HoliDes

Holistic Human Factors **Des**ign of Adaptive Cooperative Human-Machine Systems

D5.4 - Techniques and Tools for Empirical Analysis Vs1.5 incl. Handbooks and Requirements Analysis Update

Project Number:	332933
Classification:	Public
Work Package(s):	WP5
Milestone:	M3
Document Version:	V8.0
Issue Date:	July 09, 2015
Document Timescale:	Project Start Date: October 1, 2013
Start of the Document:	Month 19
Final version due:	Month 21
Deliverable Overview:	**Main document**: Techniques and Tools for Empirical Analysis Vs1.5 incl. Handbooks (Public) **Annex I**: Common Metrics Methodology and Requirements Update (confidential) **Annex II:** Analysis of requirements (confidential) **Annex III**: HoliDes Integration Plan (confidential)
Compiled by:	Elisa Landini - REL
Authors:	Morten Larsen (AWI) Marco Botta (UTO) Fabio Tango (CRF) Simona Collina (SNV) Flavia De Simone (SNV) Svenja Borchers (TWT) Denis Martin (TWT)

	Marieke Thurlings (TWT) David Käthner (DLR) Stefan Griesche (DLR) Klas Ihme (DLR) Harald Kolrep (HFC) Juliane Bürglen (HFC) Linda Onnasch (HFC) Paul Kaufholz (PHI) Adam Herout (BUT) Maros Raucina (HON) Zdenek Moravek (HON)
Reviewers:	Stefan Kaufmann (IAS) Caterina Calefato (IRN)
Technical Approval:	Jens Gärtner, Airbus Group Innovations
Issue Authorisation:	Sebastian Feuerstack, OFF

DISTRIBUTION LIST		
Copy type[1]	Company and Location	Recipient
T	HoliDes Consortium	all HoliDes Partners
E	ARTEMIS	Patrick Vandenberghe

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

RECORD OF REVISION		
Date (DD.MM.JJ)	Status Description	Author
07.04.2015	Initial Version	Elisa Landini
29.04.2015	Share structure with partners and request of contributions	Elisa Landini, David Käthner
29.05.2015	Collection of contribution from all MTT owners	Elisa Landini, all MTT owners
12.06.2015	Integration of further changes	Klas Ihme, Flavia De Simone
16.06.2015	Inclusion of integration plan and metrics methodology	Elisa Landini
24.06.2015	Sent to the internal reviewers	Elisa Landini
29.06.2015	Integration of feedbacks after the internal review	Elisa Landini, all MTT owners
01.07.2015	Final version	Elisa Landini

Table of Contents

List of Figures

List of tables

List of abbreviations

AdCoS	Adaptive Cooperative Human-Machine System
HF-RTP	Human Factors - Reference Technology Platform
ICA	Index of Cognitive Activity
MTT	Method, Technique and Tool
TY.X	one of the tasks of work package Y
UML	Unified Modelling Language
WP	work package

1. Introduction

1.1. Objectives and structure of the document

The objective of the document is to provide a general overview of the status of the activities conducted in WP5 since the last deliverable (February 2015).

In particular, after the first project review in December 2014, an intense activity has been performed to identify relevant metrics and subsequent activities to reflect on the feasibility of reaching the overall project goals. This activity has been described in details in Chapter 2 and in Annex I (respectively for the public and confidential content).

By starting from the results of this methodology and the feedback provided by the AdCoS owners, new requirements have been identified and old requirements have been reviewed in Chapter ☐ and Annex II (respectively for the public and confidential content).

A Common Integration Plan has been drafted in D1.5 to be applied for the integration of all MTTs into the HF-RTP. The description of this methodology has been included in Chapter 4, while its actual application for the MTTs in WP5 has been included in Annex III.

Finally, the objective of D5.4 was also to provide an update of the status of all MTTs developed in WP5 (compared to the version 1.0 described in D5.3). This update is described in Chapter 5.

Chapter 6 describes the conclusions of these activities and the next steps that will be carried out in the next months.

2. Common Metrics Methodology

At the beginning of the HoliDes project, we set out to develop methods, techniques and tools to increase the level of adaptiveness in Cooperative Human-Machine Systems on the Health, Aeronautics, Control Rooms and Automotive domains in order to improve the performance of these systems, in particular in terms of safety and confidence of human operators.

Therefore, if we want to evaluate the reaching of this objective (i.e. the overall HoliDes objective), we must apply quantitative measures to verify that

- **The MTTs developed in the project can support the increase of adaptiveness.**
- **The adaptiveness can bring benefits to the AdCoS compared to the existing systems.**

In the design phase of an empirical investigation, operationalization [2] is the process of defining the measurement of a phenomenon that is not directly measurable, because its existence is indicated by other phenomena. We have adapted this approach to define a methodology for the evaluation of the project goals in the different domains by identifying concrete and measurable metrics and subsequent activities to reflect on the feasibility of reaching the goals of each AdCoS and, as a consequence, reaching the overall project objective.

Even though the objectives of the AdCoS have been already described in previous deliverables (in particular in the WP6-9 deliverables that describe the uses cases), so far they have not been explicitly formalized in terms of goals that can be achieved only by increasing the degree of adaptiveness of the AdCoS [3].

[2] Shields, Patricia M., and Nandhini Rangarajan. A playbook for research methods: integrating conceptual frameworks and project management. New Forums Press, 2013.

[3] We defined the "degree of adaptiveness" as the number of independent variables that can change the behaviour of the AdCoS to adapt to the context

Moreover, in order to prove that we reached the main goals of the project, we have also to quantitatively demonstrate that the MTTs developed in the project had actually supported the increase of adaptiveness and, as a consequence, that they acted as enablers for the development of innovative systems (whose performances are improved compared to the existing systems already on the market).

Therefore, we defined a methodology that includes a common evaluation framework to:

1) **Quantify** the achievement of the project goals by measuring the **performance of the AdCoS**.
2) Define and refine a set of concrete **requirements to develop innovative AdCoS** with increased degree of adaptiveness.
3) Define a set of **metrics (indicators and success criteria)** to actually quantify the achievements of these requirements and, as a consequence, to measure the benefits brought by the adaptiveness.
4) Define and refine a set of concrete **requirements to develop MTTs** that enable the improvement of performance of the aforementioned AdCoS.
5) **Map the set of MTTs to the AdCoS' needs and goals**.

2.1. Methodology steps

The methodology foresees the definition of 4 macro steps, in order to fill in the table shown in Figure 1:

Step 1:

- **Goals**: Definition of the main goal(s) of the AdCoS
- **Problems**: problems that, so far, prevented the development of the AdCoS. They can include any technological and/or methodological barrier.
- **Solutions**: solutions that would address the aforementioned problems.
- **Adaptiveness**: description of the (increased) adaptiveness of the AdCoS.

Step 2:

- **Design**: requirements for the AdCoS in order to actually implement the solutions previously defined, and requirements for the MTTs that can facilitate and support the design and development of these solutions.
- **Evaluation**: requirements for the MTTs that can facilitate and support the evaluation of the solutions previously defined.

Step 1

					Metrics (indicator + success criteria)		
Goal		G1	Main goal of the AdCos				
Problems		P1.1	Problem 1 that, so far, prevented the development of the AdCos				
Problems		P1.2	Problem 2 that, so far, prevented the development of the AdCos				
Solutions		S1.1	Potential solution to the problem 1				
Solutions		S1.2	Potential solution to the problem 2				
Adaptiveness		A1.1	(Increased) adaptiveness of the AdCos				
Requirements	Design	R1.1	AdCos requirements 1 for the design	AdCos	Metric 1	Module	Brief description of the module that address the AdCos requirement 1 and how it contributes to reach the goal of the AdCos.
Requirements	Design	R1.2	AdCos requirements 2 for the design	AdCos	Metric 2		
Requirements	Design	R1.3	MTT requirement 1 for the design	MTT	yes/no	[Method | Technique | Tool]	Brief description of the MTT that addresses the MTT requirement 1 and how it contributes to reach the goal of the AdCos.
Requirements	Evaluation	R1.4	MTT requirement 2 for the evaluation	MTT	yes/no	[Method | Technique | Tool]	Brief description of the MTT that addresses the MTT requirement 2 and how it contributes to reach the goal of the AdCos.

Step 2 | Step 3 | Step 4

Figure 1: Table for the collection of the information to apply the metrics methodology

Step 3:

- **Metrics (indicators and success criteria):** Definition of concrete metrics (i.e. indicators and corresponding success criteria) to quantitatively measure that the requirements of the AdCoS are met [4].

Step 4:

- **MTTs/modules assignment:** by starting from the MTT and AdCoS requirements, MTTs and modules [5] (i.e. tools used within the AdCoS) developed in HoliDes will be assigned to the requirements. A brief description will be also provided to show how they contribute to the goal of the AdCoS.

The table can be navigated in both directions (from the top-goals to the bottom-MTTs and vice versa) to retrace the rationale either for the selection of the MTTs according to the initial goal and needs, or the solution according to the potential of the MTTs (to understand which issue each MTT can address).

2.2. Application of the methodology to the domains

The methodology has been iteratively applied to 4 AdCoS in order to test it with the AdCoS owners:

- Safe Parallel Transmit Scanning (WP6 - Health)
- Airport Diversion Assistant (WP7 - Aeronautics)
- Border Security Control Room (WP8 - Control Room)
- Adaptive Assistant (WP9 - Automotive).

[4] Since the preliminary objective of the methodology is to demonstrate the improvement of the performances of the AdCoS due to the application of the MTTs, this stage will not be used to evaluate the performance of the MTT (thus the metrics will be used only for the requirements of the AdCoS). Once version 1.0 of the methodology is complete, it will be extended and applied to the design and evaluation of the MTTs.

[5] A clear distinction has been made between tools and modules: tools are software to be integrated into the toolchains of the AdCoS owners to improve their development process and deal with lifecycle data such as requirements, test cases, analysis results and behavioural models, while modules are software that are meant to be used as part of the AdCoS.

At this stage, the main aim of the application of the methodology was to verify its effectiveness to correlate the project objectives with the AdCoS objectives, and to provide a means to qualitatively and quantitatively demonstrate the benefit of the MTTs for each AdCoS.

Therefore, even though the number of metrics identified per AdCoS is limited (e.g. 3-4 metrics per AdCoS), the methodology has clearly shown the ability to support the AdCoS owners in the definition of concrete indicators and success criteria to evaluate the performance of their AdCoS' (in addition to the ability to support them to identify suitable MTTs to address their needs and facilitate the evaluation of the AdCoS').

Annex I describes in details the results of this activity in terms of flows of information from the goal to the selected MTTs per AdCoS.

2.3. Lesson learned

During the project, in WP4 and WP5, we have already carried out activities for the definition of metrics to evaluate the AdCoS.

These activities started by considering the requirements defined by the AdCoS owners at the beginning of each cycle. However, even though these requirements have been defined and refined according to the SMART (Specific, Measurable, Acceptable, Realizable and Traceable) principle, they did not provide a bridge towards the project goals. Moreover, most of these requirements did not show either a clear connection with the (increased) adaptiveness of the AdCoS or with the MTTs that would be necessary to achieve a requested degree of adaptiveness.

Therefore, we started an intense discussion with the AdCoS owners to clearly state their goals and trace how they are going to select and use the MTTs to reach them. The methodology is the result of this intense collaboration, where we iteratively refined the structure of the excel file for the collection of the information (that reflected the overall methodology) and the AdCoS owners provided their feedback by applying it.

2.4. Next steps

In the next versions of the methodology (that will be detailed in the next WP5 deliverables), we plan to:

- Increase the number of metrics identified for each AdCoS
- Extend the methodology to all AdCoS
- Define specific plans for the empirical evaluation of each AdCoS (to support the AdCoS owners in the Tx.8 (where x = [6|7|8|9]).
- Analyse the goals to extract similarities (i.e. cross-domain goals), in order to identify similarities in the problems to be address and, as a consequence, in the solutions (i.e. MTTs to be used). Example: in automotive and aeronautic domain we monitor the operator (driver/pilot). How are we creating the operator's model? How can we evaluate the model? Do we need to involve real users and collect data? Which HF data are we planning to collect?

2.5. Conclusions of the Common Metrics Methodology

With the Common Metrics Methodology we have defined not only metrics to evaluate the AdCoS, but also the metrics to define the reaching of the overall project objectives.

Moreover, the methodology provides a way to clearly identify (and justify) the MTTs used in each AdCoS by starting from the goals that drive the research performed in the project. In fact, by filling in the table, we actually mapped the MTTs to the AdCoS goals (and so to the project goals), thus we qualitatively justify the selection of a set of MTTs for each AdCoS, and quantitatively demonstrate the benefit each MTT brings to the AdCoS (in terms of reaching its goals).

In other words, we created a *fil rouge* between the project goals and the AdCoS developed in WP6-9 and the MTTs developed in WP1-5 [6].

[6] We always include WP1 among the MTT provider because we consider the HF-RTP as a sort of macro-MTT, i.e. a means to facilitate the development and evaluation of our AdCoS.

Finally, the methodology provides the rationale for the selection and the use of some MTTs according to the specific and concrete needs of each AdCoS. From this point of view, it can be regarded to as a prototype concept of the Platform Builder developed in WP1, because, by starting from a need (the description of the "problem"), we can navigate the excel file up to the solution (i.e. the MTT).

According to the feedback received by the AdCoS owners, the methodology was effective to:

- Formalize the goals of the AdCoS owners that can be achieved only by increasing the degree of adaptiveness of the AdCoS
- Identify concrete metrics to evaluate the performance of the AdCoS and, as a consequence, the achievement of the HoliDes project objectives
- Describe in an intuitive and clear way the rationale for the selection of some MTTs for each AdCoS and their benefits.

3. Analysis of requirements

The requirements addressed by the method, techniques and tools developed in WP5 are listed in Annex II (the confidential part of this deliverable).They also include the new requirements identified by applying the Common Metrics Methodology (only for the MTTs used in the 4 selected AdCoS).

4. Common Integration Plan

The integration of MTTs into the HF-RTP will be based on a detailed plan and methodology defined in D1.5 section 2.4.4 "HF-RTP integration methodology". In order to define this methodology, it was important to distinguish between Methods/Techniques and Tools (as shown in Figure 2).

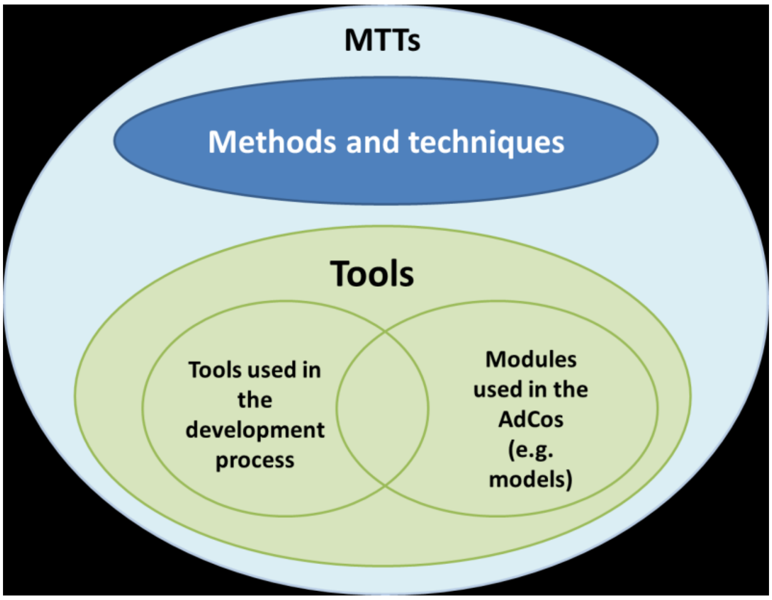

Figure 2: MTTs – methods, techniques and tools

The methods and techniques that have been developed in the HoliDes project (and in particular in WP5) have been delivered in the form of innovative algorithms, procedures and guidelines for the different phases within the AdCoS development process in the different domains.

Most of the methods and techniques are not delivered as a piece of software. Therefore, they cannot be actually integrated into the HF-RTP and no interoperability with the other tools can be achieved. However, the outcomes of these methods and techniques will be gathered in documents that will be inputs to other MTTs.

Therefore, the integration of the methods and techniques (including guidelines and design of empirical experiments) will be guaranteed by uploading the documents into MTTs which are part of the HF-RTP. An example of such a tool is the HF Filer tool developed in WP5.

Moreover, they will be easily accessible by using the Platform Builder developed in WP1 that allows finding the most suitable method and technique according to the specific need they are meant to address.

As regards the tools, two different categories of software have been developed:

- The tools that are meant to be integrated into the toolchains of the AdCoS owners to improve their development process and deal with lifecycle data such as requirements, test cases, analysis results and behavioural models (e.g. Magic-PED of OFF for the task modelling).
- The tools that have been developed to be embedded into the AdCoS to improve its functionalities (e.g. the Driver Distraction Classifier of TWT).

In order to avoid any confusion between them, from now on the latter will be identified as "modules" instead of "tools".

Moreover, some of these pieces of software present a double nature: according to the context of use, they can be either used as tools in the development process or as modules of an AdCoS (e.g. a module for the detection of the distraction can be included in the AdCoS to improve its functionalities and it can be employed as a tool for the evaluation of an AdCoS).

For the tools, the integration implies the development of specific piece of software (i.e. adapters) to implement the OSLC specification and allow the sharing of data with other tools.

A similar approach can be also applied to the modules: they do not deal with the lifecycle data management but they share data with other pieces of software within the AdCoS. In particular, most of these modules share real-time data with RTMaps for the implementation of the AdCoS.

Therefore, even though they will not be provided with OSLC adapters, a common methodology will be defined in order to easily share real-time data with RTMaps between multiple AdCoS, by exploiting the interoperability features of RTMaps as common exchange mechanisms, and the HoliDes Meta Model to define the structure of the input and output data.

It will allow reusing the same module within different AdCoS (i.e. to make the modules interoperable across different AdCoS and domains).The overall methodology for the tools and modules foresees the following activities to be integrated into the HF-RTP:

1) Identify their inputs and outputs, related MTTs or AdCoS. Moreover, only for tool, identify their compliance with OSCL and the estimated date for integration into HF-RTP.
2) Identify the correspondence between its inputs and outputs and the concepts of the HoliDes meta model.
3) Only for tools: develop the adapters or parsers between the concepts and the data managed by the MTT (parse numerical values and strings into RDF and from RDF to numerical values and strings).

In this particular deliverable, we will focus on Step 1 of the described methodology. The work done so far for the integration plan will be described in the Annex III HoliDes Integration Plan.

5. MTTs developed in WP5

Table 1 provides an overview of the MTTs developed in WP5.

Table 1. MTT landscape of WP5.

Partner	MTT	Application domains	Description of usage
AWI	Modelling of AdCoS from a means-ends perspective	Guided patient positioning (Health Care domain – WP6)	Modelling technique that can be used at different stages of the AdCoS design process, during system development and during the validation or evaluation phase.
AWI	HF Filer	Energy Control Room AdCoS (Control Room - WP8) HMI of the Lane (automotive - WP9) Change Assistant developed by CRF All AdCoS in the Health Care Domain - WP6 Border Security AdCoS (Control Room - WP8)	Tool to • Records the results of human factors evaluation activities • Makes the evaluation data accessible from the RTP • Provides traceability of human factors data
PHI	U-DAT: User test – Data Acquisition Tool	Guided Patient Positioning VCG Triggering (Healthcare - WP6)	Tool for structured collection of feedback and scores from user tests in different phases of an AdCoS design process.
BUT/HON	Operator state detection from implicit hand gestures	Airport Diversion Assistant AdCoS (Aeronautics - WP7)	Module to detect the state of the operator according to his/her gestures and facial expressions
BUT/HON	Detection of operators' head orientation	Airport Diversion Assistant AdCoS (Aeronautics - WP7)	Module to detect the activities the operator has carried out as well as any potential mode confusion according to his/her head

			orientation.
DLR	Methods and techniques for the driver adaptive parameterization of a highly automated driving system	AdCoS Adaptive HMI and AdCoS Adapted Assistance (Automotive - WP9)	Empirical method to conduct experiments for the collection of data to design a system able to adapt the vehicle driving style to the individual driver's preferences.
DLR	CPM-GOMS Task Analysis of a Lane Change for manual and automated driving	AdCoS Adaptive HMI and AdCoS Adapted Assistance (Automotive - WP9)	Method for task analysis to compare manual driving with automation-support driving in a lane change situation.
DLR	Theatre Technique for acceptance tests and systems variants exploration during AdCoS design	AdCoS Adaptive HMI and AdCoS Adapted Assistance (Automotive - WP9)	Method to support the collection of feedback and expectations of the human operator with respect to an adaptive system early in the design process.
DLR	Surrogate Reference Task (SuRT)	AdCoS Adaptive HMI and AdCoS Adapted Assistance (Automotive - WP9)	Tool to simulate visual distraction as caused by executing a secondary task during driving.
HFC	Human Factors Methods and Metrics for HF and Safety Regulations	Healthcare AdCoS: 3D acquisition, patient positioning, or safe patient transfer (HealthCare - WP6)	Method and guidelines to identify evaluation criteria and metrics for four selected evaluation criteria ("HF issues": usability, distraction, situation awareness, mental workload).
HFC	HF-Task Analysis Tool	Healthcare AdCoS: 3D acquisition, patient positioning, or safe patient	Tool to support task analysis procedures

		transfer (HealthCare - WP6)	
SNV	Empirical analysis and validation methods of cognitive and communicative processes in the Control Room domain	Energy Network Control Room (WP8)	Design and conduction of empirical experiment to collect data on communication processes in a control room.
SNV	Empirical analysis and validation methods of cognitive and communicative processes in the Automotive domain	AdCoS Adapted Assistance (Automotive - WP9)	Design and conduction of empirical experiment to collect data on distracted drivers.
TWT	Driver Distraction Classifier (DDC)	All AdCoS in the Automotive Domain - WP9	Module to classify the level of distraction of the driver
UTO/CRF	Classifier of driver distraction based on data on vehicle dynamics	AdCoS Adapted Assistance (Automotive - WP9)	Module classify the distraction of the driver

5.1. Modelling of AdCoS from a means-ends perspective (AWI)

5.1.1. Summary

The means-ends modelling for AdCoS can be used at different stages of the AdCoS design process, during system development and during the validation or evaluation phase. During the former, it may be used to identify parts of the human-machine interaction in which an increase or decrease of adaptation and/or cooperation is useful. During the latter stage, means-ends modelling can identify potential causes of errors produced by unclear or weakly defined states of automation, adaptation or cooperation.

The product of the technique is a hierarchical (often tree-structured) model of the goals that the operator should meet and the means available in the AdCoS and the controlled system to achieve those goals. The higher-level goals will tend to be more abstract, as they aggregate the results of the underlying, more concrete physical tasks.

5.1.1.1. Means-end analysis for AdCoS design

When used in the design phase, the method will provide functional modelling to help analyse high-level functions that should be supported by the AdCoS.

These high-level functions should be defined in such a way as to be independent in their description from the subsystems and operator procedures that are used to provide these functions.

5.1.1.2. Means-end analysis for AdCoS evaluation

When used in the evaluation phase, the method is an aid to aggregate data.

The method uses the functional modelling from a means-ends perspective to help analyse whether a given AdCoS supports the operator in achieving key goals in the operation of the system.

The modelling will therefore transform measured and observed data into information about high-level system state, thus allowing system evaluation to be reported in terms closer to operational goals, rather than at the lower action level.

5.1.1.3. Overview of the analysis method

In order to guide the analysis, the functionality of the AdCoS is classified according to a system that is a specialisation of the Goals-Functions-Components classification originally found in Lind, 1994 [[7]]:

1. Goals (Why)
2. Functions (What)
3. Behaviour (How)
4. Structure (How)

The words "why, "what" and "how" summarise the type of information along the functional hierarchy.

A goal describes *why* certain work is being done, i.e., the rationale behind a task or a sequence of tasks.

The level in the middle, the functions, describe *what* is being done, but in abstract terms. The level of abstraction to describe the functions should use terms that are independent of the nature and design of the underlying user behaviour (actions) or system structure. This is also referred to as a solution-independent description (where the actions and (sub)systems used to implement a task are seen as a solution).

The behaviour and structure (sometimes also referred to as (sub)systems or components) provide the information about the concrete means to provide a function. An example of that could be an imaginary task defined as "Press the button marked 'up'". In this little example behaviour as well as (physical) structure are referred to (the act of pressing and the actual button to press).

To make the analysis easier to map to the physical world (controlled entity and the AdCoS itself), we add *objectives*, as specialised, verifiable subgoals. As such, they still fall into the "Why" category, but should be described at a level of abstraction that corresponds to actual, verifiable states of controlled entity or AdCoS. This brings us to the final classification used in this method:

---

[7] Lind, M. Modelling Goals And Functions Of Complex Plant, AAAI '94, The Twelfth National Conference On Artificial Intelligence, 1994. http://www.iau.dtu.dk/~ml/aaai94.pdf

1. Goals (Why)
2. Objectives (Why)
3. Functions (What)
4. Behaviour (How)
5. Structure (How)

5.1.2. AdCoS Use-Cases

This method will be applied to the Guided patient positioning AdCoS of WP6.

The method will provide a structure to match the actions of the operator and the components available to the operational goals of the patient positioning procedure.

5.1.3. Input and output

The method requires that one or more purposes can be associated with the operator's tasks, in such a way that goals at a higher level of abstraction can be identified.

The more concrete subgoals (the objectives) need to have specific criteria associated, and the modeller will need to have access to these criteria, or be able to determine them independently. They are dependent on the AdCoS, the use case and the test scenario.

In most cases, the data will be observations (or log data from the AdCoS) recorded in a format that can be used to verify if a given goal has been achieved.

In certain cases, it will be possible to ascribe numerical levels of goal achievement to a given situation. An example of this could be how closely a driver follows a lane on a road, or how closely an operator is keeping the equipment within its safety margins.

As an output of applying the method, the AdCoS designers will get two different types of information about the functionality of the AdCoS being designed.

For each of the two main uses of the method, the output can be summarised as:

1. Identify system functionality by purpose
   - Help design system and UI
   - Help identify opportunities for adaptivity in the system
2. Map evaluation observations to goal trees
   - Help identify consequences in the system in case of a "failed" evaluation
   - Help structure feedback in the design process

The first type of information can be used in the system design phase to identify areas of functionality that can be achieved or implemented in different ways.

If these different ways include both human tasks and automated control, then the area is a potential target of adaptive task allocation in the AdCoS.

The second type of information will help identify which branches of their user goal tree are being achieved by the operators, and which are not. The intention of this structuring of the information is to guide the re-design phase in a cyclic design process.

5.1.4. Current status and functionality

The Means-end modelling has been used to produce simple hierarchical task models as proof of concept of this approach, with a simple model of the Guided patient positioning use case from WP6 (defined in deliverable 6.1, WP6_HEA_MRI_UC02) and associated AdCoS as the example.

The model produced can be seen in Figure 3. It splits the activities in branches regarding the actual position of the patient's arms and legs, the communication with the patient and the technical aspects of the configuration of the equipment.

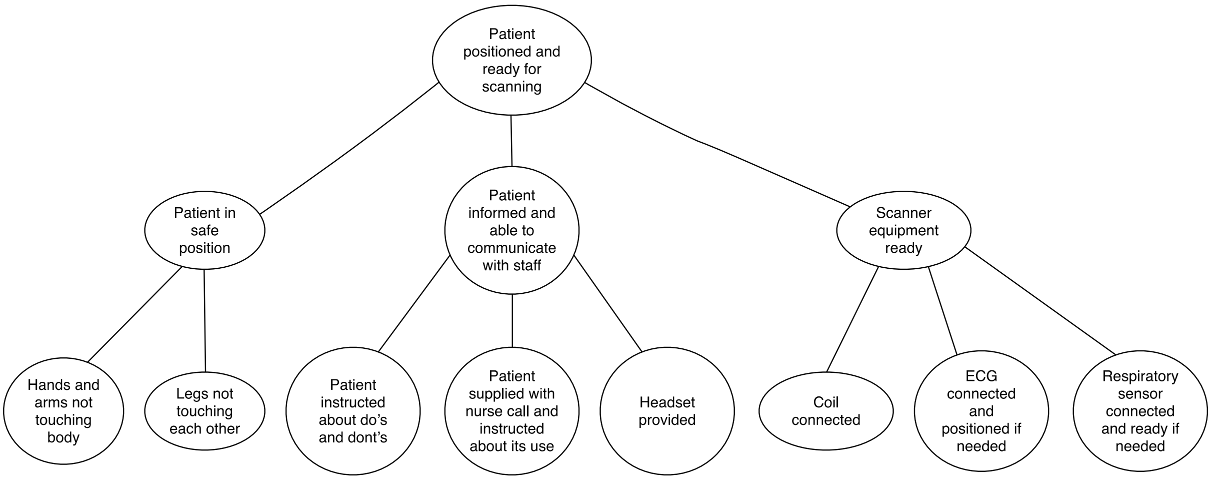

Figure 3 A simple means-end analysis of the Guided patient positioning use case from the healthcare domain.

This structuring of the tasks will help organise the data as well as the reporting of any feedback in the design phase. For instance, if an evaluation of the operation of the equipment shows that the examination takes longer time and produces imaging of varying quality due to communication problems with the patient during a breath hold scan, then the "information and communication" branch of the AdCoS needs more work. On the other hand, if the problem is due to a problem with the respiratory sensor use or positioning, then the "equipment ready" branch of the AdCoS requires more work.

5.1.5. Integration of the outputs of the MTT

The integration plan of the means-ends modelling technique will be compliant to the Common Integration Plan defined in D1.5 and reported in Chapter 4. In particular, we will upload the documentation with the description of the method and its results onto the HF-RTP as soon as the overall HF-RTP infrastructure will be completed.

Table 4 shows in details the steps as well as the estimated date to complete the integration.

Table 2: integration plan for the means-ends modelling technique

	Activity	Date
1	First version of method description	29/5/2015
2	Feedback from partners	31/8/2015
3	Model representation	31/1/2016
4	Documentation and upload on the HF-RTP	30/9/2016

5.2. HF Filer (AWI)

5.2.1. Summary

HF Filer is a simple tool that aims to achieve the following points:

- Record the results of human factors evaluation activities
- Make the evaluation data accessible from the RTP
- Provide traceability of human factors data

In the following, the generic term "evaluation" is used, but the tool will also file and track validation and verification data, provided that they are textual in nature.

The tool will support the human-factors-related workflow around a given RTP instance as outlined in the figure below, with some example tools and methods. HF Filer will make it possible to provide the functionality marked in red for models and tools that do not provide a means to file human-generated evaluation results. As such, it works as connection for MTTs that normally rely on human-authored reports as their "output" of the evaluation cycle to the HF-RTP.

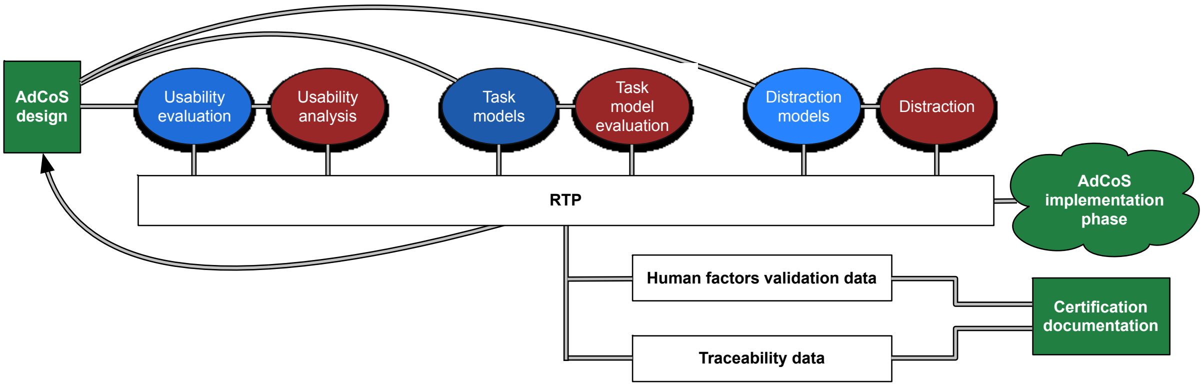

Figure 4. Schematic representation of the scenario with RTP compliant human factors tools in the HF-RTP.

It should be underlined that the intention is to develop a technically simple tool, along the line of a database program to record human factors data, and it should be seen as a proof of concept of a generic human factors tool.

The interest in providing traceability of human factors validation (or evaluation) data stems from certification procedures requiring the manufacturer to provide traceability data.

The tool is intended to be the duct tape of the HF-RTP, in that the many MTTs that normally lead to a human factors expert writing a textual evaluation document need a connection to the HF-RTP that allow to read the evaluation data out of the tool for use in the design feedback loop and possibly certification.

As such, it is intended to be used in conjunction with methods and techniques that would otherwise lead to the production of a text document, or with tools that support a human factors expert's analysis of a given aspect, but do not support the filing of the results in a manner compliant with OSLC.

5.2.2. AdCoS Use-Cases

The HF Filer is being developed to handle evaluation data from mainly healthcare AdCoS use cases, with a specific eye to use cases where task analysis or usability evaluation is employed.

It is currently being considered for the Safe parallel transmit scanning use case from WP6 (defined in deliverable 6.1, WP6_HEA_MRI_UC03) to log the user interface evaluation work. However, in general, the tool is not restricted to any specific domain.

5.2.3. Input and output

The tool is read-only from OSLC, and as such is intended to take its input from the web interface, in which the human factors expert can enter and edit an evaluation plan and subsequently enter the results for the various evaluation reports.

HF Filer is a tool that allows the filing of human factors evaluations for several development projects simultaneously, and for each project several evaluation reports can be created and filled. This work is all undertaken in the web-based user interface (UI) of the tool itself.

Exposure to the HF-RTP through OSLC is read-only, as a means to integrate the evaluation results into the workflow, reporting, etc., but not currently to be able to file evaluation results in the tool through OSLC.

HF Filer is organised around simple data structures as illustrated in Figure 5 in two main strands – the evaluation plan and the relative report.

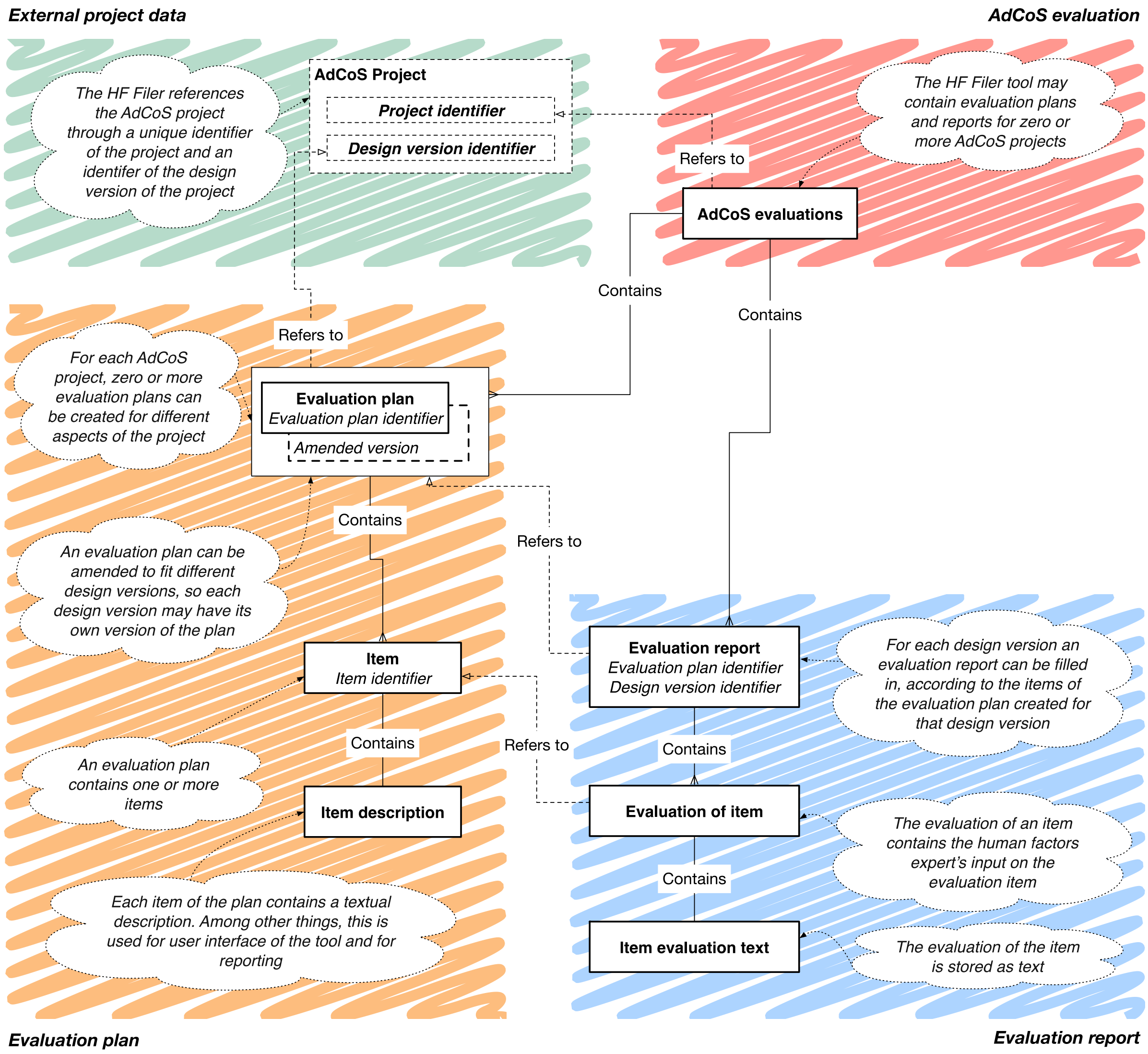

Figure 5. Annotated diagram of the data structure as seen from the RTP.

To allow linking the reports with the project, it uses a project identifier, which must be shared by other tools in the RTP instance and a version string to identify the various versions of the design in the design and test cycle. This string can for instance be "0.1", "0.2", etc.

Internally, the tool allows items of the evaluation plan to be inherited from one version to the next, but, being a read-only interface, this is transparent to the OSLC interface.

Hence, the structures appear to the RTP like outlined in Figure 5.When querying the tool, a hierarchical query structure can be used, beginning from the top level, down to the items.

For the evaluation plan, the sequence of identifiers that can be used to access the data in HF Filer (i.e., the "empty form", without any evaluation data filled in) is the following:

[AdCoS project identifier]

>> [Evaluation plan identifier]

>> "plan"

>> [Design version identifier]

>> [Item identifier]

>> [Item description]

As an example, consider a project with the following data (*please note that the final format of the item identifiers has not been decided yet):*

Table 3. Identifiers to read HF Filer data

Data path element	**Identifier**	**Comment**
Project identifier	projectx	Project name
Evaluation plan identifier	data_input_evaluation	Human-readable identifier
Data type identifier	plan *or* report	Fixed strings
Design version identifier	0.1	Version string
Item identifier	i001	Machine-generated identifier
Item description	description	Fixed string
Item evaluation	evaluation	Fixed string

With a REST-like interface to the tool, a query string to access the description of the first item of the Data Input Evaluation plan would look like this:

https://hffiler/projectx/data_input_evaluation/plan/0.1/i001/description

For the evaluation report, the sequence of identifiers that can be used to access the data in HF Filer (i.e., the "form filled in" with the textual evaluations) is the following:

[AdCoS project identifier]
>> [Evaluation plan identifier]
>> "report"
>> [Design version identifier]
>> [Item identifier]
>> [Item evaluation]

With a REST-like interface to the tool, a query string to access the evaluation of the first item of the plan would look like as follows, in line with the previously shown query string for the plan:

https://hffiler/projectx/data_input_evaluation/report/0.1/i001/evaluation

5.2.4. Current status and functionality

The HF Filer is currently developed as a beta version with a functioning web interface to create evaluation plans and evaluation reports.

Currently, the tool supports a single version, while the version with multiple version identifiers is under development.

Feedback from partners has led to the following extra features to be considered for the next version:

Feature	**Rationale**	**Status**
Add the option to tag an evaluation with elements from the HF ontology	This will allow a precise categorisation of HF issues identified during the evaluation phase	Waiting for a more stable ontology definition
Provide single running instances of HF Filer for each interested partner	This will make it easier to handle privacy issues regarding user observations	Waiting for partners to express a specific need.
Allow other tools to pre-populate evaluation reports	This will allow the HF Filer to be used to aggregate and report evaluation data collected with other tools.	Part of the U-DAT toolchain development effort. Currently waiting for more precise specs on data format.
In evaluation plans, add reference to the evaluation methods used.	This will allow the tool to be work more as a stand-alone solution for reporting.	Under consideration. The original intention was to be a part of a toolchain. Stand-alone operation was not included.
Add the option to indicate if an evaluation item has been "passed the test" or not.	This will allow the tool's output to be used in project management software to indicate if a certain user validation has been concluded successfully or not.	Awaiting more information about the relevant development use cases.

The OSLC interface is under design. There is no direct set of OSLC specifications available for Human Factors work, and it has therefore been decided to first develop a proof of concept interface to test the use of OSLC with the tool, and meanwhile define the needs for a human factors format for OSLC.

The proof-of-concept version will be based on the Change Management specification, based on the availability of test platforms and its ability to contain a challenge-response workflow.

5.2.5. Integration plan

1) The integration plan of the HF-Filer tool will be compliant to the Common Integration Plan defined in D1.5 and reported in Chapter 4. In particular, the integration of this tool will be accomplished by completing the following steps:Identify the inputs and outputs of the tool (described in Annex III), its compliance with OSCL and the estimated date for integration into HF-RTP.
2) Identify the correspondence between its inputs and outputs and the concepts of the HoliDes meta model.
3) Develop the adapters or parsers between the concepts and the data managed by the MTT

The details of the integration plan are described in the Annex III (confidential part).

Table 4 shows in details the steps as well as the estimated date to complete the integration.

Table 4: integration plan for the HF-Filer tool

	Activity	**Date**
1	Prototype for beta testing	6/4/2015
2	Feedback on prototype	15/4/2015

3	OSLC Bridge	15/6/2015
4	Data format description (preparation for IOS)	31/8/2015
5	Feedback on data format description	31/10/2015
6	First version of IOS	31/2/2016
7	Test client (lyo-based)	15/7/2015
8	OSLC Connector allowing other tools to read data from HF Filer	31/1/2016
9	Data exchange specification for data from U-DAT	31/3/2016
10	Final version of IOS	31/5/2016
11	Module to read U-DAT data	31/5/2016
12	Documentation	30/9/2016

5.3. U-DAT: User test – Data Acquisition Tool (PHI)

5.3.1. Summary

The U-DAT tool can be used for structured collection of feedback and scores from user tests in different phases of an AdCoS design process.

Throughout an AdCoS design process user tests can be performed to get qualitative feedback on the usability of tasks. The qualitative feedback can be used to improve the design and hence the usability score of the product.

For products that are subject to safety standards this tool can also collect data on the safe and effective use of tasks that have been identified as critical, or find critical tasks that may lead to a hazard.

Changing the design can now mitigate tasks that may lead to a hazard and considered a safety risk. The required application of usability engineering to medical devices is described in the IEC standard 62366 [8].

In line with this standard all tasks in the final AdCoS design that are part of risk mitigation need to be validated for safe and effective use. U-DAT can also be used to capture the required data for such a so-called summative usability evaluation. Typically, you need to have 15 end-users or more in the product evaluation to ensure safe and effective use.

The tool needs a Task Model and Scenarios to define the user test. It will collect information on the usability scoring, qualitative feedback per sub-task and a root-cause analysis for fails on critical tasks.

5.3.2. AdCoS Use-Cases

This tool can be applied to the Guided patient positioning AdCoS of WP6. Also other AdCoS can make use of this tool.

The created task analysis for the Guided patient positioning AdCoS can be used as input to U-DAT. After performing user tests, the raw data collection can be abstracted from the U-DAT output.

5.3.3. Input and output

The input task model covers all needed attributes from a task analysis such as e.g. described in Software for Use [9]. An example task map is depicted in Figure 6. A user role performs a number of 'essential tasks' – depicted in orange. Essential tasks are high-level tasks that do not imply a solution direction, but define the intention of the user. Additionally, so-called 'concrete tasks' – depicted in blue - can be part of the model, which are based on a solution direction. They define the concrete steps by which the intention can be carried out. Different arrow types model relationships between tasks like "include" (normal arrow) or "optional" (dotted arrow) sub-tasks.

---

[8] IEC 62366-1:2015, Medical devices – Part 1: Application of usability engineering to medical devices

[9] L. Constantine & L. Lockwood, Software for use, ISBN 0-201-92478-1

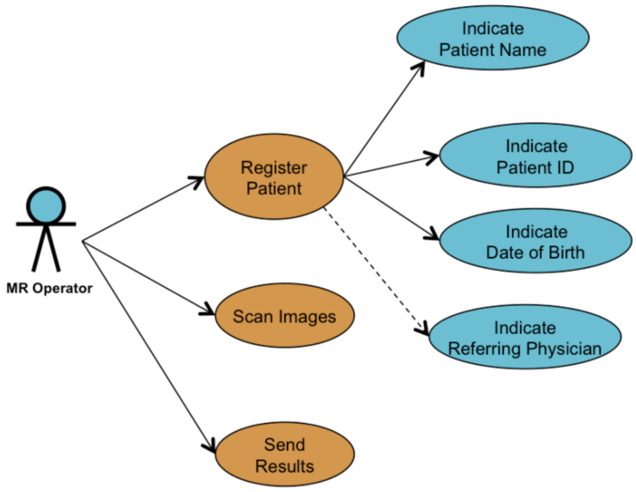

Figure 6: An example of Task Map

In addition to the tasks from the task map other attributes are important to be captured as input to U-DAT:

- User goal: each task is defined by its user goal
- Criticality: specific for medical devices: if a task is identified as critical it requires special attention and might be subject to summative evaluation
- Description: all additional information that can help understanding the task

Another input of U-DAT is scenarios. Scenarios are narrative stories that describe one specific flow through the task model. They are use as carrier for the user tests.

As regards the outputs, the U-DAT tool captures the information from a user test per participant in a structured way. Apart from the relevant information on the participant, per task or step tested, the tool captures a usability score. The scoring mechanism can eventually indicate whether a task has passed or failed. Qualitative feedback is captured for any task that does not succeed at first attempt. For failing tasks a root-cause analysis is captured that will be used to iterate the design solution.

Possibly usability categorizations can be added to the task scoring. This can enable to differentiate the scoring of task per category. For example, differentiate for effectiveness, efficiency, satisfaction, learnability, etc.

5.3.4. Current status and functionality

Currently the tool supports all characteristics as described in the previous sections. Entering the right input model, and eventually analysing and summarizing the results is done manually. As such, it can be used for performing user tests for AdCoS Guided patient positioning from WP6.

U-DAT is being further developed to support a formal abstraction level. The right abstraction level will enable the definition of a clear input and output interface by means of an agreed upon (UML) structure. Once that is in place, U-DAT can connect to other tools from the HF-RTP. For example other tools might support the creation of a structured task model that can be used as (UML) input to U-DAT. Vice versa, U-DAT can create a structured output format that can be used by other tools to create summary reports or further analysis.

5.3.5. Integration plan

The integration plan of the U-DAT tool will be compliant to the Common Integration Plan defined in D1.5 and reported in Chapter 4. In particular, the integration of this tool will be accomplished by completing the following steps:

1) Identify the inputs and outputs of the tool, its compliance with OSCL and the estimated date for integration into HF-RTP.
2) Identify the correspondence between its inputs and outputs and the concepts of the HoliDes meta model.
3) Develop the adapters or parsers between the concepts and the data managed by the MTT

The details of the integration plan are described in the Annex III (confidential part).

Table 5 shows in details the steps as well as the estimated date to complete the integration.

Table 5: integration plan for the U-DAT tool

	Activity	Date
1	First version of tool description	29/5/2015
2	Feedback from partners	31/8/2015
3	Model representation (UML interface)	31/1/2016
5	Matching between the HoliDes Meta Model concepts and the input/output.	As soon as the HoliDes Meta Model is available
6	Documentation	30/9/2016

5.4. Operator state detection from implicit hand gestures (BUT/HON)

5.4.1. Summary

In human to human communication non-verbal cues such as hand gestures can transmit valuable information [[10]]. Especially implicit gestures, i.e., gestures that are not intended to be specialised (such as in gesture-based interaction), carry semantic meaning. For example, the operator might reach towards a knob in the cockpit, but not push it for a while. Similar to a human conversation partner, the system might benefit from knowledge that does not necessarily lead to system input but still indicates important information about the operator's state of mind.

Videos of the operator (pilot, driver, etc.) during task accomplishment will be recorded. Computer vision techniques will be used to enable automated analysis of the video sequences.

[10] H. Lausberg and S. Kita, "The content of the message influences the hand choice in co-speech gestures and in gesturing without speaking", *Brain and Language,* vol. 86, no. 1, pp. 57-69, 2003.

Validation of a prototype aims at revealing flaws in the design with respect to how easily and efficiently the user uses the prototype. In addition to the objective data (e.g. number of failures compared to a baseline), information about the operator's state is usually obtained ex-post and often it is compromised by subject forgetting or being influenced by the course of the experiment.

Real-time state inference provides valuable information about how the operator uses the prototype:

- How natural it is for him to use the prototype interface (searches for elements, retracted inputs etc.)
- What confidence he has in the information provided by the interface (cross-checks, uncertainty patterns etc.)
- How the prototypes affects the ergonomics (gestures for re-focusing, annoyance etc.)

Such information can identify weak design elements or modes of use.

In addition, this tool bears the potential to be used online to classify the pilots' implicit hand gestures not only during testing of a prototype, but also during everyday interaction with the AdCoS. In that case, fatigue and attention can be estimated and the output of the tool can be used to adapt the degree of automation of the AdCoS.

5.4.2. AdCoS Use-Cases

The tool will be used in connection with Airport Diversion Assistant (DivA) AdCoS (Aeronautics - WP7) to support design development and also real-time pilot state assessment. In early development phase the cockpit based camera will provide design-related information, later pilot state information. The computer vision techniques described below have a high relevance in this AdCoS, because the freedom of motion of the subject is high. At the same time, for safety, regulatory and practical reasons, it is not possible to use structured light-based approaches, such as infrared depth sensors similar to Microsoft Kinect.

5.4.3. Input and output

Whole-body videos of the human operator while accomplishing the task are needed as input for the tool. Videos need to be stored in a way that enables linking them to certain system states, e.g., inputs from the user to the system. Thus multimodal data integration and synchronization needs to be guaranteed.

As regards the outputs, the tool provides a continuous description of the gestures of the operator over time of system usage. The outputs are signals in time when an (implicit) gesture occurs, possibly with a short latency (up to 2 seconds) induced by the processing.

5.4.4. Current status and functionality

In our solution, we aim at implicit gestures performed in aeronautic cockpit. In particular, we study pilot's implicit gestures connected with controlling of selected important cockpit elements: Control wheel, Touch screen, Navigation control panel, Throttle lever, Electronic flight bag. We define three levels of interaction with a particular element: Full interaction, Touch-and-Go and Unfinished. Full interaction is a long and rich interaction where the pilot fully grasps or touches given element and works with it for a period of time. Touch-and-go is interaction where pilot slightly touches an element and moves to another gesture. Unfinished interaction includes the cases where the pilot proceeds to interact with an element, but the beginning gesture is interrupted and not even the touch-and-go class of interaction is carried out. According to this description, our implicit gestures connected with aeronautic cockpit elements can be defined as a Cartesian product: cockpit element and level of interaction. Some combinations of cockpit element with level of interaction can be neglected, because they do not occur in practice (e.g. wheel control and unfinished).

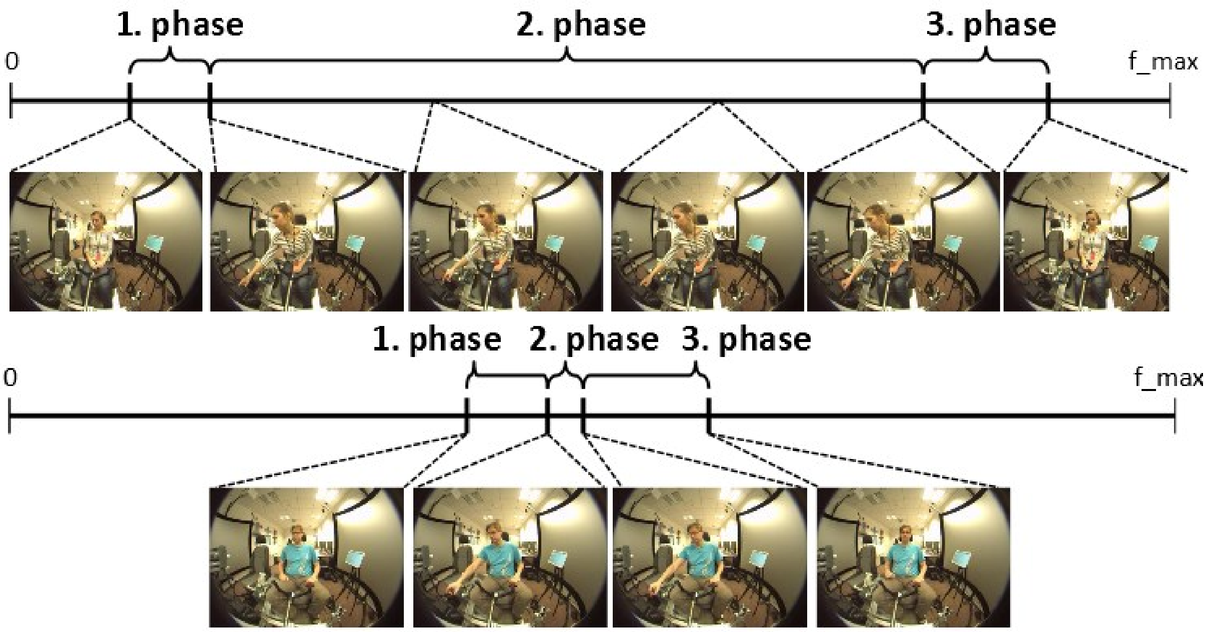

Figure 7: Phases time course of implicit gestures connected with touch screen.

Each gesture can be divided into three phases. In the first phase, pilot is moving his hand towards a particular cockpit element and this phase ends when his hand stops. The second phase is the phase of ongoing interaction with the cockpit element and in the third phase, the pilot's hand is moving back from the incriminated cockpit element and the gesture is being finished. For each level of interaction, each phase can happen during a different span of time as can be seen in Figure 7.

We focus on recognition of transition between phases of implicit gestures when interaction with cockpit element starts/ends. For unfinished interactions, we mean the moment when the pilot stops his hand because of changing his mind to interact with chosen cockpit element and the moment when he returns his hand back. These are key moments defining the temporal span of all interactions.

Our method assumes that implicit gestures are composed of a sequence of pilot's poses in time. First, we recognize 10 important upper body joints: head, shoulder centre, right/left shoulder, right/left elbow, right/left wrist, right/left hand. For human pose estimation, we use Pose Machine. The principle of Pose Machine is shown in Figure 8. Pose machine is a hierarchical method consisting of multiple stages. Each stage is modelled by multiclass random forest with j classes ($j = 10$ joints of upper body).

Output r_k of previous stage k is used as a part of the input of next stage $k+1$ via context features as an added set of features to the original feature vector x_z (features computed for image patch at position z in input image). These context features are intensities of probability for each joint in patch at position z and offset vectors from position z to K highest peaks in each probability map $r_k^1, \dots, r_k^j$. These context features help to learn relationships among body parts during the training process, which improves pose estimation accuracy for each body part.

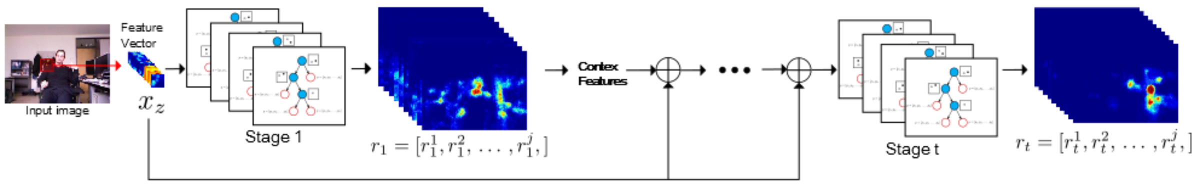

Figure 8: Human pose estimation by hierarchical inference machine named Pose Machine.

However, this human pose estimation method is not perfect and there are many cases where detection is wrong, i.e. the global maximum of certain probability map is at a wrong position. In order to compensate for this imperfection, our system for implicit gesture recognition (Figure 9) takes $n+1$ frames $f, f+k, \dots, f+nk, f+(n+1)k$ with a step k and for each frame i human pose estimation (HPE module) is applied (Pose Machine is used). It then produces a probability map r_i^j of joint position for each joint j (i.e. not only the single recognized best position). In the next step, all probability maps r_i where $i = 1, \dots, n+1$ are subsampled by local max operator. Having an input image $I(x)$, the operator computes maximal value locally$M(x) = \max_{p \in \Omega(x)} I(p)$, where $\Omega(x)$ is a local surroundings of position$x$, e.g. a 13×13 pixels block. This image $M(x)$is adequately subsampled, e.g. 7×7 times. These resulting subsampled probability maps s_i of all frames r_i from sequence $i = 1, \dots, n+1$ are concatenated into a single feature vector $x = [s_1, s_2, \dots, s_{n+1}]$ (patch of features with $(n+1)j$ feature channels) and this vector is used as input for recognition by random forest.

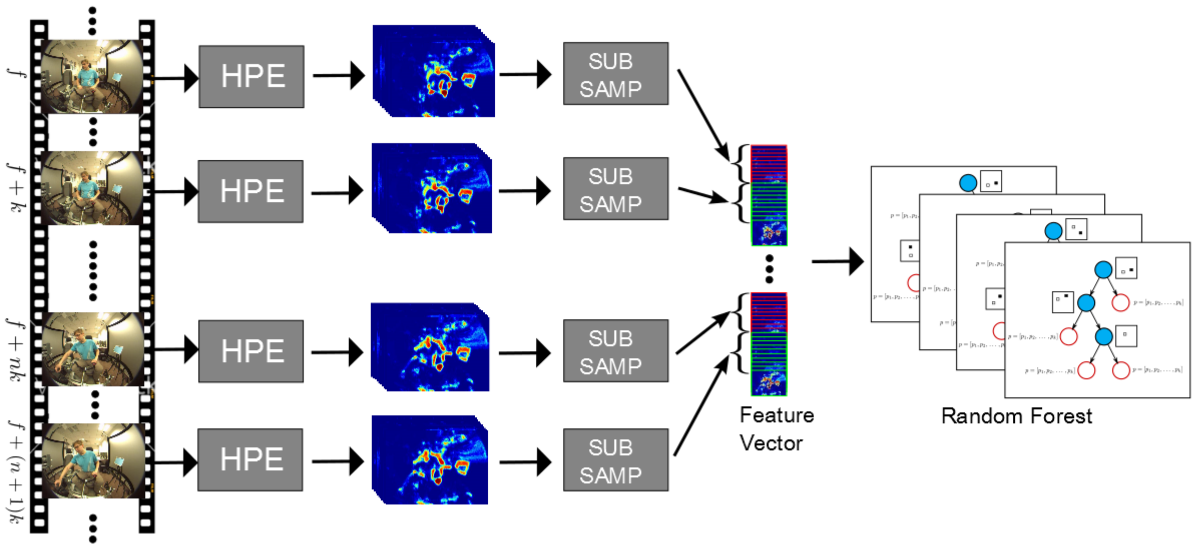

Figure 9: Our approach to implicit gesture recognition.

For each transition between two phases of a certain implicit gesture, a binary random forest is trained, where positive input vectors are generated from frames forming transition between phases of particular gesture (last frame in the vector is situated in this transition between two phases of implicit gesture) and negative input vectors are generated from other parts of video sequences. Two types of split rules are used in nodes of these random forest trees. In the first rule, the intensity of subsampled probability map s_k^i at position q is compared against threshold. In the second rule, the intensities of subsampled probability map of joint i between different frames k and l are compared at position q against threshold. The type of the split rule is chosen randomly during the training process.

The performance of this system for implicit gesture recognition is evaluated on the implicit gesture database collected as part of the project work. We collected a gesture dataset in Honeywell flight simulator. The gesture set contains implicit gestures connected with these cockpit elements: Control wheel, Touch screen, Navigation control panel, Throttle lever, Electronic flight bag. We deal with 12 implicit gestures connected to the 5 cockpit. All of these implicit gestures were demonstrated in cockpit simulator by 10 people and all of them were recorded by a camera situated in front of the pilot. Frame examples captured by the camera are shown in Figure 10. Three video sequences were recorded per 1 person.

Figure 10: Sample frames from the implicit gesture dataset collected in the cockpit simulator.

The first video sequence of each person separates the gestures fully - before starting a new gesture, the subject returns hands to the idle position on their lap. The second and third sequence contains gestures executed one immediately after the other, in random order. The third phase of one gesture can therefore overlap with the first phase of the following one. The resulting dataset is composed of these video sequences, yielding 30 video sequences of 10 people, where each video sequence contains each implicit gesture demonstrated once. These video sequences have 35 fps and resolution 1080×780 pixels. Division into training and testing set is as follows: video sequences of 8 people (24

video sequences) are included in training set and the testing set consists of video sequences of two people (6 video sequences).

The human pose estimation based on Pose Machine was trained on the dataset of different people poses performed in office environment, as well as in a simulator cockpit. This dataset consists of frames where one seated person was recorded at a time from an approximately frontal view. These frames contain poses with various backgrounds (from solid to very complex/cluttered), various lighting conditions (from dark, stable lighting, through constant office lighting, up to backlit by a window), and the people wore different clothes. The resulting dataset contains 6213 pose frames of 24 people.

Our experiments performed on implicit gesture recognition by proposed method shows that the highest responses manifest not only in the location of the given level of interaction (full, touch-and-go, unfinished), but also on the other ones. However, these high peaks are always related to the same cockpit element (e.g. navigation control panel). This behaviour is not malignant, because the borderline between the levels of interaction is unclear. A positive feature is that the gestures are detected very accurately in time - at single frame precision. Another positive result is that the unfinished gesture can be also reliably detected, without the hand even actually reaching the region of interest. The responses of one gesture detector are constantly low for other parts of the videos, including the other gestures. The recognition of gestures is thus very stable and reliable. The detector behaves very similarly for all other video sequences and all gestures from the set defined earlier.

5.4.5. Integration plan

So far the software has been used as module, therefore its integration will be compliant to the steps for the modules of the Common Integration Plan defined in D1.5 and reported in Chapter 4. In particular, its integration will be accomplished by completing the following steps:

1) Identify their inputs and outputs towards the AdCoS
2) Identify the correspondence between its inputs and outputs and the concepts of the HoliDes meta model.

The details of the integration plan are described in the Annex III (confidential part).

Table 6 shows in details the steps as well as the estimated date to complete the integration.

Table 6: integration plan for the Operator state detection module

	Activity	**Date**
1	First version of module	29/5/2015
2	Feedback from partners	31/8/2015
3	Production version of the module operating in real time	31/3/2016
4	Identification of the correspondence between its inputs and outputs and the concepts of the HoliDes meta model.	As soon as the HoliDes Meta Model is available
5	Documentation	30/9/2016

5.5. Detection of operators' head orientation (BUT/HON)

5.5.1. Summary

Orienting attention towards new locations is normally accompanied by reorientation of the head direction. When interacting with an AdCoS, the operator may orient towards other locations in the work environment (e.g. a navigation system in a car) which can indicate distraction from the main task. Thus, automatically detecting these head movements provides valuable information about the operator's current focus of attention and possible distraction. Videos of the operator's head (pilot, driver, etc.)

during task accomplishment are recorded and computer vision techniques are used to enable automated analysis of the video sequences.

Deriving knowledge about the human operator is valuable in the system validation phase. Despite the limited detection ability of a video recording, the tool can provide valuable information related to operator's visual focus. The applicability of such approach in design phase and real-time use of the tool will be evaluated in comparison with traditional methods (eye-tracking, questionnaires).

In addition, this tool is used in real time to detect the likelihood of missing significant information in the environment. Based on the head direction, the elements in the (aeronautic and-or automotive) cockpit can be identified as being or not being in the primary focus. If an element with important information does not get in primary field of view, it is considered as missed and the system should adapt to regain attention.

5.5.2. AdCoS Use-Cases

The tool is being tested in use in connection with the Airport Diversion Assistant (DivA) AdCoS (Aeronautics - WP7) to support design development and also real-time pilot state monitoring. In the first case, the applicability of the tool for the design development is being evaluated. See below for results in the lightweight flight simulator.

In the second case, the tool is being used to detect situations where adaptation of DivA AdCoS should be triggered to deliver missed information to a pilot in real situation.

5.5.3. Input and output

Videos of the human operator's head while accomplishing the task are used as the input for the tool. The head movement is being recognized in the input video and its direction is estimated. Based on a three-dimensional map of the controlling environment (Figure 11), the system is trained to recognize the control device the user interacts with or is likely to interact.

As regards the outputs, the tool provides a continuous description of the operator's head orientation over time of system usage. The head orientation is determined by pitch and yaw angles with confidence level of the estimation. A valid output is a "null output" signalling that the operator head is not visible, out of recognizable range or the operator is not present in the scene/cockpit. The outputs are sampled in time, multiple samples per second. This internal output is mapped to the functional parts of the working environment (Figure 11).

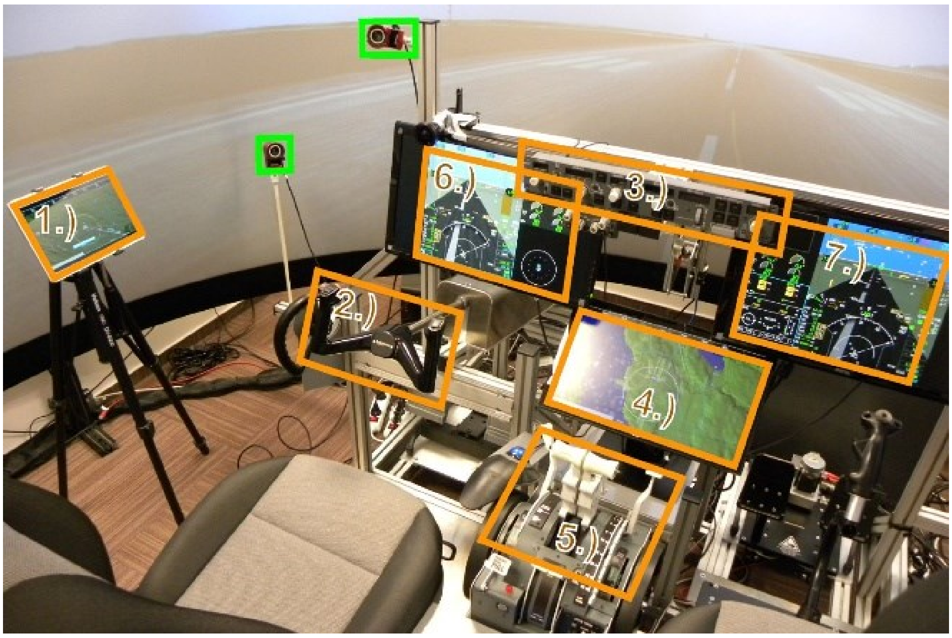

Figure 11: three-dimensional map of the controlling environment

5.5.4. Current status and functionality

A video dataset of head movement has been collected in the Honeywell flight simulator. Four subjects participated in the recording. Subjects being positioned on the pilot's seat performed the following head movements:

1. turn head from left to right and from right to left
2. turn head from up to down and from down to up
3. turn head in a circular motion clockwise and counter-clockwise (mimicking scanning the whole cockpit – overhead panel, left and right instruments and out-of window, pedestal)

4. free movement of choice (each subject different)

Video was recorded using Panasonic HDC-TM700 camcorder. Figure 12 shows several sample frames.

Figure 12: sample frames from the head movement dataset

The ground truth was collected at the same time using OptiTrack motion tracking system. Five Flex 3 motion tracking cameras were places around the cockpit, in front and on the sides of the subject. The subjects wore a TrackClipPRO headset on the left side of his/her head. It has three NIR LEDs that emit light detectable by Flex 3 cameras. During the recording, head position and orientation were computed at real time and logged into text files by the Motive software (optical motion capture software that is a part of the OptiTrack system).

For constructing the recognizer of the subject's head pose, we used the Random Forests, trained on the Annotated Facial Landmarks in the Wild (AFLW) dataset. This dataset contains over 24,000 real-world images of faces in various poses gathered from Flickr. These images are annotated with up to 21 facial landmarks and with head pose, described by roll, pitch, and yaw.

We used random regression forests to estimate the head pose in the images. The random forest algorithm should provide high accuracy, while, at the same time, be fast enough to process video in real time. From the AFLW dataset, 19,386 images were randomly chosen for training and the rest, consisting of 5,000 images, was used for testing.

To find the features providing the best results, we experimented with different feature extractors to be used as the input for the random forests. In particular, the intensity of pixels of grayscaled images, channel features by Dollár et al. (2009), and Gabor features (Daugman 1985) were tested. Figure 13 shows that the best results are obtained by using Gabor features (acquired by convolution with Gabor filters): the value on Y axis represents the fragment of samples with error less or equal to the X value, e.g. for yaw angle and blue curve, 62% of test samples had the error smaller than 0.2 radians. The graphs compare the results for Gabor features (blue curve), channel features (yellow curve) and pixel intensities (red curve), from which the Gabor features achieved the best results.

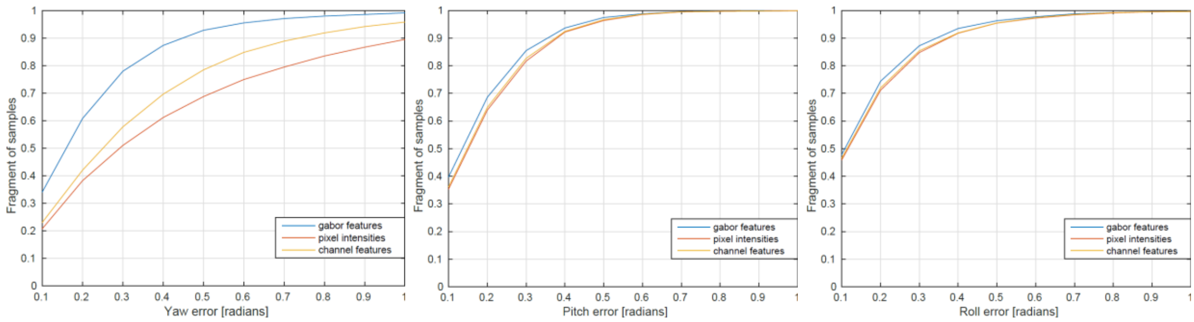

Figure 13: Cumulative histogram of error for yaw, pitch, and roll angles.

In order to increase the precision, several (in our case 50) trees are trained, each on a subset of randomly chosen training samples. The final pose for the test sample is then computed by averaging the results from all trees in the forest. An example of the results for each tree and the final pose is shown in Figure 14, where the green circle represents the ground truth value, each of the cyan `+' symbols represent the result from one tree and their mean (the final estimated pose) is pictured as a blue circle.

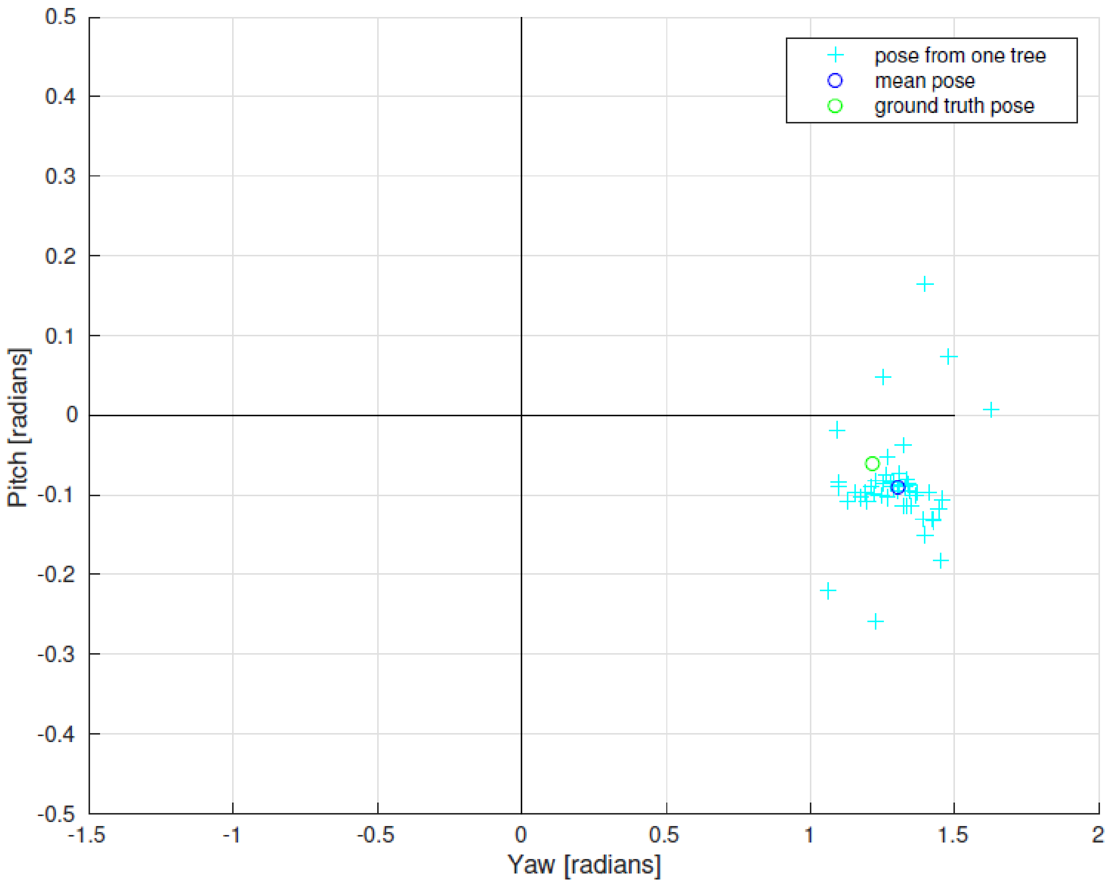

Figure 14: Resulting yaw and pitch angles for one image.

Figure 15 shows the distribution of error for the yaw angle. While the results for the near-frontal head poses are satisfactory, the results for head poses closer to the profile pose were significantly worse. This was mainly due to smaller number of training samples for the near-profile poses (yellow curve in the graph). We tried to solve this problem by not choosing the subset of samples for each tree purely randomly. Instead, we divided the training set into bins based on the samples' yaw angle value. The range of yaw value was 0.1 radians for each bin. Then, we randomly selected the same amount of samples from each bin (or less, if the bin did not contain enough samples), achieving more balanced distribution of the training samples for each tree. The result is, again, visible in Figure 15 (blue curve), where the yellow curve represents distribution of the training samples, blue curve represents the error when the training samples were selected randomly, and the red curve represents the error on samples with more evenly distributed yaw. Although the precision for the poses very near to the frontal pose decrease, for most poses, the results improve and the overall error decreases.

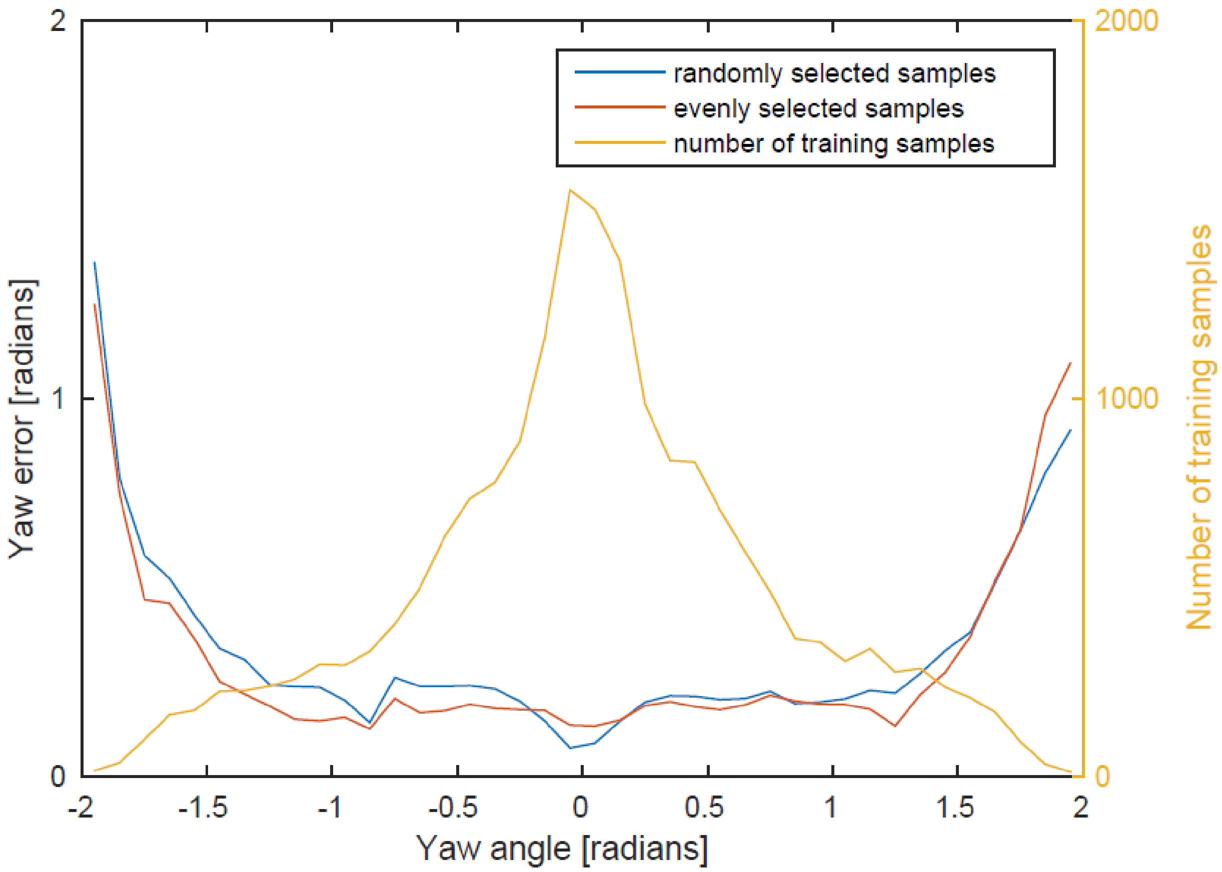

Figure 15: Error distribution over the range of yaw angle.

The yaw angle results for one video, together with ground truth data from the OptiTrack, are shown in Figure 16. The length of the video was 69 seconds and the framerate was 50 fps. The framerate of the OptiTrack system, which was used do collect the ground truth data, was approximately 11 fps, therefore the video frames between the OptiTrack frames were discarded and the data were synchronized. The *X* axis represents the sequence number of the frame and on the *Y* axis is the yaw angle. Occasional errors and inaccuracies are partially due to imperfect localization of the head. Some errors (for example in the area between frames 220 and 320 in the figure) are caused by heavy rotation of the head in the roll or pitch angle, due to which the head detector was not able to detect the head properly and even if the head was successfully detected, there were not enough training samples with these unnatural poses and the random forest was not able to provide reliable results.

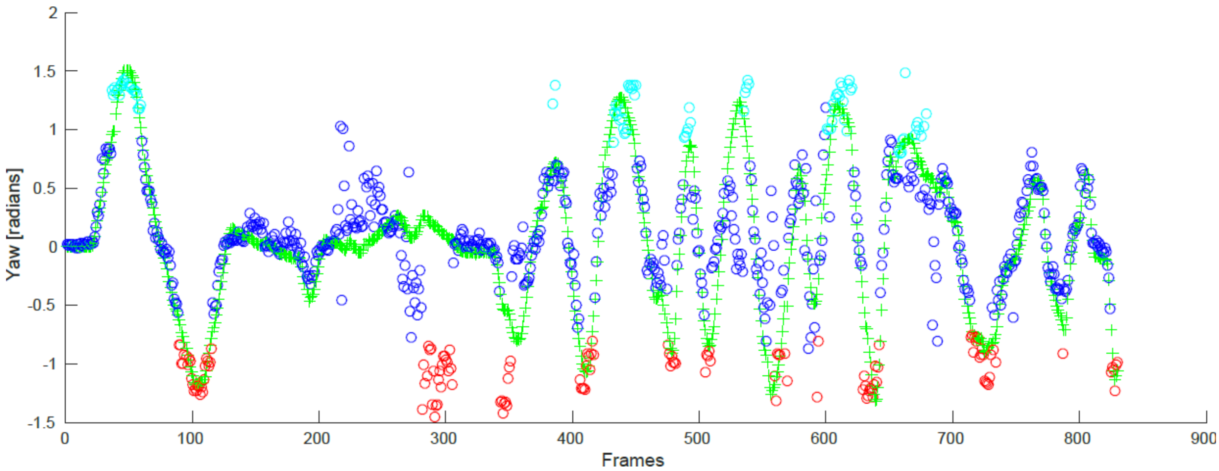

Figure 16: The ground truth (green '+') and estimated values of yaw angle (blue, cyan and red circles) for one video.

One of the main reasons for gaze tracking in the airplane cockpit is to estimate whether the pilot has or has not looked at certain information. To simulate this, we created a video dataset to establish missed event detection reliability. The dataset was made in office environment and each video was approximately 1 minute long. The test subjects were seated in a chair and were instructed to occasionally (5-10 times per video) look at a display positioned by their right hand. We annotated the frames in which the person in the video was looking at the display, basically dividing the frames into two subsets, one subset containing frames where the person was looking at the video, the other subset containing the rest.

We defined approximate yaw and pitch angle values for vector from the head to the display (angles that the head would have if the person looked straight at the display without tilting his eyes). Then, for each frame, we computed the dot product of two vectors (one vector with the previously defined yaw and pitch to the display and one with the pitch and yaw estimated by the random forests). We omitted roll, as it should not be relevant for our task. If the dot product was higher than a given threshold, it was assumed that the person was looking at the display. We compared these results to the annotations and computed miss rate and false positives.

The fraction of false positives and miss rate given the threshold is in Figure 17. For example, for threshold 0.96, the miss rate was 5.48% and 12.13% of the frames were falsely identified as positives. Certain errors were introduced due to inaccurate annotations, as it is often not clear whether the person is already (or still) looking at the display at the beginning/end of the annotated phase. If we ignored 3 frames at the beginning and end of each section, the results improved to miss rate being 4.98% and the false positive rate 9.54%. Also, the test subjects were allowed to move their heads, resulting in the change of angle to the display.

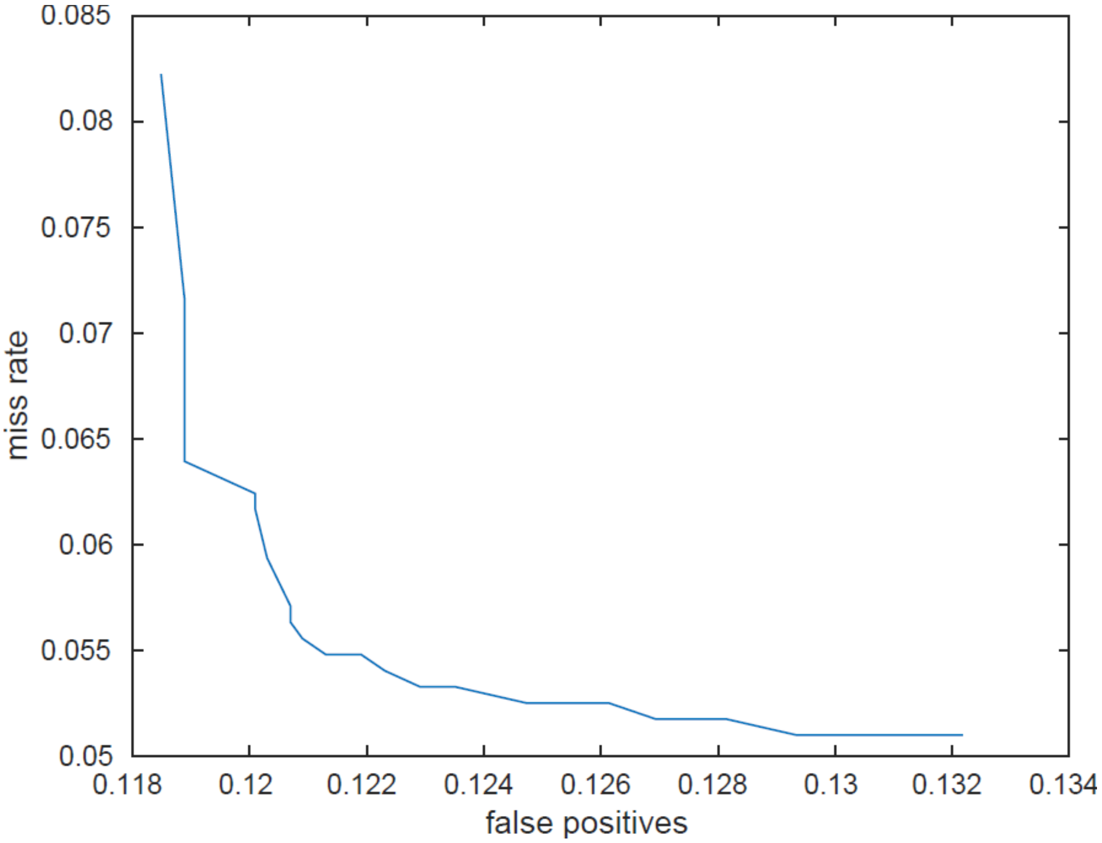

Figure 17: Comparison of miss rate and false positives rate based on the threshold for distance from the reference vector.

5.5.5. Integration plan

So far the software has been used as module, therefore its integration will be compliant to the steps for the modules of the Common Integration Plan defined in D1.5 and reported in Chapter 4. In particular, its integration will be accomplished by completing the following steps:

1) Identify their inputs and outputs towards the AdCoS
2) Identify the correspondence between its inputs and outputs and the concepts of the HoliDes meta model.

The details of the integration plan are described in the Annex III (confidential part).

Table 7 shows in details the steps as well as the estimated date to complete the integration.

Table 7: integration plan for the operators' head orientation detection module

	Activity	**Date**
1	First version of module	29/5/2015
2	Feedback from partners	31/8/2015
3	Production version of the module operating in real time	31/3/2016
4	Identification of the correspondence between its inputs and outputs and the concepts of the HoliDes meta model.	As soon as the HoliDes Meta Model is available
5	Documentation	30/9/2016

5.6. Methods and techniques for the driver adaptive parameterization of a highly automated driving system (DLR)

5.6.1. Summary

For the IAS highly automated driving demonstrator, DLR will provide a functionality to adapt the demonstrator's driving style to the individual driver's preferences. If possible, this preference shall be predicted using the driver's own driving style using manual driving data.

As described in D5.3, three research questions guide our effort:

1. How can we identify the driving style of individual drivers based on data from manual driving alone?
2. How can we measure preferences for the automation's trajectories and decisions before and during a lane change?
3. How can we predict these preferences from the individuals' driving styles?

To answer question one, a second experiment has been conducted in DLR's Dynamic Driving Simulator to validate the results from the first experiment which was carried out using a low fidelity fixed base simulator. The data from the second experiment are currently being analysed. A third experiment is in preparation to answer questions two and three. It will be conducted in July and August this year and documented in D5.5.

5.6.2. AdCoS-Use Case

This method will be applied to the automated overtaking manoeuvre use case of WP9. The necessity to conduct these experiments is threefold.

First, the design of a driving style for the automation must be based on the parameters which are used to plan and execute trajectories. A particular parameterization is the actual driving style. The driving style does not exist independently of the parameterization. It is therefore imperative to understand what constitutes different driving styles in manual driving, and how they can be expressed through an adequate parameterization.

Second, once candidate driving styles have been designed based on manual driving, a method to measure their attractiveness for individual drivers must be devised and applied. This too is a question that can be answered only empirically.

Finally, once different automation driving styles have been designed and rated by subjects, the selection of the driving styles shall be done using manual driving data. For this it is necessary to include a module in the IAS-AdCoS which recognizes driving styles from manual driving data. The development of this module also requires empirical data from a controlled experimental setting.

5.6.3. Input and output

At least three inputs are expected:

- algorithm to extract drivers' driving style developed in WP3
- questionnaires illuminating various aspects of the subject's driving behaviour to answer the question if a specific driving behaviour in the simulator is due to driving skill or to the driver's volition
- possibly a questionnaire to measure the overall attractiveness of the driving styles

At least two outputs are expected:

- an answer to research question one (see section 5.6.1)
- input for experiment three (to answer research questions two and three), a possible fourth experiment to validate the final setup, and for the study design to evaluate the demonstrator

5.6.4. Current status and functionality

5.6.4.1. Theory

The research question we are focusing on in the present experiment is how can we identify the driving style of individual drivers based on data from manual driving alone?

A literature review did not yield any well-grounded theory of driver type or driver style. Frequently terms such as "defensive driving", "normal driving" or "aggressive / dynamic driving" are used, but are only loosely tied to quantified driving data (e.g. [11][12]). To classify driving style, usually variables such as accelerator and braking pedal position, steering wheel angle, time head way, target spacing, ego velocity, differential velocity, lateral and longitudinal acceleration are used. For the classification itself researchers tend to rely on descriptive measures such as depictions of a circle of forces [13], Gaussian mixture models, or combinations of Markov and Bayes models (e.g. [14]).

All of these approaches either do not focus on driving styles or do not document possible styles in sufficient detail to use them for our current work. Further, they do not explicitly treat driving as a multivariate time series but rather model driving data on an aggregate level. However, the time series based modelling is necessary to detect possible shifts in driving style across situations. Also, the reviewed approaches do not address driving on a two lane motorway, which is our AdCoS use case.

Therefore, we decided to collect data on lane change behaviour for our specific use case in a series of experiments and apply a classification approach which focuses on the multivariate time series nature of driving data, namely CONFORM [15]. As we are only starting to explore the nature of human driving styles, the only hypothesis we are currently holding is

[11] MacAdam, C., Bareket, Z., Fancher, P., and Ervin, R., "Using Neural Networks to Identify Driving Styles and Headway Control Behavior of Drivers," *Vehicle System Dynamics*, vol. 29, no. sup1, pp. 143–160, 1998 (accessed June 3, 2014).

[12] Miyajima, C., Nishiwaki, Y., Ozawa, K., Wakita, T., Itou, K., Takeda, K., and Itakura, F., "Driver Modeling Based on Driving Behavior and Its Evaluation in Driver Identification," 2007. In Proceedings of the IEEE, ed. IEEE, 427–37. 95th ed. IEEE (accessed June 3, 2014).

[13] Wegscheider, M. and Prokop, G., "Modellbasierte Komfortbewegung von Fahrerassistenzsystemen," 2005. In Erprobung und Simulation in der Fahrzeugentwicklung - Mess- und Versuchstechnik, ed. VDI, 17–36 (accessed July 1, 2014).

[14] Eilers, M., Möbus, C., Tango, F., and Pietquin, O., "The learning of longitudinal human driving behavior and driver assistance strategies," Transportation Research Part F: Traffic Psychology and Behaviour, vol. 21, pp. 295–314, 2013 (accessed June 10, 2014).

[15] Griesche, S., Käthner, D., and Krähling, M., "CONFORM – A visualization tool and method to classify driving styles in context of highly automated driving," 2014. In 30. VDI-VW-Gemeinschaftstagung Fahrerassistenz und Integrierte Sicherheit: Wolfsburg, 14. und 15. Oktober 2014, 101–09. Nichtred. Ms.-Dr. VDI-Berichte 2223. Düsseldorf: VDI-Verl. http://elib.dlr.de/93007/.

that there are different driving styles, i.e. that not all people drive in the same manner.

As a working definition, we will define “driving style” as patterns in a multivariate time series of driving data, i.e. high and meaningful correlations between parameters of the driving data. These patterns must be consistent over situations. We do not search for driver types, defined as individuals showing patterns in their driving behaviour over extended periods of time and over different situations. Possibly driver types can be described with driving styles, but this is out of the scope of the current research. Our definition of driving styles aims at consistency over situations and context, not over persons.

Prime candidate variables for the analysis are time to collision and time headway towards cars in the own lane before and during lane changes, the angle of the trajectory of a lane change, the velocity before and during lane changes and the resulting lateral and longitudinal centrifugal forces.

Contextual factors influencing the values of these variables will be mainly the presence and differential speed of cars on the left lane, influencing driver’s decision to execute a lane change as well as the characteristics of this lane change. To address this issue our experimental design involves a scenario in which the subjects drive on a two lane motorway, being instructed to keep speed at 120 kph and stay on the right lane whenever possible. Subjects have to repeatedly overtake slower cars on the right lane. If they choose to overtake, a car on the left lane approaches with either 140 kph, 160 kph, or no car on the left lane appears at all.

5.6.4.2. Method

The basic scenario of the experiment consisted of a straight two lane German Autobahn with intermittent groups of vehicles on the right lane. Their velocity was 100 kph with a spacing of 8 sec time head way. Drivers were instructed to keep a target speed of 120 kph, stay on the right lane whenever possible, but overtake slower vehicles if necessary.

Condition A had no cars on the left lane, while in condition B and C cars were triggered to drive on the left lane if the ego car approached a leading car more than 280 m. These two conditions were designed to force the driver choosing to either change lanes rapidly in front of the approaching car on the left lane or to let it pass. The speeds of the overtaking car were 140 kph and 160 kph respectively. Figure 18 shows the scenario sequence for condition B.

In the experimental design, the only independent variable was the presence and velocity of the car on the left lane, yielding a one factor design with three levels. Dependent variables can be any of the variables the driving simulator records, among which we will focus on the following ones:

- velocity of the ego car
- time to collision and time headway to the leading car
- lane change manoeuvre in front or behind an overtaking vehicle
- lateral acceleration during the lane change

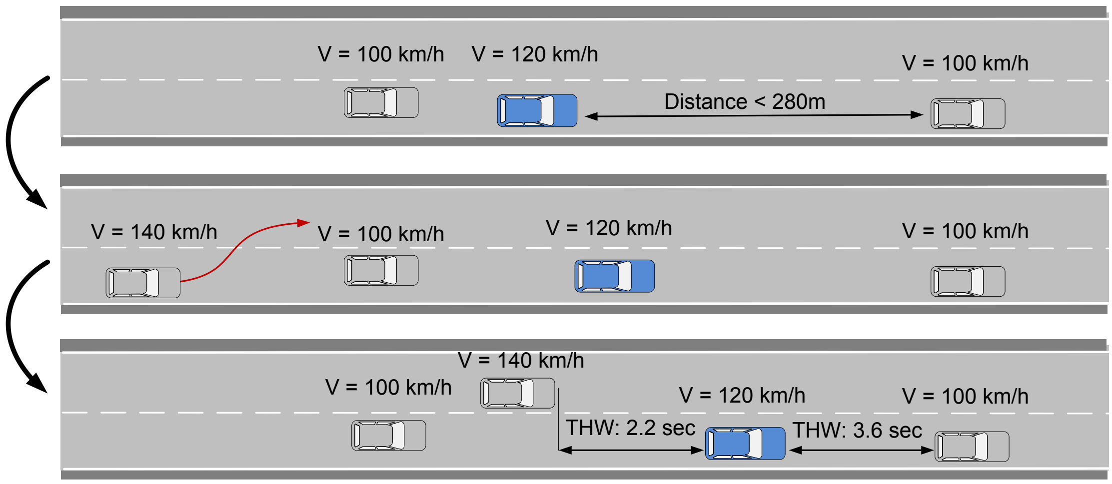

Figure 18: Scenario sequence.

The study was conducted in DLR's Dynamic Driving Simulator in Braunschweig, as shown in Figure 19.

Figure 19: DLR's dynamic driving simulator

The sample was recruited via DLR's subject data base with N = 43 (32 male and 8 female). Below is an overview over the age groups and their annual mileage (see Figure 20).

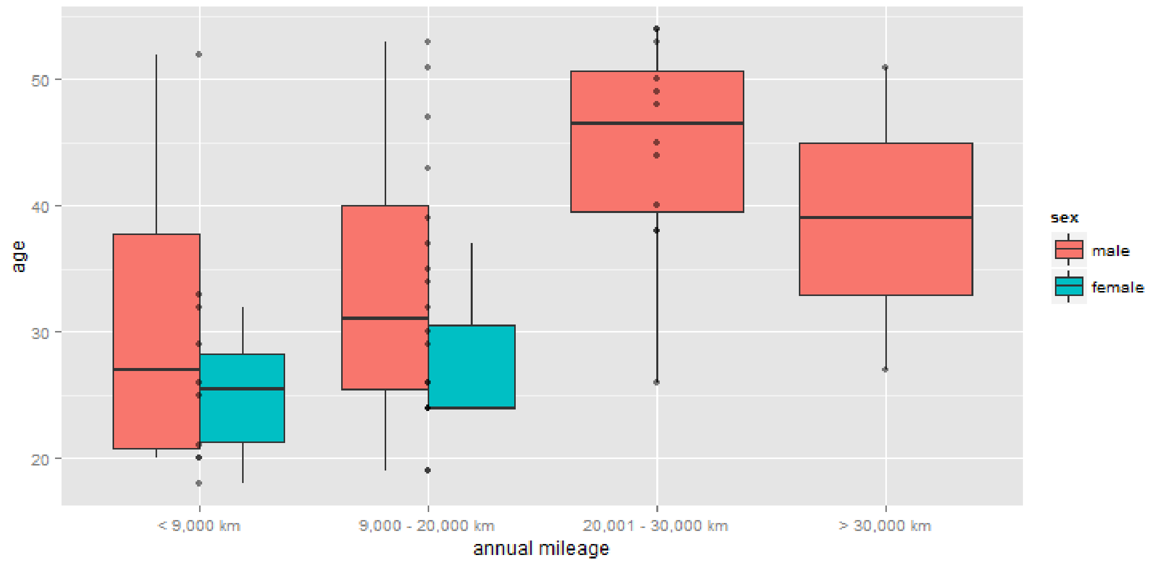

Figure 20: Age and annual mileage of the participants

To further explore the motivation behind a certain driving style, we used questionnaires. Since a given driving style can be either due to skill level or voluntary deliberations, we chose the Driver Behaviour Questionnaire [[16]] as well as the Need Inventory of Sensation Seeking [[17]].

5.6.4.3. Results

As we are still in the analysis stage, no results can be reported yet.

The outcome of the experiment will be a rating of the attractiveness of prototypical overtaking manoeuvres and the connection of this rating with the individual manual driving style of the respective drivers. This will give us the necessary know-how to produce such an automatic driving style

[16] Reason, J., Manstead, A., Stradling, S., Baxter, J., and Campbell, K., "Errors and violations on the roads: A real distinction?" *Ergonomics*, vol. 33, no. 10-11, pp. 1315–1332, 1990. http://www.scopus.com/inward/record.url?eid=2-s2.0-0025499408&partnerID=40&md5=bd78949dc4481b47c14713b895b2d1a2.

[17] Roth, M. and Hammelstein, P., "The Need Inventory of Sensation Seeking (NISS)," European Journal of Psychological Assessment, vol. 28, no. 1, pp. 11–18, 2012.

algorithmically, and the classification of manual driving data. As such, the experimental results will be integrated as part of the classification algorithm of the CONFORM module and as parameters in the IAS demonstrator.

5.6.5. Integration of the outputs of the MTT

The integration plan of the driver adaptive parameterization of a highly automated driving system method and technique will be compliant to the Common Integration Plan defined in D1.5 and reported in Chapter 4. In particular, we will upload the documentation with the description of the method and its results onto the HF-RTP as soon as the overall HF-RTP infrastructure will be completed.

Table 4 shows in details the steps as well as the estimated date to complete the integration.

Table 8: integration plan for the driver adaptive parameterization of a highly automated driving system method and technique

	Activity	Date
1	First version of method description	29/5/2015
2	Feedback from partners and experiments	31/1/2016
3	Documentation and upload on the HF-RTP	30/9/2016

5.7. CPM-GOMS Task Analysis of a Lane Change for manual and automated driving (DLR)

5.7.1. Summary

In order to gain deeper insights into manual and automation-supported lane changes, it was decided to conduct a task analysis. With that we can acquire a model that aids us understanding how the introduction of machine agents and adaptation may affect the driving task. This model can also serve as a basis for cognitive driver models in WP2. After considerable deliberation on the appropriate granularity of the analysis, we decided to model the user's interaction with the AdCoS with Critical Path Method-(CPM)-GOMS (Goals, Operators, Methods, Selection Rules) task analysis [[18]][[19]].

With our task analysis application in HoliDes, we will compare manual driving with automation-support driving in a lane change situation. As a result, we intend to provide a description of this example application in form of a tutorial or handbook and make the derived task models available for cognitive modelling in WP 2.

5.7.2. AdCoS Use-Cases

This method will be applied to the overtaking manoeuvre use case of WP9 (9.2) within the AdCoS of IAS. In principle, the applied task analysis method in its final stage shall be applicable in several stages of AdCoS design and validation across application domains.

The adapted task analysis for AdCoS can be used at different stages of the AdCoS design process, i.e., system development and validation. During the former, it can help decide on design variants which are much more promising than others, thus reducing the design space substantially. During the latter stage, task analysis can help explain why certain design variants work better than others, beyond mere quantitative statements about execution times and error rates.

5.7.3. Input and output

GOMS task analysis approaches require an intimate understanding of the task and the goals whose fulfilment the task serves. As such, we will start with a task decomposition to sketch motoric, perceptual and cognitive operations necessary to fulfil the task. This decomposition will be based on

[18] John, B. E., "Extensions of GOMS analyses to expert performance requiring perception of dynamic visual and auditory information," 1990. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 107–16.

[19] John, B. E. and Kieras, D. E., "The GOMS family of user interface analysis techniques: Comparison and contrast," ACM Transactions on Computer-Human Interaction (TOCHI), vol. 3, no. 4, pp. 320–351, 1996.

previous internal work, published literature on the subject, empirical data as well as our intuition.

In order to decompose the user's [20] task, an observation of the user while conducting the task is necessary. This may include, depending on the work environment, recording the user's head and eye movement, her feet and hands, inputs and outputs to the machine as well as other behavioural parameters such as verbal or textual communication between different users. Moreover, thinking aloud techniques may be relevant, in which the user communicates her current goals and actions while completing the task.

For improvements of already existing systems or modelling efforts during later stages of the system development, user interaction with the actual system variants can be observed. For systems under development considerable guesswork on the user behaviour with this not-yet-existing AdCoS is required. As a final result, the task analysis applied during the design process provides a (formal) description of the human agent's task execution, with and without machine agents being involved.

To model lane changes during our particular AdCoS use case (overtaking manoeuvre), we equipped a car with several sensors enabling us to observe drivers conducting overtaking manoeuvres on a two-lane highway. In particular, cameras are used to observe the drivers' head, hand and eye movements. Moreover, we record the drivers' input to the car via recording parameters from the CAN bus. In addition, the environment and surrounding traffic is observed via cameras and other sensors (e.g. for lane keeping behaviour).

5.7.4. Current status and functionality

A driving study with three participants (aged 22, 31 and 33 years, one female) has been conducted. Participants drove from the campus of the DLR at Braunschweig, Germany, on the three lane highway A2/E30 from

[20] Please note that we use the term 'user', instead of 'operator' here, although we think that 'operator' would be more appropriate in the application domains of HoliDes. However, in GOMS task analysis, the term 'operator' refers to atomic units of description, so that 'user' is less ambiguous in this section.

exit Braunschweig-Airport (exit 56) to intersection Wolfsburg/Königslutter (exit 58) and changed to the two-lane highway A39 continuing until exit 11, Braunschweig-Rautheim. Here, participants left the highway, entered again and went back to DLR on the same roads. The total covered distance was roughly 52 kilometres with about 25 km on the A2/E30 (~ 12.5 km per direction) and 27 kilometres (~ 13.5 km per direction) on the A39 (as shown in Figure 21).

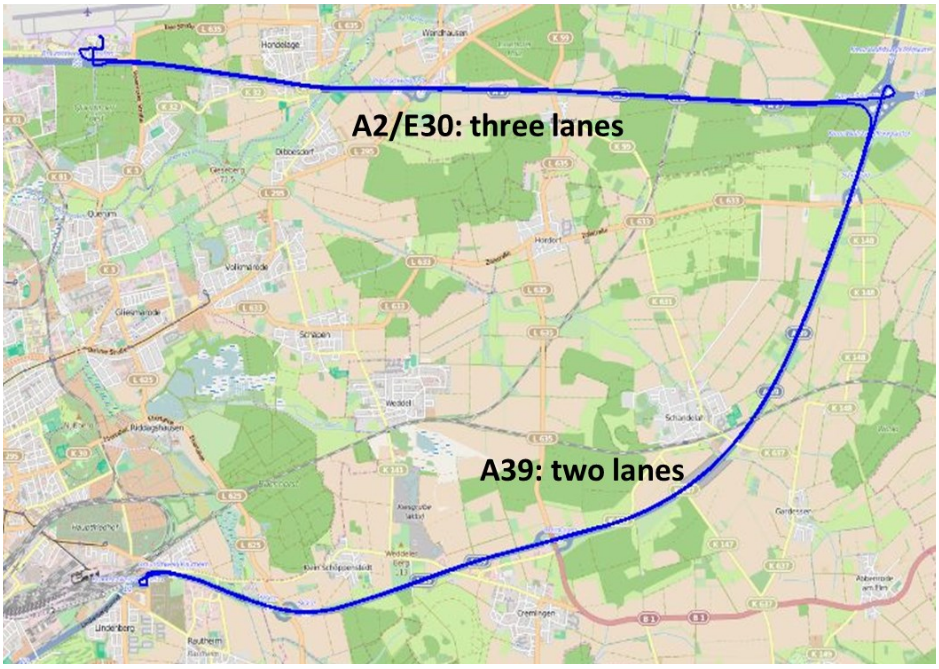

Figure 21: Road (German autobahn A2/E30 and A39) driven by the participants.

As we were primarily interested in lane changes on two-lane-highways, we considered the first three lane changes per participant on the A39 for the task analysis. This included one lane change when merging on the highway (Figure 22, A), one change from the right to the left lane (Figure 22, B) and one from the left lane back to the right lane (see Figure 22, C). Thus, in total, we analyse nine (= 3 participants x 3) lane changes.

For our analysis, a lane change was defined as starting four seconds before the car had -45 cm lateral deviation from the centre road mark of the source lane and ending three seconds after the vehicle is 45 cm away from the centre road mark in the destination lane. Based on this definition, lane changes took between 11 to 13 seconds (mean = 11.8 sec).

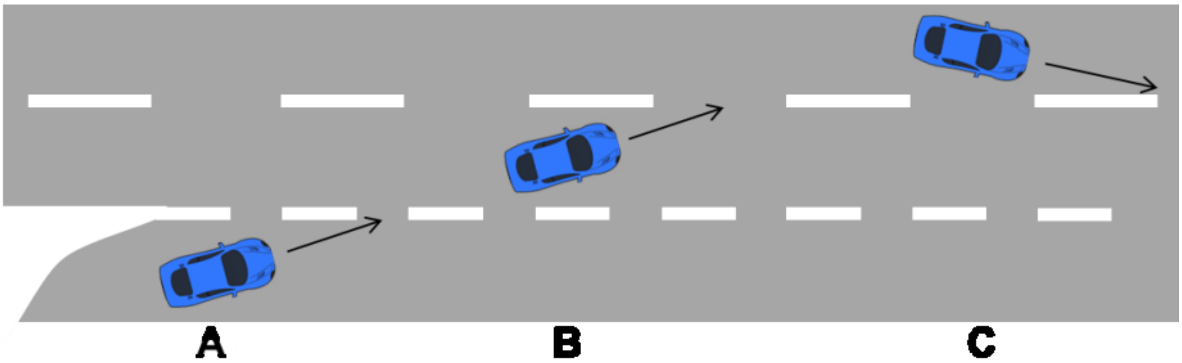

Figure 22: Lane changes considered on the two lane highway A39

The car was equipped with several possibilities to observe the driver and the surrounding environment during lane changes. Four cameras were mounted inside the car in order to record the driver's hand movements and steering behaviour (Figure 23 A, top left), face, head and upper body (Figure 23 A, top right), view in the interior rear mirror (Figure 23 A, bottom left) and the driver's feet (Figure 23 A, bottom right).

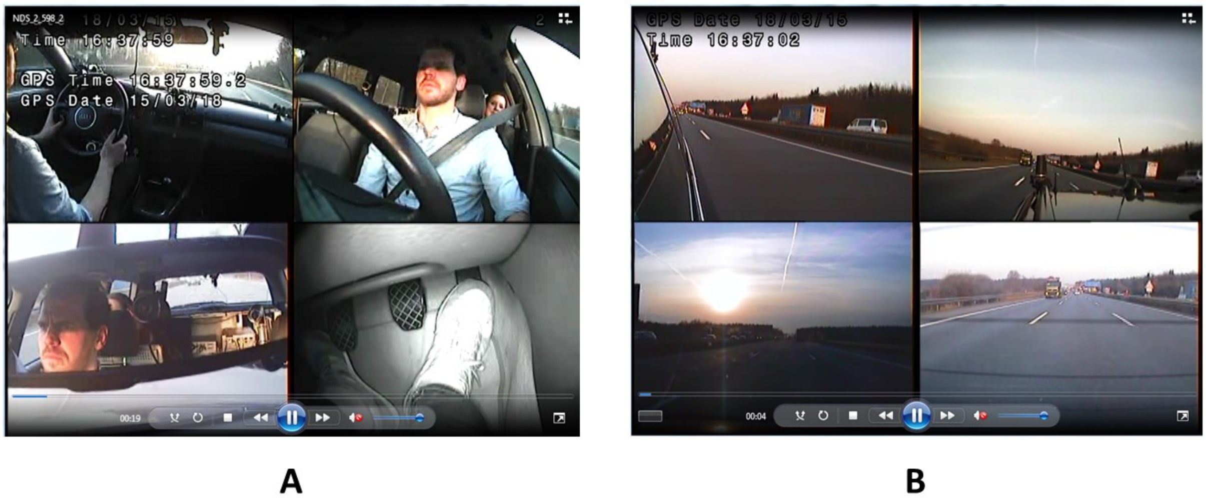

Figure 23: View of cameras for observing the driver (A) and the environment (B).

Four further cameras were installed to record the environment, i.e. on top of the left mirror facing backwards (Figure 23 B, top left), rightwards on the roof facing backwards (see Figure 23 B, top right), through the windshield (Figure 23 B, bottom left), as well as through the window in the back (Figure 23 B, bottom right). The cameras recorded with 10 fps.

Moreover, a remote eye tracker (SmartEye, Gothenburg, Sweden) tracked participants' eye movements with three infra-red cameras placed in the cockpit at a sampling rate of 120 Hz. Furthermore, additional behavioural data such as steering wheel angle (including direction), acceleration, hydraulic pressure, accelerator paddle state, lateral deviation from lane centre, blinker state, current location in GPS coordinates and velocity were recorded at 50 Hz.

For CPM-GOMS task analysis, relevant motoric, perceptual and cognitive operators for describing the task have to be defined. These operators describe atomic action units necessary for accomplishing the task on that particular level (i.e. motoric, perceptual or cognitive). Based on the recorded data, a duration (in milliseconds) is assigned to each of the operators. Finally, the flow of information through the model can be inserted, leading to a critical path for the task execution. Generally, one starts with describing the observable motoric operators, then proceeding to the perceptual and cognitive ones.

So far, we defined the motoric operators for the resources left and right foot, the left and right hand, the head and the eyes. For instance, the right foot has operators such as 'strong gas increase' or 'slight gas increase', whereas the head has motoric operators such as 'head direction to left part of the windshield' and 'move head direction from windshield to left outside mirror'. The operators are put in order and the duration was assigned to each instance of the operator as exact as possible with the information from the cameras and sensors. As an example, Figure 24 provides a screenshot depicting the motoric resources, some operators and their timing in the first two seconds of the lane change of one participant merging to the highway. Time is depicted in the top row in ms. Each row below depicts one resource with some motoric operators and their duration.

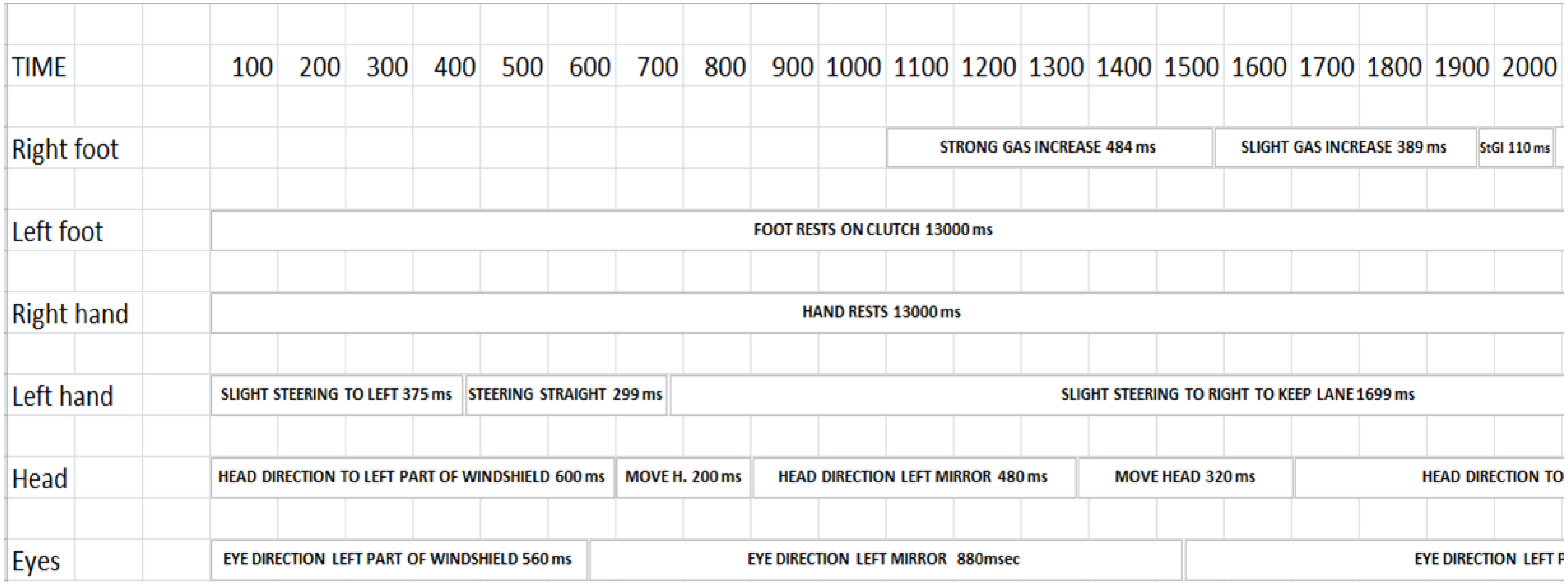

Figure 24. Exemplary screenshot of the first two seconds of a lane change when merging on the highway for one participant.

With the motoric operators, their timing as well as the collected data and knowledge at hand, the next step is to define perceptual and cognitive operators, put them into order and assign appropriate times to them. Based on the resulting decompositions of the nine lane changes on the three levels, we will attempt to aggregate the data in a meaningful way for each of the three classes of lane changes and assign a critical path to it (i.e., the time required for a particular class of lane changes). Later on, we may enrich the models and insights derived with the GOMS task analysis with a system decomposition based on a cognitive work analysis [[21]].

5.7.4.1. Usage of the method and integration with other MTTs

The GOMS task analysis can be integrated into the system design and evaluation process at several stages. For the design of IAS' automated car, we use task analysis to decompose the (manual) task of lane changing in order to understand the behavioural and cognitive operations involved.

Based on this, strategies for optimal handover-of-control between human and automation (and back) can be designed (potentially with the aid of Theatre technique, described in section 5.8). These then may be tested using another GOMS task analysis in order to compare it with the timings (or other parameters) derived from the decomposition of the manual task to evaluate the effects of introducing (adaptive) automation. Nicely, these comparisons are possible at many stages of the design process, so that unsuitable design variants can be eliminated early on.

For an efficient use of GOMS task analysis in the AdCoS development workflow an integration of the MTT into the HF-RTP is desired. This integration, however, has several components. First of all, together with partners from other work packages, we work on a classification of task analysis techniques applied in HoliDes (such as e.g. the HFC's Human Factors Task Analysis, described in section 5.10).

5.7.5. Integration of the outputs of the MTT

The integration plan of the CPM-GOMS method will be compliant to the Common Integration Plan defined in D1.5 and reported in Chapter 4. In

---

[21] Vicente, K. J., *Cognitive work analysis: Toward safe, productive, and healthy computer-based work.* Mahwah, N.J: Lawrence Erlbaum Associates, 1999.

particular, we will upload the documentation with the description of the method and its results onto the HF-RTP as soon as the overall HF-RTP infrastructure will be completed.

Table 9 shows in details the steps as well as the estimated date to complete the integration.

Table 9: integration plan for the GPM-GOMS method

	Activity	**Date**
1	First version of method description	29/5/2016
2	Feedback from partners and experiments	31/1/2016
3	Documentation and upload on the HF-RTP	30/9/2016

5.8. Theatre Technique for acceptance tests and systems variants exploration during AdCoS design (DLR)

Please note that the Theatre Technique has not yet been adapted in the context of HoliDes. Thus, the description found in this deliverable has only minor changes compared to D5.3.

5.8.1. Summary

The Theatre Technique [[22]] can be used to support the collection of feedback and expectations of the human operator with respect to an adaptive system early in the design process. Our goal in HoliDes is to demonstrate the usefulness of this technique for an AdCoS design process, adapting it where necessary. It is not used as a method to test design variants for different user groups, but as a technique to explore the design space. Therefore, participants in the design sessions will be Human Factor experts and AdCoS developers, not naive users.

[22] Schieben, A., Heesen, M., Schindler, J., Kelsch, J., and Flemisch, F., "The theater-system technique: Agile designing and testing of system behavior and interaction, applied to highly automated vehicles," 2005. In *Proceedings of the 1st International Conference on Augmented Cognition*, 43–46.

With the Theatre Technique a researcher or human factors expert (termed the confederate) mimics the intended system's behaviour in a Wizard-of-Oz like fashion. This is particularly useful when planned automation functions and interaction concepts are to be tested before implementation, reducing time and costs for re-design. The Theatre Technique can also be used to validate and refine task analysis models without a fully functioning prototype early in the AdCoS design process.

When the designer or developer has a concept of an adaptive AdCoS behaviour, he or she can use the Theatre Technique to mimic and evaluate this behaviour with test participants. With the data collected and feedback from the participants, potential problems in the adaptation as well as undesired consequences in the interaction can be detected early in the design process and compared with the requirements. Thus the extended Theatre Technique can be used as method for early, low-cost validation of adaptive AdCoS functions and as validation for task analysis models.

In HoliDes, one aim is to formulate a concept for the handover-of-control from machine to human (and back) of the IAS AdCoS demonstrator using the task analysis. This task allocation concept described is planned to be validated using Theatre Technique. Specifically, the confederate simulates the machine agent (i.e., the highly automated car), so that we can determine whether the predictions of the model on task allocation are realistic.

The final product will encompass qualitative design recommendations, task execution times for the overtaking use case and a handbook detailing the application of this technique for design and validation use cases.

5.8.2. AdCoS Use-Cases

We will apply the theatre technique to the overtaking manoeuvre use case of WP9 within the AdCoS of IAS. It is planned to use the technique for testing different handover-of-control operations between the human driver and the automation early in the design process. The application of the theatre technique is planned for the last third of the project.

5.8.3. Input and output

The Theatre Technique needs a concept of certain functions of the AdCoS and human-factors-relevant requirements with respect to these functions. Human factors experts/researchers need to be trained to produce the desired AdCoS functions in a laboratory setting. In HoliDes and our specific use case, this input will come from the task model and the AdCoS variants proposed in collaboration with IAS in WP9. We will then hold a design session, where the Human Factors experts from DLR together with the IAS partners will explore different design variants. During the design session, a trained confederate will act in place of the automation and follow a script of actions to carry out in the prepared scenarios.

As regards the outputs, the Theatre Technique provides feedback whether or not certain system functions adhere to Human Factors relevant requirements, and if the requirements should be changed (extended, refined, abandoned). It can also be utilized to validate automation concepts generated in early stages of the design process, e.g., from task analysis. In case data are recorded during Theatre Technique experiments (e.g. by using multiple cameras), these can be used to create a detailed, updated version of the task model.

5.8.4. Current status and functionality

We have not adapted the Theatre technique in HoliDes, yet. As mentioned above, it is planned for the last third of the project. Results will be documented from D5.5 onwards as well as in WP9 deliverables. Figure 25 shows the general setup of the Theatre Technique, with the driver on the left side and the confederate on the right side.

Figure 25: Setup of the Theatre Technique

5.8.5. Integration of outputs of the MTT

The integration plan of the Theatre technique will be compliant to the Common Integration Plan defined in D1.5 and reported in Chapter 4. In particular, we will upload the documentation with the description of the method and its results onto the HF-RTP as soon as the overall HF-RTP infrastructure will be completed.

Table 10 shows in details the steps as well as the estimated date to complete the integration.

Table 10: integration plan for the Theatre technique

	Activity	Date
1	First version of technique description	29/5/2015

2	Feedback from partners and experiments	31/1/2016
3	Documentation and upload on the HF-RTP	30/9/2016

5.9. Surrogate Reference Task (SuRT) for inducing driver distraction (DLR)

5.9.1. Summary

Driver distraction is one of the main causes of car accidents as it can result in a delay in the processing of information necessary to perform the driving task [[23]]. In order to counteract this, researchers and developers have envisioned designing adaptive driver assistance systems capable of recognizing when the driver is distracted. This endeavour requires research tools that can be utilized to reliably induce distraction in simulator and real traffic studies. Here we present a touch screen version of the Surrogate Task (SuRT) as a tool that can be applied in the automotive context and other domains as a secondary task to induce visual and manual distraction. The SuRT is specified in the ISO/TS 14198 [[24]].

5.9.2. AdCoS Use-Cases

Currently the tool SuRT is applied to the overtaking manoeuvre use case of WP9 within the AdCoS of CRF. In principle it may be applied in other use cases and domains as well. In the CRF use case, this tool has been used for the empirical studies by SNV that were run in order to acquire data of distracted drivers (as described in section 5.14). This data form the ground truth for the classification algorithms for the MTT Detection of driver distraction based on vehicle dynamics by UTO and CRF (as described in section 5.16).

---

[23] Horberry, T., Anderson, J., Regan, M. A., Triggs, T. J., and Brown, J., "Driver distraction: the effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance," *Accident; analysis and prevention*, vol. 38, no. 1, pp. 185–191, 2006.

[24] 14198:2012. Road vehicles - Ergonomic aspects of transport information and control systems - Calibration tasks for methods which assess driver demand due to the use of in-vehicle systems.

5.9.3. Input and output

The SuRT tool collects error rates and reaction times from participants during experiments. These data are then transferred to RTMaps and can be acquired from there for further processing and detailed analysis.

5.9.4. Current status and functionality

The SuRT is fully functional and has been used already in the studies from SNV (as described in section 5.14). Moreover, an android version is available for download in the google play store [[25]], without integration into RTMaps.

The SuRT consists of visually and manually demanding parts. Participants are presented with a set of stimuli on a touch screen (e.g. a tablet or a smart phone) which can be mounted on the right side of the steering wheel in reach of the driver's right arm. The stimuli on the screen consist of a target in a set of distractors. Whenever the target stimulus is presented, the participant is asked to touch the half of the screen containing the target as fast as possible with the right arm. In this way visual (looking at the screen) and manual (touching the screen) distraction is induced. With respect to metrics, participants' performance in this secondary task can be evaluated in terms of error rate (percentage/amount of incorrect responses) and reaction times (time between target presentation and response.

The appearance of target and distractors can be set with the parameters shape (circle, square, cross), colour (RGB), radius and line width. Another parameter that needs specification is the time between two tasks, i.e. the latency of presentation of a new target after a participant's response. Moreover, the background colour of the two halves can be specified. Figure 26 shows an example appearance of the SuRT with a large pink circle as target and small violet circles as distractors. In this configuration the participant would have to touch the left half of the screen.

---

[25] https://play.google.com/store/apps/details?id=de.lapoehn.surt

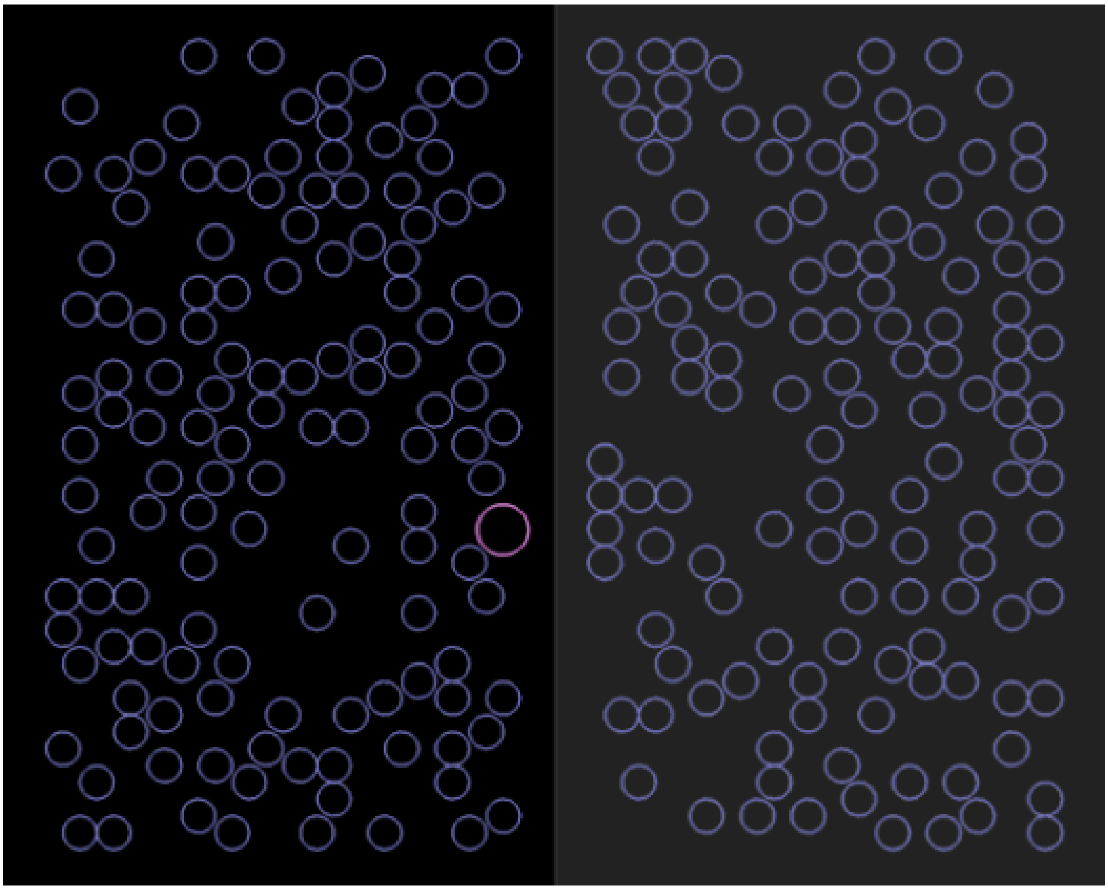

Figure 26: Example of SuRT appearance

5.9.5. Integration plan

The integration plan of the SuRT tool will be compliant to the Common Integration Plan defined in D1.5 and reported in Chapter 4. In particular, the integration of this tool will be accomplished by completing the first two steps:

1) Identify the inputs and outputs of the tool (described in Annex III), its compliance with OSCL and the estimated date for integration into HF-RTP.
2) Identify the correspondence between its inputs and outputs and the concepts of the HoliDes meta model.
3) Develop the adapters or parsers between the concepts and the data managed by the MTT

The details of the integration plan are described in the Annex III (confidential part).

Table 5 shows in details the steps as well as the estimated date to complete the integration.

Table 11: integration plan for the SuRT tool

	Activity	Date
1	Tool available to other partners for experiments	30/1/2015
2	Feedback from partners	31/8/2015
3	Matching between the HoliDes Meta Model concepts and the input/output.	As soon as the HoliDes Meta Model is available
4	Development of an OSLC-compliant version and adapter in RTMaps	31/12/2015
5	Documentation	30/9/2016

5.10. Human Factors Methods and Metrics for HF and Safety Regulations (HFC)

5.10.1. Summary

There are two objectives of the work. In the first rank we analysed the standards and regulations reviewed in WP1 systematically. We identified the evaluation criteria which need to be taken into account for each regulation and standard. The second aim was to define methods and possibly metrics for our four selected evaluation criteria ("HF issues": usability, distraction, situation awareness, mental workload). This proceeding also shows many uncovered or poorly covered, but crucial HF aspects of AdCoS, for which HF methods and metrics are obsolete or insufficient. It is expected that new measurement methods are needed for the specific aspects of an AdCoS, namely cooperation and adaptation.

The results could be applied in several stages of the system development. However the difference between HF issues (criteria), methods, and metrics has to be pointed out. The HF issues (evaluation criteria) are usually more universal in use than metrics. A criterion/issue clarifies required properties of the system under development, or which requirements it should meet (e.g. usability). Ideally, these issues/criteria should be considered and evaluated throughout the whole system life-cycle.

An HF method is a specified procedure, intended to allow empirical investigation and measurement of the degree to which the system complies with the HF issue. As HF issues cannot be measured directly (i.e. expressed in numbers), each method description includes one or more metrics that are measurable. These metrics are then defined as to indirectly measure the issue, that is, it can be defined what any output data means for the issue.

5.10.2. AdCoS Use-Cases

The four selected HF issues are related to all four domains: aeronautics, health, automotive and control rooms. Some of the proposed methods with their respective metrics, however, will be applied concentrating on the health care domain (WP6; deriving metrics proposals for all criteria seems impossible due to empirical and working framework limitations.). They should enable the operationalization of an evaluation criterion or criteria relevant for an MRI system (MRI UC01 safe patient transfer or MRI UC02: Guided patient positioning) as well as for 3D acquisition scans (iXR UC01).

5.10.3. Input and output

The inputs of the method are:

- Human Factors and Safety regulations and guidelines systematically analysed in WP1 (as described in D1.2).

- Human Factors Methods and Techniques: Review and selection of MTTs appropriate for measurement of the selected criterion/criteria. Inter alia:
  - Stanton, A. N., Salmon, P. M., Walker, G. H., Baber, C., & Jenkins, D. P. (2005). Human Factors Methods: A Practical Guide for Engineering and Design.
  - Diaper, D. & Stanton, N. (2004) The Handbook of Task Analysis for human-Computer-Interaction.

As regards the outputs, HFC will deliver two documents. In the table with the set of HF issues and HF methods derived from HF and Safety Regulations, the following information will be provided: domain, system, reference to regulation and year, HF issues (if present), evaluation methods or techniques (if present), and metrics (if present).

In the second document, the "HF Method Library", methods and corresponding metrics will be categorised and described. For all HF methods, descriptors are given in order to tag them according to HF issue, corresponding phases of the system development process, type of resulting data, type of empirical method, resources needed, and time of data collection. It contains written summaries of the method procedure, and hints and requirements for the interpretation of output, to conclude whether the criterion is met.

To provide a short example of the contents (with *descriptors* and "tags"):

The HF method with the *ID* "NASA Task Load Index" regards the *HF-issue* "mental workload", and is of the *empirical method type* "questionnaire". The *minimal resources* needed are "participants" and "pen&paper", the method can be *applied by* a "non-expert", the *costs* are "low" […]. Given you use HoliDes platform builder and want to find empirical methods to measure mental workload that do not cost much, this would be one of the possibilities.

A manual for the usage of the HF Method Library will also be provided (for a detailed description, see D1.5, section 2.3.2).

We plan to determine how the HF issues/criteria can be measured when applied to the use cases using Human Factors MTTs (option 1- appropriate task analysis method; option 2 - other established methods and techniques or a set of them).

5.10.4. Current status and functionality

The first version of the HF Method Library has been completed and discussed with partners in WP1. It is intended to expand the content by adding more of the human factors methods used in HoliDes by the AdCoS owners.The systematic review of the HF and Safety regulations and guidelines from D1.2 is completed.

The HF Method Library and corresponding references to regulations and domains are currently being integrated into the HF-ontology (for the ontology, as described in D1.5).

Name of Regulation / Concept	Adressed system	domain	which of the 4 selected HF issues?	year	HF METHODS	METRICS ("data output format of methods")
Medical Device Use-Safety: Incorporating Human Factors Engineering into Risk Management	all medical systems with human interaction	healthcare	**usability**	2000	task analysis, usability testing, function analysis, heuristic analysis, expert reviews, walk-through	"unformalised"

Figure 27: extract from table of HF regulations and corresponding HF methods as well as possible metrics

5.10.5. Integration of the outputs of the MTT

The integration plan of the HF Methods and Metrics for HF and Safety Regulations will be compliant to the Common Integration Plan defined in D1.5 and reported in Chapter 4. In particular, we will upload the documentation with the description of the method and its results onto the HF-RTP as soon as the overall HF-RTP infrastructure will be completed.

Table 12 shows in details the steps as well as the estimated date to complete the integration.

Table 12: integration plan for the HF Methods and Metrics for HF and Safety Regulations

	Activity	Date
1	First version of method and regulations description	29/5/2015
2	Feedback from partners	31/8/2015
3	Finalization of the inclusion into the HF-ontology	31/01/2016
4	Documentation and upload on the HF-RTP	30/9/2016

5.11. HF-Task Analysis Tool (HFC)

5.11.1. Summary

The aim of this MTT is to test the analysis questions/criteria, resulting from HF regulations and standards. A software tool for task analysis is currently under development to support this approach.

Task analysis is a method with a very wide field of application. It can be used in the different stages of AdCoS development and supports an investigation of several Human Factors questions, e.g. identification of usability/safety weaknesses of the human-machine interaction. The conduction of the task analysis is unfortunately very time and cost intensive, challenging for the researcher, and the presentation of the results is often problematic. The main objective of the task analysis tool is to support the task analysis procedure in different ways, i.e. data gathering, visualisation of the collected input and modelling. The method should be applicable to potentially every task with cognitive elements. However we will concentrate on task models used in chosen healthcare use cases.

5.11.2. AdCoS Use-Cases

It is intended to work with one or two of the following Use Cases from WP6 (PHI): MRI UC1 safe patient transfer, MRI UC2: Guided patient positioning, iXR UC01_3D_acquisition. Task analysis will be beneficial for these cases since it is expected to help analysing especially the cognitive and critical parts of a task model.

5.11.3. Input and output

Empirical data collected by task analysis: We intend to carry out a task analysis to investigate the MRI or iXR system regarding the selected HF issues. We will collect data using documentation, structured observation and interviews, and audio/video recordings (if allowed). The obtained empirical data will be used as the first input for the task representation and modelling provided by the tool. This step will also enable to examine whether created data structures need to be adjusted.

This MTT provides methods and techniques for testing predictions of task models for adaptive systems. The tool is intended to support the task analysis procedure in different ways and make it more efficient and easier to apply for researchers. The output shall provide a structured visualisation of the results. It will include the detailed task description ordered according to the analysed task features and the relations between the tasks.

For example, the tool might highlight subtasks which are cognitively demanding for the user, or show where the intended use of the system differs from how users really use it. The results can be used to restructure tasks and improve system design, in a more efficient way than task analysis normally can.

5.11.4. Current status and functionality

Development of the tool is under way. The essential internal C++ data structures were designed and implemented. Based on this there were a XML-parser (for reading files) and a XML-generator (for saving new and modified projects) created. The first version of the GUI architecture is now defined, the data model is work in progress.

5.11.5. Integration plan

The integration plan of the HF-Task Analysis Tool will be compliant to the Common Integration Plan defined in D1.5 and reported in Chapter 4. In particular, the integration of this tool will be accomplished by completing the following steps:

1) Identify the inputs and outputs of the tool, its compliance with OSCL and the estimated date for integration into HF-RTP.
2) Identify the correspondence between its inputs and outputs and the concepts of the HoliDes meta model.
3) Develop the adapters or parsers between the concepts and the data managed by the MTT

The details of the integration plan are described in the Annex III (confidential part).

Table 5 shows in details the steps as well as the estimated date to complete the integration.

Table 13: integration plan for the HF-Task Analysis tool

	Activity	Date
1	First version of tool	29/5/2015
2	Feedback from partners	31/8/2015
3	Development of an OSLC-compliant version	31/12/2015
4	Matching between the HoliDes Meta Model concepts and the input/output.	As soon as the HoliDes Meta Model is available
5	Documentation	30/9/2016

5.12. Behavioural Validation Tool (REL)

The Behavioural Validation Tool (BVT) addressed a specific problem in the validation of Android mobile applications. Since the tool is developed by REL and no other partners but REL is developing an AdCoS including Android mobile applications in HoliDes, the AdCoS owner would have been the same as the tool provider. Therefore, we decided to discontinue the development of this tool, and redistribute the effort of REL on other activities in WP5, mainly on the definition of the Common Metrics Methodology and the Common Integration Plan that have been addressing key issues of the HoliDes project.

5.13. Empirical analysis and validation methods of cognitive and communicative processes in the Control Room domain (SNV)

5.13.1. Summary

SNV purpose is to address, through the use of psychological and psycho-physiological techniques as reaction times, EEG, eye tracker, the need of investigating human performance to assess communication and load processing (WP8). The data obtained will be implemented in the AdCoS in order to help the operators to achieve the goals planned (WP8).

5.13.2. AdCoS Use-Cases

In the control room domain (WP8) the effort will concentrate on the use cases concerning the communication between the operator and the operational teams in the field; the communication with non-Italian speaking caller; the peak of incoming calls due to an exceptional event.

The aim of this method is to detect and analyse HF issues status in order to improve the AdCoS' adaptivity in WP8 domain.

5.13.3. Input and output

Appropriate modelling needs data which capture information about HF issues. Particularly, in the control room domain by means of an ad-hoc prepared questionnaire we have been able to investigate operator/customer interaction. For example, operators have been asked to describe how an emergency call usually proceeds, what kind of information is fundamental in order to face the emergency, how the operator interacts with the costumer to gain as more details as possible.

As regards the output, the technique has provided information about the operator status to be used during the design process of the AdCoS. Particularly, the results of the questionnaire have been used to define an emergency lexicon with the more frequent words that can be useful in the design of the adaptive interface.

5.13.4. Current status and functionality

Ad-hoc questionnaire has been presented to the operators of the Iren call-centre to the aim to assess lexical and pragmatic aspects of by phone interaction between technicians and users. It includes 21 questions: 18 multiple answer questions and 3 open questions to let the operators express their comments.

SNV contribution in the framework of WP8 started by analysing common adaptive features to be applied to cognitive and communication processes. Communication processes started to be investigated by collecting the interactions (state of the art) between control room's operators and customers calling for an emergency by means of an ad hoc questionnaire. The questionnaire has been created to assess the lexical and the pragmatic aspects of the interaction. The preliminary results evidenced the need of a more common, formalized grid to interact with costumers to speed up the process with a minor effort. The same kind of baselines will also be collected in a cross-cultural perspective. Once collected the baseline, subsequent experiments will aim at assessing a method to improve the interactions in the control room domain. Finally, workload will be studied under operators stress conditions. Experiments will be designed to measure and reduce the workload of operators in the control room. By manipulating the number of incoming calls and the level of emergency of the event and recording the EEG data we expect to measure the level of workload of the operator. Specifically the analysis of the correlation between workload measures and EEG data will provide information to re-allocate tasks in a cognitive efficiently fashion. A review of the literature is considered the first step to clarify which qualitative and quantitative methods can be efficiently applied to the workload.

5.13.5. Integration of the outputs of the MTT

The integration plan of this method will be compliant to the Common Integration Plan defined in D1.5 and reported in Chapter 4. In particular, we will upload the documentation with the description of the method and its results onto the HF-RTP as soon as the overall HF-RTP infrastructure will be completed.

Table 14 shows in details the steps as well as the estimated date to complete the integration.

Table 14: integration plan for the empirical analysis and validation methods of cognitive and communicative processes in the Control Room domain

	Activity	Date
1	First version of method description	1/11/2014
2	Feedback from partners (after additional tests)	31/8/2015
3	Consolidation of results	31/1/2016
4	Documentation and upload on the HF-RTP	30/9/2016

5.14. Empirical analysis and validation methods of cognitive processes in automotive domain (SNV)

5.14.1. Summary

SNV purpose is to address, through the use of psychological and psycho-physiological techniques as reaction times, EEG, eye tracker, the need of investigating human performance to assess distraction processes (WP9). The data obtained will be implemented in the AdCoS in order to prevent visual and cognitive distraction (WP9).

5.14.2. AdCoS Use-Cases

In the automotive domain (WP9) the use cases defined by CRF that will be considered is the Lane-Change Assistant (LCA).

The aim of this method is to assess the operator's status in order to improve the AdCoS' adaptivity in WP9 domain.

5.14.3. Input and output

Appropriate modelling needs data which capture information about the operator status. Particularly, in the automotive domain by means of a proper vehicle equipment configuration (internal camera, vehicle CAN bus and secondary task), we have been able to collect data about the driving behavior (like the steering angle, vehicle speed, position of the vehicle in the lane, etc.), the driver's head orientation (off-road, if the driver is looking the road or not), and the execution of the secondary task (e.g., response times, correct answers, etc.).

As regards the output, the techniques described will provide information about the operator status to be used during the design process of the AdCoS. For example, in WP9 we have exploited the results of the experimental campaign to derive the model of the driver's visual distraction based on vehicle dynamics. We have identified the driving behaviour under visual distraction to be used to predict the driver' status, as described below.

5.14.4. Current status and functionality

5.14.4.1. Theory

Behavioural quantitative methods as reaction times in experiments with secondary tasks (distraction), analysis of errors, eye tracker methods and EEG will be the main methods to be used in experiments (quantitative) and observational studies (qualitative). The methods will be applied to experimental designs to investigate cognitive and communication processes.

In the WP9 framework two series of experiments have been designed. The first series of experiments investigates cognitive distraction. Results coming from the literature are not homogeneous and do not provide a unified explanation of the phenomenon (Bock, Dell, Garnsey, Kramer & Kubose, 2007). For example, in the classic situation of someone talking on the phone while driving, it has not been clarified, which aspects of the conversation act as distracters to the main task of driving. So far, in the first series of experiments, predictions have been formulated taking into account which aspects of the secondary task can really interfere in the primary task of driving. The experiments have been designed to be applied in the lab and in ecological situations. In addition, visual distraction has been object of many debates among researchers. Two aspects have been considered central in the study of visual distraction: first, under which conditions a driver can be considered distracted by visual stimuli (choice of a sensitive experimental paradigm); second, how cognitive and visual distraction can be considered as two separate phenomena.

5.14.4.2. Method

Empirical analysis of cognitive and communicative processes is a procedure aimed at answering particular research questions. It is based on a rigorous methodology going from hypothesis to observations or data, passing through different steps. The first step is the definition of the topic of interest and the main research question. The research hypothesis needs to be declined into variables, events or dimensions of a particular behaviour that will be matter of observation or measurement. Given a research hypothesis, the next fundamental step is the selection of the paradigm, a task specifically defined to induce a certain human behaviour that will be observed or measured. Tasks need to be controlled in term of conditions (for example the introduction of different levels of complexity of the same task), order and time of presentation to prevent fatigue effects in participants. Another critical aspect concerns the choice of the sample of participants that has to be fully representative of the population it is extracted from. In addition, the dimension of the sample has to be adequate respect to the aim of the research: for example in case of statistical inferences on data the number of participants have to be higher than 20 in order to obtain statistically significant information.

In the following part these steps will be described by means of the concrete example of the experiment on visual distraction realized in the automotive domain (De Simone, Presta, Collina & Tango, 2015)

1 - Objectives and measures: in the automotive context the main aim of the experiment was to answer the question "how does the driving behaviour change under distraction effects?". To answer this question it was fundamental to assess the status of the driver by means of psycho-physiological measures and to collect and compare data about the driving behaviour with and without distraction.

2 – Paradigm: To induce visual distraction in driving we realized an experiment based on the interfering secondary task paradigm. Participants were involved in a primary driving task in-vehicle. While driving, they were asked to perform a secondary task, a visual search task: participant had to localize a target between distractors and to indicate the portion of the display where the target appeared. In the specific case, the target was a red circle and distractors were green circles and red squares.

3 – Scenarios: The test site consisted of two different scenarios: participant drove for 30 km in highroad and 30 km in extra-urban way. The length of the text-site has been chosen to guarantee an amount of data adequate for statistical comparisons.

4 – Equipment: The vehicle in which the experiment took place was equipped with devices and sensors to collect physical and psycho-physiological measures in order to assess the status of the driver:

- vehicle CAN: steering angle, vehicle speed, direction;
- internal camera: head position, distraction ON/OFF;
- external camera: position of the vehicle in the lane, road curvature.

5 – Procedure: The stimuli were presented on an 8" touch screen monitor, similar to a navigator device display, collocated in the central part of the dashboard to cause participants to move eyes away from the road. The display were divided into two parts, one with black background and the other one with dark grey background. Participants were asked simply to touch the portion of the display where the target appeared. The time between the onset of the stimulus and the touch of the screen were collected. In addition, a driving session without distraction was implemented to collect the baseline. The order of presentation of the conditions varied across participants to prevent fatigue effects.

Thirty participants volunteered in the experiment Participants were instructed by the experimenter about the tasks they were called to perform. They were presented with a document containing information about the procedure and the aim of the study and the law about the treatment of data. They were asked to carefully read the document and to sign it; in case of doubt they were invited to ask the experimenter. Before the experiment proper, each participant was familiarized with the secondary task and the vehicle in a practice session in which no data were collected. The experiment started when participant felt confident with the tasks.

After the experiment proper the subjective workload level of each participant in the different driving sessions has been assessed by means of the NASA TLX questionnaire. This questionnaire has been chosen among others due to its multidimensional structure that highlights different dimensions of the workload. The Italian version of the questionnaire (Bracco, Chiorri, Mortola, Bottiglieri, Piccinno, D'Anna, 2008) has been adopted.

The entire experiment lasted approximately 90 minutes.

5.14.5. Integration of the outputs of the MTT

The integration plan of this method will be compliant to the Common Integration Plan defined in D1.5 and reported in Chapter 4. In particular, we will upload the documentation with the description of the method and its results onto the HF-RTP as soon as the overall HF-RTP infrastructure will be completed.

Table 15 shows in details the steps as well as the estimated date to complete the integration.

Table 15: integration plan for the empirical analysis and validation methods of cognitive and communicative processes in the Automotive domain

	Activity	Date
1	First version of method description	1/11/2014
2	Feedback from partners	31/8/2015
3	Consolidation of results	31/1/2016
4	Documentation and upload on the HF-RTP	30/9/2016

5.15. Driver Distraction Classifier - DDC (TWT)

5.15.1. Summary

Distraction during driving can result in a delay in recognition of information that is necessary to safely perform the driving task [[26]].Thus, distraction is one of the most frequent causes of car accidents [[27]][[28]]. Four different forms of distraction are distinguished while they are not mutually exclusive: visual and auditory (sensory), biomechanical (physical), and cognitive distraction. Human attention is selective and not all sensory information is processed (consciously). When people perform two complex tasks simultaneously, such as driving and being involved in a demanding conversation, there is a competition for resources available in the brain. This affects the focussing of attention and attention shifting might also occur unconsciously. Driving performance can thus be impaired when filtered information is not encoded into working memory and so critical warnings and safety hazards can be missed [[29]]. Sources for distraction of the driver can be located within and outside of the car.

A cognitive and a computational distraction model are being developed with the aim to detect the level of distraction of the driver, employing in-car recordings. The cognitive model helps us to understand which factors influence cognitive distraction and therefore it helps us focus on the relevant types of data we should measure and select significant features. This knowledge is applied in the computational distraction classification model, which is aimed to analyse the selected data real-time, and provide the interpreted level of distraction as an output. To this end we investigate

---

[26] M. A. Regan und K. L. Young, „Driver distraction: a review of the literature and recommendations for countermeasure development.," in *Proc. Australas. Road Safety Res. Policing Educ. Conf.*, 2003.

[27] J. Artho, S. Schneider und C. Boss, „Unaufmerksamkeit und Ablenkung: Was macht der Mensch am Steuer?," Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, 2012.

[28] T. Horberry, J. Anderson, M. A. Regan, T. J. Triggs und J. Brown, „Driver distraction: the effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance.," *Accident Analysis Prevention,* Bd. 38, Nr. 1, pp. 185-191, 2006.

[29] L. M. Trick, J. T. Enns, J. Mills und J. Vavrik, „Paying attention behind the wheel: a framework for studying the role of attention in driving.," *Theoretical Issues in Ergonomics Science,* Bd. 5, Nr. 5, pp. 385-424, 2004.

standard machine learning algorithms (work-in-progress). The cognitive model can be seen as the theoretical framework (developed in WP2), and the classification model as the applied framework (developed in WP5).

The models are based on both literature and on scientific experiments. For assessing predictive parameters for cognitive distraction during driving, we run several experiments using a driving simulation and compare parameters between concentrated driving and distracted driving induced by parallel tasks such as conversations or calculation tasks. These measures will be based on an acoustic analysis, including for example the detection of the number of speakers, the level of emotional content, and information about the driver's involvement in the conversation (e.g., whether the driver him/herself is speaking). In addition, face-tracking signals such as eye blinks, head pose and mouth movements will add to the reliability of distraction prediction.

On the one hand, we hope to get new insights about the correlation between auditory signals inside the car and cognitive distraction of the driver from our experimental results. On the other hand, the overall aim of the cognitive distraction model is the development of a mobile user profile computing the individual level of distraction. Additionally, the model can be applied in other systems.

5.15.2. AdCoS Use-Cases

The cognitive distraction model is going to be integrated into the following WP9 AdCoS systems: the TAK Simulator AdCoS, the IAS Test Vehicle, and potentially the CRF Test Vehicle. A detailed description of those systems can be found in D9.2.

The benefit of the MTT for these AdCoS is twofold: It can be used during system validation phase and it can be integrated into the final product. In both cases, the MTT will be connected to the AdCoS, but the usage of its output differs.

In the system validation phase, deriving knowledge about the human operator can be very valuable. While interacting with a prototype or some modules of the AdCoS, the operator's level of distraction can be evaluated. The tool provides feedback whether or not a new system (module) increases or decreases the operator's level of distraction. The output of the MTT addresses in this case the system developer and thus must be part of the development workflow. Here, the multi-modal nature of the distraction estimation plays an essential role since it may provide the system developer with more details about the cause of distraction.

In addition, the distraction model bears the potential to be used online in the final product to classify the driver's distraction not only during testing of a prototype, but also during everyday interaction with the AdCoS. Here, the systems using the level of distraction often only need a single parameter as input. Within these AdCoS' the distraction signal is used for triggering the adaptation of the system with respect to the driver's attentional status. When the driver is distracted, the system will warn the driver giving a warning tone (TAK AdCoS). Is the car autonomously driving and the driver is distracted, the system will not perform manoeuvers that enhance the risk that the driver needs to intervene. Thus, the system will for example not induce overtaking manoeuvers and will instead keep more distance to the pace car (IAS AdCoS).

5.15.3. Input and output

In-vehicle information is needed as input. This includes, but it is not limited to in-car audio recordings, face-tracking data from the driver, and behavioural driving parameters. Audio data involve, e.g., the number of speakers, the amplitude of the noise, speech durations and pauses. By combining face-tracking signals like mouth movements of the driver, it can be identified whether the driver him/herself is speaking or a co-passenger. Other face-tracking data like the blinking rate might give hints about the level of distraction of the driver, since it was found that cognitive distraction increases the blinking rate [[30]] [Behavioural driving parameters as another source for inducing distraction of the driver can be used, as for example the distance to the pace car increases when the driver is distracted. These data will be weighted according to their technical quality and to their correlating strength for distraction, and will

30 Liang, Y., Lee, J.D., 2014. A hybrid Bayesian Network approach to detect driver cognitive distraction. Transport. Res. Part C: Emerg. Technol. 38, 146–155.

further be integrated in order to compute a temporal level of distraction of the driver. Thus multimodal data integration and synchronization needs to be guaranteed.

As regards the output, the tool provides an online continuous measure of the level of distraction provided by a regression analysis, meaning that the level of distraction will be identified using a time window of about 3-10 seconds of the past. The metrics used to quantify the driver's distraction based on in-car information are developed in T5.2. The different measurements will be integrated in RTMaps provided by INTEMPORA.

5.15.4. Current status and functionality

Our approach for the implementation and validation of the distraction model will be based on a number of experiments. We started with a pilot study and with some improvements continued with this setup to measure subjects for our first experiment. The collected data are currently analysed.

In particular, for the pilot study (

Figure 28 a), we compare in-car measurements between a continuous distraction condition and a control condition of concentrated undistracted driving. We implemented a car following paradigm with the driver's task to keep the same distance to the pace car by ensuring readability of the number on the back end of the pace car. At the start of the experiment, subjects performed a practice session of three minutes driving without distraction in order to get used to the experiment and the driving simulator. The pace car drives with varying speeds between 30 and 100 km/h and brakes or accelerates 39 times during a 10 minute drive at randomly distributed locations. After the practice session, all subjects performed both conditions of 10 minutes. For the continuous distraction condition, subjects had to carry out a mathematical calculation task while driving. These tasks were simple mathematical problems (e.g., 22+46 or 9-5) presented through headphones and subjects were asked to give the answer verbally. After eight seconds, the next mathematical problem was presented. All verbal answers were recorded.

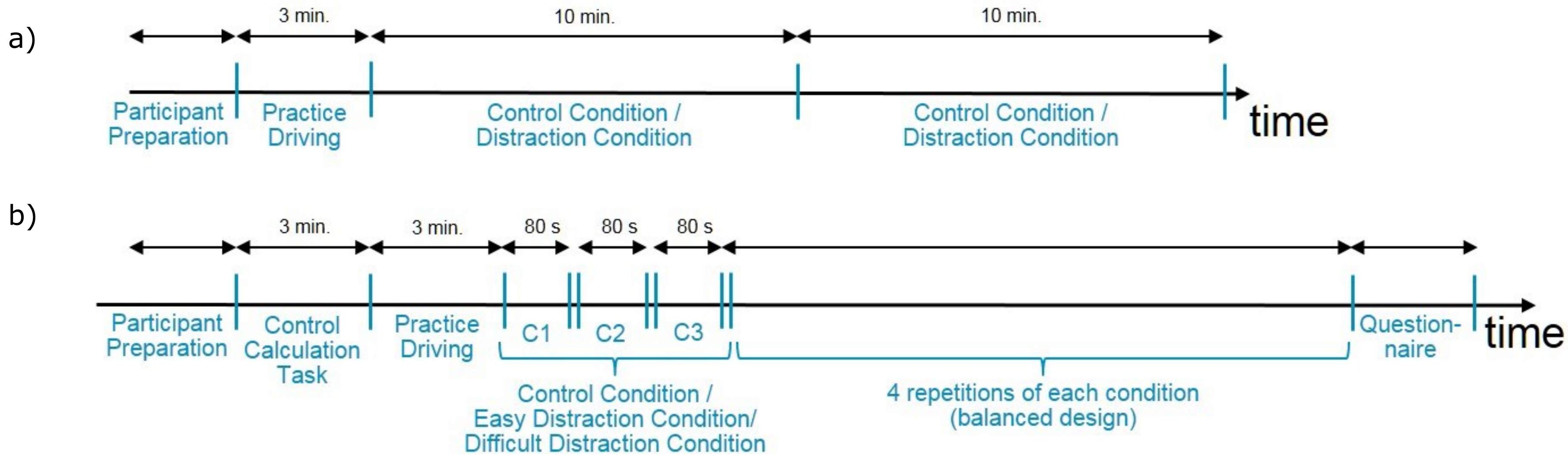

Figure 28. Procedure of the pilot experiment (a) and of experiment 1 (b).

In experiment 1 (

Figure 28 b), subjects started with -in addition to the pilot study- performing a control calculation task in order to verify their ability to solve mathematical problems when focusing only on one task, before carrying out the training driving task and the experimental conditions. Subjects had to achieve at least 70 % answering accuracy. Subsequently, subjects performed a practice driving session similar to the one in the pilot study. The subjects took part in three experimental conditions, one control condition (driving only), and two dual-tasking conditions (driving and calculating simultaneously, with two different levels of mathematical difficulty. All conditions were presented in a run in each of the five blocks in balanced order, and each run lasted 80 seconds. For the math task, within those 80 seconds, every 10 seconds a new problem was presented which the subjects had to solve within 8 seconds. The easy mathematical problems consisted of the summation or subtraction of two numbers between 1-20, with at least one number a single digit (e.g., 4+5), while for the difficult mathematical problems both number were double-digits. At the end of experiment 1, subjects had to fill out a questionnaire assessing the subjective experience of task difficulty, the interference effect on the driving task, and task prioritization.

Analysing these results, we assess parameters responding reliably to cognitive distraction. These parameters include: distance to pace car, reaction times (both for braking and speed recovery), steering wheel jitter, and lateral position jitter. Further parameters will be evaluated for their potential use as features of the cognitive model and will be included step-by-step, e.g. head orientation (which will be relevant in conversation tasks), eye blink, and facial expressions (for emotion recognition). For conversation tasks, audio analysis will be included in the feature set of the cognitive model. Here features used in voice and speech recognition such as pitch and Mel-Frequency Cepstral Coefficients (MFCC) are suitable candidates [[31]] as well as derived features such as emotional content of the utterances.

These parameters will be used as input for the cognitive model computing the level of distraction of the driver. Since features used for our cognitive model will eventually come from different sources (car data, video, audio), synchronisation plays an important role. One tool allowing acquisition of

[31] C. Xu, S. Li, G. Liu, Z. Zhang, E. Miluzzo, Y. Chen, J. Li und B. Firner, „Crowd++: Unsupervised Speaker Count with Smartphones.," in *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'13)*, Zurich, Switzerland, 2013.

multi-modal sensory data is RTMaps, which will be used as platform for implementing our driver distraction estimation component. Preliminary results of driving parameters such as the mean distance to the pace car and its variance as well as the mean reaction time for breaking indicate the effectiveness of the induced distraction through the mathematical problem solving task. Moreover, the blinking rate increases when subjects are cognitively distracted as mentioned before in the literature [[32]].

For the architectural design of the simplified cognitive distraction estimation model, we take the driver's auditory and visual perception into consideration and compute his/her level of distraction based on a resource allocation model (Figure 29). This model from Wickens (2002) [[33]] states that the more a secondary task takes up the same or similar sensory modalities (auditory vs. visual), codes (visual vs. spatial) and processing stages (perceptual, cognitive, responses), the more the secondary task leads to distraction from the primary task. The measured parameters derived from in-car audio recordings, face-tracking information of the driver, behavioural car information (e.g. driving parameters) and environmental information like the distance to the pace car to be followed will lead to conclusions about the allocation of the driver's resources and therefore enable the computation of his level of distraction. The cognitive model can also be useful to help design the feedback system. When for example the driver's visual channel is already heavily loaded, it would be more efficient to provide (feedback) information in a different sensory channel, such as the tactile [[34]], or auditory one.

[32] Liang, Y., Lee, J.D., 2014. A hybrid Bayesian Network approach to detect driver cognitive distraction. Transport. Res. Part C: Emerg. Technol. 38, 146–155.

[33] Wickens, Multiple resources and performance prediction, Theor. Issues in Ergon. Sci., Vol. 3, No. 2, 159-177, 2002

[34] Thurlings, M.E.; Van Erp, J.B.F.; Brouwer, A.-M.; Werkhoven, P., "Controlling a Tactile ERP–BCI in a Dual Task," Computational Intelligence and AI in Games, IEEE Transactions on , vol.5, no.2, pp.129,140, June 2013 doi: 10.1109/TCIAIG.2013.2239294

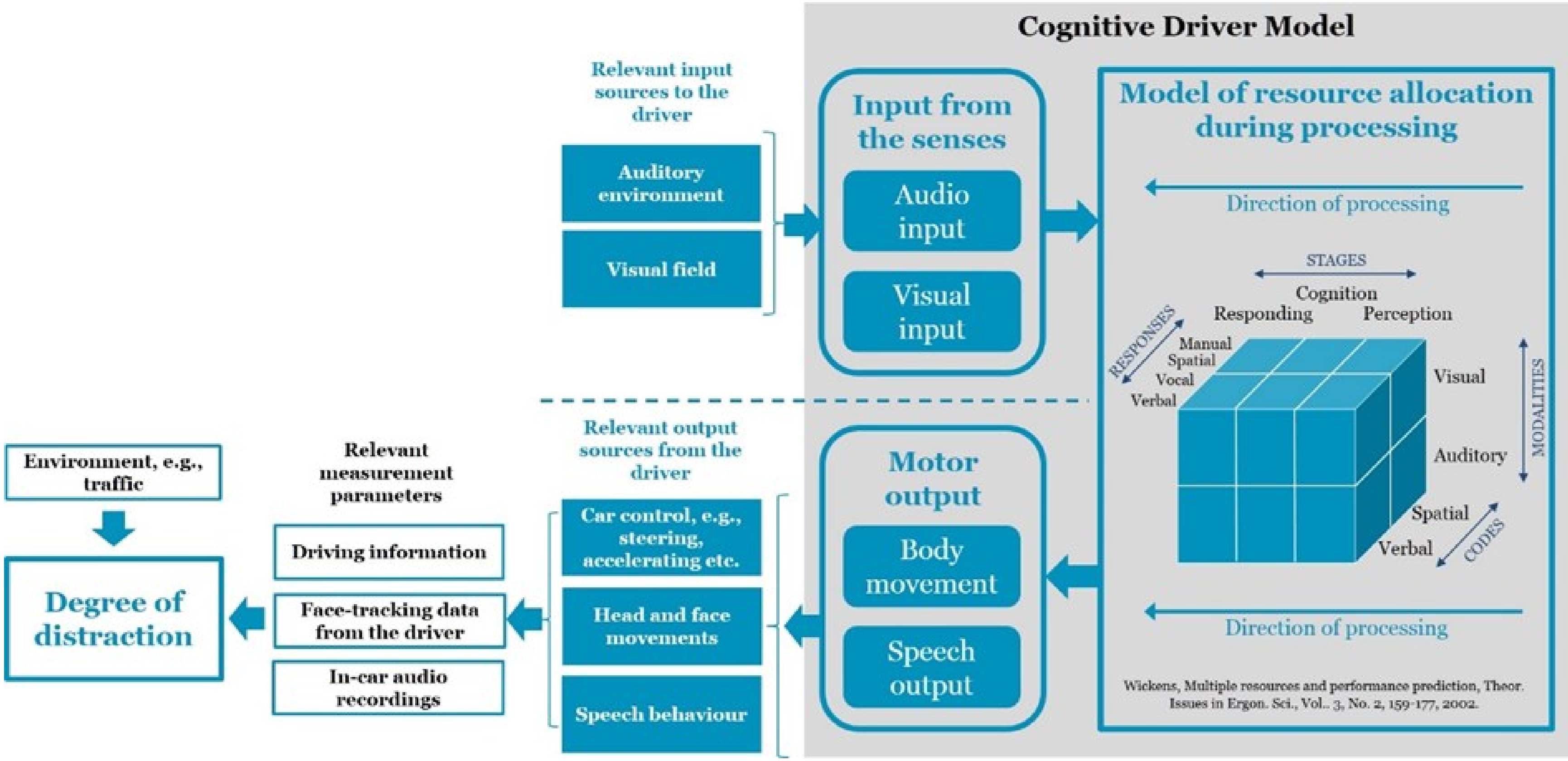

Figure 29. Architectural design of the cognitive model predicting the level of distraction of the driver.

We are currently planning the second experiment. Here, we will induce a more natural conversation condition leading to varying levels of distraction of the driver. Our computational and empirical cognitive model will be trained and tested in the course of this experiment. An acoustic analysis including the detection of the number of speakers, information about the driver's involvement in the conversation (e.g., whether the driver himself is speaking), and possibly the level of emotional content, is used for the prediction of the driver's level of distraction. In addition, face movement information can be exploited to increase the reliability of the distraction prediction.

5.15.4.1. Usage as a tool for the development of the AdCoS

Figure 30 shows the individual steps of the workflow integrating the distraction estimation MTT into the development of an AdCoS using the HF-RTP. Activities specific for the HF-domain are those related to the experimental design, the testing procedure, data analysis as well as the identification of data predicting the level of distraction of the driver. Based on these data the cognitive model will be implemented and evaluated and validated using simulator experiments. Some steps still lack proper tool-support. With the HF-RTP, we expect further refinements of this workflow regarding potential tools to cover these steps.

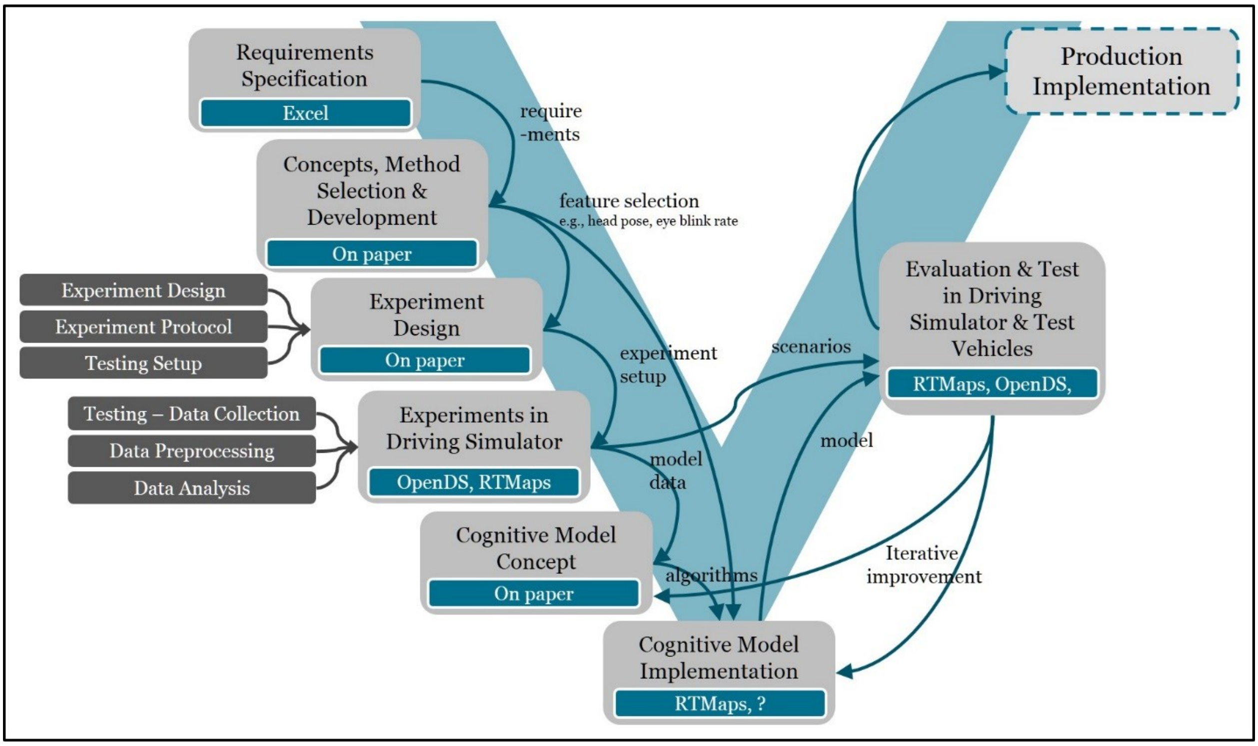

Figure 30. Workflow with potential tools to be used for the development of the AdCoS containing the driver distraction estimation.

In the system validation phase, deriving knowledge about the human operator can be very valuable. The distraction detection MTT may be used either during the design of the system when eliciting parameters for the implementation or when performing preliminary examinations investigating an appropriate setup for the design, e.g. which factors influence the level of distraction of an operator. While interacting with a prototype or some modules of a system during the development process, the operator's level of distraction can be evaluated. The tool provides feedback whether or not a new system (module) increases or decreases the operator's level of distraction (Figure 31). The output of the MTT addresses in this case the system developer and thus must be part of the development workflow. Here, the multi-modal nature of the distraction estimation plays an essential role since it may provide the system developer with more details about the cause of the distraction.

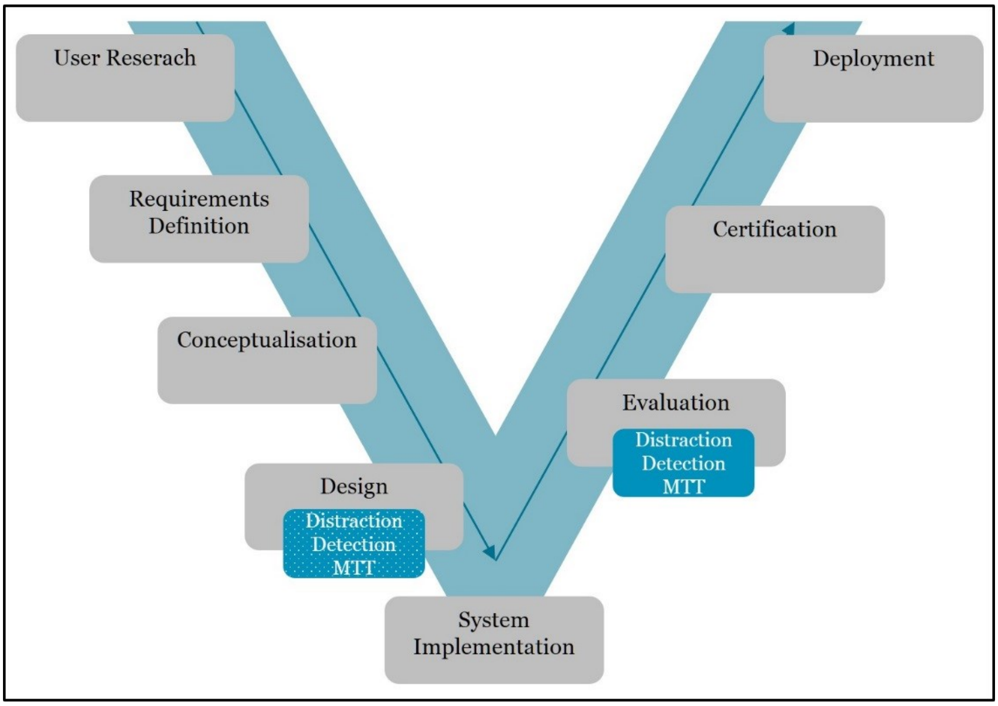

Figure 31. General workflow of a system development process with indications in which process phases the distraction detection MTT may be used.

5.15.5. Integration plan

So far the software has been used as module, therefore its integration will be compliant to the steps for the modules of the Common Integration Plan defined in D1.5 and reported in Chapter 4. In particular, its integration will be accomplished by completing the following steps:

1) Identify their inputs and outputs towards the AdCoS
2) Identify the correspondence between its inputs and outputs and the concepts of the HoliDes meta model.

The details of the integration plan are described in the Annex III (confidential part).

Table 16 shows in details the steps as well as the estimated date to complete the integration.

Table 16: integration plan for the Driver Distraction Classifier module

	Activity	Date
1	First version of module	29/5/2015
2	Feedback from partners	31/8/2015
3	Production version of the module operating in real time	31/3/2016
4	Identification of the correspondence between its inputs and outputs and the concepts of the HoliDes meta model.	As soon as the HoliDes Meta Model is available
5	Documentation	30/9/2016

5.16. Classifier of driver distraction based on data on vehicle dynamics (UTO / CRF)

5.16.1. Summary

Driver distraction and inattention are an important safety concern [[35]]. Deriving knowledge about the human operator can be very valuable in the system validation phase. While interacting with a prototype or some modules of the AdCoS, the operator's degree of distraction can be evaluated.

The purpose of this system is to classify driver distraction based on vehicle dynamics using machine learning techniques.

[35] M. A. Regan, C. Hallett und C. P. Gordon, „Driver distraction and driver inattention: Definition, relationship and taxonomy," *Accident Analysis & Prevention,* Bd. 43, Nr. 5, 2011.

The module provides feedback whether or not a new system increases or decreases the operator's degree of distraction, and this information can be used to design how to adapt the interface of the AdCoS.

In particular, it consists of two modules: the first module works offline and learns a classifier from sensory data. The second module works online and makes predictions on the status of the driver using the knowledge acquired offline.

Therefore the module can be used online to classify the driver's distraction not only during the testing phase of a prototype, but also during everyday interaction with the AdCoS. This online measure of distraction could in turn be used to adapt the degree of automation of the AdCoS to the driver's state.

A combination with the tools developed by BUT (as described in sections 5.4 and 5.5) and TWT (as described in section 5.15) is possible to increase the module's predictive power.

5.16.2. AdCoS Use-Cases

This module will help the lane change assistant functionality and it will be applied in the frontal collision use case from WP9, and also used in the overtaking use case of the same WP9.

In the system validation phase, we should derive knowledge about the human operator cognitive states, in particular those related to visual distraction. In fact, while interacting with the AdCoS or with some devices of the prototype (e.g. the SURT), the operator's degree of distraction can be evaluated. The module provides feedback whether or not the operator's is distracted and, possibly, about his/her level of distraction (therefore, it also provides an indirect measure if a new system (module) increases or decreases the distraction level). Here, the multi-modal nature of the distraction estimation plays an essential role since it may provide the system developer with more details about the level of the distraction.

5.16.3. Input and outputs

Data from the system dynamics during driving are needed. System dynamics need to be stored in way that enables linking them to certain system states, e.g., inputs from the user to the system. Thus multimodal data integration and synchronization needs to be guaranteed.

As regards the outputs, the tool provides a qualitative description of the driver's degree of inattention. At first, we decided to focus on just two degrees, distracted and not distracted, and we are working on revising this to three or more degrees of distraction, depending on how long the driver stays distracted.

5.16.4. Current status and functionality

Both theoretical developments and experimental validation and comparison have been carried out during the first year of the project. We focussed our attention toward the neural network paradigm in order to build a suitable tool for driver distraction detection.

Neural networks are computational models made up of an interconnected group of simple units, called neurons, which processes information coming from the external environment to identify complex relationships and provide consistent output signals. They are used in various disciplines such as neuroscience, mathematics, statistics, physics or engineering to solve real problems of classification, regression, diagnosis, clustering, control, automation, etc.

In particular, we studied and compared the results obtained by two different learning algorithms: the first is the well-known and largely used Back-Propagation, which is an iterative method; the second is the Extreme Learning Machine algorithm, developed more recently, that uses matrix pseudo-inversion techniques. Of the latter we also deepened regularization methods to improve its stability.

Single Layer Feedforward Neural Networks (SLFN) training was mainly accomplished by iterative algorithms involving the repetition of learning steps aimed at minimising the error function, over the space of network parameters; such methods are slow, computationally expensive and can easily lead to poor local minima.

Recently some new techniques based on matrix inversion have been developed, becoming the basis of a complete and exhaustive machine learning theory with the work by Huang and colleagues [[36]]. Their results state that SLFNs with randomly chosen input weights and hidden layer biases can learn distinct observations with a desired precision, provided that activation functions in the hidden layer are infinitely differentiable.

Besides, output weights are determined by Moore-Penrose generalised inverse (or pseudo-inverse) of the hidden layer output matrix, so iterative training is no more required.

Extreme Learning Machine (ELM) is an algorithm characterized by the fact that input weights are randomly assigned, while output weights are computed using the analytical procedure of pseudo-inversion.

With this method the training reaches the result in one step: ELM can find the minimum training error without using an iterative procedure, notably reducing the computational costs and with good generalization. The only parameter that needs to be kept under control is the choice of input weights during the first phase: error, in fact, depends on this random choice and therefore more attempts are required to reach a good result.

We have compared the Back-Propagation learning algorithm with our Matlab implementation of Extreme Learning Machines to predict the state of drivers' distraction, in particular the visual one, due to external disturbances, in terms of both prediction accuracy and learning running times.

For our comparisons, we used data related to distraction and vehicle dynamic, collected by means of dedicated experiments using a static driving simulator, derived from a previous study ([[37]] for more details). We are currently designing experiments and we will collect similar data on a real car in the near future of the project. The used data are a very good

[36] G.-B. Huang, Q.-Y. Zhu und C.-K. Siew, „Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks.," *Extreme Learning Machine: Theory and Applications,* Nr. 70, pp. 489-501, 2006.

[37] F. Tango und M. Botta, „Real-Time Detection System of Driver Distraction using Machine Learning," *Intelligent Transportation Systems,* Bd. 14, Nr. 2, 2013.

approximation of the real ones that will become available in the next months.

In particular, the vehicle dynamic data considered are the following:

- Speed [m/s]
- Time To Collision [s]
- Time To Lane Crossing [s]
- Steering Angle [deg]
- Lateral Position [m]
- Position of the accelerator pedal [%]
- Position of the brake pedal [%]

These values are directly available on the prototype vehicle CAN bus (the same one installed on the real vehicle). The frequency of data collection was 20 Hz (1 data-point each 0.05s), which is the output rate of the simulator. Values are then averaged over a period of 1.8s in order to be consistent with the target variable (distracted or not-distracted). For the time being, we just considered only two possible levels of driver distraction.

Because of the way the experiment was designed, we consider here the visual distraction (eyes off the road). Although we cannot directly address other types of distraction (e.g. cognitive) by this experiment, nonetheless visual distraction is associated with greater odds to crash-relevant conflict than cell phone conversation (cognitive distraction).

We made a large number of experiments by varying the learning algorithm parameters, such as the number of neurons, learning rates, number of training instances, etc. Anyway, here we only report the best results obtained by building and testing a model on a given subject, namely Subject 2, comparing the two training algorithm Back-Propagation and Extreme Learning Machine. Table 17 reports training times and correct classification rates for a FeedForward Neural Network trained with Back-Propagation, while

Table 18 reports training times and correct classification rates when trained with ELM.

The best performance has been obtained for 1660 hidden neurons with a correct rate equal to 94.985 %; in this case the training time was 64.8652 seconds. We note that, although the number of neurons is very high, the training time is acceptable as the ELM algorithm dramatically reduces the learning time.

Compared to the common approach employed by gradient-descent method that iteratively adjusts weights and biases, the performances are similar but computational time is significantly lower.

Table 17: Results for Back-propagation training algorithms

# Neuron	Learning Rate	# epochs	Training time (s)	% correct predictions
5	0.3	500	34.46	83.3613
100	0.3	500	511.42	90.042
100	0.5	500	525.53	90.7773
100	0.3	1000	1163.25	91.3025
100	0.5	1000	1224.82	92.605
100	0.5	3000	3636.78	94.1807
100	0.5	4000	4878.34	94.4118
500	0.5	4000	24429.46	**94.8319**

Table 18. Results for ELM training algorithms.

# Neuron	Training time (s)	% correct predictions
100	0.8424	83.905
401	4.9764	89.713
778	18.8917	92.156
1120	25.5686	94.106
1660	64.8652	**94.985**

The next steps will be to apply the method on the newly collected data, check the results and further improve the classification accuracy. Moreover, at present we just considered two levels of distraction, and we would like to extend the approach to consider more than two levels of distraction.

5.16.5. Integration plan

Its integration will be compliant to the steps for the modules of the Common Integration Plan defined in D1.5 and reported in Chapter 4. In particular, its integration will be accomplished by completing the following steps:

1) Identify their inputs and outputs towards the AdCoS
2) Identify the correspondence between its inputs and outputs and the concepts of the HoliDes meta model.

The details of the integration plan are described in the Annex III (confidential part).

Table 19 shows in details the steps as well as the estimated date to complete the integration.

Table 19: integration plan for the Classifier of driver distraction based on data on vehicle dynamics module

	Activity	Date
1	First version of module	29/5/2015
2	Feedback from partners	31/8/2015
3	Production version of the module operating in real time	31/3/2016
4	Identification of the correspondence between its inputs and outputs and the concepts of the HoliDes meta model.	As soon as the HoliDes Meta Model is available
5	Documentation	30/9/2016

6. Conclusions and outlook

In WP5, 15 MTTs have been designed and updated.

Since the last deliverable (February 2015), one new tool has been added (U-DAT), a tool has been removed (BVT) and a few MTTs have significantly increased scope and purpose thanks to the tight collaboration with the AdCoS they are meant for.

Moreover, an intense activity has been carried out to identify relevant metrics and subsequent activities to reflect on the feasibility of reaching the overall project goals. A Common Metrics Methodology has been defined and applied to 4 AdCoS'. It allowed assessing that it is effective to:

- Correlate the project goals with the AdCoS' goals
- Identify concrete metrics to evaluate the performance of the AdCoS and, as a consequence, the achievement of the HoliDes project objectives.
- Describe in an intuitive and clear way the rationale for the selection of some MTTs for each AdCoS and their benefits.

By starting from the results of this methodology and the feedback provided by the AdCoS owners, new requirements have been identified and old requirements have been reviewed.

A Common Integration Plan has been also drafted in collaboration with the other scientific WPs (WP1-WP4) in order to define a common methodology for the integration of all MTTs into the HF-RTP. It has been described in D1.5, summarized in this deliverable (as well as the other WP2-WP4 deliverables) and applied for the integration of all MTTs into the HF-RTP.

In the next months, we planned to focus our activities mainly in the following tasks:

- Improve the MTTs thanks to the feedback of the AdCoS owners we are intensively collaborate with.
- Extend the methodology to all AdCoS and increase the number of metrics identified for each AdCoS.

- Define specific plans for the empirical evaluation of each AdCoS (to support the AdCoS owners in the Tx.8 (where x = [6|7|8|9]).
- Explore the opportunities for re-using MTTs in other AdCoS or domains, also by exploiting the results of the Common Metrics Methodology to find similarities between needs (and thus solutions) of different AdCoS and/or domains.