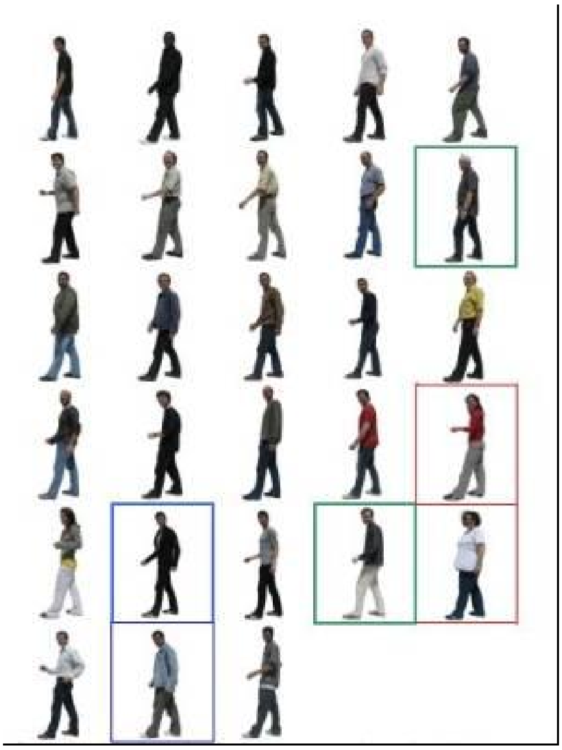


Figure 26: Pedestrians database and outlined limit cases for the different criteria.

Perspective: based on a wider data base and some other adjustments, it could be possible to use this methodology to propose a collection of dummies that could be representative of a high variety of pedestrians. The representativeness should be expressed with respect of a basic criteria used in algorithms (gradients, contrast, colour...), and with respect to 2 aspects: nominal cases (the most common = centres of classes) and difficulties (the limit of the classes). The surroundings of the target may also play an important role in the detection process. It should also be taken into account.