	HoliDes **Holi**stic Human Factors **Des**ign of Adaptive Cooperative Human-Machine Systems	

D5.3 - Techniques and Tools for Empirical Analysis Vs1.0 incl. Handbooks and Requirements Analysis Update

Project Number:	332933
Classification:	Public
Work Package(s):	WP5
Milestone:	M2
Document Version:	V2.0
Issue Date:	January 15, 2015
Document Timescale:	Project Start Date: October 1, 2013
Start of the Document:	Month 13
Final version due:	Month 15
Deliverable Overview:	D5.3
Compiled by:	Paul Kaufholz, PHI
Authors:	Borchers, Svenja (TWT) Botta, Marco (UTO) Collina, Simona (SNV) Dotzauer, Mandy (DLR) Gardas-Schmidt, Dora (HFC) Griesche, Stefan (DLR) Herout, Adam (BUT) Ihme, Klas (DLR) Käthner, David (DLR) Kolrep, Harald (HFC) Landini, Elisa (REL) Larsen, Morten (AWI) Martin, Denis (TWT) Moravek, Zdenek (HON) Raucina, Maros (HON) Tango, Fabio (CRF)
Reviewers:	Ian Giblet, CAS-UK Jan-Patrick Osterloh, OFF
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
14.05.2014	Initial Version	DK, KI
12.11.2014	First compiled version	PK
15.01.2015	Update to V2 after review V1	PK
21.01.2015	Changes according to the reviewer	DK
22.01.2015	Changes in the contributions	DK
23.01.2015	added references	DK
23.01.2015	split D5.3 in public and confidential part, update to V3	DK
27.01.2015	included Ian's comments / suggestions	DK
28.01.2015	added commitment for prototypes (outlook); added UCs in the UC landscape table; deleted authors' names from titles	DK
03.02.2015	added author's names on the title page	DK

Table of Contents

List of abbreviations

Terms

AdCoS	Adaptive Cooperative Human-Machine System
(HF-)RTP	(Human Factors-) Reference Technology Platform
ICA	Index of Cognitive Activity
MTT	methods, techniques and tools
T5.X	one of the five tasks of WP5
UML	Unified Modelling Language
WP	work package

Project partner

AWI	AnyWi
BUT	Brno University of Technology
CRF	Centro Ricerche di Fiat
DLR	German Aerospace Center
ERG	Ergoneers
HFC	Human-Factors-Consult GmbH
HON	Honeywell
IAS	Ibeo Automotive Systems
PHI	Philips
REL	RE:Lab
SNV	Universita Degli Studi Suor Orsola Benincasa
TWT	TWT GmbH Science and Innovation
UTO	Universita Degli Studi Di Torino

People

DK	David Käthner, DLR
KI	Klas Ihme, DLR
PK	Paul Kaufholz, PHI

1 Introduction

This deliverable reports the progress of the HoliDes consortium to develop methods, techniques and tools (MTTs) for the Human Factors-Reference Technology Platform (HF-RTP), version 1.0. For work package 5 (WP5), it concludes project cycle I. During this cycle, as a first step we received the requirements from the application work packages WP6-9. After an analysis of these requirements (cf. D5.1), we selected those metrics and methods to be developed in WP5, which could best meet the AdCoS owners' needs. Having documented those MTTs as HF-RTP 0.5 in D5.2, the first instantiations of these techniques and tools were made. The result of this work is described in this document.

For each method, technique or tool, a detailed description is provided concerning data the MTTs receive, data they provide, their current functionality as well as specific and five definitive and further potential use cases. These use cases (see Table 1) originate from the four HoliDes domains Health, Aerospace, Control Room and Automotive.

In our definition, a method is a general way to solve a problem. This could be the use of task analysis to answer a general design question. A technique is a concrete instantiation of such a method, as would be the application of a specific form of task analysis to the development and evaluation of an adaptive system. Finally, a tool is a technique, which has been realized as either hard- or software. Such a tool could be a program that aids the collection and organisation of observations during the task analysis.

The MTTs created in this work package follow the objective to "Develop techniques and tools for empirical analysis of Adaptive Cooperative Human-Machine Systems (AdCoS) against human factors and safety regulations." To achieve this objective and provide the application work packages 6–9 (WP6-9) with methods that best suit their needs the starting point of our work has been the AdCoS requirements from WP6-9. Some of these requirements describe genuine AdCoS functionality, while others relate to MTTs necessary to develop AdCoS functionality or aspects of the design process itself.

Consequently, the purpose of WP5's MTTs is to enable, aid, and assist the empirical analysis of adaptive, cooperative systems, or to act as part of these systems functionality. The actual outcome of the work presented here will be software tools, empirical results as the basis for system functionality and design decisions, but also procedures and algorithms and their implementation. All of these efforts help realize the human centred design process as e.g. described in ISO 9241-210, both during design and evaluation.

For a quick overview over WP5's MTT landscape, Table 1 shows both names and short descriptions of all methods, techniques and tools as well as the use cases they are being applied to.

Table 1. MTT landscape of WP5.

partner	domain	use case	contribution title D5.3
AWI	HEA	safe parallel transmit scanning use case	Modelling of AdCoS data from a means-ends perspective
AWI	HEA	guided patient positioning	HF Filer
BUT, HON	AER	cockpit	Operator state detection from implicit hand gestures
BUT, HON	AER	DiVA	Detection of operators' head orientation
DLR	AUT	automated overtaking manoeuvre	Methods and techniques for the driver adaptive parameterization of a highly automated driving system
DLR	AUT	automated overtaking manoeuvre	CPM-GOMS Task Analysis of a Lane Change for manual and automated driving Author
DLR	AUT	automated overtaking manoeuvre	Theatre Technique for acceptance tests during AdCoS design
HFC	HEA	MRI UC1 safe patient transfer, MRI UC2 Guided patient positioning	Human factors and safety regulations and guidelines for metrics
HFC	HEA	MRI UC1 safe patient transfer, MRI UC2 Guided patient positioning	Tests for cognitive task models
REL	AUT,	IRN control room	Behavioural Validation Tool

	HEA		
SNV	AUT, CTR	IRN control room	Empirical analysis and validation methods of cognitive and communicative processes in automotive and control room domain
TWT	AUT	automated overtaking manoeuvre, frontal collision warning	Detection of driver distraction based on in-car measures
UTO, CRF, SNV	AUT	frontal collision warning	Detection of driver distraction based on data on vehicle dynamics

2 MTTs for Integration

2.1 AWI: Modelling of AdCoS data from a means-ends perspective

Author: Morten Larsen

2.1.1 Summary

As an aid to aggregate data, this method provides functional modelling from a means-ends perspective to help analyse whether a given AdCoS helps the operator achieve key goals in the operation of the system. This type of modelling will help transform measured and observed data into information about high-level system state, thus allowing system evaluation to be performed in terms closer to actual operation.

The means-ends modelling for AdCoS can be used at different stages of the AdCoS design process, namely during system development and validation. 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 tree-structured (and hence hierarchical) 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 tree structure will effectively split the description of the AdCoS, along with the associated empirical data, into branches related to specific functionality of the AdCoS.

This split helps structure the work in the design-development-evaluation cycle and can be useful in reporting, whether with internal or external partners – or with public authorities in the course of approval and certification processes. As examples of this can be mentioned that branches of the model tree that have passed all evaluation items can be marked as completed in the design project management plan, or feedback for improvement of an AdCoS design can be structured by functionality rather than, say, the sequential steps of the user actions.

In an evaluation setting, the hypothesis is that the means-end modelling of tasks will help verify if single, identifiable goals for the operation of the controlled entity (car, hospital workflow, energy distribution network, etc.) are achieved.

2.1.2 AdCoS Use-Cases

This method will be applied to the use cases of the healthcare domain (WP6). No final decision on the AdCoS has been taken. Currently, the method is under consideration for the Guided patient positioning AdCoS.

This method provides a framework to structure the matching of behaviour to procedures (whether official or not) for the purpose of providing feedback to the AdCoS design process. The method will require that the procedure be re-modelled into a means-end tree, and as such is only feasible for procedures where the actions can be associated with a purpose.

The addressed requirements can be found in the confidential part of this deliverable.

2.1.3 Input

The means-end (e.g. [1]) model must include the criteria for a goal to be considered as achieved, these criteria must be either quantifiable or otherwise verifiable with satisfying precision. It is up to the modeller to determine these criteria, as 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) that 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 far an operator is keeping the equipment from its safety margins.

2.1.4 Output

As an output of applying the method, the AdCoS designers will get information about 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.

2.1.5 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 1. 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.

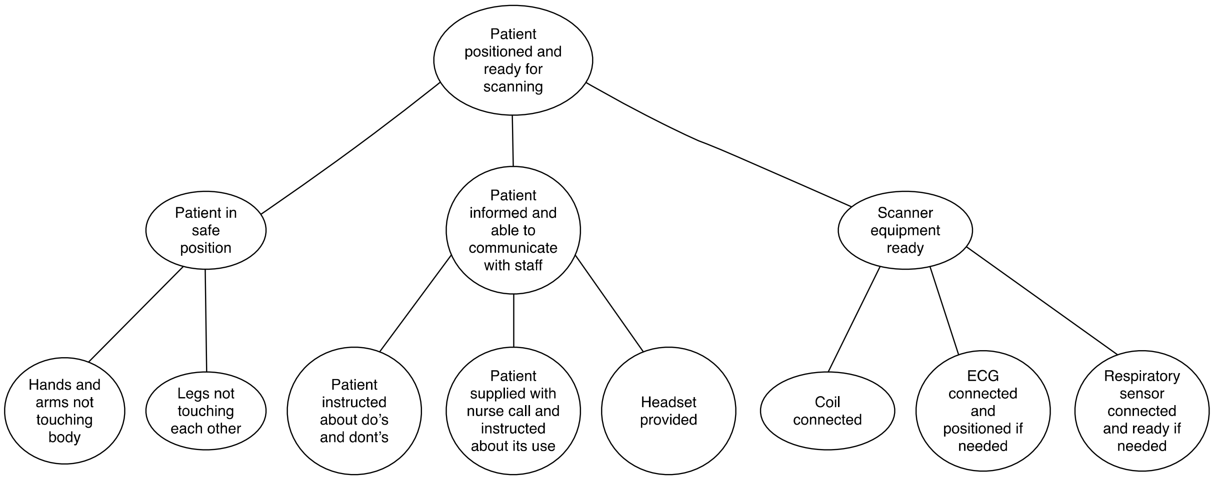


Figure 1 A simple means-end analysis of the Guided patient positioning use case from the healthcare domain.

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.

Further work in this direction will require definitions of actual AdCoS' and situations for their use as well as realistic data, whether obtained from simulations or from real use. For this reason, will the means-end analysis not be developed further until this necessary input is available.

2.2 AWI: HF Filer

Author: Morten Larsen

2.2.1 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

The description here uses the generic term "evaluation", 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.

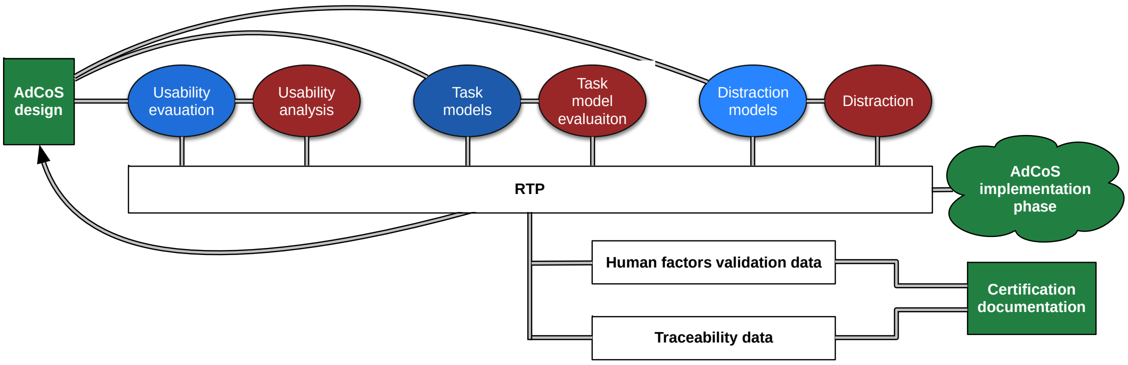


Figure 2. Schematic representation of the scenario with RTP compliant human factors tools in an RTP instance.

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.

2.2.2 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.

The addressed requirements can be found in the confidential part of this deliverable.

2.2.3 Input

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.

2.2.4 Output

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 3 in two main strands – the evaluation plan and the relative report.

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 below in Figure 3.

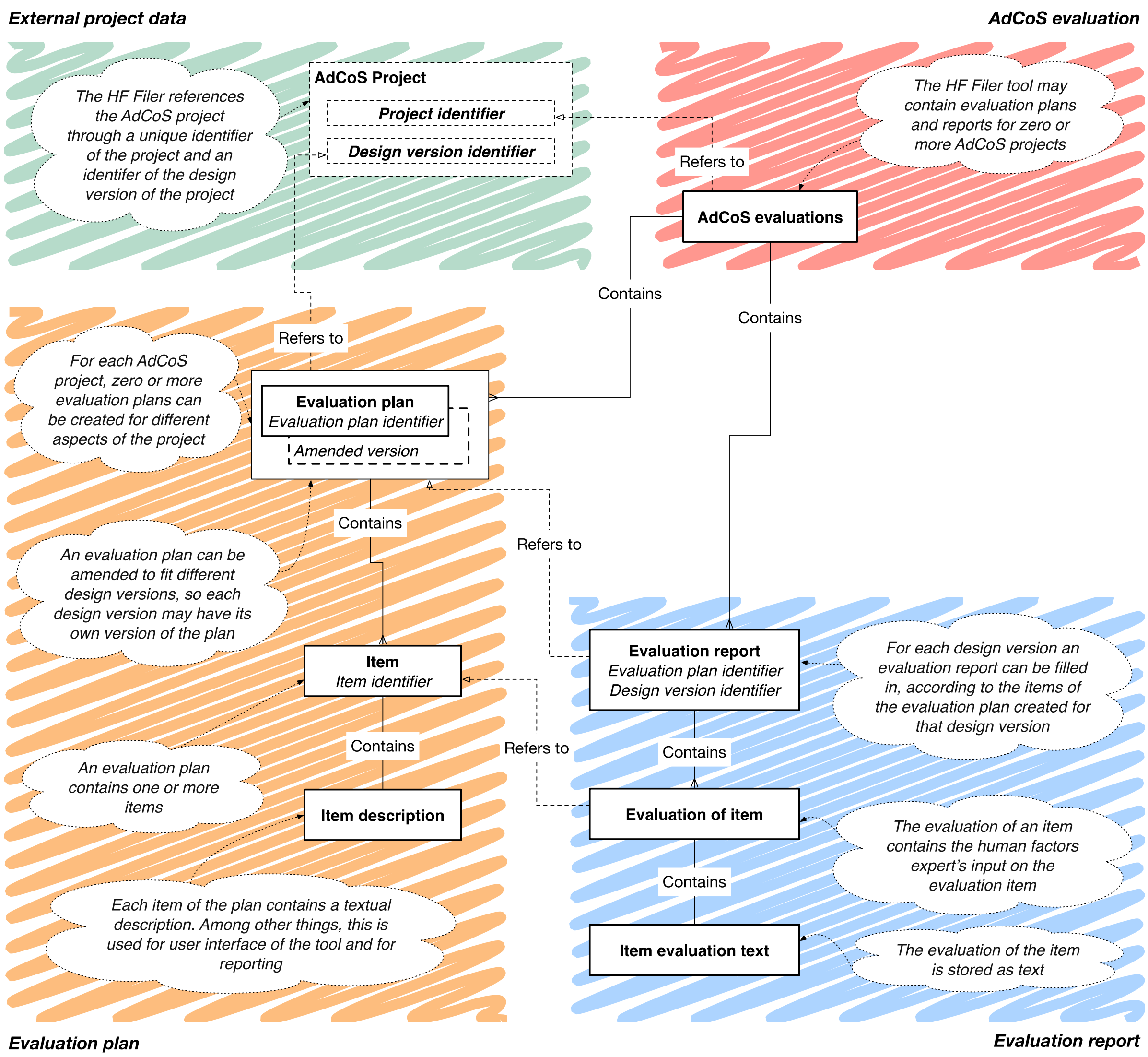


Figure 3. Annotated diagram of the data structure as seen from the RTP.

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 2. 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
```

2.2.5 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.

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.

2.3 BUT / HON: Operator state detection from implicit hand gestures

Authors: Adam Herout (BUT), Maros Raucina, Zdenek Moravek (HON)

2.3.1 Summary

In human to human communication non-verbal cues such as hand gestures can transmit valuable information [2]. 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.

Validation of a prototype aims at revealing flaws in the design with respect to how easily and efficiently the user uses the prototype. Besides 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.

2.3.2 AdCoS Use-Cases

The tool will be used in connection with DivA AdCoS 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.

The requirements addressed by the tool are listed in the confidential part of this deliverable.

2.3.3 Input

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.

2.3.4 Output

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.

2.3.5 Current status and functionality

The human body is a deformable object. Therefore, we use a pictorial structure to represent pose and in order to decompose the upper body into a collection of parts in a deformable configuration. Each part encodes local visual properties of the object, and the deformable configuration is characterized by spring-like connections between certain pairs of parts. Ten body parts were used. Using the pictorial structure representation, the best match of a learned pose model to a novel image is found by minimizing an energy function that measures both, a match cost for each part and a deformation cost for each pair of connected parts.

We use a 2-layer Random Forest as part detector for each part. The output of these detectors is then input to the Pictorial Structure for pose estimation. The performance of this system is evaluated on the pose database collected as part of the project work. We collected a gesture dataset in Honeywell flight simulator. The gesture set contains hand gestures designed for controlling cockpit displays, e.g. object control, map/list/chart control, map control, activate and decline. Thirteen subjects attended the recording session. They sit on the pilot's seat in the cockpit. They were asked to consider the display right in front of them as the multifunctional display to be controlled. Towards the display, they performed the gesture set once with the right hand and once with the left hand. A Kinect camera was used to record RGB frames along with the depth information. The resolution of each frame is 640×480 pixels. Figure 4 below in this section shows some sample frames in this dataset. To speed up the pose annotation process, we use a semi-automatic approach. We first let Kinect create an initial annotation of the pose. However, Kinect annotations are often imprecise and/or unreliable – there are often small shifts of body parts from their correct positions. To overcome this limitation, all annotations were manually corrected. The result is 6322 frames of precisely annotated pose images.

From the 6322 frames in the pose image database we collected, we use 4632 frames for training and rest 1690 frames for testing. Training is done separately for body parts and connections. Two layers of Random Forests are trained as detectors for each body part, where the second layer works like a Hough Forest. A Gaussian distribution model is trained for each connection.

We run pose recognition on 1690 test images from the testing image database and compared the results to the pose annotation. People do not overlap in test images and training images. To quantify the recognition accuracy, we measure the joint localization error as a fraction of the upper body size. This measurement is well established for other computer vision tasks, e.g., fiducial point detection. It is independent of the actual size of the image and more precise than common measures derived from bounding box-based object detection.

Comparing body parts, recognition accuracy is the highest for head and shoulder parts, followed elbows, wrists and hands. Head and shoulder centre are reliably detected because they look similar in the whole image dataset (people all face the camera and show the frontal view of their shoulder centres). When shoulder centre is detected, the possible positions for shoulders are relatively certain. In contrary, hands, elbows and wrists have a large field of possible positions. They also have a large variance of appearance – they look different due to movement, clothing and viewing angle.

Figure 4: Sample frames from the gesture dataset collected in the cockpit.

2.4 BUT / HON: Detection of operators' head orientation

Authors: Adam Herout (BUT), Maros Raucina, Zdenek Moravek (HON)

2.4.1 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 will be recorded. Computer vision techniques will be used to enable automated analysis of the video sequences.

Deriving knowledge about the human operator can be 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 bears the potential to be used in real-time to detect a likelihood of missing significant information in the environment. Based on the head direction, the elements in the 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 may be considered as missed and the system should adapt to regain attention.

2.4.2 AdCoS Use-Cases

The tool will be used in connection with DivA AdCoS to support design development and also real-time pilot state monitoring. In the first case, the applicability of the tool for the design development will be evaluated. Results should define situations in which the tool can be used instead of more expensive and intrusive eye-tracking.

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

The requirements addressed by the tool are listed in the confidential part of this deliverable.

2.4.3 Input

Videos of the human operator's head while accomplishing the task are needed as input for the tool. Videos 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.

2.4.4 Output

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.

2.4.5 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 2 below in this paragraph shows several sample frames. 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).

We used Random Regression Forests with simple node tests based on intensity in patches around SURF keypoints (Speeded-Up Robust Features, a standard feature detector in the computer vision domain) and used information gain for the selection of binary tests. We tested this solution on Pointing'04 and AFLW datasets, but the results were unsatisfactory. Our further work will be to improve this implementation of Random Forests by using differently generated node tests and experimenting with its parameters. According to [3], using Gabor features should provide more accurate results while still be computationally inexpensive. Also, implementing LDA (Linear Discriminant Analysis, a standard statistical tool used in machine learning) as the node test should improve discriminative power of the trees. The improved algorithm will be applied on the dataset collected in the cockpit environment. Further, depending on the speed and accuracy of the Random Forests, we plan to employ tracking techniques to improve the accuracy of the estimated head position.

Figure 2: sample frames from the head movement dataset.

2.5 DLR: Methods and techniques for the driver adaptive parameterization of a highly automated driving system

Authors: Stefan Griesche, David Käthner, Mandy Dotzauer

2.5.1 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.

To achieve these two goals, we need to answer the following research questions:

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?

Question 1 will be answered in WP3, question 2 and 3 are entirely or partially the subject of WP5. To answer question 2 we will have to devise a method to measure the attractiveness of trajectories.

The method of measuring the attractiveness of trajectories will be in the form of a procedure, including techniques such as questionnaires and experiment designs. Further, there will be empirical findings answering the research questions.

It will be applicable to any highly automated driving system. This developed attractiveness measure is especially useful for the evaluation of the final system (or other existing systems).

2.5.2 AdCoS-Use Case

This method will be applied to the automated overtaking manoeuvre use case of WP9.

The addressed requirements can be found in the confidential part of this deliverable.

2.5.3 Input

- Algorithm to extract drivers' driving style from WP3
- A set of possible trajectories is required. In a first step, these trajectories are generated by the DLR's highly automated driving system. Possibly trajectories generated by the IAS-system will be used later on.
- Driver Behaviour and Sensation Seeking questionnaires to explore possible relationships between driver personality, driving style and preference for certain trajectories or manoeuvres

2.5.4 Output

- Answers to questions 2 and 3 (see 2.5.1) including method and techniques of how to assess the attractiveness of trajectories of highly automated driving systems

2.5.5 Current status and functionality

As of the end of October 2014, a first study has taken place, with at least one more being scheduled by April 2015. The purpose of the study is to evaluate the possibility to assess individual drivers' driving style from manual driving data.

Our study's methodological approach follows WP9's lane change use case. While driving on a two lane highway, subjects are instructed to generally keep on the right lane but overtake slower cars if necessary. The presence of cars on the left lane and their speed difference to the ego car were varied. All subjects were presented with a baseline condition at the beginning and the end of the study, and two trials of each of two conditions in random order.

<u>Baseline condition</u>

In the baseline, we used a traffic set on the right a lane with a fixed time headway of 8 seconds between the vehicles. The velocities of the other vehicles were set to 100 km/h. Participants were instructed to drive with a speed of 120 km/h and to overtake slower vehicles. The runtime of the scenario was 500 seconds (see Figure 5).

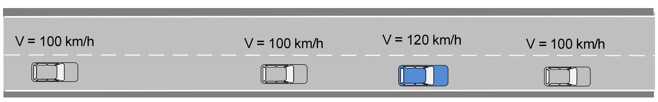


Figure 5: Scenario for the baseline.

Condition A

For condition A the time headway of the vehicles in the traffic on the right lane was set to 12 seconds. The velocity of those vehicles was again 100 km/h. Participants were instructed to drive with a speed of 120 km/h and to overtake slower vehicles. If the distance between the ego vehicle and the lead vehicle was smaller than 280 m another vehicle popped up 250 m in behind the ego vehicle. This vehicle was directly performing a lane change to the left lane. We therefore call this vehicle in the following “lane changer”. The velocity of the lane changer was set to 140 km/h. The lane changer is used to disturb the overtaking manoeuver of the participant. Depending on the participants’ action two cases were possible:

First Case: The lane changer overtook the participant earlier than the participant overtook the lead vehicle. In this case the lane changer accelerated to 170 km/h to avoid an influence in other overtaking manoeuvers later in the scenario.

Second Case: The participant overtook the lead vehicle before the lane changer overtook the participant. In this case the lane changer decelerated to 120 km/h and performed a lane change to the right after 4 seconds. The intention is the same as in the first case. This was used to avoid any influence of the lane changer in other overtaking manoeuvers later on in the run.

After the participant overtook the lead vehicle the procedure started from the beginning for the next lead vehicle. The runtime of the scenario was about 500 seconds. In this period participants experienced nine overtaking manoeuvers.

Condition B

Condition B is similar to condition A. The only difference is the velocity of the lane changer, which is set to 160 km//h. To have the similar disturbance of the lane change as for condition A the distance criteria to the lead vehicle was set to 220 m. Moreover the lane changer popped up 300 m behind the ego vehicle instead of 250 m. Therefore, at a time headway of 3.6 seconds to the lead vehicle and a time headway of 2.2 sec to the lane changer, conditions A and B were identical beside the velocity of the lane changer (see Figure 6: Overview of experimental scenario.

.

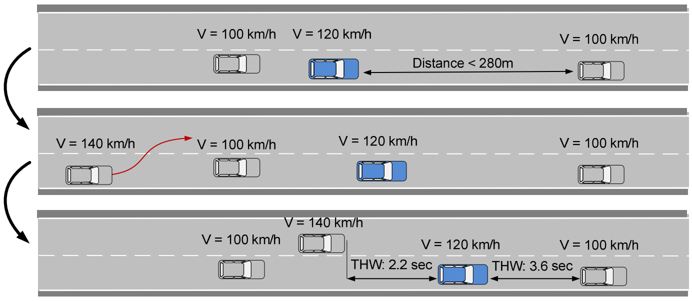


Figure 6: Overview of experimental scenario.

2.5.6 Deviation from initial plan (i.e., from D5.2)

These methods had not yet been introduced in D5.2. Initially, we had not envisioned the need for such an explicit method development. It arose only as a need during the system development in WP3/9. The deviation from D5.2 therefore consists in the additional development of this method to our work plan.

2.6 DLR: CPM-GOMS Task Analysis of a Lane Change for manual and automated driving

Authors: David Käthner, Klas Ihme

2.6.1 Summary

Psychology's understanding of lane changes is cursory at best and not suited to accurately predict human behaviour, especially when mixing manual driving and automated driving, as is planned in the IAS/DLR lane change use case in WP9. With traditional theories and methods, we cannot give well founded advice concerning design decisions of handover-of-control between a human and a cognitive agent. We decided therefore to conduct an in-depth task analysis, yielding a model which allows prediction of human behaviour during manual lane change. Such a model lets us simulate how the introduction of machine agents and adaptation changes the driving task. This model can also serve as a basis for cognitive driver models in WP2.

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. With our task analysis application in HoliDes, we will compare manual driving with driving with an adaptive automation in a lane change situation. We will provide a description of this example application in form of a tutorial or handbook.

This MTT's output will be the description of how to apply CPM-GOMS to the lane changing use case, empirical results

2.6.2 AdCoS Use-Cases

This method will be applied to the automated overtaking manoeuvre use case of WP9, including its decomposition (e.g. lane change, passing of vehicles, approach of vehicles).

In principle, the developed task analysis method in its final stage shall be applicable in several stages of AdCoS design and validation across application domains.

The addressed requirements can be found in the confidential part of this deliverable.

2.6.3 Input

Goals, operators, means, selection rules (GOMS) task analysis approaches are based on 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 with which goals and sub-goals the task of changing lanes should be modelled. This decomposition will be based on previous internal work and published literature on the subject as well as our intuition. To adequately model the resources of the entire system (human *and* machine agents), a system decomposition in a Cognitive Work Analysis-manner (Vicente 1999) can be necessary.

The next step will be data collection in the field with an instrumented vehicle. In addition to the usual CAN-data we will record the drivers' head and eye movements, their feet and hands, the surrounding traffic. If feasible we will also record the drivers' vocal explanations of their current goals and actions, while they are driving.

As the system under development does not exist yet, we necessarily have to do considerable guesswork on the user behaviour with a not-yet-existing AdCoS. For improvements of already existing systems or modelling efforts during later stages of the system development, user interaction with actual system variants can be observed.

2.6.4 Output

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. For integration into the HF-RTP, this description may be transformed into a formal language that is compatible with the aspired OSLC standard.

2.6.5 Current status and functionality

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 (John 1990; John & Kieras 1996). With the synchronized data from the data sources video, thinking aloud, driving data as well as eye and head tracking the task model can be built. Starting out with observable operators (e.g. hand and foot movements, eye and head movements), we will insert parameters for these operators and not-observable operators (mental operators). Finally, the flow of information through the model can be inserted, leading to a critical path for the task execution.

With the task model for manual driving, we can simulate different AdCoS-variants. For selected situations, task changes introduced by the automation variant can be traced to specific AdCoS-properties, allowing for well-founded advice on the AdCoS-design. We will validate the modelled AdCoS execution times using Theatre Technique (see section 2.7).

2.7 DLR: Theatre Technique for acceptance tests during AdCoS design

Authors: David Käthner, Klas Ihme

2.7.1 Summary

The Theatre Technique [4] 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. 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 (see [4] for further explanations). This is particularly useful when planned 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, as the one developed in 2.6.

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 transition from machine control to human control (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 (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.

2.7.2 AdCoS Use-Cases

This method will be applied to the automated overtaking manoeuvre use case of WP9.

The addressed requirements can be found in the confidential part of this deliverable.

2.7.3 Input

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.

2.7.4 Output

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 detailed 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.

2.7.5 Current status and functionality

In our current workflow, Theatre Technique is used after the first iteration of the task analysis. Since this first iteration is not yet completed (see above), we have not started to work on the adaptation of the Theatre Technique, yet. Thus, it is planned to start with the advancement of this MTT in the next project cycle.

2.8 HFC: Human Factors and Safety regulations and guidelines for metrics

Authors: Dora Gardas-Schmid, Harald Kolrep

2.8.1 Summary

There are two objectives of the work. In the first rank we derive a list of analysis questions and evaluation criteria from standards and regulations reviewed in WP1. The systematic analysis of the standards and regulations helps to identify the crucial HF aspects and evaluation criteria needed to be taken into account within the development process of AdCoS. The second aim is to define metrics for selected evaluation criterion or criteria. This proceeding also shows the uncovered or poorly covered, but crucial HF aspects of AdCoS, for which metrics/measures are obsolete, insufficient or do not even exist. It is expected that new metrics are needed for the specific aspects of an AdCoS, namely collaboration and adaptation.

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

A metric is more detailed and system specific, its form and validity also depends on the development stage. A metric is intended to allow the empirical investigation and measurement of the degree to which the criterion is fulfilled.

2.8.2 AdCoS Use-Cases

The collected criteria are related to all four domains: aeronautics, health, automotive and control rooms. The proposed metrics however are dedicated to the health care domain (WP6). They should enable the operationalization of an evaluation criterion or criteria relevant for MRI system (MRI UC1 safe patient transfer, MRI UC2: Guided patient positioning). There is the possibility, that one of the MRI Use Cases will be replaced with iXR UC01_3D_acquisition.

2.8.3 Input

Human Factors and Safety regulations and guidelines systematically analysed in WP1 (D1.2). These include inter alia:

- IEC 60601-1-6 General requirements for basic safety and essential performance. Collateral standard: Usability; IEC International Electrotechnical Commission; 2010
- IEC 60601-1-8 General requirements for basic safety and essential performance. Collateral standard: General requirements, tests and guidance for alarm systems in medical electrical equipment and medical electrical systems; IEC International Electrotechnical Commission; 2012
- IEC 60601-2-33 Particular requirements for basic safety and essential performance of magnetic resonance equipment for medical diagnosis; IEC International Electrotechnical Commission; 2013
- IEC 60601-2-43 Particular requirements for basic safety and essential performance of X-ray equipment for interventional procedures; IEC International Electrotechnical Commission; 2010

Human Factors Methods and Techniques: Review and selection of MTTs appropriate for measurement of the selected critereion/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. & Standton, N.(2004) The Handbook of Task Analysis for human-Computer-Interaction.

2.8.4 Output

HFC will deliver a document/table with a set of analysis question and evaluation criteria derived from HF and Safety Regulations. The criteria will be grouped by domains and provided with additional information: system, issuer, reference and evaluation according to i.a. their operationalization potential/possibilities. In the second part of the document the metrics will be described. Metrics will be derived to determine how the 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).

2.8.5 Current status and functionality

The intended output is a work in progress and planned to be accomplished it in February 2015.

- The systematic review of the HF and Safety regulations and guidelines from D1.2 is completed and HFC compiled a list of analysis questions and evaluation criteria. These have been reviewed, evaluated and selected during internal workshop. There are some modifications needed to formalize the list.
  It has been decided to focus on two healthcare use cases by defining the metrics, because deriving metrics proposals for all criteria seems impossible due to empirical and working framework limitations. It is planned to concentrate on two MRI Use Cases (UC1 & UC2), but possibly one of them will be replaced with the iXR UC01.
  The final results will be presented in detail in the next deliverable (D5.4).

2.9 HFC: Tests for Cognitive Task Models

Authors: Dora Gardas-Schmid, Harald Kolrep

2.9.1 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, 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.

2.9.2 AdCoS Use-Cases

It is intended to work with one or two of the following Use Cases from WP6: MRI UC1 safe patient transfer, MRI UC2: Guided patient positioning, iXR UC01_3D_acquisition.

2.9.3 Input

Empirical data collected by task analysis:

We intend to carry out a task analysis to investigate the MRI or iXR system regarding the selected quality criterion/criteria. We will collect data using 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 structure need to be adjusted.

2.9.4 Output

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. 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.

2.9.5 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 basic GUI architecture and data models will be defined by the end of January 2015.

2.10 REL: Behavioural Validation Tool

Author: Elisa Landini

2.10.1 Summary

The Behavioural Validation Tool (BVT) addresses a specific problem in the validation of adaptive embedded systems with a physical human-machine interface (HMI). In fact, even if the AdCoS and its HMI have been correctly developed, the final product might not have the expected behaviour on all physical devices the HMI has be deployed on, because a number of these devices could be flawed or they could not support some feature of the HMI. In this case, the behaviour of every single device must be validated to ensure it meets quality standards: ensuring that these systems perform as intended prior to release them on the market is utmost importance for the producers.

Relevant examples could be the dashboards of a car or a cockpit, as well as medical equipment with buttons and panels.

The original prototype of the BVT was developed by REL some years ago, to address a real industrial need of a top Italian car manufacturer: the validation of every dashboard took at least 4 hours, and they aimed at reducing the validation time and costs, and assessing the compliance to industry standards and regulations.

The solution developed by REL consisted in a system that could compare the actual behaviour of the dashboard (its HMI) with the expected behaviour, by using a camera to record the actual behaviour. The prototype developed by REL is shown in Figure 7.

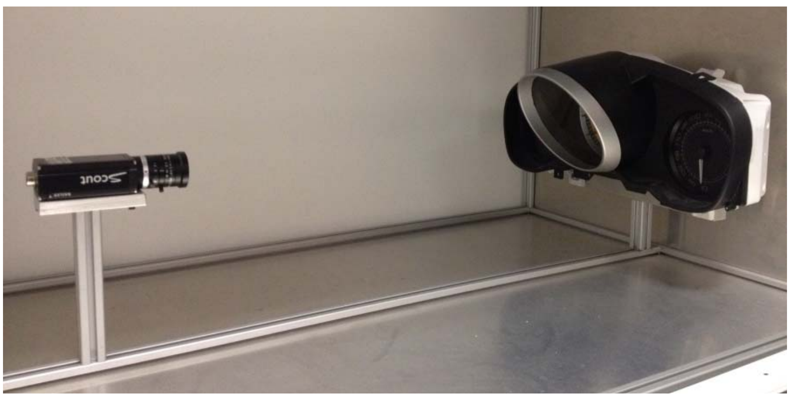


Figure 7: Initial BVT prototype

As a result of the lessons learned, REL started to investigate the possibility of using the same approach on the issues experienced while validating Android mobile app companies. In fact, since the number of Android devices is increasing very fast, testing the application on all potential mobile phones is extremely time-consuming.

Nowadays these systems are mostly tested by engineers/developers manually pressing buttons while going through a test procedure. It is a very monotonous task, which gets the operator very tired and increases the probability of making mistakes by not repeating the procedure correctly. (e.g. by skipping parts or distraction while checking the results)

Moreover, even when the operator identifies a problem in the HMI, it is hard to re-create it to understand which combination of inputs caused it.

If the behaviour of the application can be represented as a finite-state machine, a relevant part of the validation can be conducted by forcing the inputs of each state and then automatically checking whether the graphical layout matches the expected behaviour.

So far some tools have been developed to simulate the inputs of a system and then validate its status against the initial requirements. However, this approach does not allow easily validating the HMI of the AdCoS, that requires a graphical match between the expected behaviour and the actual behaviour of the device.

Testing the correctness of the AdCoS installed on each device implies the testing of each HMI (that is the observable representation of the behaviour) according to the current context (i.e. inputs and current state of the finite-state machine representing the AdCoS, where each state can be associated with a specific HMI).

Therefore, to test the correctness of the HMI of each state, the BVT graphically compares the expected HMI with the actual HMI of the device, in order to understand if any discrepancy occurred.

When validating the HMI of an Android application, the images to be compared are the screenshots of the applications. With a physical device (e.g. a dashboard of a car) another mean of acquisition is needed, thus a camera can be used to collect the information on the status of the AdCoS (by taking pictures or videos of the HMI in each state).

At the end, the BVT provides a test report with relevant information to easily recreate the issues identified in the validation process.

By simply comparing the screenshots of the HMI, it does not need to install third-party software on the target device that could modify or impact on the behaviour of the software to be validated.

The BVT developed in HoliDes will focus on the validation of Android mobile applications (because it mainly addresses the original needs defined by WP8 (Control Room) and WP9 (Automotive) domains, where Android applications will be developed), even though the underlying overall concept could be also applied for the validation of physical interfaces, such as the dashboard of a car.

Moreover, the BVT is a cross-domain tool: in fact it can be used for the validation of the behaviour of any AdCoS installed on a physical device with an HMI (and any Android application), regardless of the domain it belongs to.

2.10.2 AdCoS-Use Case

The BVT will be applied in the WP8 (Control Room) and WP9 (Automotive) domain, for the uses cases that foresee the development of mobile applications.

In particular, for WP8, the main use case where the BVT will be used is **"The Communication between the operator and the operational teams in the field".**

However, the Android mobile application will be also used for the following additional uses cases:

- The Peak of incoming calls - Emergency for an exceptional event;
- The Collection of relevant information for the correct interpretation of the malfunctioning;
- The Collection of historical information about intervention of each installations for future events

For WP9, the main use case where the BVT will be used is the "Lane Change".

The BVT will support the developers of the Android applications to validate them on a great number of different Android devices.

The purpose of the BVT is to provide the developer with a tool to perform repetitive (and iterative) testing on Android mobile apps, ensuring accurate and measurable results and eliminating the need for time-consuming and labour-intense manual test plans.

The BVT is meant to be included in the development tool chain and process, to speed up the validation process, reduce human errors and human factors by ensuring the compliance of the process to a pre-defined test protocol.

Moreover, it provides a test report with relevant information to easily identify the HMI issues and recreate them to understand which combination of inputs generate them and how to solve them.

During the iterative cycles of the project, the requirements will be further elaborated and reviewed in order to assess if the BVT developed in WP5 would actually cover them all.

The addressed requirements can be found in the confidential part of this deliverable.

2.10.3 Input

The input of the BVT will be a formal XML representation of the finite-state machine diagram that includes the states of the AdCoS as well as the events that trigger each state.

This approach should facilitate and support the developers because:

- They should already have created the state diagram for the overall development of the AdCoS, thus they may easily re-use it.
- The states of the diagram will match the states of the application, thus the developer should be easily cover all states and understand which graphical layout correspond to each state (by taking a screenshot of the application in each state).

2.10.4 Output

The output of the BVT will be a report (in .xlsx format) showing the results of the comparisons carried out in the tests and, to easily identify the problems in the validation process.

In particular the report will include the following information:

- Date of the test
- Time of the test
- ID of each state of the HMI

- Validation outcome ({PASS|FAIL})
- In case of FAIL, a screenshot of the expected HMI and the current HMI

Figure 8 shows an example of a potential report.

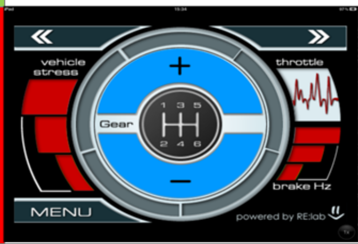
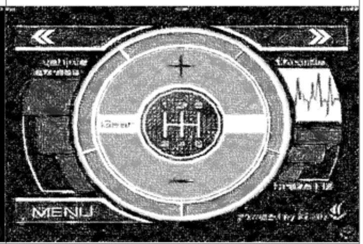
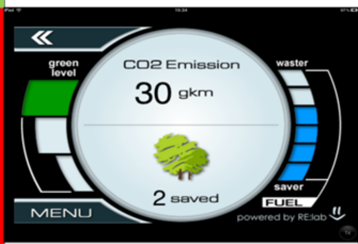
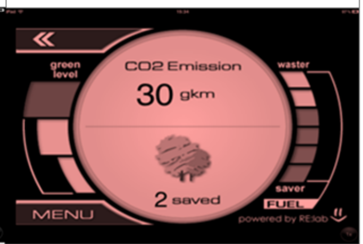
Data	Time	State	Validation outcome	Expected HMI	Current HMI
10/06/2014	11.01	A1	PASS		
10/06/2014	11.01	A2	PASS		
10/06/2014	11.03	A3	PASS		
10/06/2014	11.04	A4	FAIL		
10/06/2014	11.04	A5	PASS		
10/06/2014	11.04	A6	PASS		
10/06/2014	11.07	A7	FAIL		
10/06/2014	11.08	A8	PASS		
10/06/2014	11.09	A9	PASS		
10/06/2014	11.10	A10	PASS		
10/06/2014	11.11	A11	PASS		
10/06/2014	11.12	A12	FAIL		
10/06/2014	11.12	A13	PASS		
10/06/2014	11.12	A14	PASS		
10/06/2014	11.12	A15	PASS		

Figure 8: Example of the output report.

2.10.5 Current status and functionality

The development of the BVT has been started by identifying the functionalities to be implemented according to the requirements listed in the confidential part of this deliverable, yet they have been analysed and elaborated to be included in the list of HF-RTP requirements. Since it provides a cross-domain support for the validation of any AdCoS provided with an adaptive HMI, the BVT is planned to be integrated into the RTP platform.

The BVT can be considered a validation tool-chain including 3 macro blocks (as shown in Figure 9):

- Input: A tool to automatically generate the test cases by exploiting the information included in the finite-state machine diagram used by the developer to create the AdCoS and its HMI.
- Image processing: A tool to elaborate and compare the images to identify any discrepancy.
- Output: A tool to generate a test report to show at a glance the outcome of the validation process.

Figure 9: BVT macro-block diagram.

2.11 SNV: Empirical analysis and validation methods of cognitive and communicative processes in automotive and control room domain

Author: Simona Collina

2.11.1 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) and communication and workload processing (WP8). The data obtained will be implemented in the AdCoS in order to help the operators to achieve the goals planned (WP8) and to prevent visual and cognitive distraction (WP9). In the automotive domain (WP9) the use cases defined by CRF that will be considered refer to the Lane-Change Assistant (LCA). In the control room domain (WP8) the effort will concentrate to 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.

2.11.2 AdCoS Use-Cases

The aim of this method is to assess the operator's status in order to improve the AdCoS's adaptivity in WP8 and WP9 domains.

The main working research questions for WP8 were:

1. How is the state of the art in communication processes among the operators in a control room domain?
2. How can the communication performance be improved among the operators in a control room?
3. How can an operator handle the cross-cultural differences emerging from the relationship with a foreign customer?

For WP9 the main working research questions are:

1. Is it possible to dissociate between visual and cognitive distraction?
2. When and under which conditions can a driver be considered distracted?
3. How are the variables acting in visual distraction?
4. How are the variables acting in cognitive distraction?

Particularly, the method will fit the requirements of the energy control room and collision scenarios.

The addressed requirements can be found in the confidential part of this deliverable.

2.11.3 Input

Appropriate modelling needs data which capture information about the operator status. Psycho-physiological data, as eye-tracker and electroencephalography (EEG) data, and reaction times in task execution will be collected.

2.11.4 Output

The techniques described will provide information about the operator status to be used during the design process of the AdCoS.

2.11.5 Current status and functionality

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. 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. Specifically, the content of the conversation but also the syntactic complexity of the used sentences has been considered as the independent variables to take into account. 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, as some experimental paradigms for visual distraction involved more central cognitive processes. In this respect, a series of experiments have been designed to test visual distraction under the presentation of different stimuli involving different stages of processing.

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.

2.12 TWT: Detection of driver distraction based on in-car measures

Authors: Svenja Borchers, Denis Martin

2.12.1 Summary

Distraction during driving leads to a delay in recognition of information that is necessary to safely perform the driving task [5]. Thus, distraction is one of the most frequent causes for car accidents [6] [7]. Four different forms of distraction are distinguished while they are not mutually exclusive: visual, auditory, 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 having a demanding conversation, the brain shifts its focus. This kind of 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 [8]. Sources for distraction of the driver can be located within and outside of the car.

A computational and empirical cognitive distraction model will be developed in order to analyse different signals from in-car measures with the purpose to detect the distraction degree of the driver. For assessing predictive parameters for cognitive distraction during driving, we run several experiments using a driving simulation and comparing parameters between concentrated driving and distracted driving induced by parallel tasks like conversations or calculation tasks. These measures will include an acoustic analysis including e.g. the detection of the number of speakers, the degree of emotional content and information about the driver’s involvement in the conversation (e.g., whether the driver himself is speaking). In addition, face-tracking signals such as of the blinking of the eyes, 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 for the application of the cognitive distraction model is the development of a mobile user profile computing the individual distraction degree and being applicable also to other systems.

2.12.2 AdCoS Use-Cases

The cognitive distraction model is going to be integrated into the following WP9 AdCoS systems: the TAK Simulator AdCoS, the CRF Test Vehicle, and the IAS 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 degree of distraction can be evaluated. The tool provides feedback whether or not a new system (module) increases or decreases the operator's degree 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 the 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. This online measure of distraction could in turn be used to adapt the degree of automation of the AdCoS to the driver's state. Here, the systems using the level of distraction often only need a single parameter as input.

The addressed requirements can be found in the confidential part of this deliverable.

2.12.3 Input

In-vehicle information is needed. This includes, but is not limited to, in-car audio recordings and eye-tracking data from the driver. These data 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.

2.12.4 Output

The tool provides a temporal description of the driver's degree of distraction. We will thus use a continuous measure provided by a regression analysis. 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. Personal components of the cognitive model and computations are intended to be used as a user profile that potentially can be used by other systems with the same model. This user profile could, for instance, be transferred to other cars.

For integration of the tool into the HF-RTP, its usage during system validation phase plays an essential role and focusses on its output parameters which, in this case, still need to be detailed.

2.12.5 Current status and functionality

Our approach for the implementation and validation of the distraction model will be based on two experiments. So far, we have set up our first experiment and have started testing subjects and collecting preliminary results.

In particular, we have set up a continuous distraction condition comparing in-car measurements with a control condition of concentrated un-distracted 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.

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. Some of the subjects started with the control condition, i.e. driving without distraction, of ten minutes, while other subjects started with the distraction condition, i.e. driving while performing a mathematical calculation task. After the first condition, subjects continued with the other condition, so that each subject performed once the control and once the distraction condition.

During the distraction condition, subjects were presented with simple mathematical problems (e.g. 22+46 or 9-5) via headphones and subjects were asked to give the answer verbally. After eight seconds, the next mathematical problem was presented. All verbal answers were recorded.

Analysing these results, we will 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 [9] 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 degree of the driver's distraction. 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 multi-modal sensory data is RTMaps, which will be used as platform for implementing our auditory 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. During our upcoming experiments, we will use these parameters to evaluate the suitability of auditory and facial features for predicting distraction of the driver.

In a second experiment, we will then induce a more naturalistic conversation condition leading to varying degrees of driver distraction. 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, the degree of emotional content, information about the driver's involvement in the conversation (e.g., whether the driver himself is speaking), is used for the prediction of the driver's degree of distraction. In addition, eye-tracking signals and face movement information can be exploited to increase the reliability of the distraction prediction.

2.13 UTO / CRF / SNV: Detection of driver distraction based on data on vehicle dynamics

Authors: Marco Botta (UTO), Fabio Tango (CRF), Simona Collina (SNV)

2.13.1 Summary

Driver distraction and inattention are an important safety concern [10]. 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.

The tool developed here provides feedback whether or not a new system (module) 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 tool 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 (see sections 2.3 and 2.4) and TWT (see section 2.12) is possible to increase the tool's predictive power.

2.13.2 AdCoS Use-Cases

This method 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.

The addressed requirements can be found in the confidential part of this deliverable.

2.13.3 Input

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.

2.13.4 Output

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.

2.13.5 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 [11]. 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 (see [12] 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 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 conducted 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 3 reports training times and correct classification rates for a FeedForward Neural Network trained with Back-Propagation, while

Table 4 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 3: 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 4. 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.

3 Summary and Outlook

In WP5, 13 MTTs have been defined and updated. So far no significant deviations from the original plan have been noted. In addition, one new tool has been added and a few tools have significantly increased scope and purpose.

A clear relation has been made to the requirements as documented in the confidential part of this deliverable. Requirements can relate to the AdCoS itself, for which an MTT could provide a concrete solution. Alternatively, a requirement can relate specifically to an MTT that should contribute to the AdCoS. These two kinds of requirements have been labelled “AdCoS” or “MTT” in the requirements attribute “Type” throughout this document.

An MTT can have different relations with the AdCoS.

- An MTT can be a missing tool for the development of AdCoS-functionality, be part of the AdCoS itself, or both.
- An MTT can be part of the development-process. This could be requirements gathering, design, evaluation, or a combination of thereof.

Currently, most MTT’s only relate to one specific AdCoS. While addressing the broader set of requirements the challenge is to find other application areas or AdCoS in potentially other domains. In that way an MTT adds more value and has more impact being part of the HF-RTP.

In 2015 WP5 will explore opportunities for re-using MTTs in other AdCoS or domains. To achieve that, closer interactions will be set up with all domains. By creating a better understanding of the domains and AdCoS WP5 can propose a possible use of an MTT. Vice versa, a domain or AdCoS can be triggered to use an MTT by getting a better understanding of the MTTs potential. By doing this, our goal this year is to create and demonstrate prototypes and implementations of appropriate tools and techniques, applied to one or more AdCoS, at milestone 3.

Matching an MTT to a new AdCoS or domain will require tuning to the requirements for the specific AdCoS. This can imply deleting unwanted, changing existing, or adding new functionality. These changes should be identified by close cooperation with AdCoS and reflected in updated requirements.

4 References

[1] K. J. Vicente, Cognitive work analysis, Mahwah, N.J.: Lawrence Erlbaum Associates, 1999.

[2] 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,” *Brain and Language,* vol. 86, no. 1, pp. 57-69, 2003.

[3] C. Huang, X. Ding und C. Fang, „Head Pose Estimation Based on Random Forests for Multiclass Classification,“ in *20th International Conference on Pattern Recognition (ICPR)*, 2010.

[4] A. Schieben, M. Heesen, J. Schindler, J. Kelsch und F. Flemisch, „The theater-system technique: Agile designing and testing of system behavior and interaction, applied to highly automated vehicles.,“ in *Proceedings of the First International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, Essen, 2009.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

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

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

[14] B. John und D. E. Kieras, „The GOMS family of user interface analysis techniques: Comparison and contrast.,“ in *Transactions on Computer-Human Interaction*, 1996.