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

Subproject

Deliverable type: IP Deliverable

Deliverable name: IP_D12/D16.4 Project final report and recommendations for future assessments

Version number	1.10
Lead contractor	VTT
Confidentiality status	Public
Due date of deliverable	31.1.2008
Actual submission date	31.1.2008
File name	PR-16100-SPD-080131-v110-D16_4_Final_Report

Authors

Johan Scholliers, VTT
Ralf Kutzner, University of Hannover
Juha Luoma, VTT
Niina Sihvola, VTT
Susanna Leanderson, Volvo Technology
Mariana Netto, LCPC
Jean-Marc Blosseville, LCPC
Martijn van Noort, TNO
Fabio Tango, CRF
Sami Koskinen, VTT
Somya Joshi, ICCS
Gerhard Noecker, Daimler
Mikael Ljung Aust, Volvo Cars
Jian Chen, ika
Maria Gemou, CERTH

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	15.10.2007	table of contents, generated
	3.12.2007	1st draft of safety assessment added for comments
0.5	17.12.2007	Susanna: Preliminary contribution from WP4 added.
0.6	20.12.2007	Integration with Ralf's contribution
0.7	27.12.2007	Second version of safety assessment
0.10	08.1.2008	version to be discussed at the consortium meeting
0.11	13.1.2008	structure changes according to the discussions at the consortium meeting.
0.12	25.01.2008	Changes in section 3.1 (MN). Introduction of section 5.2.1. (technical evaluation (WP16.300)) (MN)
1.00	31.1.2008	Version for peer review
1.02	1.2.2008	Results of CMS safety analysis added (MN) & conclusions updated (JL)
1.04	6.2.2008	Chapters 2 (Review of Methods), 6-8 (Project Results, dissemination) added.
1.05	6.2.2008	Comments of peer reviewers included
1.10	31.3.2008	Comments review included.

Table of contents

Executive summary

The PReVAL project addresses the possible safety impacts of applications developed and demonstrated in the PReVENT integrated project.

This deliverable describes the main results obtained in the PReVAL project, and reports the work which is performed in the last phase of the project, i.e. the quantitative safety assessment, the feedback on the framework and the recommendations in more detail.

The deliverable is both a deliverable for the PReVAL subproject and for the PReVENT IP (IP D12 “PReVENT impact assessment”).

The work in PReVAL has been concentrated along two lines of research:

- the analysis of the PReVENT evaluation results and the assessment of the safety potential of PReVENT functions. The technical and human factor evaluation results of the PReVENT subprojects have been analysed. A safety assessment has been performed on selected PReVENT functions;
- the development of an integrated framework for the assessment of active and preventive safety functions. The framework consists of procedures for technical, human factors and safety assessment.

Based on the analysis of the PReVENT evaluation results and the experiences from the framework development, recommendations for the system development process and the assessment of future safety systems are derived.

Assessment serves two purposes: to assure the functionality of the system, and to assess and quantify the system’s impact on the traffic system. The systems developed in PReVENT are mainly research prototypes, and the main objective of the performed evaluation was to assure that the system works as expected. All PReVENT subprojects followed the CONVERGE approach for the technical evaluations. All subprojects achieved good results for the reliability indicators (correct, false and missed alarm rates). The PReVENT project has hence demonstrated the feasibility of the demonstrated concepts and hence brought the technologies a step forward towards market introduction.

The human factors evaluation of six subprojects has been analysed. All the analysed subprojects report positive results on driving performance and driver behaviour, as well as for acceptance and usability, however with a variation in the significance and distribution of the results. Most projects emphasize the needs for further experiments to achieve statistically significant results and to optimise the HMI solution, as well as long term studies for evaluating e.g. effects due to behavioural adaptation.

PReVAL has assessed the safety potential of APALACI/COMPOSE, INTERSAFE (left turn assistant), MAPS&ADAS, SAFELANE and SASPENCE, using the

behavioural affect approach, which has been developed and used by the eIMPACT project, which is assessing the remaining PReVENT functions, as well as some other safety functions. All investigated safety functions have positive effects on fatalities and injuries, but in varying degrees. In part, this has to do with the use cases for which the safety function is relevant, and the frequency with which corresponding accidents occur. The safety assessment conducted in PReVAL has produced low, most probable and high estimates of injuries and fatalities for 100% fleet penetration, and for estimated low and high penetration rates in 2010 and 2020.

Starting from the experiences of the PReVENT subprojects and from the work of other related projects, procedures have been developed for technical and human factors evaluation, which have been integrated in a single framework. The proposed evaluation procedures have been applied by the INSAFES project to their validation plan. INSAFES provided feedback on the feasibility of the methodology. The framework was sent to selected evaluation experts for feedback, and discussed at the PReVAL workshop. The feedbacks received have been taken into account, and D16.3 has been updated. The result is Annex E. The expert evaluation method for human factors has been tested for the MAPS&ADAS functions. The safety potential assessment method has been developed by the eIMPACT project.

Based on the analysis of the evaluations and the experiences with the framework, a set of recommendations for the system development process and the needed research in evaluation are derived. Regarding the development of safety systems, a main recommendation is that the functional specifications should be based on the identified and relevant accident types, taking into account the status of the technology for detection of objects and control of the vehicle. In order to be able to verify the performance of similar systems, a common set of high-level scenarios is needed, which could also be used for homologation. Simulation tools and hardware-in-the-loop tests allow optimising the use of resources.

Evaluation in the PReVENT project was mainly targeted to assure that the systems work as expected. More statistically meaningful data on the effect of the functions on driver behaviour is needed. Data is needed for two purposes: to improve the technical performance of the systems, and to have more reliable data for safety assessment. Specifically, more data is needed on the long-term behavioural adaptation, which is caused by preventive safety system introduction. This demands for naturalistic driving and Field-Operational Tests.

1 Introduction

1.1 Overview of PReVAL objectives

The PReVAL project addresses the possible safety impacts of applications developed and demonstrated in the PReVENT integrated project. The main objectives of PReVAL are:

- to identify best practices for the assessment of IP PReVENT safety applications;
- to define a framework for estimating the system's safety impact taking into account technical performance and human factors;
- to apply the framework to estimate the potential safety impacts of selected PReVENT applications;
- to make recommendations for future assessment and development of preventive safety applications

1.2 Project Workflow

The work in PReVAL is concentrated on two lines:

- the analysis of the PReVENT evaluation results. The results of the evaluation work in the different PReVENT functions are gathered and analysed. The work is concentrated on the analysis of the technical performance and the human factor evaluation results. The safety potential of selected PReVENT functions is assessed;
- the development of an integrated assessment framework for the assessment of active and preventive safety functions. This framework takes the experience gained in the technical and human factors related evaluation activities in the various PReVENT subprojects – all of these are conducting evaluation activities, which, obviously, are designed with respect of the particular needs of the respective subproject – as basis and in parallel investigates the activities done in related projects. The aim is, on the one hand, to identify what can be called "best practice in evaluation" and, on the other hand, to find out, how the very application specific evaluation activities done in the PReVENT subprojects or related projects can be generalized in a way, that they produce the expected, comparable results.

Figure 1 shows the workflow of the PReVAL work:

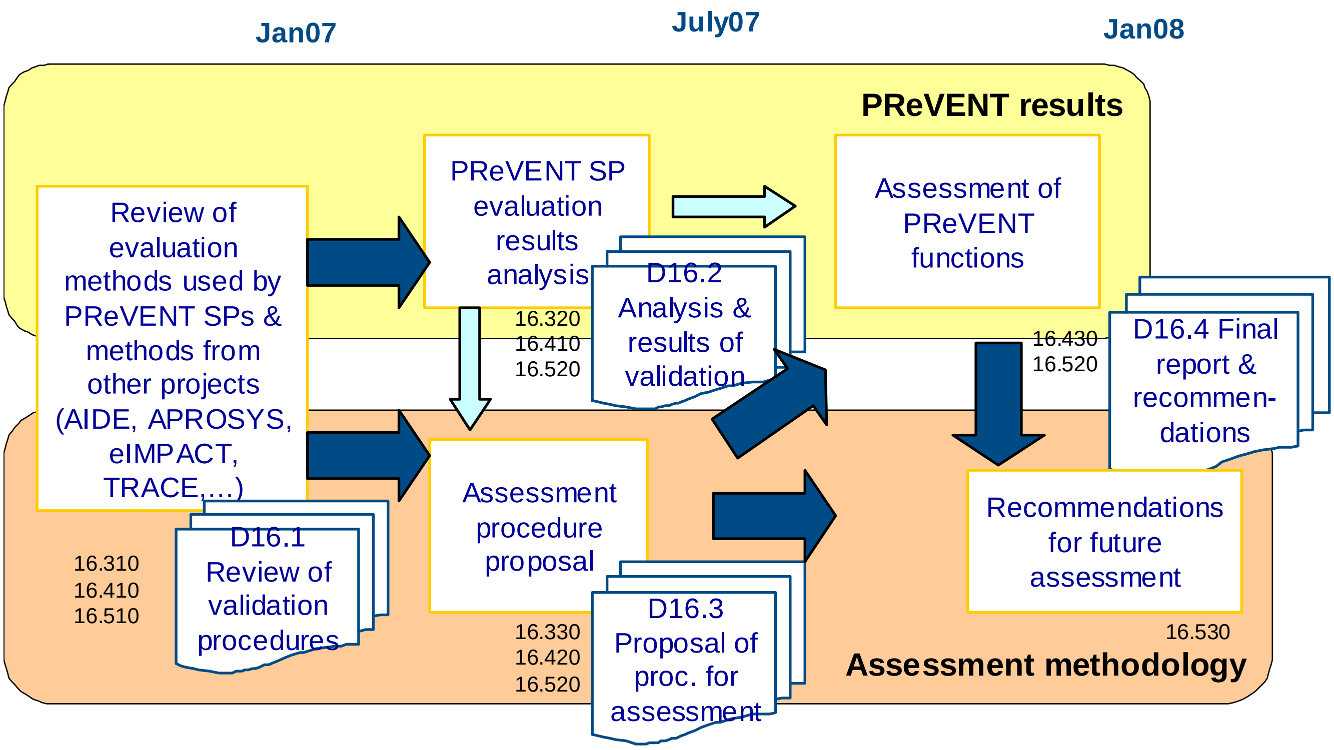

Figure 1: PReVAL project phases

For the results, the work goes through the following steps:

- analysis of the evaluation plans of the different PReVENT subprojects;
- analysis of the evaluation results of the PReVENT subprojects, more specifically technical and evaluation results;
- safety assessment of PReVENT functions. The safety assessment is performed in two phases: a qualitative safety assessment, reported in D16.2, and a quantitative safety assessment, reported in this deliverable. PReVAL uses the methodology, which has been developed and is used by the eIMPACT project.

For the assessment framework, the work goes through the following steps:

- review of the evaluation methods used by the PReVENT subprojects and related projects, such as APROSYS, AIDE and eIMPACT;
- development of the framework, consisting of procedures for technical and human factors evaluation. For safety potential assessment, the method developed and used by eIMPACT is selected.
- the framework is applied within PReVENT and the methodology is discussed with experts. The methods for human factors evaluation with respect to expert evaluation is applied to selected PReVENT functions;
- the different approaches (technical, HMI, safety potential) are integrated to obtain one holistic approach. The updated framework is included as Annex E.

1.3 Structure of the deliverable

This deliverable has two purposes:

- to give an overview of the PReVAL results and achievements;
- to describe the work in the last phase of the PReVAL project, which has not been reported previously.

This chapter provides an overview of the project objectives and of the work methodology for the last phase of the project. This work includes the quantitative safety assessment, the application of the framework and finalising the framework, and setting up guidelines for future assessments.

Chapter 2 gives an overview of the project's first phase: the review of the evaluation methods.

Chapter 3 gives an overview of the results from the analysis of the PReVENT evaluations. This includes the results of the technical and human factors evaluations, and the quantitative safety assessment.

Chapter 4 gives an overview of the assessment framework. In deliverable D16.3 the procedures proposed for technical, human factors and safety assessment were reported. The procedures have been applied to selected PReVENT functions, and feedbacks obtained and implemented. The expert analysis method, proposed in the human factors evaluation procedure, has been applied to the MAPS&ADAS functions. The results of this analysis are reported in Annex D. Additionally feedback has been obtained by assistance to the INSAFES validations. The INSAFES team applied to PReVAL methodology to their validation plan. Feedback on the framework is received from INSAFES, from external experts and through discussions at a workshop, organised by the PReVAL project together with TRACE and eIMPACT. The updated version of the framework is described in Annex E.

Chapter 5 describes the recommendations for the system development process of preventive and active safety functions, and the guidelines for future assessment programs.

Chapter 6 gives an overview of the project results, and describes how PReVAL has achieved its objectives. Chapter 7 describes how PReVAL is linked to the other PReVENT projects. Chapter 8 gives an overview of the project deliverables and dissemination.

In chapter 9 the conclusions of the project are presented.

1.4 Methodology

This section describes the methodology used for the work during the last phase of the work, which has not been reported in the other deliverables. This work includes:

- quantitative safety assessment of PReVENT functions. This work is reported in Section 3.3;
- updating the framework by application of the framework and through feedback on the methodology. This includes expert evaluation of the MAPS&ADAS functions, the main purpose of which was to gain practical experience and identify

weaknesses in the methodology. This work is reported in Section 4.2.1;

- making recommendations on the assessment of preventive safety applications.

1.4.1 Safety assessment

1.4.1.1 Scope of assessment

The background of safety assessment of PReVAL is reported in D16.1 and the qualitative safety assessments are given in D16.2. As the safety assessments of PReVAL and eIMPACT were coordinated, PReVAL focused on five functions that were not covered by eIMPACT (Table 1). Specifically, the assessed systems and their functions were collision mitigation system of APALACI/COMPOSE, left turn warning of INTERSAFE, MAPS&ADAS, SAFELANE and SASPENCE.

Table 1. Assessed PREVENT functions by project.

System, function	PReVAL	eImpact
APALACI/COMPOSE, collision mitigation system	X	
INTERSAFE, Left turn assistance	X	
INTERSAFE, Right of way assistance		X
INTERSAFE, Traffic light assistance		X
LATERAL SAFE		X
MAPS&ADAS	X	
SAFELANE	X	
SASPENCE	X	
WILLWARN, Detection and warning of obstacles on the road vehicle		X
WILLWARN, Detection of reduced friction or reduced visibility through bad weather		X

The analyses were carried out using the behavioural effect approach developed in eIMPACT, which is described in detail in Annex E, Section 5.

Apart from other functions, the safety assessment of APALACI/COMPOSE was conducted differently. The assessment focused on the mitigation system of APALACI/COMPOSE, which is expected to affect road safety by modification of accident consequences.

1.4.1.2 Overview of the safety impact analysis calculation

The aim of the work was to provide estimates of safety impacts for the selected functions in terms of percent changes of fatalities and injured persons in two target years, 2010 and 2020, and in two penetration scenarios which were business as usual and an enhanced, promoted business scenario.

The method used for the safety potential estimation is described in Annex E, Section 5.

The safety assessment goes through the two phases. First is a qualitative phase, consisting of:

- system and function definition;

- literature survey of the effects of the function;
- relevant safety impact mechanisms.

This qualitative phase was reported in D16.2, however some of the impact mechanisms were necessary to complement during quantitative analyses. The second phase is the quantitative assessment, in which the effects of the mechanisms are estimated. The calculation of the efficiency method is explained in Annex E, Section 5.5.

For Collision Mitigation Systems (CMS), which only affect road safety by modification of accident consequences, the potential of accident reduction with CMS is estimated by means of calculations that take as the basis the kinetic energy reduction due to CMS introduction. A high number of use cases is considered for the identification of the targeted accidents. The reaction of the driver to the imminent collision (braking correctly, braking not enough and not braking at all) is taken into account and a sensitivity analysis is carried out including sensitivity with respect to CMS parameters (CMS activation TTC (Time to Collision), CMS braking capability) and with respect driver reaction. The method has been proposed and carried out by LCPC.

Penetration rates

Penetration rates were provided by eIMPACT. Firstly, the estimates were provided, based on expert evaluation of different stakeholders, for new vehicle penetrations. It was expected that it will take approximately 5 years until a system is available in all new car models from a specific year after the decision to implement (even mandatory) has been made. In addition, it will take several years before the whole vehicle pool is renewed and the systems will be in more wide use in fleet.

For some systems some technical difficulties may be to be solved before a wider use. For the cooperative systems, also the infrastructure should be implemented.

1.4.2 Evaluation framework

Deliverable D16.3 described the framework, which is proposed by the PReVAL project. During the last phase of the PReVAL project, one of the aims has been to apply and review the methodology. Experience on how to apply the methodology in practice has been achieved both by external feedback and from internal work, by application of the methodology.

1.4.2.1 Application of the methodology

To identify ways in which the evaluation methodology can be improved or what further information and knowledge that is needed was of great importance for defining the final framework and also for providing recommendations for future research and assessment programs.

The work in PreVAL included an expert evaluation of the MAPS&ADAS functions. Through this work, further results related human factors related was achieved for the MAPS&ADAS system, as well as practical experience on how to apply the human factors methodology.

Expert-based evaluation was performed on the two MAPS&ADAS functions speed limit warning and hot spot warning, the functions have been integrated in a demonstrator car at University of Hannover. The results, complementary to the already achieved results within MAPS&ADAS are presented in Annex D.

The collaboration with INSAFES with respect to the human factor evaluation procedure has provided important feedback on the methodology. The human factors evaluation procedure presented in D16.3 was delivered to INSAFES in July 2007, with aim that it should found the basis for their human factors evaluation. The INSAFES consortium has used parts of the PReVAL methodology when developing the evaluation plan for the INSAFES functions. External feedback from this work has been delivered to the PReVAL in terms of review comments and recommendations for potential improvements of the methodology with intention to assist the progressive work with the methodology.

Valuable experience on human factors evaluation was also achieved when partners participated in the INSAFES user tests on the Volvo Cars demonstrator car. The aim with this work was to a) evaluate the HMI in these applications and b) to evaluate the human factors evaluation methodology in practical use. The results from the INSAFES test will not presented in this report, but is included in the overall user test results presented within INSAFES.

1.4.2.2 Feedback on the framework

D16.3 has been sent to selected experts for comments, and their feedback is taken into account for the updated framework.

A final workshop, in which the results of the project and the method are explained, has been organised in Brussels on 10.1.2008.

The discussions of the workshop have, as far as possible within the remaining resources of the project, been taken into account for the final framework. Suggestions, which could not been integrated, have been included in the recommendations for further research section.

1.4.3 Recommendations for further research

One of the objectives of PReVAL is to make recommendations for further work in the development of effective preventive safety functions.

This work comprises:

- recommendations for research in safety functions development
- recommendations on evaluation of preventive safety functions
- recommendations on evaluation research

Research in safety functions development. Assessment is a fundamental part of the development process of safety functions.

By examining the assessment process, also considerations on the complete development process can be made. The recommendations are given in Section 5.1.

Recommendation on evaluation of preventive safety functions. The framework, reported in Annex E, gives guidelines on the evaluation of preventive safety systems, including the selection of methods and tools.

Recommendations on evaluation research. Starting from the review of methods and the development of the framework, gaps and areas for improvement in evaluation can be identified. Examples are the need for long-term research and the lack of data to support safety analysis. A list of research items is compiled, starting from:

- gaps and problems identified;
- comments from experts on the methodology;
- discussions at the final workshop.

The latter work is reported in Section 5.2.

2 Review of methods

The aim of the first phase of the PReVAL project is to make an overview of the evaluation work performed in the PReVENT subprojects and of the methods developed in related projects, which deal with evaluation. The work has been reported in D16.1.

2.1 Technical evaluation

The validation in the PReVENT subprojects is mainly pure technical, which is understandable since most systems are early prototypes, which consist of different advanced technological components which all have to work together. The main assessment objectives in the validation are linked to checking if the system works as supposed.

The technical validation follows the CONVERGE approach, as has been previously agreed within the PReVENT Consortium. There are differences in the way the approach is implemented, e.g. if only the complete system or also subsystems are evaluated, the accuracy of the assessment objectives. Due to limited resources for testing and the non-destructive character of testing, only a subset of the possible real-life situations can be evaluated. The use of simulation, using accurate sensor models, including vehicle-hardware-in-the-loop simulation can assist in reducing the number of tests required. Areas for improvement in the technical evaluation are a clear definition of the reference measurement, the way in which the environment is simplified (target representativity), and the way in which environmental factors are handled (e.g. adverse weather conditions.

2.2 Human factors evaluation

As most of the PReVENT systems are early prototypes, the design of the HMI has often not received major attention. Not all sub-projects have made HMI-evaluations. Since there is a lack of material for doing a satisfactory assessment, work from related projects (RESPONSE 3, AIDE, INVENT, HUMANIST, ADVISORS, ISO 17287:2003, TC22/SC13/WG8 WI023) has been studied.

Based on the review of PReVENT user studies and existing methodologies for ADAS HMI and driver behaviour evaluation, some general conclusions may be drawn. First, it is clear that well established methods for ADAS (Advanced Driver Assistance Systems) HMI evaluation are scarce. Most existing work on automotive HMI evaluation, e.g. the European Statement of Principles [36] and HASTE [35], has focused on IVIS (In-vehicle Information Systems) which generally require entirely different tools and measurements. One key reason for the lack of a general ADAS evaluation methodology is probably that, by contrast to IVIS, the effects of interest (and the associated hypotheses) are quite specific for each ADAS function.

There seems to be is little consensus on terminology. Thus, it is not always entirely clear what is meant e.g. by “HMI evaluation” or “impact evaluation”. Sometimes, HMI evaluation is even viewed as synonymous solely to acceptance evaluation! It is thus an important task for PReVAL to create a common general conceptual framework that can be used for clearly describing the general scope of an ADAS user test and stating clear and specific hypotheses about the expected effects of the system.

Finally, it is clear from the review that the definition of an ADAS human factors evaluation methodology is far from a trivial task. Thus, the expectations for the methodology development work should be kept realistic. It seems advisable to start from a top-down perspective and resolve the general conceptual issues first, before going into the details of specific methods and tools. A key issue is to find a way to connect clearly stated hypotheses (derived from the functional specifications of the system as well as empirical results of potential behavioural effects) with the selection of appropriate methods and tools.

2.3 Safety potential

The review of PReVENT material showed that although the technical descriptions of the systems were very detailed, the functional and traffic situation related (scenarios) descriptions were quite often in very general level. To be able to do safety assessments, more detailed descriptions including for example the system (warning) strategy, a detailed description of relevant scenarios and use cases and system limitations is needed. The more detailed analyses (annexes G-N in D16.1) of the existing evaluations showed that the safety related impact evaluations were somewhat limited and cover only some of the functions and the scenarios.

The safety impact is affected by both the technical performance, the HMI (interaction between the driver and the vehicle, behavioural change of the driver), and the traffic safety level (safe operation of the traffic system, interaction between users and non-users). Traffic safety level evaluations are often quite demanding both on resources and time, and lack of information is problematic in some cases. The use of existing research (literature review) and expert opinions plays therefore an important role in the safety assessments. Also simplifying the safety assessment is sometimes needed.

3 Analysis of PReVENT functions

This chapter gives an overview of the results of the analysis of the PReVENT technical and human factor evaluation results. This section also contains the results of the quantitative safety assessment of selected PReVENT functions.

3.1 Overview of technical evaluation results

3.1.1 Introduction

PReVENT subprojects can be classified in different function fields, which correspond to different types of interactions. The "time to risk" is the major key differentiated parameter. Directly related to the "time to risk", the horizon (in time or distance) to be considered is also a key differentiated parameter. Those function fields are consequently fed by technologies showing complementary information access capabilities in term of time and distance. A distinction is made in three function fields.

1. The first function field deals with tight (<1s) **and short interactions related to collision mitigation and avoidance**. There is a risk of immediate collisions with obstacles, which are - for collision mitigation systems - unavoidable. The time and distance related parameters are very short. The technologies able to address those circumstances are based on perception sensors located on-board. The explored concepts are based on pre-crash, collision prevention and mitigation systems. Three PReVENT subprojects address this function field: APALACI, COMPOSE (frontal imminent collisions, tight interactions), INTERSAFE (side and frontal collisions especially those related to intersections, short interactions)

2. The second function field addressed by PReVENT relates to **short interactions with other moving objects** constantly managed by the driver (1 to 5s). The time and distance related parameters are short and can be addressed again by on-board technologies. Maps and vehicle-to-interface communications can be used as additional sensors. Three PReVENT subprojects address this function field: SASPENCE (frontal interactions), LATERAL SAFE (lateral and rear interactions) and SAFELANE (lateral interactions with the lane and road sides).

3. The third function field corresponds to **more distant interactions**. The time and distance related parameters are longer **(~> 5 s).** The decision time is not critical; the implied correction operations can be assimilated to precaution more than accident prevention. The concept is based on the creation of an electronic horizon that enables foresighted driving. A distinction is made between two classes of addressed risks:

- A first class of risks, addressed by MAPS&ADAS, is permanent and located on the road. The technologies are localisation techniques and digital maps. The PReVENT subproject covering this function field is MAPS&ADAS.
- A second class of risks, addressed by WILLWARN, is highly dynamic and non permanent and cannot be

accessed through maps alone. Dynamic access to the information is needed, and telecommunications and localisation techniques are the “natural technologies” to be considered in this function field.

An additional remark has to be made. PReVENT subprojects address many new concepts and technologies, for which no well established standard assessment procedures exist. So, besides of the challenge to build such innovative concepts, the evaluation procedure in itself is innovative as well, specially taking into account the tight time constraints, since PReVENT is not an assessment dedicated project. In the following a summary is given of the analysis of the evaluation results of the PReVENT subprojects.

3.1.2 Overview of Results

Function field 1: Tight and short interactions related to collision mitigation and avoidance: APALACI, COMPOSE, INTERSAFE

APALACI and COMPOSE have dealt with risk of crash: the risks consist of immediate collisions with obstacles. Since the assistance activates when the accident is unavoidable, the cooperation level between the assistance and the driver is very low. COMPOSE concentrates on autonomous braking systems, APALACI applications address semi-autonomous braking.

The main challenges of APALACI (pre-crash & Collision Mitigation System, CMS) and COMPOSE (Vulnerable Road User protection & CMS), are the detection of potential obstacles in very short time intervals (for both projects) and the classification of obstacles - making a difference between vulnerable road users and other obstacle types - (for COMPOSE). APALACI and COMPOSE have innovated by the use of several combinations of sensors that have been explored and tested: ultra-sonic sensors, short range radars and cameras for short distance, lidar and long range radars for long distance obstacle detection. Fusion and tracking techniques have proved to be powerful: no missed alarms; weak rate of false alarms (0% false alarm rates for APALACI; and 2/253 situations; 0/5h18 driving, for COMPOSE); good classification rate (>98%). Excellent detection rate (100%).

INTERSAFE extends the collision assistance systems towards intersection contexts: INTERSAFE can be considered as the PReVENT subproject that deals with less mature technologies, since collision prevention in intersections involves the development of a highly complex detection system (fusing information brought by precise maps and on-board sensors) that has to be able to locate and classify objects in the intersection. Quite good results (0% missed alarms, 7% FAR) have been obtained and demonstrated through very precise assessment tools.

Function field 2: Short interactions with other moving objects (1-5 s): SASPENCE, LATERAL SAFE, SAFELANE

These PReVENT subprojects, while addressing each one a privileged direction, can be characterized by the ambition to cope with a high variety of risks factors. Consequently, fusion of information is a keyword. Since larger time constants are present, the cooperation level between driver and assistance is higher demanding a stronger effort in the HMI conception. Also, since the available time before the accident is larger, an additional challenge concerns the conception of a decision system able to choose which type of assistance suits better to each driving scenario.

In **SASPENCE (support for safe speed and safe distance)** sensors (long range radar, digital maps, camera) provide data which are fused at multiple levels to provide an enhanced view of the environment ahead (obstacle, vehicle in front, road geometry). SASPENCE has innovated in the computation of information and warning criteria, based on the comparison between the optimal manoeuvres and the actual ones (travelled by the ego-vehicle), also considering various risk factors. In SASPENCE, much effort has been put on searching a suitable HMI. Technical results (Missed alarm rate: 7.5%, 4% FAR) indicate that further work of function tuning (e.g. threshold alarms setting) and longer tests are needed. The validation phase of SASPENCE included the use of innovative and promising tools: digital and hardware-in-the-loop simulation (PRESCAN and VEHIL environments).

Still in group 2, **LATERAL SAFE (Lateral collision warning and lane change support)** addresses the risks of collisions with vehicles or objects being on the side or approaching behind the equipped vehicle. Complementing a frontal support, lateral & rear 270° bird view monitoring is the concept. This is achieved through three information and warning systems based on short range radar (side), long range radar (rear) and cameras (in side and back mirrors). At the technical level, fusions have proved to be powerful. Technical assessment was based on subjective perception of the driver and by on-board technicians. Actual false and missed alarm rates can be attributed partly to threshold alarm setting. The obtained rates are listed in the table below. LATERAL SAFE addresses very complex problems, as the driver has to be warned continuously of events happening all around the vehicle. This can explain the obtained rates.

Table 2: LATERAL SAFE false and missing alarm rates - LCA: Lane Change Assistant, LCW: Lateral Collision Warning, LRM: Lateral and Rear Area Monitoring, SRR: short range radar, BSD: blind spot detection.

	LCA	LRM (3 cameras)	LRM (BSD-based)	LCW-DC (SRR based)	LCW-CRF (fusion based)
False alarms/ total incidents (%)	0.7	4.5	1.3	2.8	2
Missing alarms/ total incidents (%)	6.1	10.5	0.7	3.6	2.4

Longer technical assessment of LATERALSAFE-like functions is needed in order to validate the correct (or false) function activations from a purely technical point of view. The results are affected by two factors: the threshold settings and the use of subjective perception during evaluation. The activation temporal threshold should not be too small such that risk can be avoided, but should not be so high to not be activated all the time, compromising the system acceptance. Since the LATERAL SAFE function addresses interactions in different areas around the vehicle, the choice of the threshold is not an easy task. The indicators are based on subjective perception of the driver. It is very hard to decide whether the system has been activated correctly (i.e. if the TTC has been calculated correctly), and therefore PReVAL suggests to use appropriate reference measurements. There is also a risk that observers have difficulties to discriminate between the pure technical assessment (i.e. is the message given according the TTC) and the driver perception of the risk – which is a more user acceptance related issue.

SAFELANE (Active lane keeping support aids drivers in staying in their lane through warning and corrective steering) has innovated first in a robust lane tracker that uses data fusion (map data, radar object trails, vehicle dynamics sensors and camera) that has been developed and thoroughly tested. A second innovation of SAFELANE is the decision system that comprises a situation model based on the knowledge of different elements (driving manoeuvres, road conditions, and situation characteristics). The decision model decides which type of assistance the system is able to provide to the driver: Nothing, information, warning, correction. Through an extensive evaluation, the system showed very good results (~0% False Alarm Rate and Missed Alarm Rate). The results show as well that all lane keeping support modes clearly improved the drivers' global performances, resulting in a significant and large reduction in the duration of lateral excursion and in the steering reaction times both in bends and in straight lines.

Function field 3: More distant interactions (>5s): MAPS&ADAS, WILLWARN

MAPS&ADAS brought mainly in the PReVENT basket of technologies, a CAN interface for map horizon data (that can be a standard one for the future) and functions capable of providing hot spot and speed limit warning. Extensive testing of all horizon providers implementations has been carried out: no errors in the CAN usage occurred, 100% of integrity of the map horizon was observed. In terms of positioning, the performance of existing positioning systems used for navigation has shown to be sufficient. The map quality governs the system performance.

WILLWARN has enriched the creation of an electronic horizon through telecommunications. WILLWARN warns drivers early whenever a safety related critical situation occurs ahead, especially obstacles, adverse road and weather conditions or hazardous construction sites. WILLWARN complements the PReVENT capabilities with several modules: Hazard Detection Module (sensor data collection and processing), Hazard Warning Module (HWM), Warning Message Management Module (relevance check of the incoming messages, birth and death of relevant messages…) and Vehicle to Vehicle Communication Module. The complete WILLWARN system was successfully tested and presented in various use cases with four cooperating cars and a road side unit on public roads. One should point out the contribution of WILLWARN to the position relevance check that verifies if the host vehicle is concerned by the receiving message.

Moreover, further tests are needed to study the effects of more communicating partners through ad-hoc networks.

To conclude, a fundamental outcome in PReVENT is that each PReVENT subproject has looked at different technologies in order to improve the performance. PReVENT sub-projects make then use of a large set of different sensors in a global sense (proprioceptives & extereoceptives, maps, telecommunications, …) that enter the data fusion module in order to constitute reliable detection/positioning inputs (with an enlarged detection area as a result of this combination) for the proposed assistances. PReVENT constitutes then a large "basket of new technologies", fundamental main bricks for preventive safety systems. Then, PReVENT detection results are very significant, shown by the excellent rates obtained. Some projects, like LATERAL SAFE and INTERSAFE address very complex problems and need further evaluations. In general, longer evaluations are needed in order to give more statistically significant (and blocking) results. Many PReVENT evaluations included a wide variety of test conditions but, due to time restrictions, results have been averaged under the set of these different conditions. This demands for "*Blocking":* in future long duration tests groups of tests in the same conditions (road, lane markers, meteorological, etc) should be grouped.

3.1.3 Best Practices

All validation plans were good and follows the CONVERGE methodology. In this section, we extract some important features that can be illustrated through the actual validations performed in PReVENT subprojects.

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

<u>A central point in the methodology: ways to measure the indicators of success: the "reference measurement":</u> 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. Examples are given in Annex E, section 3.2.4.

<u>Combination of simulation environment with hardware-in-the-loop tests</u>: in a first phase of the validation a simulation study on a high number of scenarios allows identification of the most critical cases, which are subsequently tested in a hardware-in-the-loop environment. SASPENCE follows this process.

<u>From technical assessment to HMI assessment</u>: for a very HMI closely linked function like SASPENCE, it becomes necessary to exploit more than simply reliability indicators. Indeed, in SASPENCE a large set of indicators have been taken into account. This set contains: reliability indicators (MAR, FAR), comfort indicators (RMS value of the longitudinal acceleration), safety effect indicators (minimum TTC during the scenario),

appropriateness indicators (is warning level appropriate? Linked to HMI) and timing indicators (is the warning in time, too soon, too late?).

3.2 Overview of human factor evaluation results

3.2.1 Introduction

One of the goals with PReVAL was to survey and summarize the results from the human factor evaluations from PReVENT functions. Based on an initial survey on the evaluations done within PReVENT, a subset of six PReVENT subprojects (INTERSAFE, LATERAL SAFE, MAPS&ADAS, SAFELANE, SASPENCE, WILLWARN) were retained for further analysis. The evaluations made on human factors have been more or less extensive in the subprojects; in general human factors have not been the main focus within PReVENT. The deliverable D16.1 describes the methods used, while the deliverable D16.2 focuses on analysing the results achieved, categorized in terms of effects found on driving performance, driver behaviour, acceptance and usability. Except for summarizing the results, the clarity of the results was reviewed in PReVAL, based on how they were reported in the subprojects. In general the results could not be compared between the subprojects since the results reported are very heterogeneous and the scope of evaluations differs between the subprojects.

In this deliverable a very brief and simplified overview of the human factors related results is provided. An overview of the studies that have been performed and thereby found the basis for the results is provided as well.

The summary is based on the results and findings that were concluded in the subprojects with respect to driving performance, driver behaviour, acceptance and usability and also takes into account how extensive and complete the results were, which is an important aspect when studying the results. The reliability, significance and practical meaning of the results is directly influenced by the scope of the evaluation; the type of tests carried out, the tools and equipment used and the extent of the test in terms of sample sizes of test groups and duration and repetition of the tests.

Table 3: Overview on the final human-factor-related results and the methodologies used in the final evaluations as reported in the PReVENT subprojects.

Subproject	Conclusions on human factor related results*	Subjective data evaluation	Objective data evaluation
INTERSAFE	+	Sim.** (47) Test track (16)	Sim**. (47)
LATERAL SAFE	+	Sim**.(21) Test track(12)	Sim**. (21) Test track (12)
MAPS&ADAS	++	Pub. roads (67)	Pub. roads (67)
SAFELANE	++	Sim**. (20) Test track (10)	Sim**. (20)
SASPENCE	+	Pub. roads (20) x 2	Pub. roads (20) x 2
WILLWARN	+	Sim**. (40)	Sim**. (40)

* The results reflect system effects on driving performance, driver behaviour, acceptance and usability.
** Driving simulator environment
++ Positive and evident results concluded in subprojects.
\+ Positive results concluded in subprojects and/or with documented need for further tests to establish results.

3.2.2 Results

There is a quite large variation in the methodologies used for evaluation. Partly this variation is natural due to the different characteristics of functions but also the scope of evaluations differs between the subprojects, and also the way the results are interpreted and reported.

Due to the differences in functions, in scope of the evaluation and the methodology used within PReVENT the reported results can not easily be compared between subprojects.

The evaluation performed in PReVENT has focused on *short term* testing on *intended* (desired) system effects. Long term effects such as behavioural adaptation effects have not been addressed within PReVENT. In some projects undesired effects such as high workload and driver distraction have been addressed, but the focus has mainly been on intended effects.

All subprojects that have been reviewed in PReVAL report positive results on driving performance and driver behaviour, as well as for acceptance and usability. However, as mentioned, there is a large variation in the methodologies used and by that in the significance and reliability of the results. Variation is found with respect to

- the definition of the evaluation objectives
- the extent and outlines of the hypotheses made
- the type of studies used
- the number and distribution of test subjects in each study
- the amount and type of data collected

- the data analysis performed
- the interpretation of results

One subproject where human factors had a major focus was in the MAPS&ADAS project. The hypotheses used in the evaluation addressed both intended and unintended effects and significant positive results were concluded, based on real road test with a large test group of subjects.

Most subprojects emphasize the need for further experiments with a larger amount of scenarios and a larger group of test subjects for achieving statistically significant results. Also further tests for optimising the HMI solutions are mentioned in some subprojects as well as the need for assessing the long term behaviour of the driver

3.2.3 Best practices from PReVENT

In this section a discussion is provided on the best practices for human factors evaluation based on a review of methodologies used in the subprojects.

In most subprojects human factor related evaluations and potential system impacts on the traffic system have been done at an early stage of the projects, during the development phase

Different types of simulations and HMI pre studies are often used prior to the final evaluations. For example traffic simulations for investigating relevant traffic scenarios and potential system impacts. Software simulations and hardware-in-the-loop simulations have also been used for evaluating potential system effects. In several projects, extensive pre-studies for defining suitable HMIs were performed, which can be concluded as positive for having a good status of the HMI in the final user tests.

In the final evaluations, **different types of studies complement and supplement each other with respect to data collection**. Driving simulator studies are used in most projects as well as tests on test tracks and/or public roads.

Objective data collection and **subjective data collection** have been used for evaluation of human factors related aspects, sometimes in a complementary way. All projects have evaluated, to various extents, driver behaviour and driving performance, primarily based on **statistical analysis of recorded (logged) data** but also sometimes in combination to subjective data collection on the driver’s own view on his/her driving performance during the test drive.

For a few projects objective data has been collected and used for evaluating driver distraction.

In addition to driver behaviour and performance, most projects have addressed **acceptance and usability**, by subjective data collection via questionnaires. Slightly different definitions of these concepts and different methods for acceptance assessment were used. **The Van der Laan scale seems to be the most common for evaluating acceptance** [20]. Usability has been addressed in most projects, even here with slightly different definitions. Often

Brooke's usability scale [19] has been used for evaluating this data.

Except for acceptance and usability subjective data collection has also addressed other subjective related items such as workload, willingness to pay for systems and subjective experience on the system's impact on safety.

While similar evaluation techniques for evaluating subjective issues are found in the subprojects there seems to be little consensus on specific indicators reflecting driver performance and behaviour. This is partly a natural consequence from the different functionalities in the subprojects.

The system effects reflected by changes in driving performance and driver behaviour and the interpretation of these to impact on safety is not made in a generalized way, which show that it is not easy to quantify and translate the test results onto impact on road safety.

One interesting method that was used in the SASPENCE project was the **definitions of different driver types** used in a simulated hardware-in-the-loop environment for evaluating the system effects. Different results were concluded for different driver types, which emphasizes the need of a further developed test methods that includes different drivers, where further research on driver models would be valuable.

From further study of the methodologies, it can be concluded that **clear hypotheses prior to the evaluation are of importance**, in order to prioritize among the often large amount of recorded data available from the trials. User tests and consequent data analysis are resource consuming, thus clear hypotheses and indicators will help to prioritize how defining and setting up a test; at the end, this will make data analysis after the trials easier.

A key for achieving significant and reliable results is the use of **a scientifically calculated value of the needed sample size** in user tests. Collection of data from large test groups is essential in order to achieve results that has a significant effect and also implies practical meaning. Handling of large test groups is resource demanding both with respect to carrying out the tests but also to analyze the data afterwards, which puts high demands on prioritization when planning the evaluation. Large sample sizes might imply a need for reducing the number of hypotheses addressed in the evaluation and a reduction of the amount of data recorded in the tests. This in turn means that it is of great importance to focus on the most essential things in the evaluation and to **prioritize among all possible aspects of the evaluation**.

It is also of importance to consider what kind of experiment design and method is the most appropriate for the evaluation purpose and how the group of users should be composed. This depends on the timing of the evaluation in the development phase, the resources available, and the kind of realism that is desired. For example, user tests run in a simulator environment might be appropriate when testing situations, which are safety critical for the driver, but the tests might not provide realistic limitations in the functionality. A problem with simulator studies is that the functions are often "too good"; it is, for example, very hard to simulate a radar as it works generally in real environment.

An important, but also difficult aspect, is to define the baseline conditions in tests, to which the achieved results will be compared when evaluating them. Often test drives performed with and without system will provide the baseline condition but ideally it also requires that the test drive without the system represents the "natural driving situation" for the driver.

In order not to have bias in the results when performing tests with and without a preventive safety system it is important to have a **balanced experiment design** for avoiding familiarisation effects in the results, for instance originating from the drivers being unaccustomed to the test vehicle at the beginning.

For a complete evaluation basis the **short term studies should be followed up with long term studies.**

The best practices for final validations of PReVENT can be summarized according to below:

- **Extensive pre studies of different HMI solutions and simulations with different driver types** have been used within PReVENT which should serve as important input to the final evaluations.
- **Various types of studies have been used** that complement and supplement each other with respect to data collection.
- **Objective data collection** has been used for evaluating primarily driving performance and driver behaviour and in some cases, driver distraction.
- **Subjective data collection** has been used for evaluating acceptance and usability, but also other subjective aspects such as experienced workload and experienced system impact on traffic safety and willingness to pay for the system.
- **Clear hypotheses** addressed, for focusing the evaluation to the important parts.
- Numbers of test persons are based **on a scientific calculation of the sample size** needed for achieving significant results with practical sense.
- A **balanced experiment design** is of importance to avoid familiarization / learning effects.
- The results from short term testing should be followed by long term evaluation of long term effects, when possible.

An additional discussion on human factors evaluation in general is provided in Chapter 4.4.2.

3.3 Quantitative safety assessment

For the safety assessment, the behavioural effect method, developed and used by the eIMPACT project is used. The calculation of the efficiency method is explained in the Framework

Annex E, Section 5.5. The safety mechanisms, through which ADAS functions affect safety, are a key element in the study. The safety mechanisms are briefly:

1) Direct in-car modification of the driving task
2) Direct influence by roadside systems
3) Indirect modification of user behavior
4) Indirect modification of non-user behavior
5) Modification of interaction between users and non-users
6) Modification of road user exposure
7) Modification of modal choice
8) Modification of route choice
9) Modification of accident consequences

All the estimates in the analysis are based on previous knowledge about function, other telematic applications or knowledge about driver reactions and behaviour. Therefore, a state of the art review concerning the known safety impacts has been carried out. The literature searched is classified as follows:

- Empirical evidence on safety impacts (verified results e.g. experimental design)
- 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 systems, e.g. road side telematics (potential results). These are usually referred as “assumptions”.

For Collision Mitigation Systems (CMS), which only affect road safety by modification of accident consequences, the potential of accident reduction with CMS is estimated by means of calculations that take as the basis the kinetic energy reduction due to CMS introduction.

3.3.1 INTERSAFE (left turn assistance)

3.3.1.1 Overall effects

INTERSAFE includes three functions: left turn assistance, right of way assistance and traffic light assistance [3]. The following assessment deals with the left turn assistance, while other functions are assessed by eIMPACT.

Table 4 presents overall effects of the left turn assistance of INTERSAFE on proportion of fatalities and injuries by target year and penetration level (fleet penetration of passenger cars).

Table 4. Overall safety effects of the left turn assistance of INTERSAFE by target year and penetration level.

	Penetration rate for light/heavy vehicles (%)[2]	Reduction (%) in	
		Fatalities	Injuries
Impact low[1]	100/100	-0.2	-0.6
Impact most probable[1]	100/100	-0.6	-1.5
Impact high[1]	100/100	-1.1	-2.9
2010 low	0.0/0.0	0.00	0.00
2010 high	0.0/0.0	0.00	0.00
2020 low	0.3/0.4	0.00	-0.01
2020 high	0.5/0.7	-0.01	-0.02

[1] These figures represent the expected impact if all vehicles would be equipped, regardless of the year.

[2] Penetration levels were provided by eIMPACT.

3.3.1.2 Impacts at 100% penetration rate by mechanism

It was assessed that the safety impacts of the left turn assistance are limited to mechanisms 1 (direct modification of the driving task) and 3 (indirect modification of user behaviour) (Figure 2 and Figure 3). Other effects were assumed to be relatively small (although it may be a conservative estimate). The effect of mechanism 1 was much more substantial than that of mechanism 3.

The assessment of mechanism 1 was based on German data showing that 18.7% of the intersection injury accidents occur when left-turning driver collides with the oncoming vehicle [2]. The same reference indicated that the corresponding proportion for severe injuries and fatalities is 16.6%. Based on this percentage, it was assumed that the proportion for fatalities is 16%.

Furthermore, PReVENT [3] indicated contributing factors of target accidents. Specifically, the left-turn driver misinterpreted the situation in 38% of accidents and the left-turn driver was inattentive in 5% of accidents. Those accidents (43% in total) are the target accidents of the left turn assistance [2]. However, given that the system provides only a warning (no intervention), it was assumed that in 60±20% of cases the accident could be avoided (the most probable impact ± high/low estimate). Consequently, these figures resulted in a reduction of 4.8% (0.187 * 0.43 * 0.60 = 0.048) for injuries and a reduction of 4.1% (0.16 * 0.43 * 0.60 = 0.041) for fatalities. No direct effect was assumed for accidents that occur in link sections.

In case of mechanism 3 (indirect modification of user behaviour), it was assumed that the driver learns to rely on the system which results in delegation of responsibility to the system. Therefore, he or she might become more careless, for example. The magnitude of these long-term effects is difficult to assess. In line with the safety assessment of right-of-way assistance of INTERSAFE [32] the magnitude of this negative effect was assumed to be +1.5%. This effect was also limited to intersections.

Consequently, the total impact was a 2.7% reduction for fatalities and a 3.4% reduction for injuries that result from accidents that occur at intersections (see calculation principle presented in Annex E, Section 5).

Given that the proportion of fatalities resulting from intersection accidents in the database was 20.6% in EU25 and the corresponding percentage for injuries was 44.7%, the overall safety impacts were -0.6% and -1.5%, respectively.

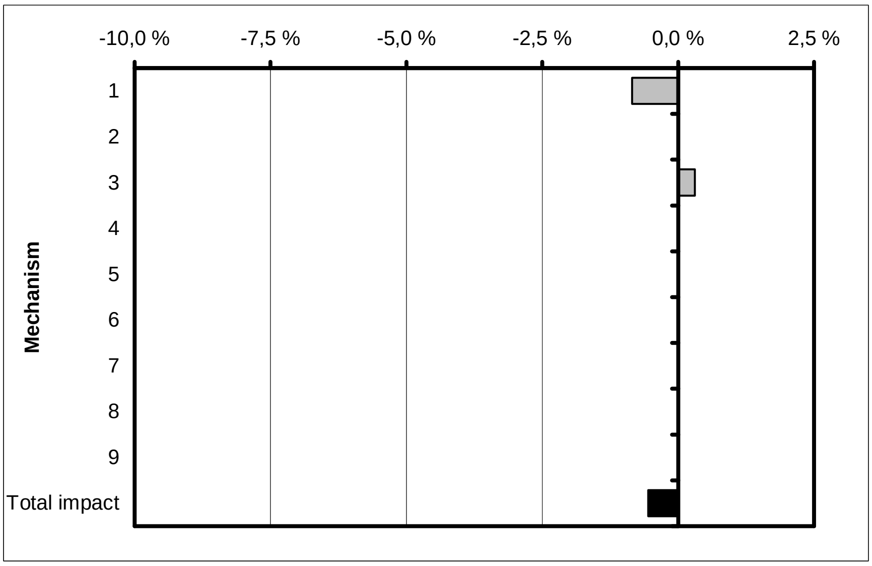

Figure 2. Safety impacts of the left turn assistance of INTERSAFE by mechanism for fatalities.

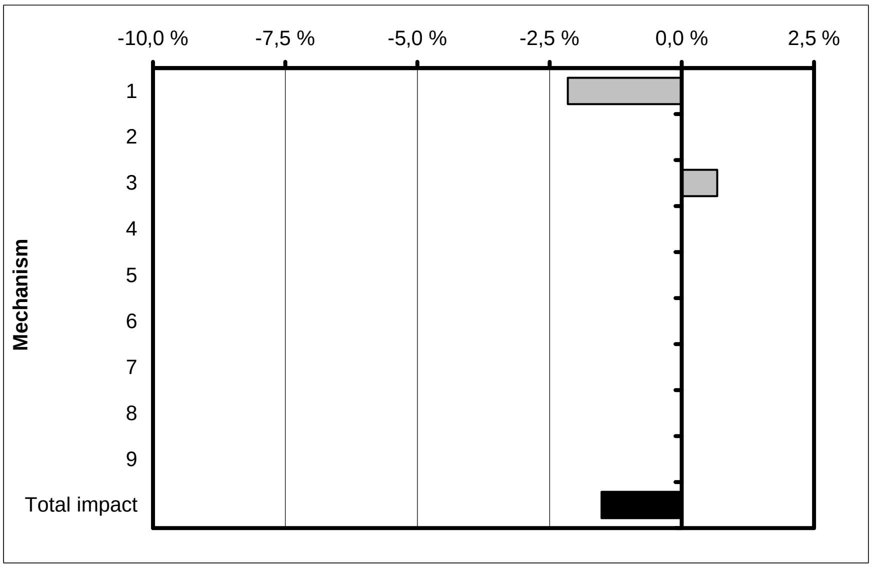

Figure 3. Safety impacts of the left turn assistance of INTERSAFE by mechanism for injuries.

3.3.1.3 Additional remarks

While the main safety analysis was based on data given by link/intersection, the reasoning of mechanism 1 (direct modification of the driving task) for other variables was as follows (note that these assessments did not affect the overall safety effects shown in Table 4):

- The target accident types of the function included frontal and angle accidents. The function was assumed to have a maximum effect on these accidents. In addition, it was assumed that the function reduces collisions on the road with all other obstacles, other accidents with two vehicles and side-by-side collisions as the outcome of the error of the left-turning driver and the classification of the accidents may vary.

- The magnitude of the effect was assumed to be similar on rural and urban roads and no effect was assumed for motorways.
- The magnitude of the effect was assumed to be similar on passenger cars and heavy vehicles.
- In comparison with adverse weather conditions, it was assumed that the effect is somewhat smaller in normal weather conditions. This assessment was based on the following assumption: in adverse weather conditions drivers might have more problems to estimate velocity and distance of oncoming vehicles which is one of the typical contributing factors of intersection accidents.
- In comparison with night driving, it was assumed that the effect is smaller in daylight because drivers might have more problems to estimate velocity and distance of oncoming vehicle at night.

3.3.2 MAPS&ADAS

3.3.2.1 Overall effects

The safety assessment included two functions of MAPS&ADAS: Hot Spot Warning (HSW) and Speed limit Warning (SLW). Both functions intend to reduce the driving speed to increase road safety. HSW addresses possibly dangerous locations which have been accident prone in the past. SLW addresses speeding in general. Table 5 presents overall effects of the left turn assistance of MAPS&ADAS on proportion of fatalities and injuries by target year and penetration level.

Table 5. Overall safety effects of MAPS&ADAS by target year and penetration level.

	Penetration rate for light/heavy vehicles (%)[2]	Reduction (%) in	
		Fatalities	Injuries
Impact low[1]	100/100	-11.5	-7.3
Impact most probable[1]	100/100	-13.1	-8.2
Impact high[1]	100/100	-14.7	-9.2
2010 low	2/4	-0.4	-0.3
2010 high	3/7	-0.6	-0.4
2020 low	30/42	-5.1	-3.6
2020 high	46/61	-7.4	-5.2

[1] These figures represent the expected impact if all vehicles would be equipped, regardless of the year.

[2] Penetration levels were assumed to be the same as for SpeedAlert in eIMPACT.

3.3.2.2 Impacts at 100% penetration rate by mechanism

It was assessed that the safety impacts of MAPS&ADAS include mechanisms 1, 3, 5 and 8 (Figure 4 and Figure 5). The effect of mechanism 1 (direct modification of the driving task) was dominating in comparison with that of others (indirect modification of user behaviour, modification of interaction between users and non-users and modification of route choice). Other effects were assumed to be negligible.

Within MAPS&ADAS user tests were carried out to estimate the impact of the system on user behaviour. Specifically, 64 test drivers drove a specified route on rural roads with and without the system. It was found that the mean speed reduced by 4.65% [see D16.2]. However, Hjälmdahl & Dukic [31] showed that this type of speed effects are short with no motivation and bonus schemes. Therefore, these estimates were reduced by 30±10% that is a cautious assumption in comparison with the results of Hjälmdahl & Dukic [31]. These estimates were used for the assessment of mechanism 1 (direct modification of the driving task) on rural roads. The safety impacts were computed by the power model [17] that resulted in most probable reduction of fatalities by 13.8% (high 15.7% and low 11.9%) and in most probable reduction of injuries by 8.5% (high 9.7% and low 7.3%).

Furthermore, it was assumed that the most substantial effects (above percentages) would be obtained on rural roads (with lots of hot spots) by improved awareness of hot spots and posted speed limits and especially by improved awareness of speed exceeding, followed by urban roads (20% lower) and motorways (40% lower). In urban areas, the frequency of hot spots might be lower but, on the other hand, importance of speed is high because of frequent encounter of vulnerable road users. On motorways, the hot spot density is relatively low and speed effect lower as well. Consequently, the most probable reduction of fatalities was 11.0% on urban road and 8.3% on motorways, for example.

In case of mechanism 3 (indirect modification of user behaviour), it was assumed that the drivers get used to drive more carefully because of increased situation awareness and as they prefer to avoid situations when warnings are issued by the system. On the other hand, it is possible that drivers get used to higher speed when possible (as close as possible to the posted speed limit). In summary, it was assumed that the total effect is positive but the magnitude of this effect is 10% of that with mechanism 1. Specifically, the effects on fatalities were -1.4% on rural roads, followed by -1.1% on urban roads and -0.8% on motorways, for example.

For mechanism 5 (modification of interaction between users and non-users), it was assumed that some of the following drivers would reduce the speed when encountering a vehicle that obeys the posted speed limit. The most substantial effect was assumed for urban roads with limited possibilities to overtake (-1.5%), followed by urban roads (-1.0%) and motorways with relatively limited need to overtake (-0.5%).

For mechanism 8 (modification of route choice), it was assumed that in some cases there might be changes in route choice –

drivers might want to avoid routes with low speed limit, if possible. Consequently, it was assessed that this positive effect would be most substantial for rural roads (-0.5%), followed by urban roads (-0.25%) with limited chances to effective rerouting in terms of road type change and no effect for motorways.

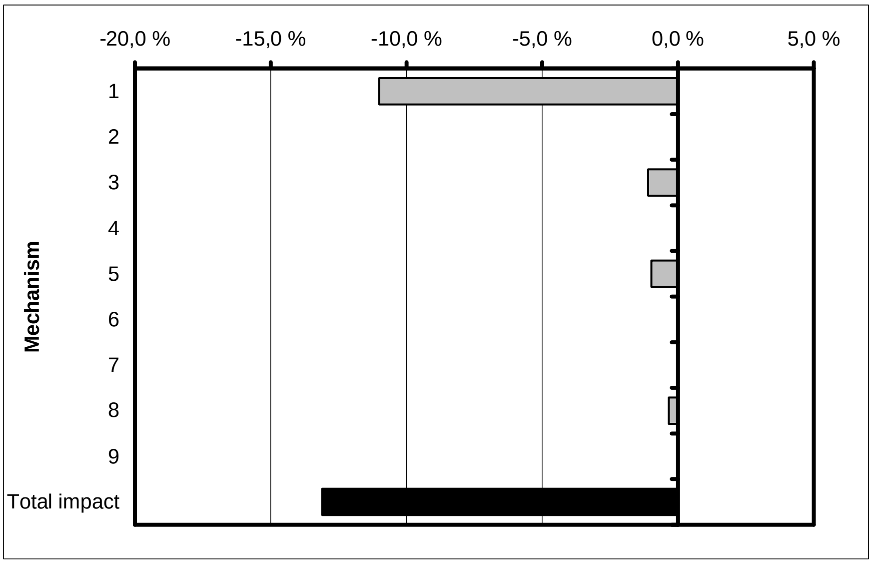

Figure 4. Safety impacts of MAPS&ADAS by mechanism for fatalities.

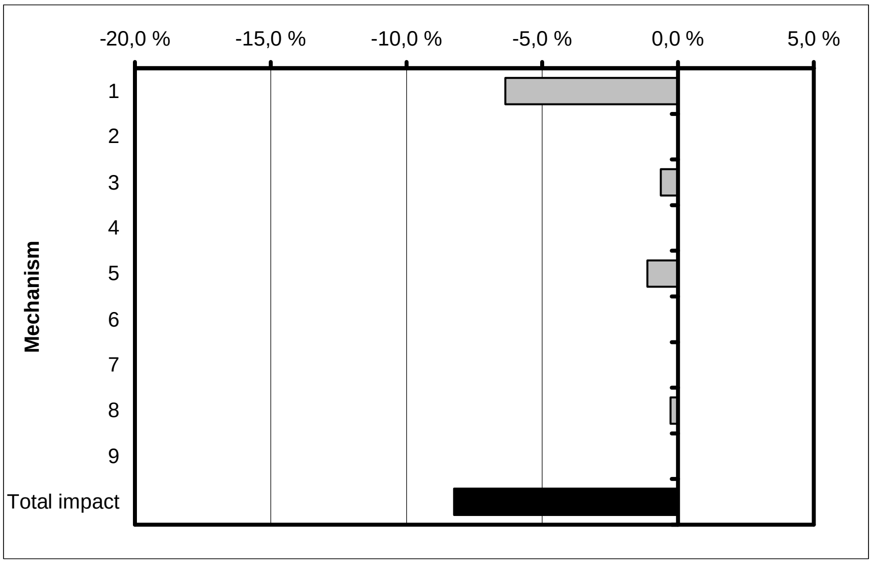

Figure 5. Safety impacts of MAPS&ADAS by mechanism for injuries.

3.3.2.3 Additional remarks

For other variables apart from road type, the main reasoning of mechanism 1 (direct in-car modification of the driving task) was as follows:

- The maximum effect was assumed for collisions on the road with pedestrians (especially SLW), collisions besides the road with pedestrians or obstacles or other single vehicle accidents and frontal collisions (especially SLW), followed by collisions on the road with all other obstacles and angle collisions (lower because of no HSW effect), side-by-side collisions and rear collisions (lower because of no HSW effect and potentially relatively lower speed effect) and other accidents with two vehicles.

- The effect of HSW was expected to be most substantial in link sections as the system is designed to warn drivers of accident prone locations such as sharp curves. SLW was expected to have the same effect on all roads.
- The impact was expected to be most substantial at night when speeding is more frequent in free-flow conditions and visibility is lower.

3.3.3 SAFELANE

3.3.3.1 Overall effects

The SAFELANE system provides lane keeping support in critical lane departure situations by acoustic, visual and haptic feedback to the driver. Table 6 presents overall effects of SAFELANE on proportion of fatalities and injuries by target year and penetration level.

Table 6. Overall safety effects of the SAFELANE lane keeping support system by target year and penetration level.

	Penetration rate for light/heavy vehicles (%)	Reduction (%) in	
		Fatalities	Injuries
Impact low[1]	100/100	-9.0	-7.3
Impact most probable[1]	100/100	-13.5	-9.5
Impact high[1]	100/100	-18.0	-12.4
2010 low	1.1/0.3	-0.2	-0.2
2010 high	2.9/1.4	-0.6	-0.4
2020 low	6/6	-1.3	-0.9
2020 high	21/23	-4.1	-3.0

[1] These figures represent the expected impact if all vehicles were equipped, regardless of the year.

3.3.3.2 Impacts at 100% penetration rate by mechanism

The results showed that the safety impacts of SAFELANE were limited to mechanisms 1,3 and 6 (Figure 6 and Figure 7). Other effects were assumed to be relatively small (although this may be a conservative estimate). Again, mechanism 1 was dominating.

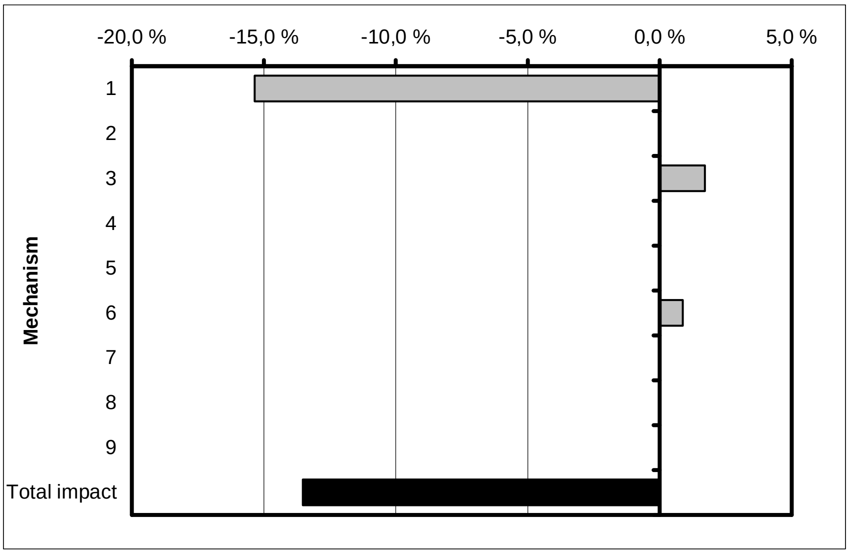

Figure 6. Safety impacts of the SAFELANE lane keeping support system by mechanism for fatalities

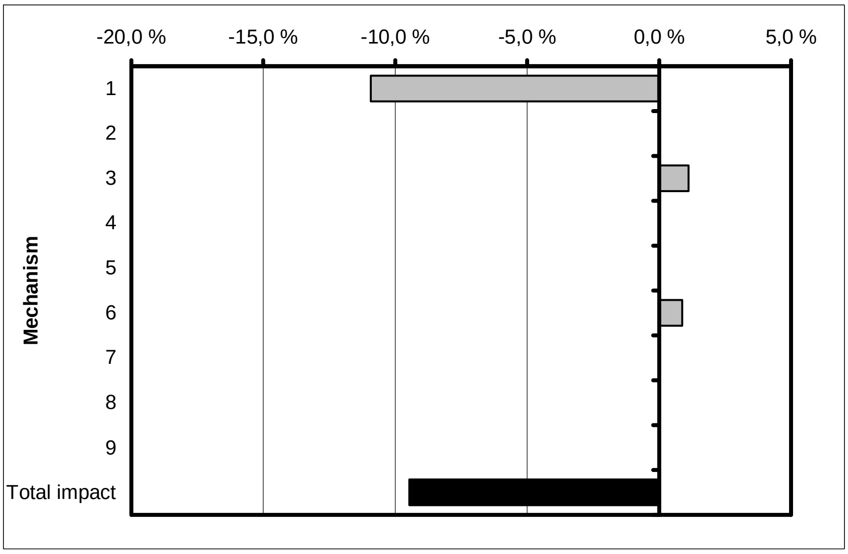

Figure 7. Safety impacts of the SAFELANE lane keeping support system by mechanism for injuries.

Although a number of studies that have assessed potential effects of lane keeping support systems are available (see D16.2 [2]), the assessment of mechanism 1 was primarily based on results of Enke [29] showing that the system could result in the following accident and severity reductions: (a) a 25% reduction of “left roadway” accidents and an additional 15% accident severity reduction, (b) a 25% reduction of frontal accidents and an additional 25% accident severity reduction and (c) a 60% reduction in the number of side-by-side accidents and an additional 10% reduction in accident severity. However, because these estimates show the maximum effects, it was assumed to be reasonable to make some reductions for the most probable effects. Specifically, the fatality and injury reduction estimates for mechanism 1 by accident type were 30±10% for “collisions besides the road with pedestrian or obstacle or other single vehicle accidents”, 40±10% for frontal collisions and 60±10% for side-by-side accidents.

For mechanism 3 (indirect modification of user behaviour), it was assumed that drivers delegate responsibility to the system to some degree. The magnitude of the effect was assumed to be 4% for each target accident types (collisions besides the road with pedestrian or obstacle or other single vehicle accidents, frontal collisions and side-by-side collisions).

For mechanism 6 (modification of road user exposure), it was assumed that drivers somewhat increase driving because of the increased driving comfort. However, it was assumed that this negative effect on safety is minor, i.e. 1% on all accident types.

3.3.3.3 Additional remarks

For other variables than accident type, the main reasoning of mechanism 1 (direct in-car modification of the driving task) was as follows:

- The maximum effect was assumed for link sections, while somewhat lower at intersections where the drivers might be more alerted.
- The maximum effect was assumed for rural roads, while somewhat lower for motorways (no possibility of frontal collisions) and much lower for urban roads (much lower speed).
- The maximum effect was assumed for normal weather conditions and somewhat lower for adverse weather conditions because the adverse weather conditions include conditions, such as slippery roads, in which the system is not able to support the driver.
- The maximum effect was assumed for night driving and less for daylight, since road markers are less visible at night.

3.3.4 SASPENCE

3.3.4.1 Overall effects

Expert evaluations of the safety impacts of SASPENCE were based on the accident analysis and user tests. For situation “distance keeping” the indirect evidence was based on studies on collision warning and ACC, keeping in mind that SASPENCE is purely advisory. For situation “speed advice” the studies on advisory ISA and SpeedAlert were used as reference. It was assumed that the speed limit advice operates for fixed speed limits only in 2010 and for fixed, variable and dynamic speed limits in 2020.

Situations “distance keeping” and “speed advice” were assessed separately. However, Table 7 presents overall effects of SASPENCE on proportion of fatalities and injuries by target year and penetration level.

Table 7 Overall safety effects of SASPENCE by target year and penetration level.

	Penetration rate for light/heavy vehicles (%)[2]	Reduction (%) in	
		Fatalities	Injuries
Impact low[1]	100/100	-4.1	-2.5
Impact most probable[1]	100/100	-6.5	-3.8
Impact high[1]	100/100	-9.2	-5.1
2010 low	0.01/0.01	0.0	0.0
2010 high	0.01/0.01	0.0	0.0
2020 low	4/15	-0.3	-0.3
2020 high	13/25	-1.2	-0.8

[1] These figures represent the expected impact if all vehicles were equipped, regardless of the year. All impacts are given for system including speed advice for fixed, variable and dynamic speed limits, except 2010 low and high that are for speed advice for fixed limits.

[2] Penetration levels are adopted from ACC FSR assessed by eIMPACT.

3.3.4.2 Impacts at 100% penetration rate by mechanism

The results for situation "distance keeping" showed that the safety impacts of SASPENCE were limited to mechanisms 1,3 and 8 (Figure 8 and Figure 9). Other effects were assumed to be negligible.

The assessment of mechanism 1 (direct modification of the driving task) was based on results of Ammerlaan et al. [4] showing that head-tail collisions will be reduced by 20-30%. However, the most probable effect was reduced by 50% as SASPENCE is not full speed range, but only 30 km/h and up. Specifically, no impact was assumed for urban areas where approximately 50% of rear-end accidents occur (Statistics Finland 1995-2002). Moreover, the current system is not installed in trucks. However, because SASPENCE is possible to use in heavy vehicles as well, the safety assessment included both light and heavy vehicles. In summary, the estimate for the injury reduction was 12.5±5% ((20%+30%)/2*0.5).

The estimate for fatal accidents was based on the above assessment and ACC assessment conducted in eIMPACT [32]. For SASPENCE, it was assumed that the effect on fatalities would be proportional to those on injuries which led to the effect of 18.8±5%.

For mechanism 3 (indirect modification of user behaviour), there was no apparent estimate although it was assessed that the situation awareness might reduce, the system might divert driver attention, the system might result in shorter headways on average. On the other hand, it was assumed that drivers might learn to avoid warnings (i.e. short headways) and the speed in platoons might be more homogeneous. Given that there was no explicit reason to assume any difference between the impact on fatalities and injuries, it was assumed that the impact would be +0.5% on fatalities and injuries.

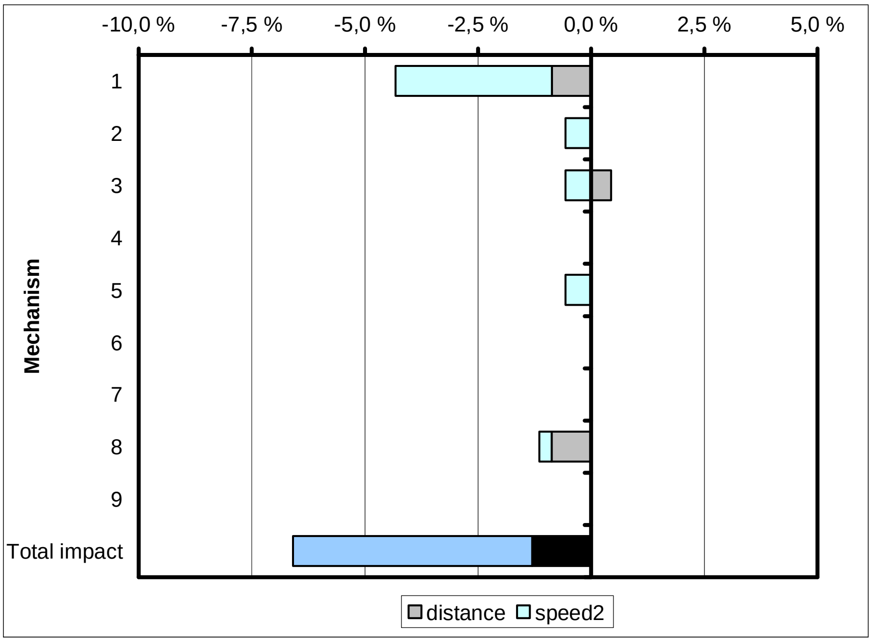

Figure 8. Safety impacts of SASPENCE by mechanism for fatalities.

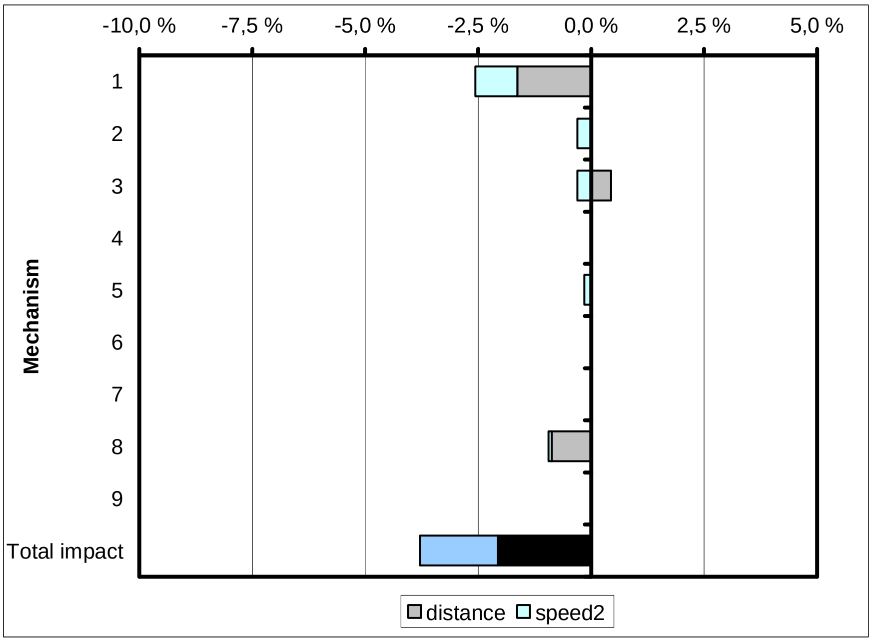

Figure 9. Safety impacts of SASPENCE by mechanism for injuries.

The assessment of mechanism 8 (modification of route choice) assumed that there could be some positive effects on safety in the long run. Specifically, ACC is likely to make motorways more appealing and thereby motorways are used more frequently. However, the effect was assumed to be rather small, -1.0% on fatalities and injuries.

In summary, these estimates resulted in overall reduction of 1.5% in fatalities and 2.4% in injuries.

The results for situation "speed advice" were obtained from the safety assessment of SpeedAlert 1 and SpeedAlert 2 conducted in eIMPACT [32].

3.3.5 APALACI/COMPOSE (Collision Mitigation Systems)

In this chapter a method and application for the evaluation of CMS safety benefits is presented[1]. The analysis focuses on an automatic CMS applying a 1g braking deceleration at TTC=0.5s, during 0.5s. A global model of the CMS is assembled, from technical specification to market penetration, with appropriate hypotheses. This model enables a discussion of safety benefits and a sensitivity analysis with respect to the main parameters, as well as of error causes and margins.

Based upon "use-cases", data coming from traffic and accidentology (France, 2005), the proposed method evaluates the probability of individual severity shift associated with the above-defined CMS, resulting from the collision speed reduction through automatic braking. CMS is evaluated with respect to three incremental options [2]:

a. CMS-level-1 is capable to deal with frontal collisions based on object detection in an area defined by a wide angle Ar at a short perception distance Sr (e.g. Ar=60-90°, Sr=20m)
b. CMS-level-2 includes the CMS-level-1 perception capabilities added to the capability of object detection in a narrow angle (Ar <10°) at a long perception distance Xr (e.g. Xr =100m)
c. CMS-Level-3 adds to CMS-level-2 the capabilities of medium distance perception (e.g. X= 40m) in a wide angle (Ar=130-180°).

The method starts with the identification of these use cases and the corresponding targeted accidents for each use case.

The use cases take into account the network (motorway, rural, urban), the type of collision (front-to-rear, front-to-side and front-to-front) and the collided road users (vulnerable or not road users).

On the basis of the kinetic energy reduction (see listed hypothesis below) by use of CMS, the efficiency of the CMS is computed in percentages for each use case. These percentages are applied to the targeted accidents (French data, 2005), for each use case in order to obtain the accidents reductions.

The precise method is described in the following. By using Nilsson's formula for the effect of speed on safety [17], the number of targeted accidents is distributed along speed classes for each use case. A central point is the use of experimental curves [33],[34] that give the probability of individual injury severity (three curves are given: lightly, severely or fatally attained) as a function of the speed - more precisely the Energy Equivalent Speed (EES). The use of CMS will shift these curves, reducing this probability. With the distribution of the casualties over the speeds, we know, for a given speed, the number of accidents targeted and we know as well the reduction shift due to introduction of CMS. This

[1] With many thanks to other contributors, Sébastien Glaser (LCPC), Yves Page (LAB), Risto Kulmala (VTT) in particular.

[2] These incremental options consider CMS systems with incrementally enlarged field-of-view. The calculations in terms of casualties reduction, by considering systems with these capabilities, will correspond then to a potential in reducing accidents.

combination gives the reduction of accidents for each severity class, for each use case.

Two types of models are compared: a discrete and a continuous one. The use of two models allows checking of the methods based on results comparison. Nevertheless, both models use hypotheses and mechanisms that relate to the assistance function and the driver behaviour. They are listed below:

H1. Segmentation through light injuries (LI), severe injuries (SI), and fatalities (F) are considered.

H2. Only accidents that relate to shocks are considered: they include vehicle to vehicle collisions, vehicle to objects and vehicle to vulnerable users.

H3. Target object categories, includes light vehicles, heavy vehicles (such as commercial, buses, other), 2-wheel vehicles, pedestrians.

H4. Accidents occur at certain speed called Accident Speed The accident speeds are estimated as the observed travel distributed speed modified by a risk function based on Nilsson empirical formulae.

H5. Travel speed distributions depend on network characteristics. Three types of networks are considered: motorway, urban, rural

H6. Considered accidents end in a crash that is characterised by a shock severity based on EES.

H7. We assume that for some use cases (e.g. light vehicles) a significant part of kinetic energy does not contribute to the crash severity. This part of the kinetic energy is “vented” through various schemes, essentially friction.

H8. The relation between the crash violence and the probability to be injured can be investigated through empirical curves plotting the individual probability to be injured versus EES.

H9. Three categories of driver behaviour are introduced: the drivers that brake with the right force; a%; the drivers that brake with insufficient forces; b%; the drivers that do not brake at all: 1-a-b.

The work includes as well a thorough sensitivity analysis with respect to a variation in the following parameters: CMS activation TTC (reference value: 0.5sec, sensitivity tests done for TTC=0.3sec and TTC= 0.7sec), CMS deceleration capability (reference value = 1g, sensitivity tests done for 0.8g and 1.2g), proportion of drivers that brake with right force, with insufficient force or do not brake at all (reference value: 1/3,1/3, 1/3, sensitivity tests done for 7/10, 1/10, 2/10 and for 2/10, 1/10, 7/10), modelling errors in empirical curves plotting the individual probability to be injured versus EES (shifts of 5km/h and -15km/h in the original curves) and EES ponderation (reference value: sqrt(2/3), sensitivity tests done for 1 and 0.5). This analysis shows a significant sensitivity with respect to variation in the CMS activation TTC and on drivers braking force for the tested values. A lower sensitivity is observed when varying the CMS braking capacity and the empirical severity injury probability curves within the above cited values. Table 8 gives the results of the sensitivity analysis.

Table 8: CMS efficiency in %: variation interval corresponding to sensitivity analysis.

	CMS activation TTC [-0.7s, -0.3s]	CMS deceleration [0.6g, 1.2g]	Braking reaction (.7,.1,.2), (.2,.1,.7)	Shifting in empirical curves [+5,-15km/h]	EES ponderation [1, 0.5]
Urban	[-54, -32]	[-32, -50]	[-22,-63]	[-43,-55]	[-45,-59]
Rural	[-43, -24]	[-24, -39]	[-17,-50]	[-33,-39]	[-27,-36]
motorway	[-23, -8]	[-8,-19]	[-8,-24]	[-12,-25]	[-5,-28]

Table 9 gives the best and the worst CMS theoretical efficiencies in relation with the use cases, for each network type.

Table 9: best and worst CMS theoretical efficiencies

	CMS worst case	CMS best case
	dec=-0.8g, TTC=0.3s, shift+5km/h	dec=-1g, TTC=0.7s, shift-5km/h
Urban	-25.0%	-56.7%
Rural	-18.3%	-45.0%
motorway	-5.1%	-26.9%

Application (France, 2005) of this method given the vehicle park and supposing a CMS 100% market penetration and maximum technical efficiency (100% detection rate), the result obtained with the more advanced technical option (option3) would be a global macro-economic saving of some 2.200 Mio € (2.228) per year (fatalities, severe and light injuries altogether), or -18.6% of total road insecurity costs (-19.6% fatalities, -14.3% injuries) in the hypothesis of all vehicles are equipped.

Based on two horizon deployment hypotheses starting from 0% 2012 (low: 10% in 2020, high: 25%) and linear rates, the safety benefits become -1.4% (low hyp.), -1.9% (high hyp.).

Table 10. Overall safety effects of CMS by target year and penetration level.

	Penetration rate for light/heavy vehicles (%)	Reduction (%) in	
		Fatalities	Injuries
Impact[1]	100/100	-19,6	-14,3
2010 low	0/0	0	0
2010 high	0/0	0	0
2020 low	10/10	-0,7	-0,5
2020 high	25/25	-1,9	-1,4

[1] These figures represent the expected impact if all vehicles would be equipped, regardless of the year.

4 Evaluation framework

The framework is attached to this deliverable as a separate Annex E. This chapter describes the framework and the work performed to finalise the framework, which was proposed in deliverable D16.3. The following work was performed to finalise the framework:

1. application of the HMI expert analysis tool to MAPS&ADAS and assistance to INSAFES;
2. the framework was sent to the INSAFES project, which applied it to their validation plan;
3. D16.3 was sent to selected evaluation experts for feedback;
4. discussions at the final PReVAL workshop.

This chapter reports the work and the results in a concise form.

4.1 Overview of the methodology

The safety potential of a preventive safety function is determined by several factors. These include the technical reliability and performance, the ability to improve driver response and the impact these factors taken together, have on the traffic safety level (safe operation of the traffic system, interaction between users and non-users) [5]. Assessment is organized according to these three aspects: technical and human factors evaluation, followed by safety potential assessment.

The *technical evaluation* focuses on the technical performance and reliability of the system. Technical evaluation is performed in two phases: "Verification" to test the individual components and subsystems towards the technical specifications and "Validation" to test whether the goals and specifications of the complete system are met. The main goal of the *human factors evaluation* is to assess the extent to which the system succeeds in generating the intended behavioural responses from the driver in target situations, i.e. once the risk for loss of control is detected, hence to assess the ability of the function to affect situational control *through the driver* by providing information and/or warnings. The goal of the *safety potential assessment* is to make an aggregate-level assessment of the preventive system's effects on relevant harm metrics (usually number of fatalities) in target situations. The impact assessment is based on the assessments of technical performance and behavioural effects making use of accident statistics, estimations of fleet penetration rates, and other relevant tools. For safety assessment, PReVAL uses the procedure developed and used by the eIMPACT project.

The first purpose of assessment is to evaluate whether the system works as required, i.e. if it achieves the desired improvement of situational control. Therefore, the entire design cycle (including system specifications) is considered rather than merely the evaluation process. The "V" design cycle, which is commonly used in the automotive industry, is extended by including the different steps of the evaluation process (Figure 10). The new workflow is based on CONVERGE [9], [10], the evaluation methodology used in the PReVENT subprojects, and the experiences of APROSYS and AIDE. The different evaluations go through similar steps:

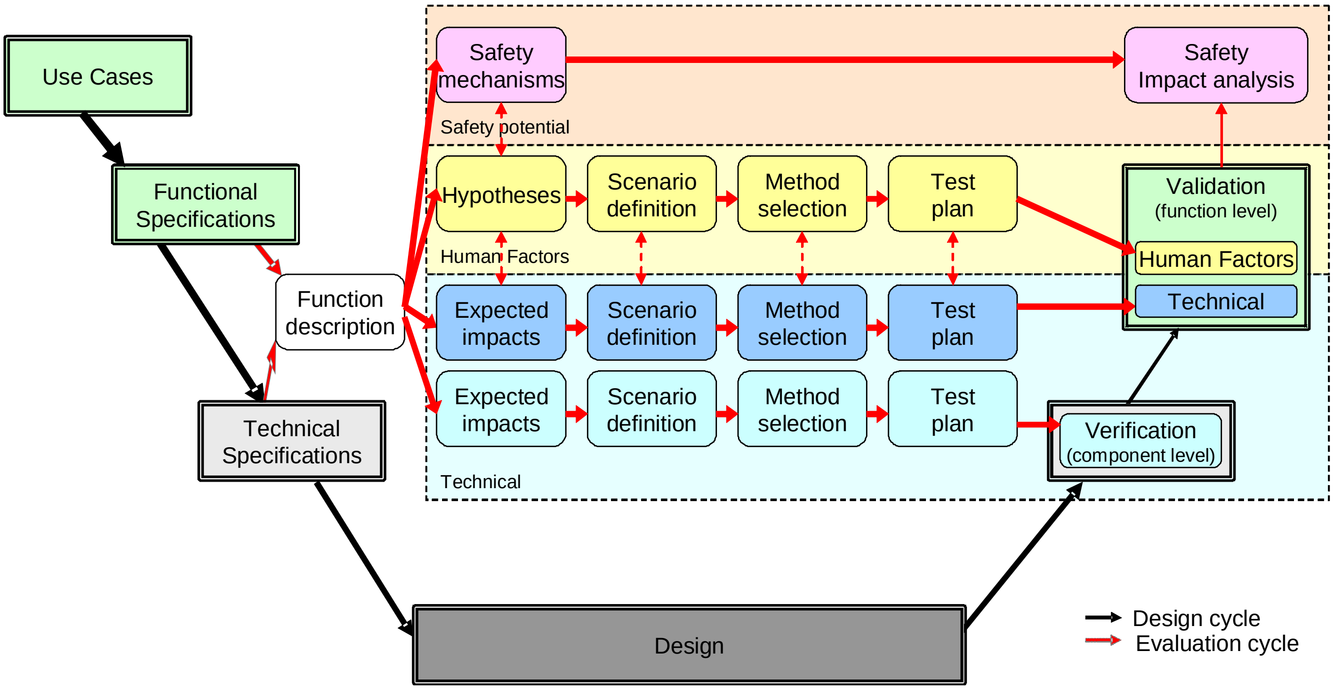

Figure 10: Adapted V-shape design and evaluation cycle, showing the relation between technical, human factors and safety potential evaluation and the different steps in the evaluation processes

0) ***System and functions description***: a function description is normally the first document produced before the functional specifications, but may not be available to the evaluators and not include all needed information or updates made during development. At the start of the validation, a sufficiently detailed function description needs to be available, which is common for all assessments and done in a consistent way to assure that all information needed for developing the evaluation plan is available and that similar systems can be compared.

1) ***Expected impacts.*** For technical evaluation, this step involves describing the technical objectives of the system in such a way that it is possible to evaluate the performance of the system. For human factors evaluation, this step involves generating hypotheses on how the system can be expected to change the driving behaviour in the target situations. This step includes definition of indicators for measuring relevant aspects of system performance in the target situations.

2) ***Test Scenario definition***. In order to verify the expected impacts and hypotheses, test scenarios are defined for the different evaluations. The scenarios are specified through a description of the maneuvers, operational conditions for the tests and the parameters of the target objects for detection.

3) ***Evaluation method selection.*** The selection of the evaluation method depends on desired level of result quality as well as availability of resources. The range of methods available include inspection methods (e.g. expert panels), inquiry methods (HMI concept simulators,

simulator studies, Computer Aided Engineering methods including hardware-in-the-loop simulations), and trial methods (professional or test drivers on test track, roads or in driving simulator).

4) ***Measurement plan.*** The test plan specifies the number of tests and the definition of independent and dependent variables. The goal should be to get statistically significant answers for all hypotheses under evaluation.

5) ***Execution and reporting.*** The verification and validation tests are executed, data are analyzed and conclusions are drawn.

4.2 Application of the methodology

4.2.1 Expert human factors evaluation of MAPS&ADAS

Introduction

An expert evaluation is a method often used for finding usability or function specific attributes that can be improved as part of an iterative design process. It involves having a small set of evaluators (combining human factors and technical expertise), examine the interface and judge its compliance with recognized usability principles. As a follow-up, the experts can go on to suggest improvements, or if needed, caution the designer towards unintended effects of the system/interface.

An expert evaluation is in many respects a complementary method to a user-based evaluation, with it's over arching objective being to highlight any issues of usability that the designer has overlooked.

An expert evaluation was performed on the MAPS&ADAS functions, and aimed to produce complementary human factor related results to the results already achieved within the subproject, by focusing on some specific aspects, not brought up during the evaluation done within the subproject.

The expert evaluation performed on MAPS&ADAS aimed at addressing both the general HMI design of the functions, assessed in a stationary vehicle, as well as the experience of the functions such as experiences of driving performance and driver behaviour while driving.

Tools such as checklists and questionnaires for performing expert evaluation of ADAS are currently not well established; expert evaluation methodology addressing in-vehicle information systems is in a more advanced stage.

During the past years there have been efforts in several projects for deriving different types of checklist tools as help for designing and evaluating driver assistance systems, of which a few examples are the projects RESPONSE, ADVISORS and AIDE. In ADVISORS, a checklist for checking the HMI of driving assistance systems (*ADAS Quick Check)* was developed [37]. This checklist aimed at addressing general items on a quite high level at different stages of the development phase. The RESPONSE project derived a checklist for assessment of driver assistance system for assessing ADAS, intended to serve as a support during the development phase of ADAS [27]. In AIDE an extensive survey of existing methodologies for assessment of ADAS is available.

In the expert evaluation made within PReVAL, checklists and questionnaires were developed, on a detailed level, addressing ADAS in general and MAPS&ADAS in particular. The work was influenced by available material addressing assessment of both IVIS and ADAS, but with a slightly new scope. Once the scope and material for the evaluation was defined, the methodology was applied to MAPS&ADAS and both the methodology as well as the results from applying it was evaluated.

The expert evaluation of the MAPS&ADAS functions took place in Hannover with a group of in total 7 persons, with experience in different areas related to preventive safety functions.

The procedure and the results are discussed in more detail in Annex D.

Expert Analysis

The MAPS&ADAS subproject provided detailed information about hypotheses and scenarios, the scenarios applicable for subjective evaluation were used also for the expert evaluation done in PReVAL, with some modification; a few additional hypotheses were added that was considered of interest.

The opinions of the experts were gathered by using questionnaires, which were based upon the hypotheses, to largest possible extent. It was decided that one part of the evaluation should be performed in a stationary vehicle focusing on HMI design in simulation mode. A second part of the evaluation should be performed while driving with the system, focusing on the system-driver interaction and potential effects of the functions on driver performance.

The evaluation offered some deep and insightful feedback that could indeed acts as an added value to the extensive user-based evaluation already carried out by MAPS&ADAS. The final diversity of expertise brought forward by the panel was rich, and the design of the evaluation (i.e. orientation, driving on an open track, questionnaires) was sound, as well as representative of real driving conditions.

For each question/scale, a judgment can be given on the level of agreement among the experts, where there are roughly three levels:

- Total, or almost total, agreement
- Some, not extreme, agreement/disagreement
- Total, or almost total, disagreement

The questionnaires and the results are shown in more detail in Annex D.

Results

In this section a brief overview on the results from the expert evaluation is provided. A more detailed analysis of the results are provided in Annex D.4.

In summary, the expert evaluation carried out for MAPS&ADAS, posed some initial challenges, i.e. the timing of the evaluation with regard to the project phase as well as the number of human factors experts that was available. However, the evaluation gave some valuable feedback that could indeed act as an added value to the extensive user-based evaluation already carried out at time for the expert evaluation.

The speed limit warning function was generally judged as positive by the experts, but the hot spot warning function (as well as the combination of these two functions) was less appreciated. There were some concerns regarding the symbol of a “guardian angle” used for the hot spot warning function, for instance there were some worries for higher-order effects, like drivers being distracted or ignoring the warnings. These comments provide directions for improving future versions of the MAPS&ADAS system with respect to the HMI design.

Analyzing the distribution of answers from the experts, and the agreement of these answers, the experts agreed in no more than 1/3 of the questions/ratings.

This proves that we recognize and acknowledge that a number of highly relevant aspects simply do not lend themselves to being captured by expert judgment, and there are questions that should be eliminated when doing this kind of evaluation.

Conclusions

On the basis of these results we can conclude that there is room for improvement in the HMI of the MAPS & ADAS system, in particular, the hot spot warning function.

The experts agreed only on 1/3 of the items which makes us realize that expert analysis for evaluating preventive safety functions, not only focusing on the HMI design but also on more complex areas like driving performance and driver behaviour, are difficult to cover. However, this can be a valuable qualitative tool to use early in the design phase to collect *indications* on potential problematic human factors related issues, which needs improvement prior to larger user tests.

4.2.2 INSAFES

A draft version of the framework deliverable D16.3 was sent to the INSAFES project, who implemented it into their validation plan. The feedback has been taken into account in the update of the deliverable.

Experts in human factors from PReVAL also assisted to the HMI evaluation tests with the VTEC instrumented truck in Gothenburg. The main purpose was to test the human factors evaluation methodology, as well as to provide INSAFES with additional test persons. The experiences with these evaluations are directly reflected in section 4.4.2.

4.3 Discussion of the framework

4.3.1 Feedback from Experts

The deliverable has been sent to selected experts in evaluation, with the request for feedback. Feedback has been received from Volpe Center.

The feedback from Volpe was mainly concentrated on the technical evaluation. The feedback has been taken into account in the update of the framework.

4.3.2 Final workshop

The PReVAL project held a final workshop on 10.1.2008 in Brussels. There were about 40 participants to the workshop. The programme of the workshop and the list of participants is given in Annex C. The annex contains also an overview of the discussions on the evaluation results and the panel discussion.

The presentations of the workshop are available on the PReVENT website[3].

4.4 Discussion of the framework

This section discusses the feedback on the framework, resulting from the application of the methodology and the discussions with experts.

4.4.1 Technical evaluation

From the analysis of the technical evaluation results from the PReVENT subprojects, some main points that deserve special attention have been identified and are discussed below.

- A compromise has to be found between addressing the full complexity of the driving context and the limited resources allowed in the evaluation process (identification of use cases, test scenarios - are of prime importance in this process)
- Statistical aspects: blocking tests of the same nature, avoiding bias, considering disturbing identified factors…
- 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.
- Tools and methods used. The automotive sector is presently facing the same difficulty as the aeronautic sector 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 Hardware-in-the-loop test benches or pure digital simulations for example.

---

[3] with the exception of the presentation on the safety assessment results, which were not yet final at the moment of the workshop.

- Representativeness of the tests, that 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 concerns the definition of the dummy targets against which the perception systems are evaluated. Due to the absence of standards, the diversity of target object parameters (shape, colours…) is very high in PReVENT subprojects. 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. We discuss this issue further on the chapter 4.

4.4.2 Human factors evaluation

In this section a general discussion on methodology is provided, in which some items have been brought up in the discussion held on best practices in evaluations made within the PReVAL final workshop (Annex C)

PReVAL focuses on the functional verification and validation, thus the validation of functional performance. A fundamental difficulty in testing of preventive safety functions is how to define the functional performance, as well as the level of acceptance for the results. There is today no obvious and established definition of functional performance of an ADAS or preventive safety function.

For a preventive safety system the aim is accordingly to increase traffic safety, but it is not trivial in what way traffic safety can be interpreted directly from a set of test results. The *situational control* concept, which links the system effects to safety, and also links the different dimensions in evaluation, is introduced in PReVAL. Situational control addresses functional performance as an overall concept, including both technical and human factors performance. 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 [27]. Situational control has another perspective, searching for the safety impact of the functions on traffic safety as a whole.

The situational control concept needs to be further developed by for example investigating how it applies to different situations and systems. A question for future research would be how to define and generalize the situational control indicators and in what way they differ between different type of system and traffic situations.

One important aspect in evaluation is also how to define a base line or a reference level in user tests to which the results from the tests carried out can be compared. What is the driving situation like, without the ADAS system in the vehicle? And what test environment is the most appropriate to choose for your test?

One way of creating a base line for short term testing is to test a scenario with and without the system. This puts high demands on repeatability and control of the test scenarios, in order to limit all variations that are not due to system effects. Thus an appropriate testing environment for these tests would be a driving simulator. A problem with simulator studies is that the functions are often “too good”, it is e.g. very hard to simulate e.g. a radar as it works generally in real environment. You will have high repeatability but, less realism in terms of system performance.

A problem in the physical environment is the difficulty of having repeatable tests, for comparing situations with and without the system active. In addition, there might be a safety risk for the driver to drive in critical situations. In a real environment it might also be necessary to give instructions to the driver to get similarities in the test scenarios, but instructions to the driver during the tests will influence the outcome of the test in one way or another. Giving too many instructions to the driver is not desired since it implies an unrealistic driving situation, and test subjects are mostly always aware that they were being tested which already influence their way of reaction. The most appropriate testing environment should be selected, depending on what kind of realism that is preferred and depending on what kind of system that should be tested.

Further information on base line conditions is of great importance, and might be achieved from naturalistic field operational tests.

Another critical aspect of testing is how to optimize among available resources. User tests are often very resource demanding and it is therefore of great importance that the data collected are good for the purpose and well reflect the hypotheses defined. Also data analysis is very time consuming and a well defined and limited scope of the evaluation is essential for saving valuable time when performing the data analysis and interpretation of the results.

When doing user tests with subjective data collection it is of importance that the drivers have time to get a feeling for the system. To be able to answer questions like “do you find the system helpful or assisting when driving”; you need the driver for some time to be able to judge that in a good way. Test routes of about one hour (for example) with the same system are preferred, or driving a scenario several times. Questionnaires should be carefully developed and adapted so that they are suitable for the tests performed in order to avoid the risk that questions are misunderstood or not relevant for the situation.

The aim of the evaluation is either to to verify that the system works as expected or to provide data for safety impact assessment. A distinction between evaluation for making a safety impact assessment and evaluation to test a system towards is specification is necessary at current stage. Ideally these two approaches would go hand in hand, if you have done large research on accidentology and real life safety needs when designing you system and writing its specifications.

The PReVAL human factors procedure will deliver data primarily to mechanism 1 for the safety impact assessment, which also can be used as input for accident consequence, mechanism 5.

Defining an evaluation framework that produces all the input needed for safety impact estimation needs to include a long-term evaluation for collecting data for the other mechanisms.

4.4.3 Safety impact assessment

The framework used in PReVAL aimed to assess the safety impact mechanisms exploring firstly how the functions affect driver behaviour and travel behaviour. Based on earlier research results concerning the relationships between driver behaviour and accident risk and/or consequence or desktop estimates based on expert judgments, these behavioural changes were projected into changes in fatality and injury frequencies. In summary, the mechanisms covered direct modification of the driving task, indirect modification of user behaviour and non-user behaviour, modification of interaction between users and non-users, modifications of road user exposure and modification of accident consequences.

In comparison with other available approaches, it is assumed that this type of behavioural effect approach is less likely to miss any important effects and especially behavioural adaptation effects. The same holds also for systems and functions affecting exposure. As the number of fatalities and injuries is a product of three factors, namely accident risk, risk of fatal injuries in a crash and exposure, it is important to consider each of these dimensions. One important aspect of this approach is that there is no a priori definition of the accident type that is affected by a given system or function. Specifically, the target accident types are determined based on intended effects of the system/function but the potential other effects are covered as well, including both positive and negative effects. These other effects often cover also other accident types, frequently even all accident types.

This approach was adopted from another active evaluation study, namely eIMPACT. Although the general framework has been developed in the 1990s, the approach was operationalised, widely applied and demonstrated in eIMPACT and PReVAL for the first time. This resulted, for example, in the development and testing of needed tools in these projects.

The application of the behavioural effect approach has been encouraging. This approach provides a comprehensive and valid assessment of safety effects of most systems and functions. However, it is noteworthy that the analyses require qualified and experienced experts from engineering, psychology and sociology disciplines; moreover, the performance of analyses is time-consuming. In addition, it is acknowledged that the safety assessments conducted in PReVAL are based on insufficient data in many respects. Finally there is a general limitation of ADAS safety assessments. The estimation of effects of combined functions is much more demanding and sometimes even impossible with the current knowledge. Nevertheless, the basis of the assessment is valid and it is strongly suggested that the safety assessments of any ADAS should be based on this type of approach. There is no need to reject this approach. In contrast, the application of the approach suggests that this type of analysis is doable with practical and valuable results. In the future, when more accurate data is likely to be available, the safety estimates can be further improved.

5 Recommendations for future assessment programs

Further development and implementation of the PReVENT or similar systems will require the knowledge of system performance and system impacts on driver behaviour and finally traffic safety in order to reduce uncertainties for the stakeholders involved and to obtain market-exploitable project results. Therefore the different stakeholders have to agree to a common assessment framework. The goal of PReVAL was to analyse the assessment methods used in the different PReVENT subprojects and develop a general guideline for the assessment of preventive safety functions.

This chapter provides recommendations for future assessment programs. This includes recommendations for the safety functions development and recommendations for future research on evaluation.

5.1 Recommendations for safety functions development

The start of the development process

The functional specifications should be based on use cases, which correspond to identified and relevant accident types.

Starting from the accident data, use cases which will have the most effect on traffic safety can be derived. A set of common high-level use cases can be defined as reference.

However, suitable technology may not yet be available to address these use cases. The connection to real world accident scenarios in terms of how they happen and most importantly *why* they happen is essential and should ideally be made clearer, when deriving use cases. However, this puts high demands on knowledge in accident statistics and causation.

Vehicle manufacturers have a long-term view on the development of intelligent vehicles. They will not start from scratch to develop systems which specifically address these use cases. New systems are built on existing systems. An example is the development from ACC over stop-and-go and pre-crash to collision avoidance systems. The technological development follows from systems which are on the market and the technology available.

When further knowledge in this field of accident causation and statistics has been achieved, the technology under development can be tuned to the improved use cases. The use cases can act as guidance for tuning the long-term development process.

The goal and customer of assessment

When defining an evaluation plan of for a preventive safety system it is of great importance to clearly state the goal with the evaluation and to define who the user of the methodology is. This influences the demands that are put on the evaluation and in what way the methodology is applied.

Assessment can serve two purposes: to assure the functionality of the system, i.e. that the system works as required, and to assess and quantify the system's impact on the traffic system. The goal with evaluation in the PReVENT subprojects has been to assure the functionality according to specifications, the goal of PReVAL has been to derive a framework for evaluation, and to address the potential safety impacts of the systems.

In order to be able to calculate the impact, the efficiency of the system for preventing accidents with respect to the different circumstances has to be determined. To do safety impact estimation puts hence high demands on the amount of data achieved from evaluation. The link between outcome of evaluation and safety impact is not trivial. To provide a broad set of data for safety impact assessment is not necessarily the same thing as providing test results that can assure the functionality of the system with respect to the uses-cases defined during the development phase. Ideally there should be a clear link between need for safety impact assessment and evaluation objectives, but this is not always possible to achieve, depending on the resources available and also the knowledge of traffic scenarios and their mapping to real life safety.

Input from technical and human factors evaluation for safety assessment

Accident data are classified according to the accident type and different background variables, such as vehicle type, collision type, road type, weather conditions, lighting and location type. The safety assessment calculates efficiency coefficients for each type of accident. The technical and human factors evaluation contribution to the safety assessment hence has to determine the "efficiency coefficient" for the respective accident type, mechanism type and background variable. The concept of situational control provides a mechanism for this purpose.

The evaluation process should be designed so, that it supports these calculations. The scenarios for testing the systems should hence correspond to identified accident types and sets of background variables. Since resources are limited, from the accident database relevant combinations of accident types and background variables can be determined for testing.

A basic set of scenarios, which correspond to the common high-level use cases and are based on relevant accident cases, can assist in designing the evaluation process; furthermore, it also would allow to compare different systems which each other. In order to be able to compare different results, a common set of indicators has to be agreed. This could also offer possibilities for use in homologation processes.

An important next step in future work would then be to develop test methods for functional verification and validation of systems where systems with similar functionalities and characteristics are grouped into one "system cluster." Instead of developing test methods and use cases that are unique for every system, similarities in functions and systems are considered when selecting indicators and scenarios, which is a step towards harmonization in test methods.

As a future step, test methods for addressing the performance on a vehicle level, where the functional performance regardless of safety system is addressed, would be of interest. This would be a similar approach that currently exists within passive safety.

A study that has recently been performed and that addresses the topic of functional performance testing is the *Feasibility Study for the Setting-up of a Performance testing Programme for ICT – based Safety Systems for Road Transport* [38].

The evaluation process as part of the development process

In order to fulfil the main target of the evaluation and in order to use the resources optimally, the evaluation process should be integrated in the development process.

Ideally evaluation should be integrated in the development process already from the beginning and the project plan should allow for an iterative procedure of development and validation for improving the design and functionality. It is also crucial that functional specifications contain testable requirements and attributes.

The end point of a research project

The final result of a research project should not be a working prototype. There should at least be one evaluation cycle.

In research projects like PReVENT the goal in evaluation is to search for the largest problems and the natural consecutive step is to improve the functionality and design followed by another evaluation cycle.

Evaluation objectives

The objective of the evaluation has to be stated clearly at the beginning of the evaluation. Often it is not possible to have a team which is specialised in evaluation. Therefore a methodology should be recommended, based on best practices, and the quality of the evaluation process should be monitored.

The framework, included as Annex E, provides a first guideline towards this purpose. More work is needed on agreeing on a common basic set of scenarios and indicators to measure the efficiency of the system.

For a large integrated project like PReVENT it would be useful to have a commonality in use-cases, for example a set of basic use-case classes, so that even though each technology addresses different subset of the use-cases they are based on the same use-case classes. This would facilitate a comparison on how different technology options for each class perform in relation to each other. However, again, further research on accidentology is needed to achieve this.

Quality control of the evaluation process

Quality control should be implemented for the evaluation process, which should be lightweight and not add too many requirements. A standardised set of use cases and scenarios for testing could assist to this purpose.

Link to homologation

The RESPONSE Code of Practice [27] gives guidelines for the design of safety systems, and all systems have to follow these guidelines. There is already a process described for homologation of some systems e.g. ACC. But today's homologation processes are probably not fit for active safety systems – due to requirement of 100% Correct Alarm Rate. PReVAL is a first step towards a detailed analysis of the validation of systems.

5.2 Recommendations on evaluation research

Within PReVAL an extensive analysis of evaluation methods was performed which lead to the development of an evaluation framework covering the three aspects of evaluation (technical performance, human factors and safety impact assessment) by using the holistic approach of the situational control concept. However, there are still several open issues, which could not be solved within PReVAL and should be subject of future research. Keywords here are: field operational test, long term effects or unintended effects. This chapter will discuss the issues which should be addressed by future research projects.

5.2.1 Technical evaluation

In Section 4.4, we have discussed some main issues concerning the technical assessment of preventive safety systems. We point out in this section some directions to be taken in technical evaluation research.

<u>Compromise between addressing the full complexity of the driving context and the limited resources</u> allowed in the evaluation process: the identification of critical and most significant scenarios is crucial. We evolve then to very specialized tests with high number of repetitions in order to allow approaching to realistic results. The use of simulation in the requirements / specification phases of the system design can assist in a more optimal use of the resources.

<u>Concerning statistical aspects</u>: the measurement plans in the evaluation process should include the following considerations:

- **Blocking**: the conditions influencing the data collection should be as far as possible controlled and homogeneous. A series of test 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).
    - o **Statistical relevance**: the number of tests performed should be related to the expected level of statistical confidence.

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

<u>The “reference measurement”:</u> that is a second source of measure used in the evaluation procedure in order to validate the system sensor output.

This constitutes a real trade that should be constructed by means of very specialized equipments. The way to have such high cost test benches is to spread out the benches in one country with each site specializing in one field (e.g. site on ADAS communication-based).

<u>Indicators:</u> the same indicator can be of technical or HMI type depending on the type of function considered (warning or intervening). As an example, for a lane departure (intervening) trajectory correction system, the lateral acceleration is a technical indicator, whereas, if we consider a lane departure warning system, it is not anymore a technical indicator since the correction depends on the driver’s reaction. It is hence very important that the choice of indicators is coordinated with the human factors evaluation team through the concept of situational control. Although both technical and HMI indicators are of main importance for evaluation, testing objectives are different and a clear distinction between these two types of indicators should be made, such that evaluation objectives are very clear.

<u>Tools and methods</u>: for this purpose, the combination of simulation environment with hardware-in-the-loop tests is a valid alternative: the validation phase (evaluation of the entire system) combining a high number of scenarios that can be performed in a first stage through a simulation study with as purpose to identify the most critical cases, which are subsequently tested in a hardware-in-the-loop environment.

<u>Targets and representativeness of the tests, which lacks of standards</u>: 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.

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

- **Detectability specifications for Lidar and Radar (ISO 15622:2002).** The test targets are defined as possessing a *CTT* (Coefficient for Test Target) for lidars or a RCS for Radars.

  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 = $(1 \pm 0{,}1)$ 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.

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

<u>Tests representativeness and simulation environment </u>: This last point links the above ones. If we have talked about limited resources leading to delimitation of scenarios, number of repetitions of tests, simulation environments and targets representativeness, all these points are linked together. The simulation is a key point but it includes large future work in order to approach at the best the real situations. In particular, if targets representativeness is a key point on test tracks, it is as well a key point while carrying out evaluation by simulations. It is of main importance, after the real targets definitions, to implement in the simulation environment such representative (and a big work is needed to define standards) targets.

5.2.2 Human factors evaluation

In this section some recommendations for future research is briefly summarized.

Further research on how to define functional performance of active safety system and ADAS is desired; today there is no well established definition.

It is of importance not to separate the technical verification and validation and the human factor evaluation, but address the performance as an overall concept; including both these aspects, as done with the introduction of the situational control. Future research should address the question how this can be further developed, with respect to indicators and test scenarios.

How can the functional performance of an “ADAS”, active or preventive safety system be defined? There is today no obvious definition of functional performance of an ADAS. Compare with e.g. passive safety where the performance can be defined by level of injury in a crash and the aim with passive safety systems is to reduce the injuries in the event of a crash, for which there a re established measures. This founds the basis for the tests; how can we define the reference and level of acceptance for the tests carried out. It has to be defined what variables should be measured and how this has to be done.

It is of importance not to separate the technical verification and validation and the human factor evaluation, but to address the performance as an overall concept; including both these aspects, as done with the introduction of the situational control. The concept of situational control has been proposed in PRevAL, but is not yet complete and defined in detail, and hence needs further development. Future research should address e.g. the indicators and test scenarios.

Further research on how to select and prioritize among scenarios and indicators would also be of great interest for optimizing your test. A first step might be to study common indicators and traffic scenarios for systems with similar characteristics.

A general issue is how future tests can be planned and optimized among available tools and test facilities. For example, further research on how to choose among potential test facilities when evaluating you preventive safety system would be of interest, but it is difficult to provide guidelines on this issue. It is also often a question of resources.

Another unsolved issue related to test facility is how to test systems in critical scenarios, in a physical environment in a representative way, such as lane departure warning systems.

There are scarce guidelines in the current human factors evaluation procedure on how to develop questionnaires for the subjective tests. Tools for collecting and analysing subjective data; answering forms and rating scales are available but there are no generally established guidelines on how to use them. Further work for developing questionnaires for ADAS would be of interest.

Research on driver models and drivers behaviour in general is of great importance as input to human factors evaluation.

5.2.3 Safety impact assessment

The availability of reliable and comprehensive accident databases is critical for analyses. However, the used approach showed indisputably how insufficient the current accident information is. The general trend is that easily available databases include aggregated data (the data that can not be cross-tabulated according to variables describing accident type and location, vehicles and persons involved etc.), but disaggregated data with more specific data is rare. Fortunately, CARE database provides sufficiently disaggregated data at the European level (that was utilised in the analyses). However, much work is needed to have a harmonised, reliable and comprehensive data for each European country. For example, it is a well-known fact that the information about the fatalities is usually reliable, but there is much room to improve information about the injury accidents and especially about the severity of injuries (i.e. separation of severe and less severe injuries). Consequently, the research would greatly benefit from expanded availability of disaggregated accident data (the data that include specific information for each accident, vehicle and person involved) for each European country. In addition, the accident information should be comprehensive, detailed and harmonised.

Another substantial challenge concerns the information about the driver behaviour. There is no comprehensive European database describing the driver behaviour by circumstance and situation (e.g. road type, vehicle type, lighting condition, weather condition etc.). This behavioural information is needed for normal driving and accidents in order to find reliable and accurate safety results. This type of information would be especially important about the effects of various ADAS and their functions. The same applies to travel behaviour and the impacts of different functions on trip making, mode and route choice. Consequently, it is strongly recommended to conduct studies such as “Naturalistic Driving” and “Field Operational Tests” that could provide required information.

A further topic requiring exploration is the combined effects of several intelligent functions. Currently the analyses usually assume that the effects of several functions operating at the same time are independent, and hence, their effects can be "multiplied" to get their combined effect. As many functions target the same aspects of driver and travel behaviour, it is likely that the effects of such functions are not mutually independent. The relationships between the effects of various functions utilised at the same time should be studied with methods of sufficient validity in proper experimental designs, again possible within e.g. Field Operational Tests.

6 Projects results and achievements

6.1 Meeting the project objectives

Table 11 shows that the PReVAL project has met the S&T objectives which are stated in the Description of Work.

Table 11 How PReVAL meets the S&T objectives

Sub-project objective	Meeting the objective	Explanation
Assess and make conclusions of the technical performance of PReVENT functions	Yes	The PReVAL project has analysed the evaluation results reported by the PReVENT subprojects. The analysis and the conclusions are reported in D16.2. A summary of the results is given in Section 3.1.
Evaluate the potential of PReVENT subprojects HMI results – user acceptance, preferences and behaviour. Use existing data accumulated in subprojects, if possible	Yes	The PReVAL project has analysed the human factor evaluation results reported by the PReVENT subprojects. The analysis and conclusions are reported in D16.2. A summary of the results is given in Section 3.2.
Estimate the potential safety impacts of PReVENT functions in cooperation with ongoing other relevant projects such as AIDE, eIMPACT and TRACE.	Yes	PReVAL has assessed the safety impacts of selected PReVENT functions, i.e. the functions which are not assessed by the eIMPACT project. The same approach as eIMPACT is used, allowing comparing the results. PReVAL has used accident data, provided by TRACE. AIDE did not performed safety impact assessments, but the HMI methodology of AIDE is taken into account in the PReVAL methodology.
Create an overall view of the potential PReVENT safety impacts and make recommendations for further work in the development of effective preventive safety functions	Yes	A summary of the safety impacts of PReVENT functions analysed is given. PReVAL has developed an assessment framework for preventive safety systems, which allows producing comparable and reproducible results. The framework has been described first in D16.3, and is updated to the Annex E of this deliverable. PReVAL has made recommendations for the development process of preventive safety functions (Section 5.1) and for needed research in the evaluation of safety systems (Section 5.2).

6.2 Scientific & technological quality and innovation

The PReVAL project has analysed the evaluation results of PReVENT functions and assessed the safety impact of PReVENT functions. For the safety assessment, the method developed and used by the eIMPACT project has been used.

The project has first reviewed the existing methods and selected the best practices, based on which procedures for technical and human factor evaluation have been derived. These procedures have been applied to the INSAFES project, and human factor procedures have been tested to the MAPS&ADAS functions and the project assisted in the INSAFES validation. Feedback on the method was requested from selected experts, and the methodology was discussed, with experts from eIMPACT, TRACE, EuroNCAP/BeyondNCAP and APROSYS at the final workshop. The framework is hence based on approved methods and provides guidelines for future assessments.

6.3 Economic development and S&T prospects

The project has not directly created any commercial products.

The evaluation framework will be used by the partners in their own projects. The framework will be among others used in the evaluation in the SAFESPOT IP in 2008.

The framework will be promoted to other institutions, and discussions are ongoing with Beyond NCAP for the use of parts of the methodology in their approach.

7 Contribution to PReVENT integration

The PReVAL project has intensively cooperated with the other PReVENT subprojects. The evaluation results of the other subprojects (APALACI/COMPOSE, INTERSAFE, LATERAL SAFE, SAFELANE, SASPENCE, MAPS&ADAS, WILLWARN) have been analysed. Safety assessments have been performed for APALACI/COMPOSE, INTERSAFE, MAPS&ADAS, SAFELANE and SASPENCE.

PReVAL has forwarded a draft version of the framework to INSAFES, which has applied the framework in the validation plan. The INSAFES team has provided PReVAL feedback on the applicability of the method. PReVAL has assisted in the validation of INSAFES, in order to get additional feedback on the methodology.

A human factors expert analysis method, proposed by PReVAL, has been applied to MAPS&ADAS functions, in order to get feedback on the feasibility of the method.

PReVAL has also extensively collaborated with other projects:

- eIMPACT: PReVAL uses the safety assessment methodology developed and used by eIMPACT. PReVAL addresses the PReVENT functions, not analysed by eIMPACT. Collaboration at workshops and at the PReVENT IP Exhibition.
- TRACE: provision of accident data of TRACE to eIMPACT and PReVAL. Collaboration at workshops and at the PReVENT IP Exhibition.
- AIDE, INVENT, HUMANIST, ADVISORS: human factors evaluation method. Experts involved in AIDE methodology have been very active in PReVAL.
- APROSYS: technical evaluation procedure. APROSYS representatives collaborated to the workshops.
- EuroNCAP/BeyondNCAP: participation to PReVAL workshop
- Volpe Center: feedback to the methodology.

8 Project outputs

8.1 Deliverables

The PReVAL project produced four deliverables:

Table 12 INSAFES' deliverables

Del. Number	Del title	Description of contents and result
D16.1	Review of validation procedures for preventive and active safety functions	The report contains the review of the validation plans of the PReVENT subprojects. The evaluation methods used by PReVENT subprojects and related projects (APROSYS, AIDE, INVENT, HUMANIST, ADVISORS, eIMPACT) are analysed. The report contains also annexes with a detailed system description as basis for the safety assessments.
D16.2	Analysis and results of validation procedures for preventive and active safety functions	The technical and human factors evaluation results of the PReVENT subprojects are analysed. The report also includes a qualitative safety assessment of PReVENT functions
D16.3	Proposal of procedures for assessment of preventive and active safety functions	This report contains the framework, consisting of procedures for technical, human factors and safety impact assessments. The Annex E to D16.4 is an updated version of this document, in which the different procedures are integrated in one framework.
D16.4	Project final report and recommendations for future assessments	This is the final report of PReVAL, and contains the results of the project. The report contains also the quantitative safety assessment of PReVENT functions, the work on the application of the framework, and the final recommendations for the assessment of safety functions.

8.2 Dissemination and other outputs

PReVENT as all PReVENT SPs followed a specific dissemination strategy based on the guidelines and the framework of the IP itself. This strategy engaged a number of means in order to disseminate the subproject concepts, objectives, methodology and results to the wider possible audience. These means included:

- Web site: a pubic website was created linked to the PReVENT site and following the same graphical template. Information on the workshops organised by PReVAL was put on the website and the presentations of the workshop are made available on the website. The public deliverables of the subproject were made available for download by the website viewers.
- Leaflets: Leaflets on PReVAL were designed to be identical to the PReVENT graphical concept for easier recognition. These leaflets were disseminated to all important events and in all PReVENT occasions.
- Posters: four posters were developed for the PReVENT IP Exhibition. The posters include a general poster, and one poster for the different workpackages: technical evaluation, human factors evaluation, safety assessment.

- Organisation of a final workshop on evaluation in collaboration with eIMPACT and TRACE.
- Participation to major conferences. PReVAL has been presented at the ITS Europe conference in Aalborg and ITS World conference in Beijing.
- Participation to Special Sessions in important events related to PReVENT.
- Participation to PReVENT events including the PReVENT IP Exhibition

8.2.1 Subproject website

The PReVAL website is hosted by the PReVENT website and can be found at:

http://www.prevent-ip.org/en/prevent_subprojects/horizontal_activities/preval/

The webpage contains information about PReVAL, its objectives key concepts and results. The presentations of the workshops and the public deliverables are accessible and downloadable.

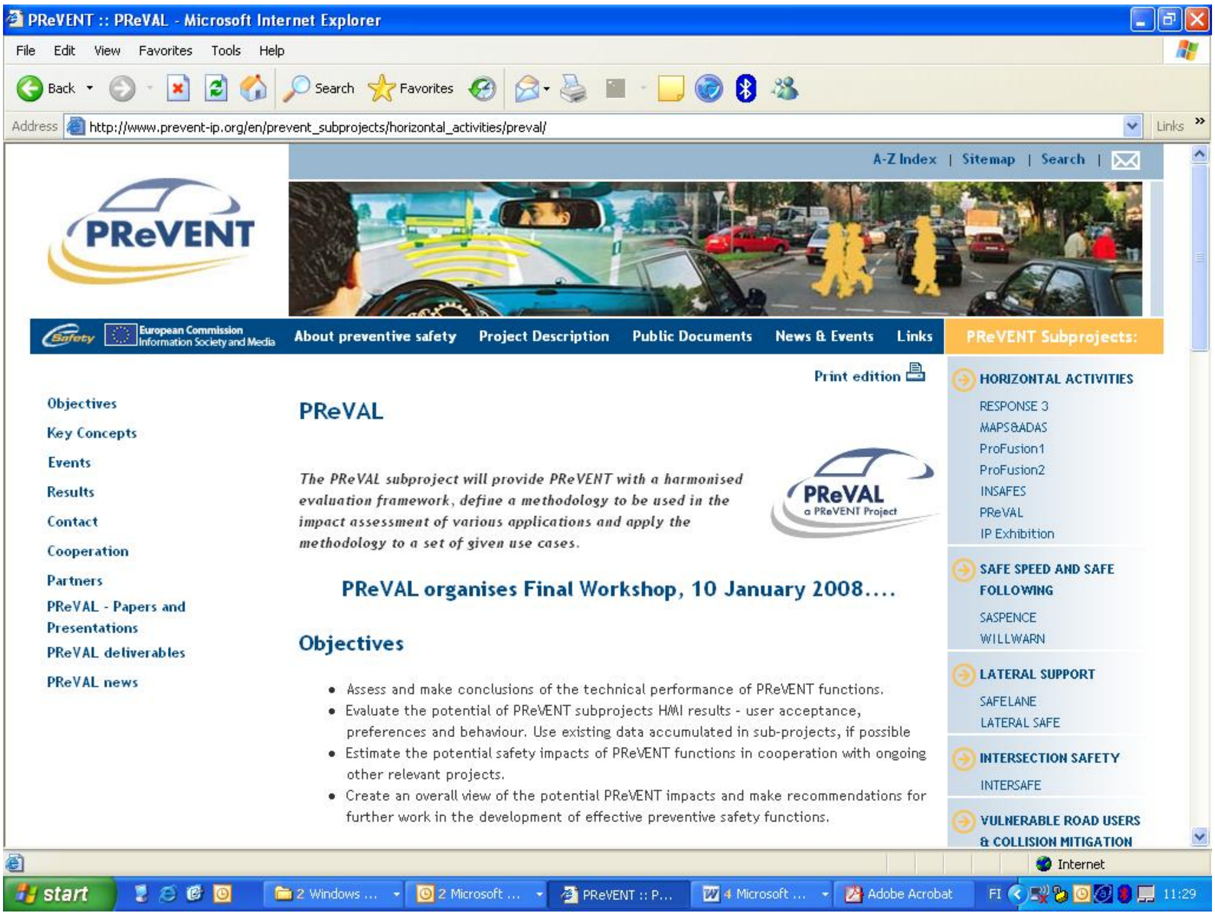

Figure 11: PReVAL website

8.2.2 Workshops

PReVAL organised two workshops: a first on 22-23.01.2007 in Delft. The discussions and conclusions of this workshop are included in D16.1.

A final workshop was organised in Brussels on 10.1.2007, in collaboration with the TRACE and eIMPACT projects. The programme and list of participants is included in Annex C. In the annex also an overview of the discussions is given. The presentations are available on the PReVAL website.

8.2.3 Other dissemination material

PReVAL was present at the PReVENT IP Exhibition. PReVAL had a tent in collaboration with TRACE and eIMPACT (Figure 12). The material in the tent included the four posters, mentioned above. PReVENT also contributed to the PReVENT IP conference.

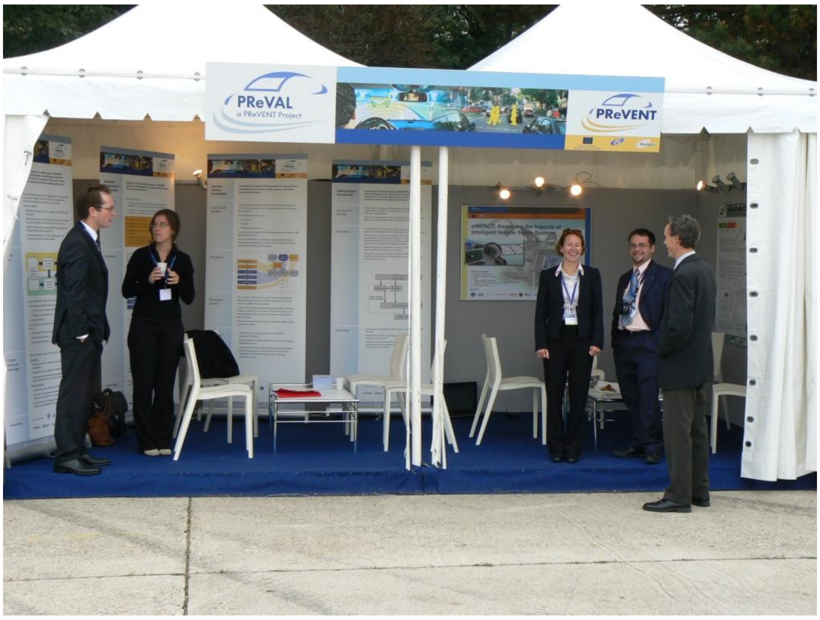

Figure 12: PReVAL tent at the PReVENT IP Exhibition

The following table gives an overview of the dissemination activities of the PReVAL project:

Table 13: Overview of articles and presentations on the PReVAL project

Type / title	Date and where published	Details
First workshop	Delft, 22.1.2007	
TRACE meeting	Prato, 17.1.2007	presentation by J. Engström, VTEC of PReVAL at TRACE general meeting
PReVAL Evaluation of Safety Functions	ITS European congress in Aalborg 18-20.6.2007	Presentation by A. Hiller, Daimler
Tent, flyer, poster, presentation	PReVENT IP exhibition, 18-20.9.2007	
Evaluation framework for preventive safety applications, J. Scholliers, F. Hendriks, M. Ljung, V. Anttila, M. Netto, J. Engström, K. Heinig, A. Amditis	ITS world congress in Beijing, 9-13.10.2007	paper presented by A. Amditis, ICCS
PReVAL and first results	ITS world congress in Beijing, 9-13.10.2007	presentation in PREVENT session by M. Netto, LCPC
PReVAL project presentation	"Älykäs liikenne" (Intelligent transport) day in Helsinki, 30.10.2007	Presentation by J. Scholliers, VTT, about 50 persons in session

Type / title	Date and where published	Details
An Integrated Evaluation Framework for Preventive Safety Applications, J. Scholliers, F. Hendriks, M. Ljung Aust, J. Luoma, M. Netto, J. Engström, S. Leanderson, R. Kutzner, F. Tango	ITS Europe, Geneva, 4-6.6.2008	Draft paper submitted
Some technical results related to the prevent project and methodological aspects, J. Scholliers, M. Netto, J.M. Blosseville, J. Chen, K. Heinig, F. Hendriks	ITS Europe, Geneva, 4-6.6.2008	Draft paper submitted
An Integrated Evaluation Framework for Preventive Safety Applications, J. Scholliers, F. Hendriks, M. Ljung Aust, J. Luoma, M. Netto, J. Engström, S. Leanderson, R. Kutzner, F. Tango	Intelligent Vehicles 08 (IEEE, IV08), 4-6.6.2008	Paper submitted
PReVAL final workshop	PReVAL final workshop, 10.1.2008	Organized by PReVAL project, in collaboration with TRACE and eIMPACT
New framework for evaluating preventive safety functions: focusing on technical evaluation, M. Netto, J.M. Blosseville, F. Hendriks, J. Scholliers, J. Chen, K. Heinig, M. Ljung Aust	ITS World congress, New York, 16-20.11.2008	Draft paper submitted
New framework for evaluating preventive safety functions: focusing on human factors evaluation J. Scholliers, S. Leanderson, M. Ljung Aust, J. Luoma, R. Kutzner	ITS world Congress, New York, 16-20.11.2008	Draft paper submitted
APROSYS Workshop on Test methods for pre-crash systems	Helmond, 30.1.2008	presentation of PReVAL by R. Schram, TNO

9 Conclusions

The PReVAL project addresses the possible safety impacts of applications developed and demonstrated in the PReVENT integrated project. The objectives of the PReVAL project, which are discussed in this deliverable, are,

- to analyse the evaluation results of the PReVENT subprojects and to assess the safety potential of PReVENT functions;
- to develop a framework for the assessment of preventive and active safety functions;
- to make recommendations for future assessments of preventive safety functions.

Assessment serves two purposes: to assure the functionality of the system and to assess and quantify the system's impact on the traffic system. The systems developed in PReVENT are mainly research prototypes and the main objective of the performed evaluation was to assure that the system works as expected.

Such systems make use of a large set of sensors (environmental sensing, maps, telecommunications etc) and data fusion techniques in order to provide reliable detection and positioning inputs for the proposed assistance functions. PReVENT subprojects provide a large "basket of new technologies", which are fundamental main bricks for preventive safety systems.

All PReVENT subprojects followed the CONVERGE approach for the technical evaluation of the developed prototypes. All subprojects achieved good results for the reliability indicators (correct, false and missed alarm rates). The PReVENT project has hence demonstrated the feasibility of the demonstrated concepts and hence brought the technologies a step forward towards market introduction.

The human factor evaluation of six subprojects has been analysed. The amount and nature of the tests performed by the subprojects is heterogeneous. All the analysed subprojects report positive results on driving performance and driver behaviour, as well as for acceptance and usability, with a variation in the significance and distribution of the results. Most projects emphasize the needs for further experiments to achieve statistically significant results and to optimise the HMI solution.

PReVAL has assessed the safety potential of APALACI/COMPOSE, INTERSAFE (left turn assistant), MAPS&ADAS, SAFELANE and SASPENCE. The safety assessment used the behavioural affect approach (except APALACI/COMPOSE that was assessed by another methodology), which has been developed and used by the eIMPACT project, which is assessing the remaining PReVENT functions, as well as some other safety functions. The results of the eIMPACT assessment will be published in eIMPACT Deliverable D4 [32].

All investigated safety functions have positive effects on fatalities and injuries, but in varying degrees. In part, the magnitude of the safety impacts is affected by the circumstances in which the safety function is relevant, and the frequency with which corresponding accidents occur. The safety assessment conducted in PReVAL has produced low, most probable and high estimates of injuries and fatalities for 100% fleet penetration. The most probable estimates are reproduced in Table 14. Moreover, it has produced most probable estimates of injuries and fatalities for estimated low and high penetration rates in 2010 and 2020.

Table 14: Most probable effects of safety functions on injuries and fatalities, at 100% fleet penetration.

Safety function	Effect on fatalities	Effect on injuries
APALACI/COMPOSE	-19,6%	-14,3%
INTERSAFE	-0.6%	-1.5%
MAPS&ADAS	-13.1%	-8.2%
SAFELANE	-13.5%	-9.5%
SASPENCE	-6.5%	-3.8%

The results given in Table 14 show that the most substantial safety impact were found for APALACI/COMPOSE, followed by SAFELANE, MAPS&ADAS, SASPENCE and INTERSAFE (left turn assistance). However, this type of comparison should be interpreted with caution since the effects of APACI/COMPOSE were obtained by the different methodology. Nevertheless, the overall conclusion is that the most probable safety effects are substantial in comparison with many traditional road safety measures **Error! Reference source not found.**]. This suggests that one of the main challenges dealing with the safety benefits of these functions concerns the implementation. Specifically, the results showed that the expected penetration rates in 2010 and 2020 were relatively low for most functions and thereby the expected safety impacts in 2010 and 2020 were much lower than presented in Table 14.

Starting from the experiences of the PReVENT subprojects and from the work of the other related projects, such as AIDE and APROSYS, an integrated framework for the assessment of preventive safety functions has been developed. A key concept in the framework is situational control, i.e. the degree of control that a Joint Driver-Vehicle System (JDVS) exerts over a specific traffic situation. The framework consists of procedures for technical, human factors and safety potential assessment. The procedures for technical and human factors evaluation go through the same steps: (1) expected impacts; (2) test scenario definitions; (3) evaluation method selection; (4) measurement plan; (5) execution and reporting. The safety potential assessment method has been developed by the eIMPACT project, and uses the following ingredients: (1) an accident database; (2) identification of relevant accidents for the safety function; (3) estimation of changes in various aspects of driver behaviour (both users and nonusers), caused by the safety function; (4) estimation of the impact of various parameters such as road type, weather condition and lighting condition; (5) estimation of the (future) penetration rate of the system; (6) assessment of the overall effectiveness of the safety function.

The procedures, reported in D16.3, have been applied to the INSAFES validation plan. INSAFES evaluation experts provided feedback on the application of the methodology, and feedback was requested from other experts. The expert evaluation method for human factors has been tested for the MAPS&ADAS functions. The feedbacks received have been taken into account in the final version of the framework.

Starting from the analysis of the PReVENT evaluations and the experiences and feedback from the framework, a set of recommendations for the system development process and the needed research in evaluation are derived. Regarding the development of safety systems, a main recommendation is that the functional specifications should be based on the identified and relevant accident types, taking into account the status of the technology for detection of objects and control of the vehicle. In order to be able to verify the performance of similar systems, a common set of high-level scenarios is needed, which could also be used for homologation. The evaluation process should start at an early phase of system development, and use common scenarios and indicators. Simulation tools and hardware-in-the-loop tests allow optimising the use of resources.

The evaluation in PReVENT has mainly been concentrated on the intended effects of the systems. All the effects of the preventive safety system on driver behaviour cannot be detected with short term testing. Also safety assessment requires information on the behavioural adaptation. Field operational tests are required to analyse the behavioural adaptation, and to provide data for safety assessment.

Further research is also needed to investigate the effects of multiple functions in a single vehicle. Also, for this purpose, Field Operational Tests provide a solution.

In order to carry out such activities, more experiments on current PREVeNT applications have to be taken into account. The goal is twofold: on one side, it is necessary to achieve results more statistically meaningful and so to improve further the technical performances of these systems; on the other hand, HMI solutions (proposed up to now) have to be optimised.

Therefore, all these phases shall co-operate and interact each other, in order to reach a full and effective safety assessment framework, as well as good and valuable results.

Acknowledgements

The work presented in this document has been funded by the European Commission through the PReVENT IP subproject PReVAL. The PReVAL Consortium wants to express thanks to the European Commission and the PReVENT IP Management for their support. The PReVAL Consortium also wants to thank the PReVENT projects for their collaboration and providing the requested information to PReVAL. Special thanks also to the eIMPACT project for their support in the safety assessment analyses and, together with the TRACE project, for the collaboration in dissemination activities. The PReVAL consortium wants to thank all workshop participants for the fruitful discussions and Volpe Center for their feedback.

National Accident Data for Great Britain is collected by police forces and collated by the UK Department for Transport. The data are made available to the Vehicle Safety Research Centre at Loughborough University by the UK Department for Transport. The Department for Transport and those who carried out the original collection of the data bear no responsibility for the further analysis or interpretation of it.

National Accident Data for France is collected by police forces and collated by the French Ministry in charge of Transport. The data are made available to the Laboratoire d'Accidentologie, de Biomécanique et d'études du comportement humain PSA Peugeot Citroën - Renault (LAB) by the Ministry which therefore bear no responsibility for the further analysis or interpretation of it.

References

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

[2] PReVAL Project Deliverable IP_D10/D16.2, Analysis and results of validation procedures for preventive and active safety functions, December 2007

[3] INTERSAFE Project Deliverable D40.75, Final Report (Project Evaluation and Effectiveness of the Intersection Safety System).February 2007.

[4] Ammerlaan, Driever and van Campen (2003). Optiedocument duurzaam veilig voertuig, final report with factsheets (in annex).

[5] 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.

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

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

[8] 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

[9] D. Maltby, S. Morello, K. Perrett and J. Hopkin, "Checklist for preparing a validation plan: updated version," CONVERGE project deliverable 2.4.1, 1998

[10] 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.

[11] 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.

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

[13] 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,

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

[15] 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.

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

[17] 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.

[18] 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.

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

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

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

[22] 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.

[23] 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

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

[25] 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.

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

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

[28] R. Großpietch, PhD thesis at the Institut für Kraftfahrwesen der RWTH Aachen, Aachen, 2004, to be published (in German)

[29] Enke, K. 1979. *Possibilities for improving safety within the driver-vehicle-environment control group*. Proceedings of the 7th International Technical Conference on Experimental Safety Vehicles, Paris, France, 789-802. Washington, DC:U.S. Department of Transportation, National Highway Traffic Safety Administration.

[30] Elvik, R. and Vaa, T. (2004). The handbook of road safety measures. Amsterdam: Elsevier.

[31] Hjälmdahl, M., and Dukic, T. (2007). IT's to improve speed behaviour for truck drivers: short term effects of a speed support system. In, Proceedings of 14th International Conference on Road Safety on Four Continents. Ministry of Transport Thailand, VTI, GRSP, FERSI, TRB, ECTRI., Bangkok, Thailand, November 14-16, 2007

[32] Wilmink et al, eIMPACT Deliverable D4, to be published in 2008.

[33] European Enhanced Safety Vehicle Committee, WG19, Version 3 report, January 2006

[34] Priorities for EU motor vehicle safety design, Brussels 2001

[35] Carsten, O. M. J. & Brookhuis K. (2005). The relationship between distraction and driving performance: towards a test regime for in-vehicle information systems Transportation Research Part F: Traffic Psychology and Behaviour 8, pp.75-77

[36] European Commission. 2006. Commision recommendation of 22 December 2006 on safe and efficient in-vehicle information and communication systems: Update of the European Statement of Principles on human machine interface.

[37] Brook-Carter, N. et al. (1999) Development of a multiparameter criteria and a common impact assessment methodology. ADVISORS Deliverable D4.1

[38] Lindholmen Science Park AB, Jacobson, J. et al (2007), Feasibility Study for the Setting-up of a Performance testing Programme for ICT-based Safety systems for Road Transport, DG INFSO, unit ICT for Transport, 20 September 2007.

.

Annex A Keywords

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

Annex B Abbreviations

Abbreviation	Explanation
ACC	Adaptive Cruise Control
ADAS	Advanced Driver Assistance System
CAN	Controller Area Network
CAR	Correct Alarm Rate
CMS	Collision Mitigation Systems
EES	Energy Equivalent Speed
ESC	Electronic Stability Control
FAR	False Alarm Rate
FOT	Field Operational Test
HMI	Human Machine Interaction Human Machine Interface
HSW	Hot Spot Warning (MAPS&ADAS function)
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
RMS	Root Mean Square
SLW	Speed Limit Warning (MAPS&ADAS function)
TTC	Time to collision

A glossary is included in the Annex E on the Framework

Annex C Final Workshop

C.1 Invitation Text

The PReVAL Consortium is pleased to invite you to attend the PReVAL Final Workshop that will take place at Foundation Universitaire, Brussels, on the 10th of January, 2008. The workshop is organized together with the ongoing eIMPACT and TRACE projects.

PReVAL is a subproject of the PReVENT IP and has as main objectives:

- to identify best practices for the assessment of IP PReVENT safety applications.
- to define a framework for estimating their safety impact taking into account technical performance and human factors.
- to apply the framework to estimate the potential safety impacts of selected PReVENT applications.
- Make recommendations for the future assessment and development of preventive safety applications

Substantial work on safety assessment is also being performed in eIMPACT and TRACE.

The key aims of the Final PReVAL workshop are:

- To present the assessment framework developed by PReVAL and the safety assessment methodologies developed by eIMPACT and TRACE.
- To promote and disseminate the assessment work of PReVAL on technical, HMI and safety potential of preventive and active safety functions;
- To report on the assessment of PReVENT applications.
- To examine the application of above findings towards arriving at Best Practice within the evaluation of active and preventive safety functions.

The workshop consists of two parts:

- *Morning: Parallel break-out sessions on* devoted to fine-tuning internal and external assessment methodology links. At the start of the sessions, the approaches will be presented, followed by a discussion on selected issues.*:*
  - § *Safety assessment methodologies*. Presentation and discussion of the eIMPACT/PReVAL and TRACE methodologies.
  - § *Technical and human factors evaluation.* Presentation and discussion of the evaluation methodologies proposed by PReVAL.
- *Afternoon: Presentation and discussion of analysis results of PReVENT subprojects, and of future work in evaluation.* PReVAL will present the work on the analysis of the technical, human factors and safety impact evaluation of PReVENT functions. The analysis work in TRACE and eIMPACT is still ongoing. TRACE and eIMPACT projects will be presented.

C.2 Agenda

<table>
<tr><td>10:00</td><td>10:15</td><td colspan="2">Welcome by the project coordinator and introduction of the 2 breakout sessions</td></tr>
<tr><td>10:15</td><td>10:25</td><td colspan="2">Welcome by the EC representative (Fabrizio Minarini)</td></tr>
<tr><td>10:25</td><td>12:00</td><td>Breakout Session 1: Safety assessment methodology
Introductory presentations:
a) TRACE methodology (Yves Page)
b) PReVAL/eIMPACT methodology</td><td>Breakout Session 2: Technical and human-factors evaluation framework
Introductory presentations:
a) PReVAL technical evaluation
b) PReVAL HMI evaluation
Moderator: Magnus Rilbe</td></tr>
</table>

		+ example safety assessment (Pirkko Rämä) Moderator: Kerry Malone	
13:00	13:10	Welcome by the chairman (Matthias Schulze)	
13:20	13:30	PReVAL: Overview of the project (Johan Scholliers)	
13:30	13:50	PReVAL: Results of PREVENT technical performance evaluations (Jean-Marc Blosseville)	
13:50	14:10	PReVAL: Results of PREVENT human factors evaluation (Susanna Leanderson)	
14:10	14:30	Discussion on results of PREVENT evaluation (Matthias Schulze)	
14:40	15:00	TRACE: project overview, illustrated with preliminary results (Yves Page)	
15:00	15:15	eIMPACT: project overview (Kerry Malone)	
15:15	15:35	PReVAL safety assessment of PREVENT functions (Juha Luoma)	
15:35	16:00	Overview of discussions in the breakout sessions	
16:00	17:00	Final panel discussion	

C.3 Overview of the discussions in the workshop

C.3.1 Discussion on the technical and human factors evaluation results

Are the results too good to be true?

According to the analysis of the evaluation results by PReVAL, the reported results reported are very good. This means that the subprojects have achieved their objectives. The evaluation focused on short-term effects, and long term effects have not been included. They were not in the scope of PReVENT, and should be researched further. Also, test subjects were aware they were being tested, and what the purpose of the system was, which may have an impact on the results. The significance of the HMI tests is that they weed out the really bad systems. The significance depends on the number of tests.

Evaluation in new projects

The evaluation has to be better connected to safety assessment. High-level use cases/ scenarios have to be defined. Testing needs to take place early in the development. A common ground is needed for interpreting the safety effects. High level indicators are needed, which can be translated into safety assessment input. . Recently, algorithms such as the German I-TSA procedure and the AIDE algorithm have become available for extrapolating behavioural results to expected accident risk effects. The ITS-A-procedure and the AIDE project provide tools for this purpose. However, it is in general not possible to validate this. Developing a prototype is not the end point, but the mid point of the development.

There is a need for common (standardized) tools. Interaction between FOTs, TRACE and future PREVENT projects will be beneficial. TRACE provides information on accident data, but does not include driver behaviour. FOTs provide the baseline for driver behaviour analysis. These data can be used in order to assess how development improves driver behaviour. There is a concept needed to quantify driver behaviour. Long term tests do not immediately mean the need for FOTs: e.g. 5 weeks can be sufficiently long for long-term testing, which gives room for optimization of the resources.

Cross-cultural tests

The question was raised if the differences in driving style between different countries were taken into account. Different driver types (conservative, assertive, etc) were seen in the testing in PReVENT subprojects, but they were not mapped on nationality. In the HUMANIST project, comparisons between different countries were made.

PReVAL within PReVENT

PReVAL would have had more impact in PReVENT, if it would have started in the beginning of the PReVENT project. Then it would have been possible to interact with and give guidelines to the different PREVENT subprojects, in order to harmonize the approaches and calculation of common indicators.

C.3.2 Panel discussion

Use of the methodology

The Commission representative, Mr. Fabrizio Minarini, was asked how the Commission intends to use the methodologies developed in eIMPACT, PReVAL and TRACE. The methodology can be first used in the FOTs, which will start in the middle of 2008, and for which assessment is very important. Accidentology is also an important issue, and more work on this issue is needed. There are a lot of databases, but not synchronised. Extension of the databases to the whole EU is important, even if difficult. The EC wants to push the take-up of safety systems very fast, as (predicted) penetration rates are very low. Quantitative results are very important for the Commission, in order to allow a political push. Safety systems have a high potential, but impact data which can be trusted is needed. Legislation is not an option, consumers have to be convinced to buy safer vehicles.

Euro NCAP is researching how active and preventive safety systems can be verified. They are looking at the quality of the development process, and a possibility would be to give a quality mark to the development process. Specific scenarios can be defined again which the system is tested. The PReVAL methodology could be part of the process.

Hypotheses

Safety assessment depends very much on hypotheses. The question was raised if it would make sense to have standardised hypotheses. The answer by the experts was that more standardised test scenarios are needed, which would give an overview picture of how the systems complement each other. According to Yves Page, coordinator of the TRACE project, data and methods need to be standardised. This is a dynamic process where the methods developed now can be used by later projects. TRACE tries to limit the number of hypotheses. Furthermore, the methodology has to correspond to the data, in order to be useful.

Accident data

The question was raised if it is possible to have standardized data collection (police with handheld PDA). The CARE database covers a broad range, but is shallow. Deeper information on accident causes is needed. This is one of the objectives of the SafetyNet project. An accident theory is needed to decide which data has to be collected, also to make collection economical. Different objectives for use of data have different categorisations, so there is no agreement on a single accident causation categorisation.

C.4 Participants

First Name	Last Name	Association
Johan	Scholliers	VTT
Sami	Koskinen	VTT
Juha	Luoma	VTT
Pirkko	Rämä	VTT
Somya	Joshi	ICCS
Mariana	Netto	LCPC
Jean-Marc	Blosseville	LCPC
Susanna	Leanderson	VTEC
Magnus	Rilbe	VTEC
Mikael	Ljung Aust	VCC
Falke	Hendriks	TNO
Martijn	van Noort	TNO
Wiel	Janssen	TNO
Richard	Schram	TNO
Kerry	Malone	TNO
Fabio	Tango	CRF
Ralf	Kutzner	UHA
Gerhard	Noecker	Daimler
Jian	Chen	IKA
Dimitris	Margaritis	CERTH
Matthias	Schulze	Daimler
Maxime	Flament	ERTICO
Stephanie	Chaufton	ERTICO
Fabrizio	Minarini	EC
Irina	Patrascu	EC
Menelaos	Pappas	UPatras
Pete	Thomas	Loughborough University
Pierre	Castaing	UTAC
Thomas	Wohllebe	Volkswagen AG
Yves	Page	LAB
Thierry	Hermitte	LAB
Cyril	Chauvel	LAB
Peter	Morsink	SWOV
Christoph	Kessler	Ford
Aria	Etemad	Ford
Dieter	Willersin	FhG/IITB
Pierre	Van Elslande	INRETS

.

Annex D Expert human factors assessment of PReVENT functions

This annex discusses the expert evaluation performed in the PReVAL project on the MAPS&ADAS functions.

D.1 Introduction

An extensive HMI assessment was done during the MAPS&ADAS subproject within PReVENT. The expert evaluation on the MAPS&ADAS functions aimed to produce complementary human factor related results to the results already achieved within the subproject, by focusing on some specific aspects, not brought up during the evaluation done within the subproject.

Since there is currently no well established method for performing expert evaluation of an ADAS or preventive safety function, the first step was to define how to approach the assignment.

A description of expert evaluation and an analysis of the potential gain from it were discussed at the beginning of the task, of which a summary is provided in next section.

D.1.1 Introduction

An expert evaluation is a method often used for finding usability or function specific attributes that can be improved as part of an iterative design process. It involves having a small set of evaluators (combining Human Factors and technical expertise), examine the interface and judge its compliance with recognized usability principles. As a follow-up, the experts can go on to suggest improvements, or if needed, caution the designer towards unintended effects of the system/interface.

An expert evaluation is in many respects a complimentary method to a user-based evaluation, with its over arching objective being to highlight any issues of usability that the designer has overlooked.

D.1.2 Discussion

When, in the development chain, is expert evaluation of most value?

Ideally expert evaluations should take place both at the early stages of development and towards completion. The early stage evaluation would enable the designers to take on board suggestions and warnings, thereby allowing them to test the improved version on users. The latter expert evaluation would offer a comparative angle on the design processes, as well as enable any fine-tuning that remains. The danger of having an expert evaluation too late in the developmental chain of events, is that valuable input could come in at a time when most functionalities and interface features have been fixed. Also the lack of resources towards the end of development, further impede any realisation of meaningful changes that might be required.

What should be the focus in an expert evaluation of ADAS and how should an expert evaluation questionnaire be designed in an expert evaluation?

Any Human Factors oriented evaluation should be user-focused in its appraisal of system design. One should keep in mind established considerations such as cognitive and sensory overload; simple design and clarity of information given or received.

Traditionally, expert evaluation of IVIS has focused on such HMI design considerations to which there are standards or well established rules and recommendations for comparison. In addition to these HMI design criteria such as modality of warnings, size of icons, colours etc., ADAS evaluation calls for focus on a few additional variables. These concern usability and behavioural adaptation. For instance the ways in which an ADAS interacts with existing in-car systems, the driving task, as well as other drivers within the road environment, emerges as a critical area of investigation. Also risk-compensation, and over-reliance on the driver's part, along with the requirements (costs) on the part of the driver to learn and adapt to the new system, become the focus of the evaluation.

How should the questionnaires and answering forms in an expert evaluation be structured? What are the possibilities/limitations of an expert evaluation group?

Much like a user-based evaluation, there should be minimum repetition and redundancy in the questions asked. The structure adopted should be simple and immediately relevant to the specific functions being evaluated.

In the case of an expert evaluation, higher level questions regarding predictions on driver behaviour can be asked, especially with reference to usability and adaptation. Design improvements can be sought within the remit of the questionnaire, based on the expert's knowledge.

With regard to limitations of an expert evaluation group, the main issue is validity. No matter how knowledgeable the group is about a given system or domain, the evaluation input cannot be representative of the entire user base. Thus an expert group should be extra vigilant for diversity in user groups and stakeholder needs, when recommending changes. The strength of an expert evaluation in my opinion lies, in its accompanying a thorough user based evaluation and not in it being a stand-alone. Also the earlier on in the development cycle of a system, that an expert evaluation is carried out, the more effective will be its input.

How to analyze results from an expert panel consisting of 6-8 persons? It can not be subjected to a quantitative analysis and found a basis for statistical significant results.

No. Indeed the analysis will be qualitative in nature. The main idea being to draw from the in-depth feedback of the experts and to find co-relations between it and the findings of the previous user-based study; to find areas of resonance and areas of conflict. Thereby providing an analysis that combines both levels of input and thus provides added value to the evaluation framework.

What areas should be addressed (system performance, potential driver behaviour etc).

Potential driver behavioural adaptation should form the primary focus of the questionnaire. Of course the ability of the system to perform up to standard, has a direct and strong impact on user acceptance, however the main object of the expert evaluation (especially given it's a Human Factors evaluation) should remain on driver behaviour. This will add value to existing feedback and evaluation of functionality, thereby avoiding duplication of efforts.

What answering alternatives should be selected, how to balance between true/false questions and multiple choice questions and comment fields and different areas of interest?

In addition to polarised answering options (i.e. true/false) as well as grading options (where the subject can rate the level of agreement/denial); there should also be questions where the subject is allowed to elaborate/ comment in depth, thereby adding valuable input that the designer of the questionnaire might not have anticipated.

Should the different competence area of each expert found the basis for the questions considered by that person, so that each expert has different questions to assess? Or should the complete questionnaire include questions of various types, aligned to the overall expertise in the group?

In addition to a common set of questions pertaining to system performance and usability, there can be a supplementary set of questions that are designed in conjunction with the expert's particular area of knowledge and expertise. These expert-specific questions would have the advantage of allowing the designer to elicit higher level input on various aspects of the system (e.g. technical design features, HMI), without burdening all the experts with irrelevant questions, or those that they are not qualified to offer an expert opinion on.

D.2 Evaluation plan for expert evaluation

D.2.1 Applications

The system under test is called "Driver Warning System" and consists of the two applications "Speed Limit Warning" and the "Hot Spot Warning". Both applications are map based and use only a limited set of additional sensor data, like temperature and rain sensors.

The **Speed Limit Warning (SLW)** informs the driver constantly about the legal speed limit, and warns him in two levels: The speed limit icon starts blinking when the speed is exceeded by more than 6 kmph and an acoustical warning is issued when the speed limit is exceeded by more than 20 kmph.

The concept of the **Hot Spot Warning (HSW)** is to warn the driver in case of a potentially dangerous situation ahead in case a similar driving situation at the same location has already led to an accident in the past. The factors of these situations have been derived from the analysis of accident data recorded by the police and stored at statistical state offices in Germany. From this analysis accident prone locations have been identified and stored in the in-vehicle ADAS map as optional road attributes "Hot Spot". For each Hot Spot certain warning speed thresholds are defined depending on moisture (road dry vs. wet), temperature (below or above 3 °C) and lighting conditions (daylight, twilight, darkness), as well as data about the nature of the Hot Spot (like for example "curve", "slope", "deer animal", etc). The application uses functionalities of the navigation system to get its current position, heading and most probable path and evaluates constantly the data from the vehicle (rain and temperature sensor, time) to compare it with the attributes of the Hot Spots. A warning is then issued to the driver when the warning speed thresholds are exceeded.

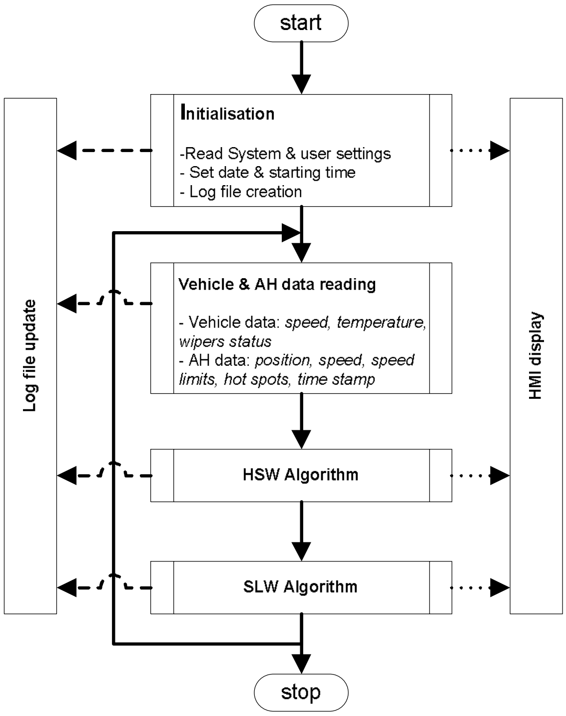

Figure 13: High-level description of the HSW and SLW algorithms

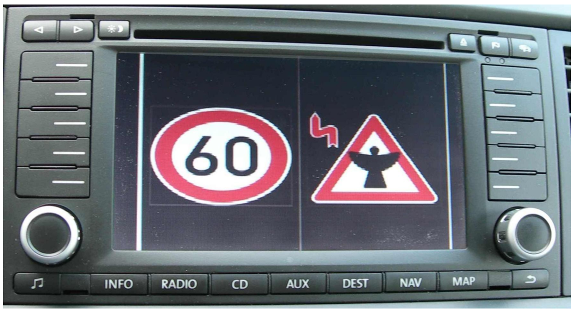

Figure 14 HMI of the SLW and HSW

D.2.2 Objectives

General objectives: Analysing the influence of the system on driver behaviour and the user acceptance/ induced mental workload of the Driver Warning System.

Research questions – Generation of hypotheses:

Hypotheses for the evaluation have been derived. They describe the expected influences and effects that the Driver Warning System in total and its different aspects might have on driver behaviour and user acceptance/ induced mental workload. The goal of the evaluation is to estimate if these hypotheses are appropriate. Therefore indicators and methods have to be determined which allow a measurement of the effects described in the hypotheses. All hypotheses can be classified under Michon's Tactical Level of the Driving Task. Both, long time and short time effects are addressed.

<table>
<tr><th colspan="2">Hypotheses</th></tr>
<tr><th rowspan="2">Michon level addressed</th><th>Intended effects (short term)</th></tr>
<tr><th>Target situation</th></tr>
<tr><td>Strategical level</td><td>N/A</td></tr>
<tr><td>Tactical level</td><td>1. The user experiences of the HMI design is generally good with respect to:
a. user-system interaction
i. type of interaction
ii. visual information presentation
iii. auditory information presentation
b. placement of display
c. vision
d. acoustics
e. information comprehension
f. anthropometrics
2. The function has potential for increasing the situational awareness in critical situations (can be either at over speeding or/and at critical hot spots); thereby increasing the overall safety of the driver and passengers in car.
a. The driver awareness of current speed limit increases. (Also driver awareness of potential criticality, i.e. non speed related factors, such as poor visibility or room to maneuver, with regard to hot-spots)
b. The system has a low likelihood for behavioural adaptation in terms of over-reliance and misuse.
c. The function does not compromise safety in term of interaction with other functionalities/systems in the car.
d. The timing of information/warning is appropriate.
e. The information displayed has an appropriate complexity.
f. The system has a high likelihood for high user acceptance and usability. (Within this it might be useful to asses – to what extent user perceives 'added value', i.e. over and above outside information).
g. No distraction of the driver will be caused by the DWS.</td></tr>
<tr><td></td><td>N/A</td></tr>
</table>

D.2.3 Scenarios

The tests are restricted to only a small area of public roads south of Hildesheim, Germany, on which the system works, since map data is only available for this region. Hence the tests will take place out on rural and urban roads. No beforehand definition of weather or lighting conditions is given. Since the test drives take place on public roads, a definition of standard situations with respect to other vehicles, etc. is not reasonable. Due to the variety of road characteristics and the length of the test track it is expected that during one test run most of the relevant scenarios are covered anyway.

D.2.4 Methods

Subjects

For evaluation an expert panel will be used. The panel will make its assessment on the basis of professional interpretations of verbal and written descriptions of the system and of their experience of driving the prototype themselves. The expert panel is used because the availability of resources precludes a quantitative assessment based on experiments. The panel will consist of six to eight experts with experience in the field of traffic engineering and ADAS development, including two to three experts of human factors.

General methodology and experimental design

Test will begin with an assessment of the general HMI solution in the stationary vehicle. Therefore a test run is simulated, so the actual speed limit on the simulated track will be displayed and several warnings will appear according to the position and speed of the simulated driver and the environmental conditions that are simulated as well. Experts will have the chance to experience the look and feel of the HMI without having to focus on the actual driving task. The usability of the system will be ascertained by questionnaires afterwards.

The actual driving tests are then carried out on a test track consisting of 48 km of public roads south of Hildesheim, Germany. The test track consists of four different sections varying in slope, curvature (gon/km), main speed restriction and road class (cp. Figure 15) hence it includes a big variety of different driving situations (scenarios). Experts are asked to drive the instrumented vehicle (Volkswagen Bus T5, year of manufacture 2006, 128 kW, manual gear box) over one lap on the test track. The experts are free to drive the way they want to and to choose their own speed, so the system can be experienced in a real life scenario and in real life situations. It's assumed that warnings against speed limit violations and hot spots will occur occasionally.

Besides the test runs experts also will have access to all functionality specifications and descriptions and the evaluation plan so that the system behaviour in situations that aren't covered during the test runs can be assessed as well.

The independent variables are hence the support by the MAPS&ADAS Driver Warning System and the different road types of the test track. The dependent variables are the impressions of the subjects collected during the test run.

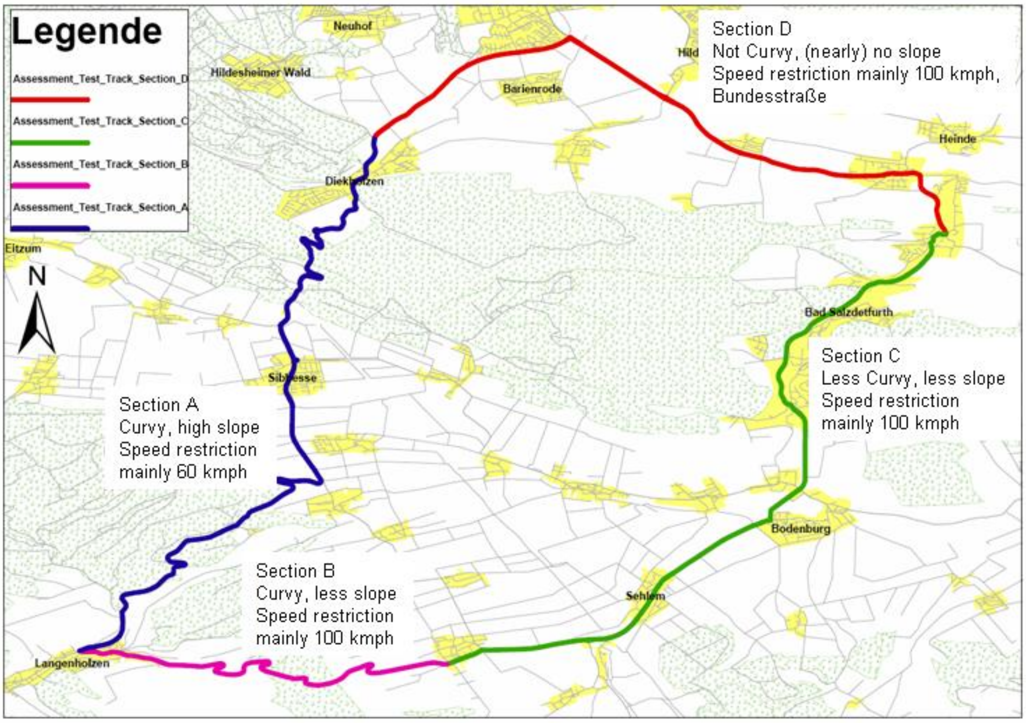

Figure 15: MAPS&ADAS Test Track

Methods and tools

The evaluation will use questionnaires which are addressed to experts. Goal is to estimate the reactions of users to the system and the probable user acceptance based on experience the experts have in this field.

Questionnaire for expert evaluation

The opinions of the experts were gathered by using questionnaires, which were based upon the hypotheses, to largest possible extent. The hypotheses generated within the MAPS&ADAS subproject were used, with some modification and additions, to adapt them to the current evaluation method – expert based.

Based on the hypotheses, questions were designed and structured in sections depending on what issue they addressed: vision, acoustics, usability, acceptance or more driver performance related issues.

It was decided that one part of the evaluation should be performed in a stationary vehicle focusing on HMI design in simulation mode. A second part of the evaluation should be performed while driving with the system, focusing on the system-driver interaction and potential effects of the functions on driver performance.

Defining appropriate questions for assessment by the experts while driving with the system was not an easy task, all experts were encouraged to provide a lot of comments besides selecting one of the answering alternatives they thought best corresponded to their opinion.

Parts of the questionnaire used in moving vehicle are provided below. These are *examples* of questions addressed in the evaluation in moving vehicle. Not all questions could be provided due to the size of the questionnaire.

SPEED LIMIT WARNING FUNCTION

Note: This part of the questionnaire is addressing the speed limit warning function only.

5.0 SITUATIONAL AWARENESS

5.1A What do you think about the function's ability of raising your awareness of speed limits restrictions and your speed, compared to not having the function available? Make a cross in the selected box that best matches your opinion.

Much less aware	☐	☐	☐	☐	☐	☐	☐	Much more aware

Comments:

___

___

5.2A What do you think about the function's ability of raising your alertness in situations where speed limit restrictions change or when your speed is above the allowed speed? Make a cross in the selected box that best matches your opinion.

Much less alert	☐	☐	☐	☐	☐	☐	☐	Much more alert

Comments:

___

___

6.0 USER-SYSTEM INTERACTION

No.	QUESTION	SCORES				
		NA	T	f	F	DN
6.1	The **visual information displayed** is adequate and appropriate for the information to be conveyed					
	Comments: ___ ___					
6.2	The **auditory information issued** is adequate and appropriate for the information to be conveyed					

	Comments: ____________________ ____________________					
6.3	The **combination of visual and auditory information** is adequate and appropriate for providing the information to be conveyed.					
	Comments: ____________________ ____________________					
6.4	If the function can be turned off it provides feedback to the driver about its status					
	Comments: ____________________ ____________________					
6.5	The quantity of information presented at any one time is not excessive					
	Comments: ____________________ ____________________					

6.6 Is there a risk that drivers will deactivate the function by shutting down the system after using it for a while because of its intrusiveness? Make a cross in the selected box that best matches your opinion.

Very high risk	☐	☐	☐	☐	☐	☐	☐	Very low risk

Comments:

____

____

6.7 The **information** is presented **sufficiently in advance** of driving decision. Make a cross in the selected box that best matches your opinion.

Strongly disagree	☐	☐	☐	☐	☐	☐	☐	Strongly agree

Comments:

____

____

8.0 USEFULENESS

8.1 The function is useful. Make a cross in the selected box that best matches your opinion.

Strongly disagree	☐	☐	☐	☐	☐	☐	☐	Strongly agree

Comments:

____

____

8.2 The function is pleasant. Make a cross in the selected box that best matches your opinion.

Strongly disagree	☐	☐	☐	☐	☐	☐	☐	Strongly agree

Comments:

____

____

8.3 The function is assisting. Make a cross in the selected box that best matches your opinion.

Strongly disagree	☐	☐	☐	☐	☐	☐	☐	Strongly agree

Comments:

____

____

11.0 SAFETY RELATED ASPECTS

No.	QUESTION	SCORES				
		NA	T	f	F	DN
11.1	The system does not compromise safety in terms of interaction with other functionalities/systems in the car					
	Comments: ______ ______					
11.2	The information and warnings given by the system is not endangering the primary driving task					
	Comments: ______ ______					

12.0 PERFORMANCE OF FUNCTION

12.1 An occasional missed activation by the function is <u>not acceptable</u>. Make a cross in the selected box that best matches your opinion.

Strongly disagree	☐	☐	☐	☐	☐	☐	☐	Strongly agree

Comments:

______

______

12.2 An occasional incorrect activation by the function is <u>not acceptable.</u> Make a cross in the selected box that best matches your opinion.

Strongly disagree	☐	☐	☐	☐	☐	☐	☐	Strongly agree

Comments:

______

______

12.4 How do you experience the overall performance of the function with respect to timing of warnings? Make a cross in the selected box that best matches your opinion.

Not good at all	☐	☐	☐	☐	☐	☐	☐	Very good

Comments:

___

___

12.5 How do you experience the overall performance of the function with respect to reliability (the consistency in warnings issued, the correlation with real world, correct information/warnings, incorrect information/warnings and missed information/warnings)? Make a cross in the selected box that best matches your opinion.

Not good at all	☐	☐	☐	☐	☐	☐	☐	Very good

Comments:

___

___

12.6 How do you experience the overall performance of the function with respect to accuracy of warnings (timing of warnings, consistency in timing of warnings)?

Not good at all	☐	☐	☐	☐	☐	☐	☐	Very good

Comments:

___

___

D.3 Realization of expert evaluation

The evaluation was carried out in two sessions on the MAPS&ADAS test track south of Hannover on 17th and 19th of October. The participants and their background can be found in Table 15. Each session consisted of two parts. First the experts got the opportunity to examine the system in simulation mode. Then they were allowed to experience the system by driving the instrumented vehicle themselves.

Table 15 : Participants of expert evaluation

Date	Participant	Background
17/10	Juha Luoma, VTT	Human Factors, Accident research
	Susanna Leanderson, VTEC	RD Engineer
	Ann-Sofi Karlsson, VTEC	Human Factors
	Anders Lindgren, VCC	Human Factors
19/10	Somya Joshi, ICCS	Human Factors
	Mathaios Bimpas, ICCS	RD Engineer
	Andreas Gazis, ICCS	RD Engineer

Tests began with an assessment of the general HMI solution in the stationary vehicle. Therefore a test run was simulated, so the actual speed limit on the simulated track was displayed and several warnings appeared according to the position and speed of the simulated driver and the environmental conditions that were simulated as well. Experts had the chance to experience the look and feel of the HMI without having to focus on the actual driving task. The usability of the system was ascertained by questionnaires afterwards.

The actual driving tests were then carried out on the MAPS&ADAS test track south of Hannover, which consists of 48 km of public roads. The test track can be divided into four different sections varying in slope, curvature (gon/km), main speed restriction and road class (Figure 15) hence it includes a big variety of different driving situations (scenarios). Experts were asked to drive the instrumented vehicle (Volkswagen Bus T5, year of manufacture 2006, 128 kW, manual gear box) over one half of a lap on the test track. The experts were free to drive the way they want to and to choose their own speed, so the system could be experienced in a real life scenario and in real life situations. Warnings against speed limit violations and hotspots occurred occasionally. During the test runs notes about unusual situations, false, missing or delayed alarms were taken by an observer.

Afterwards questionnaires especially addressed to experts had to be filled out. The focus was to estimate the reactions of users to the system and the probable user acceptance based on the impression the experts got by driving the instrumented vehicle and the experience the experts have in this field. Besides the test runs experts also had access to all functionality specifications and descriptions and the evaluation plan so that the system behaviour in situations that aren't covered during the test runs could be assessed as well. The independent variables were hence the support by the MAPS&ADAS Driver Warning System and the different road types of the test track. The dependent variables were the impressions of the subjects collected during the test run.

D.4 Results

To sum up, the expert evaluation carried out for MAPS&ADAS, met with most of the criteria identified in section D.2 above. It met with some initial challenges, i.e. the timing of the evaluation with regard to the project; as well as the number of Human Factors experts available. However, at an overall level, the evaluation offered some deep and insightful feedback that could indeed acts as an added value to the extensive user-based evaluation already carried out by MAPS&ADAS. The final diversity of expertise brought forward by the panel was rich, and the design of the evaluation (i.e. orientation, driving on an open track, questionnaires) was sound, as well as representative of real driving conditions.

D.4.1 Material and contents of analysis

There were 7 experts in all, who filled out the 'Checklist HMI Expert Evaluation'. A few data were missing, were experts had apparently overlooked a question or scale. Also, there sometimes was some apparent confusion about the judgments of 'Not Applicable' or 'Don't know'.

The analysis has looked into the following issues:

(a) On what did the experts agree?
(b) If the agreement was on a negative outcome for the system aspect considered, what was the source of their concern, and what does this mean for the MAPS & ADS system in its present form?
(c) On what did the experts not agree, and what does this tell us about the contents and the applicability of the Checklist?

D.4.2 Agreement/disagreement among experts

For each question/scale, a judgment can be given on the level of agreement among the experts, where there are roughly three levels:

- Total, or almost total, agreement
- Some, not extreme, agreement/disagreement
- Total, or almost total, disagreement

The distribution of these three categories over the four separate parts of the Checklist is given in Table 16.

Table 16: Levels of agreement among experts over Checklist parts (in %)

	Total/almost Total agreement	Some (dis)-agreement	Total/almost total disagreement
HMI in stationary vehicle (n=12)	58	17	25
Moving vehicle, speed limit warning function (n=34)	32	32	36
Moving vehicle, hot spot warning (n=36)	33	25	42
Moving vehicle, combined (n=33)	33	25	42
Overall (n=115)	*36*	*26*	*38*

We see that experts agree in a little over 1/3 of their ratings, and that this is mainly the case in the stationary HMI situation. There is real disagreement in the same 1/3 proportion, and this is therefore mainly due to the moving vehicle evaluation.

One could of course say that, for the middle category ('some (dis)agreement') it is a question of whether the bottle is half empty or half full. If this category is split halfway the percentage of agreement would rise to about 50%. However, that is a little dubious: we want experts to agree in order for their judgment to be useful for future improvements to be based on.

On what (negative) aspects did the experts agree?

If experts agree, the agreement can be that the rated aspect of the system is OK or that it isn't. The Table below sorts this out.

Table 17: Disaggregation according to aspects on which experts agreed, in terms of whether the judgment was positive or negative.

	Positive	Neutral/average	Negative
HMI in stationary vehicle (n=7)	57	0	43
Moving vehicle, speed limit warning function (n=9)	78	22	0
Moving vehicle, hot spot warning (n=9)	22	22	56
Moving vehicle, combined (n=10)	30	40	30
Overall (n=35)	*46*	*23*	*31*

NB: Questions 12.1-12.3 are excluded, because they asked a generic question of the expert not to be judged qualitatively for the system

We see that the experts find the speed limit warning function OK, but that there is little enthusiasm for the hot spot warning function and (presumably because of that) for the combined function.

The list of questions/ratings on which the experts fully agreed in their negative judgment is below.

<u>HMI in stationary vehicle</u>

Q 2.2 'The visual images, icons and symbols displayed are easy to understand.'

Q 2.5 'Information presented is consistent with the road network.'

Q 4.3 'Required actions from the driver are clearly stated.'

<u>Moving vehicle, speed limit warning function</u>

No Qs with negative judgments.

<u>Moving vehicle, hot spot warning</u>

Q 6.1 'The visual information displayed is adequate and appropriate for the information to be conveyed'

Q 6.3 'The combination of visual and auditory information provided by the function is adequate and appropriate for providing the intended information to the driver'

Q 9.2 'It is easy to understand why a warning is issued'

Q 10.5 'Is there a risk that drivers will ignore warnings after using the function for a while'? This risk was rated as considerable, on average as 4.9 on a scale of 1-7.

Q 11.2 'The information and warnings given by the function is not endangering the primary driving task'.

Moving vehicle, combined functions

Q 6.1 'The visual information displayed is adequate and appropriate for the information to be conveyed'

Q 6.3 'The combination of visual and auditory information provided by the function is adequate and appropriate for providing the intended information to the driver'

Q 10.5 'Is there a risk that drivers will ignore warnings after using the function for a while'? This risk was rated as considerable, on average as 4.6 on a scale of 1-7.

What do these results mean for MAPS&ADAS?

The speed limit warning function is generally judged as positive by the experts, but the hot spot function (as well the combination) is much less appreciated. If we look at the comments given by the experiments it appears that this is, first of all, because one particular symbol, the guardian angel, is not liked very much. Apart from that symbol itself, there are some worries for higher-order effects, like drivers being distracted or ignoring warnings. These comments provide directions for improving future versions of the MAPS&ADAS system.

What do these results mean for the Checklist?

The result that experts agreed in no more than 1/3 of the questions/ratings should not be interpreted as a straightforward disaster. It could simply mean that there are plenty of silly questions to be eliminated on which experts do not agree, so that this proportion will automatically rise. However, doing so this would probably also mean that we recognize and acknowledge that a number of highly relevant aspects simply do not lend themselves to being captured by expert judgment.

D.4.3 Sources of disagreement

It is, therefore, also informative to look at those aspects on which the experts agreed least, which could lead to a decision either not to include those items, or to replace them by better ones, in future versions of the Checklist. The most extreme instances of disagreement are when there is a dichotomy, that is, when the expert group is exactly divided into two halves having opposite opinions.

Questions/scales for which there was a sharp dichotomy are listed below.

HMI in stationary vehicle

Q 2.1 'The display is located as closely as possible to the driver's line of sight.'

Q 4.1 'It is easy to understand system functionalities and limitations'.

Moving vehicle, speed limit warning function

Q 6.2 'The auditory information issued is adequate and appropriate for the information to be conveyed'

Q 6.2 'The auditory information issued is adequate and appropriate for the information to be conveyed'

Q 6.3 'The combination of visual and auditory information is adequate and appropriate for the information to be conveyed'

Q 6.6 'Is there a risk that drivers will deactivate the function by shutting down the system after using it for a while because of its intrusiveness?'

Q 6.7 'Is the information presented sufficiently in advance of driving decisions?'

Q 9.1 'The information provided by the function is consistent'

Q 9.2 'It is easy to understand why a warning is issued'

Q 11.2 'The information and warnings given by the system is not endangering the primary driving task'

Moving vehicle, hot spot warning

Q 6.2 'The auditory information issued is adequate and appropriate for the information to be conveyed'

Q 6.5 'The quantity of information presented at any one time is not excessive'

Q 7.2 'The auditory information issued is not distracting'

Q 9.1 'The information provided by the function is consistent'

Moving vehicle, combined functions

Q 6.5 'The quantity of information presented at any one time is not excessive'

Q 10.3 'Is there a risk that drivers are disturbed by the system while having passengers in the car?'

Q 11.2 'The information and warnings given by the system is not endangering the primary driving task'

Q 12.4 'How do you experience the overall performance of the system with respect to timing of warnings?'

What do these results mean for MAPS&ADAS?

By itself, this type of result (disagreement among experts) can have no implications for the design of the MAPS&ADAS system.

What do these results mean for the Checklist?

The questions/scales listed above as leading to the sharpest possible disagreement within the expert group, a dichotomy, are important ones in the evaluation of any system. It is hard to see how they could be rephrased so that experts would agree, i.e., it is not a matter of how the questions are worded. Rather, this should lead to the conclusion that we need different methods to find the truth. In particular, it is to measurements of driver reactions and/or driving performance that we should resort to in those cases.

D.4.4 Conclusions

Conclusions for MAPS&ADAS

On the basis of these results we can conclude that there is room for improvement in the HMI of the MAPS & ADS system, in particular, the hot spot warning function.

Conclusions for the Checklist

The experts agreed on roughly 1/3 of the items. That is on the low side. One could argue about what this means, but it makes us realize that in complex areas – like driving behaviour – one can only go that far. After that, and knowing that we presently have no models of driving behaviour that permit us to make sensible predictions, we will have to resort to actual behavioural

Annex E Framework for the assessment of preventive safety functions

This is a standalone document, which is the updated version of D16.3.