HoliDes

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

D5.5 – Techniques and Tools for Empirical Analysis Vs1.8 incl. Handbooks and Requirements Analysis Update

Project Number:	332933
Classification:	Public
Work Package(s):	WP5
Milestone:	M3
Document Version:	V8.1
Issue Date:	January 22, 2016
Document Timescale:	Project Start Date: October 1, 2013
Start of the Document:	Month 21
Final version due:	Month 28
Deliverable Overview:	**Main document**: Techniques and Tools for Empirical Analysis Vs1.8 incl. Handbooks (Public)
Compiled by:	Adam Herout – BUT
Authors:	Adam Herout (BUT) David Käthner (DLR) Morten Larsen (AWI) Klas Ihme (DLR) Linda Onnasch (HFC) Juliane Bürglen (HFC) Paul Kaufholz (PHI) Flavia de Simone (SNV)

	Marieke Thurlings (TWT) Marco Botta (UTO) Zdenek Moravek (HON)
Reviewers:	Bohuslav Krena (BUT) Martin Krähling (IBEO)
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	ECSEL JU	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
22.10.2015	Initial Version	Adam Herout
03.11.2015	Collected Contents	Adam Herout (and all authors)
22.01.2016	Consolidated, Pre-finalized Version	Adam Herout
27.01.2016	Internal pre-review, incremental improvements	Adam Herout, David Käthner
01.02.2016	Integration of last contributions, last revisions	Adam Herout, Zdenek Moravek
01.02.2016	Inserted images for the TWT-contribution.	David Käthner
22.02.2016	All MTT-partner-contributions integrated in 1 document.	Adam Herout
23.02.2016	Added evaluation framework from Elisa.	David Käthner
09.03.2016	Deleted erroneous WP7 requirements (WP7_HON_AER_REQ89 and WP7_HON_AER_REQ90).	David Käthner
09.03.2016	Added information on task analysis.	David Käthner
14.03.2016	Finalizing of the document.	David Käthner
15.03.2016	Deleted HFC's confidential parts	Linda Onnasch
15.03.2016	Deleted AWI's confidential parts	Morten Larsen
31.03.2016	Finalizing of the public version.	David Käthner

Table of Contents

List of Figures

List of tables

List of abbreviations

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

1. Introduction

The objective of deliverable D5.5 ("Version 1.8 of the techniques and tools for empirical analysis") is to provide a general overview of WP5's progress since the last deliverable D5.4, June 2015. The partners present the current status of their developments, giving an overview over functionality and the connection to the use cases from WP6-9. To keep the deliverable accessible to the reader, only incremental updates from previously reported versions are being documented.

The deliverable first gives an overview over the properties of various task analysis which are being used in HoliDes. Although task analysis is an established technique of Human Factors work, there are few references on operational details of actually doing task analysis. In a workshop between WP5 and WP2 we discovered the need to first find more abstract description of what task analysis is, what it can do, and how to use it in the context of AdCoS-development and -evaluation. As this is work in progress, this part of D5.5 serves primarily to report on the current status of our discussion of the issue. In D5.6, there will be more on how the various task analysis approaches have been used in HoliDes.

The second part of the deliverable puts forward an evaluation framework for the various AdCoS in HoliDes. It defines a few basic terms and illustrates the intended cross-domain process with an example from WP8.

The following main part provides the description of the status of the individual Methods Techniques & Tools (MTTs) and updated requirements, stripped down to the basic requirements and accompanied with evaluation metrics. Both parts are structured per MTTs/tasks that are being created in WP5. For each MTT, the following items are reported:

- description
- current status and functionality
- method of evaluation and evaluation metric
- list of relevant and important requirements

All MTTs developed in this work package tackle important human factors aspects occurring in the development and evaluation of innovative adaptive cooperative systems (AdCoS) and thus form an important brick stone in the development process.

In WP5 the MTTs are a mixture of empirical work, tools, and more abstract methods. Their precise position and integration in a workflow together with MTTs from other WPs depends on the AdCoS under development, the organization that uses them, as well as individual considerations. These questions will determine the tailoring of the Human Factors Reference Technology Platform (HF-RTP) for a specific use case.

Empirical MTTs are an essential part of both early and late stages of any design process of an AdCoS and human-machine system in general, for example during requirements analysis or verification of human-factors-related non-functional requirements. However, empirical MTTs can also be an integral part of the development phase, especially when using principles of agile requirements engineering approaches.

For instance, the task analysis methods developed in WP5 (c.f. sections 2.2, 2.8, and 2.11) are used to determine how users interact with AdCoS and therefore can lead to potential improvements of this interaction. At the same time they are valuable for the evaluation of newly developed AdCoS. In line with this, WP5 also created a systematic of these task analysis techniques, which supports the selection of the appropriate method for the problem at hand and thus aids the tailoring of the HF-RTP.

Further, WP5 provides MTTs that enable adaptivity in human-machine systems. The portfolio of WP5 includes methods for operator state detection based from implicit hand gestures (2.3) and cognitive distraction classification (2.16) which are essential for creating adaptation in WP3 and the application WPs 6-9. Equally important are the experiments designed within WP5 that helped WP3 partners fine-tune their adaptive functions (e.g., 2.7, 2.12, 2.13 or 2.15).

Other WP5 further developments integrate users early in the AdCoS design process, such as focus groups (2.14) or theatre technique (2.9) Finally a tool for supporting the documentation and providing guidance for the evaluation process (2.1) is provided by WP5. Next to the mere MTT development, WP5 is a service provider with respect to the evaluation of human factors requirements and contributes an evaluation guideline for this endeavour.

2. Task analysis in HoliDes

2.1. An overview over task-analysis

This section gives an overview over the properties of various task analysis approaches that are being used in HoliDes. It owes much to a workshop on task analysis, organized by DLR and AWI and held on 21 and 22 of May 2015 in Leiden, Netherlands. Partners of WP2 and WP5 present at the workshop discussed a wide range of issues related to task analysis, both in practical and theoretical terms. Many of the concepts described here have been used extensively throughout the project, but not documented in a unified report.

There is no single definition of task analysis on which everyone working in the relevant disciplines agrees on. One view often shared in the literature and personal communication is that human behaviour can only adequately be understood in terms of goals this behaviour tries to attain (e.g. [1], [2]). A task can therefore be defined as a concept in which there exists a current state, a goal state, and operators linking the two. Applying an operator with the intention of moving from the current state to the goal state is then called an activity.

Another essential properties of the concept "task analysis" is its purpose to define the essential contents of the task to describe. This relates both to the parts which constitute the task, and the inner organization of the task. A typical approach is therefore to break down the task into its elements, to warrant an understanding both of physical and of cognitive activities. An underlying theory of human behaviour helps to tie these together and to ground them in empirical data. The result of the task analysis can be a verbal description, event trees, or a rather formal task model. The latter one usually follows a special notation.

A good overview over the workflow of a task analysis can be found in [2]. The input to a task analysis can be performance data, event reports, or design data in case of not-yet-existing systems. The analysis can be performed in a descriptive way ("how things are"), a normative one ("how things should be"), or in a formative manner (what things are possible).

A descriptive would be one which tries first and foremost to record essential events or activities as they are occurring while conducting the work. Normative approaches, however, make a statement as to how a task should be carried out. Finally, the formative kind tries to specify only the constraints under which the work is being carried out, rather than to analyse each single step.

Applying any of these approaches will lead to a certain task representation. It is not necessarily a formal or even an explicit one. Sometimes it remains a rather vague idea, nevertheless representing the essentials of the work analysed, as well as its tasks. This task representation itself can be further subject to abstractions or other changes for the purpose of presentation. While seemingly trivial, it is important to remember that any sufficiently complex task will yield a complex task model. Presenting an audience this task model can be an arduous task, necessitating simplifications.

2.2. Task analysis features

The above distinction of task analysis approaches in descriptive, normative and formative is a useful first approach to the subject. However, in reality approaches can be and are arbitrarily mixed according to what the use case demands. Aside from theoretical considerations of the nature of tasks and human behaviour, there are a number of practical considerations to judge the usefulness of any approach. In different words, there is no general theory of task analysis from which all else follows. Instead, practitioners use whatever fulfils the requirements for a use case, customizing whenever necessary.

Another great advantage of this feature based-approach is a straight-forward connection between requirements of a given system development, and those features. At the beginning of an AdCoS development will be usually a phase when requirements are analysed to determine which MTT needs to be used to work on this requirement. Having a list of features (instead of wholesale descriptions) makes for an efficient decision.

Trying to capture both practical and theoretical features of the HoliDes task-analysis MTTs led the participants of the Leiden workshop to document a large number of important task analysis features.

Guiding objectives in the feature definition were:

- Every task analysis-MTT in the project must be represented with its defining features.
- Features need to be as abstract as possible.
- Features must be a useful guide in choosing an appropriate task analysis-MTT for a given use case.

The following text is organized as follows: For each central feature which was defined in the workshop, the list with the sub-features is documented here. An explaining text after the list will explain the meaning of this feature within the task analysis context, and highlight some of the important sub-features.

purpose

- data organization
- data presentation
- knowledge elicitation
- modelling
    - cognitive modelling
    - simulation
    - prediction
- user training
- steer systems engineering
- guide design
- understand human behavior

The main feature setting task analysis efforts apart is probably their purpose. It may also be the feature where different task analysis efforts vary the most. Modelling is an obvious goal. A well worked out task model can greatly support the development of a cognitive model. At the same time, task models can be worked into computational models, provided a behaviour-producing mechanism.

Highlighting the "methods" aspect of task analysis, one of its most practical uses is the organisation and subsequent presentation of data. When analysing a given domain in the real world, it is a difficult decision which data sources to use, which data to record, and at which level of granularity this needs to be done. Task analysis is of great help here, as it focuses the analysis effort at the level of the task, restricting e.g. the range of possible data sources.

task control

- hierarchical
- procedural
- + - pre/post-conditions
    - measurable
    - abstract
    - formalized
- + - dynamic
    - parallel execution
- + - relations

    - extensible
- critical
- + - goals
    - importance
    - hierarchy
    - formalization
    - + - adaptivity
        - prioritisation
        - adding/deleting
    - conflict resolution

The second most important feature is arguably the view on the control of the task to be analysed. Some of the most well known task analysis methods assume tasks to be organized tree-style. This can work well e.g. in environments or with tasks which are exclusively determined by artefacts. More dynamic environments may require a more flexible approach to task control, which can be realized with procedural control structures.

With a central tenet being the goal directedness of behaviour, many task analysis approaches make explicit use of goals in their modelling efforts. However, only a few formalize goal control, e.g. provide formal rules for prioritisation or conflict resolution.

task model

- use
    - analysis
    - architecture
    - presentation
- + - type
    - normative
    - descriptive
    - formative
- time related layering

Not all task analysis approaches have the purpose to produce task models, and for the ones that do it usually is just one result among many. However, given the effectiveness of a task model in analysing the task, it is a desirable outcome which can serve different uses. And while it is not so clear for an entire task analysis approach whether it belongs to one of the types, it is a useful distinction for the resulting task model.

system

- + - actors

  - single
  - multiple
  - human
  - system
- + - roles
  - human
  - machine
  - cooperation
- + - environment
  - structured
  - unstructured
- + - user
  - goals
  - persona
  - role

Task analysis approaches utilize different system perspectives, especially in regard to which part of the world to care most about, to put under a microscope. One approach might allow only for single actors, others conceptualize the whole system as actors. Another concept can be roles, taken up by either human, machines, or roles resulting from interaction.

data

- + - input data
  - structured
  - unstructured
  - qualitative
  - quantitative
- + - output data
  - structured
  - qualitative
  - quantitative

The type of data used in task analysis varies wildly. Essentially any combination of input data is possible. Output data on the other hand should never be unstructured, as the structuring of data is a necessary feature of any task analysis.

abstraction

- solution independent layer
- + - task level
  - essential
  - concrete
- granularity

- + - direction
    - bottom-up
    - top-down

Abstraction refers to the extent and manner in which details from the original data are removed, in order to arrive at more general world descriptions as well as the discovery of patterns or predictions. "Solution independent layer" is a special feature that was essential to one of the task analysis approaches in the Leiden workshop which was geared towards software design. It means to have a specific layer in the task representation which can describe the task, without precluding any solutions for a given scenario.

Granularity refers to the resolution in dimensions or domains such as time, space, or cognitive operations.

The direction of abstraction refers to the way abstractions are reached. Two popular approaches are to either abstract away from empirical data, or to start with a theoretical description and relate data to this description.

decomposition

- cognition
- body
- system

It is a basic tenet of task analysis to decompose larger systems into meaningful parts. Different task analysis approaches can support the decomposition of cognition, body, and the technical system, but not all of the do. E.g. looking at actions of specific body parts in the context of the task is done only by few approaches, notably those which are concerned with the task related motor behaviour.

customization

- combination with other approaches
- omit features
- add features

Some approaches are easier to customize than others. Whereas approach one will work only in the exact specific way, others are reminiscent of a toolbox and will lend themselves naturally to customization.

third party input

- regulations
- certification
- guidelines

In the context of HoliDes, third party input is especially relevant, as it is a project focus to develop AdCoS in line with relevant guidelines and regulations. Here it can be seen how important a "featurization" of task analysis approaches is. There is no special class of task analysis approaches considering third party input, while the others do not support this feature. Instead, if a requirement exists such as "The MTT shall consider Human Factor guidelines", any candidate MTT can be easily checked in regard to its suitability. Alternatively, an extension of an existing approach not supporting this feature could be developed, with inspiration drawn from MTTs that do.

3. Cross-domain evaluation framework

Verification, validation and evaluation are fundamental activities for the deployment of reliable and acceptable adaptive cooperative systems (AdCoS).

When applied to an AdCoS , the **verification** can be seen as the task of determining that the system is built according to its specifications, **validation** is the process of determining that the system actually fulfils the purpose for which it was intended, and the **evaluation** reflects its performance in the field and/or its acceptance by the end users. In other words, the VV&E elements of the expert system are designed to:

a) Verify to show the system is built right.
b) Validate to show the right system was built.
c) Evaluate the performances to show the usefulness of the system.

Even if a system is known to produce the correct result, it could fail an evaluation because it is too cumbersome to use, requires data that is not readily available, does not really save any effort, does something that can be estimated accurately enough without a computer, solves a problem rarely needed in practice, or produces a result not universally accepted because different people define the coefficient in different ways.

The decision making related to evaluation process often requires evidence from research with practical experience and human values. Evidence can be collected by **empirical evaluation**. In performing empirical experiments with humans (such as in the “Human-in-the-Loop” approach), experimenters should cope also with the intrinsic human variability that affects dependent variables.

However, the evaluation process can also be performed by applying a **model-based approach** (i.e. a simulation) to collect the quantitative data for a comparison between an initial system (i.e. the baseline, as defined in D1.6) and the final system (that includes our AdCoS).

Therefore, the evaluation refers to the appraisal of a theory by observation either in empirical studies or simulations. The key to good evaluation is the proper design and execution of the experiments/simulations so that the particular factors to be tested can be easily separated from other confounding factors.

The aim of this framework is twofold:

1. Describe the clear connection between the baseline defined in D1.6 and the evaluation performed in the domains (WP6-9)

2. Support the AdCoS owners in the evaluation activities (tasks Tx.5, with x=6-9) by providing the MTTs developed in WP4 and WP5 (i.e. model-based and empirical approach) to concretely perform the evaluation.

3.1. Guidelines for the evaluation

General steps to guide the evaluation can be summarized in Table 1.

Table 1: steps for the evaluation

STEP	Name	Description of the activity
1)	**Performance indicator**	Identify the performance indicators that are relevant to the overall system/process we want to observe.
2)	**Quantitative objective to be achieve**	Define a quantitative objective to be achieved that represents a good performance of the overall system/process
3)	**Selection of the MTTs**	Select the most appropriate empirical/model-based MTTs to measure the performance indicator
4)	**Measure BEFORE**	Design and conduct a study/simulation and measure the performance indicators BEFORE the introduction of the AdCoS (i.e. the **baseline**)
5)	**Measure AFTER**	Conduct the same study/experiment and measure the performance indicators AFTER the introduction of the AdCoS
6)	**Results:**	Compare the results and draw conclusions

3.2. Performance indicators

This section describes in details how to identify the most appropriate indicators (step 1), according to the Human-Factor Literature.

The measurement of the performance in a task in which the human is involved in, can be defined in terms of achievement, with reference to the Cognitive Load Theory (CLT). As do many others, this approach operationalizes performance by measuring

1. **Time parameters**
2. **errors**

In order to support the AdCoS owners in identifying the most relevant performance indicators, i.e. "error types" and "time parameters" that could be used for their own evaluations, Table 2 has been defined by considering the most relevant parameters identified in Human Factors literature.

Table 2: Examples of relevant errors and time parameters for the evaluation according to HF Literature (please refer to the References)

			DESCRIPTION OF DEPENDENT VARIABLE (= performance indicator)	BEFORE ADCOS	AFTER ADCOS
TIME		Completion task time	e.g. Time to complete a task		
TIME		Elapsed time from the event	e.g. Time after a critical event occurred		
ERRORS	VIOLATIONS	Routine violations	e.g. Routine violation, which entails cutting corners whenever such opportunities present themselves		
ERRORS	VIOLATIONS	Optimising violations	e.g. Optimising violations, or actions taken to further personal rather than strictly task related goals (that is, violations for "kicks"		

			or to alleviate boredom)		
	PROCEDURAL	**Failure to monitor situation**	e.g. missing an event detection		
		Followed procedures with wrong execution	e.g. wrong entry into a program		
	COMMUNICATIONS	**Misinterpretation; Misunderstanding**	e.g. misinterpretation and misunderstanding trigger inappropriate procedures		
	PROFICIENCY	**Lack of knowledge**	e.g. information not available due to a system fault		
		Lack of skill	e.g. not properly skilled user for the task to be completed		
	DECISION	**Decision that unnecessarily increases risk**	e.g. unnecessary navigation through adverse weather		

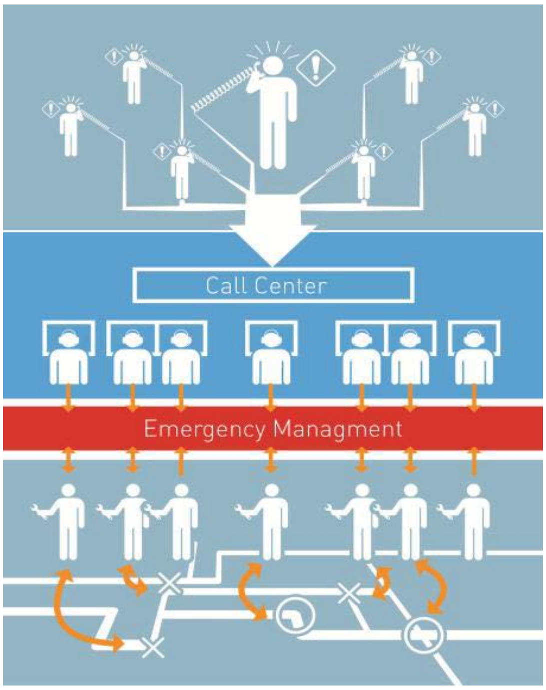
Table 2 represents a guideline for the AdCoS owners, but it can also be extended according to their specific needs (in terms of performances to be evaluated).

Once the information from all AdCoS owners is collected, WP4 and WP5 (respectively for the model-based and the empirical approach) will support the AdCoS owners for the definition of suitable studies and simulations (by using the methods and tools developed in WP4 and WP5) to measure the indicators and perform the evaluation.

Then the evaluation will be actually conducted in WP6-9.

3.3. WP8 IREN Control Room: An example application

The methodology can be applied to the evaluation of the performances of the AdCoS developed in HoliDes.



In order to understand the application of the methodology to the IRN AdCoS, it is necessary to define the baseline of the overall system (i.e. the BEFORE condition).

The Control Room of IRN collects energy emergency requests and assigns the interventions to the technicians in the field.

At present, IRN does not use any adaptive system to allocate tasks to available operational teams in the field, but the communication between the Control Room operators and the operative teams takes place only via phone calls (very time-demanding) and the allocation of tasks and responsibilities is based on the senior experience of Control Room operators.

Previous systems (e.g. mobile apps) have been introduced into the Control Room to share data between the operators in the Control Room and the technicians in the field, but none of them has been accepted by the operators and/or the technicians, and they had always switched back to the phone calls.

The aim of the AdCoS is to reduce the time spent by the operators in the overall assignment of the activities to the technicians (included the selection of the most relevant technicians and the management of the communication to share the information the technician needs to perform the intervention), in order to allow the operators to perform critical activities of management of the network that he is also in charge of.

The AdCoS includes two main functionalities to support the operators:

1. The automatic selection of the most relevant technician according to a set of dynamic parameters (zone of the intervention and actual

position of the technician, previous assignment to the operators, shift, etc..)

2. The delivery of the most important information on the intervention (address, type of malfunction, etc..) as well as the ability to accept/reject the assignment via mobile app.

Therefore, as regards the selection of the most appropriate technician (point 1) the reference is always the expert operator, that we expect to have a 100% success rate. The benefit of substituting the manual selection (with no errors) with the automatic selection (with potential errors) depends on the ability to allow the expert operator performing other critical activities in the same slot of time (even accepting a lower success rate).

Table 3: Application of the empirical evaluation methodology to the IRN AdCoS

		DESCRIPTION OF DEPENDENT VARIABLE (=performance indicator)	BEFORE ADCOS	AFTER ADCOS
TIME	Communication time (on the phone)	time spent by the operator on the phone with the technicians		
ERRORS	The errors are calculated as a comparison between an optimal condition (BEFORE) and the condition with the AdCoS (AFTER)	list of technicians selected by an expert operator compared to the list selected by the AdCoS		

Methodology (applied to the IRN AdCoS):

1) **Performance indicator:** 2 performance indicators have been identified

a. Technician selected for each intervention (this is not a quantitative indicator, but is used to calculate the correct rate of the AdCoS compared to the selection of an expert operator)
b. Time spent by the operator on the phone with the technicians

2) **Quantitative Objective:**
a. The AdCoS shall have a correct rate of at least 80% in the selection of the most appropriate technician (compared to the manual selection performed by an expert operator)
b. The AdCoS shall reduce by 70% the time spent by the operator on the phone with the technicians

3) **Selection of the MTTs:**
a. Design an empirical study with real operators and some technicians to extract the list of technician selected by an expert operator
b. Design an empirical study with real operators and some technicians to measure the time spent on the phone for the assignment of an activity

4) **Measure BEFORE:**
c. Conduct the empirical study with real operators and some technicians to extract the list of technician selected by an expert operator BEFORE the introduction of the AdCoS
d. Conduct the empirical study with real operators and some technicians to measure the time spent on the phone for the assignment of an activity BEFORE the introduction of the AdCoS

5) **Measure AFTER:**
a. Conduct the same empirical study of 4a) with real operators and some technicians to extract the list of technician selected by an expert operator AFTER the introduction of the AdCoS
b. Conduct the same empirical study of 4b) with real operators and some technicians to measure the time spent on the phone for the assignment of an activity AFTER the introduction of the AdCoS

6) **Results:** At this stage no data has been collected thus no result can be shown and discussed.

4. MTTs Developed in WP5 – Incremental Update

The updates for individual tools developed under WP5 are contained in this chapter. Second-level headings indicate individual tools, with the ID of the responsible partner in the brackets. Their descriptions are structured as similarly as possible to allow for easier orientation of the reader.

4.1. HF Filer (AWI)

Many traditional human factors methods and techniques – and especially the non-formalised ones – rely on experts making assessments of a system design or a work situation according to a list of specific human factors issues to analyse. The result is normally a textual report, delivered as a main document possibly with appendices.

Such a long textual report with feedback on a system design, for instance, is not well suited to be used by development and project management tools in the RTP.

4.1.1. Summary

HF Filer is intended to provide structure in an evaluation process, capture evaluation results in a textual form and make them available to the RTP in a granular format that can be used with other tools in the design and development process. This is done by providing three specific functionalities:

- Let users record the results of human factors evaluation activities
- Make the evaluation data accessible to other tools in the HF-RTP
- Provide traceability of the recorded human factors data

More information is available in deliverable 5.4, where the functionality is described in detail.

4.1.2. Current Status and Functionality

In the previous deliverable (5.4), the main tool has been described. It is a stand-alone web app, providing the HMI of HF Filer. The basic functionality of the tool and its user interface are complete.

In the period covered by this report, work has been concentrated on designing, developing and testing the OSLC interface for the tool. Since HF Filer is a tool to store reports from HF experts, via the HMI, the OSLC interface limits itself to publish the data it contains to other tools that need them. In OSLC-parlance, this is called a *provider* of data (a tool that reads data over OSLC is called a consumer).

Since HF Filer is a web-based tool with a REST-like API developed in Python, the most natural way to implement an OSLC interface is to build a bridge using the Java-based Lyo framework (available from eclipse.org).

The steps undertaken can be summarised as:

- An architecture has been designed for the OSLC interface based on Lyo
- The OSLC interface has been implemented as a Lyo-based OSLC bridge
- Service provider listings as well as resource queries have been implemented
- The RDF output produced has been verified

The architecture is based on the OSLC core spec, which helps reduce the complexity of the system, and allows to simply map each item type of interest in the HF Filer data structure to a resource object with a corresponding servlet to provide http transport.

The architecture is outlined in Figure 1. One resource object-servlet implementation is needed for each item type to make available over OSLC.

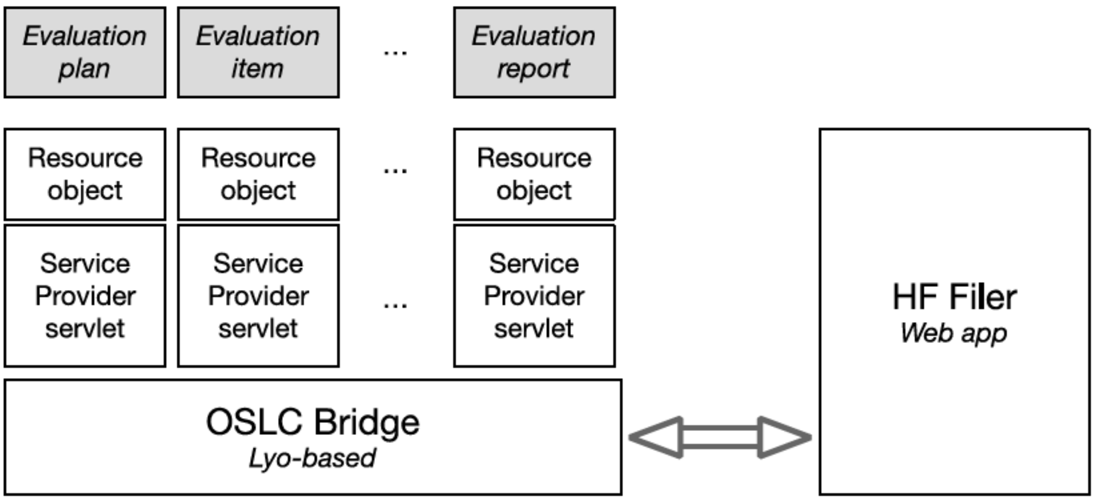


Figure 1: Overall architecture of the OSLC bridge for HF Filer.

The output of the OSLC bridge is provided in RDF format.

4.1.3. Integration of the Outputs of the MTT

HF Filer is being integrated into the RTP used to design and develop the border control room AdCoS in WP8. It is used to add human factors data to other, more technical, project data that are managed with Enterprise Architect. To this end, a specific plugin is being developed to allow EA to read, via OSLC, evaluation results from HF Filer.

4.2. Means-ends modelling of AdCoS (AWI)

Means-ends modelling of AdCoS (in the following referred to as simply "Means-end modelling") is the new, shorter name for the MTT called "Modelling of AdCoS from a means-ends perspective" in earlier deliverables.

Means-end modelling provides a functionality-oriented subdivision of the work of the operator and of the AdCoS. This helps structure the design of the AdCoS as well as the operating procedures.

4.2.1. Summary

Means-ends modelling of AdCoS can be used at different stages of the design and development process. When used in the design phase, it may help with identifying functional blocks in the system and when used in the validation stage, it may be used to identify potential causes of errors produced by unclear or weakly defined states of automation, adaptation or cooperation.

4.2.2. Current Status and Functionality

The modelling method has not undergone large changes since the previous deliverable.

However, in order to use the method to evaluate an AdCoS, an extra aspect has been added to the modelling method: classification of functions according to the type of goal they serve. Figure 2 shows an example of such a categorisation. The example is taken from the Guided patient positioning use case from WP6, and it can be seen that 6 categories of functions have been identified, each associated with a specific colour. The colours do not have any specific meaning, and can be selected by the responsible for the modelling as the prefer.

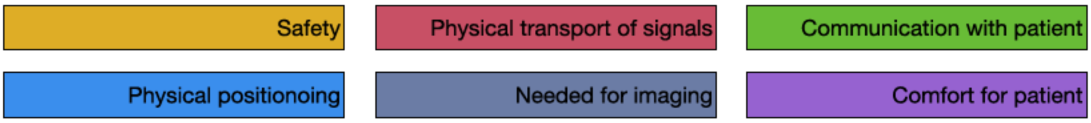


Figure 2: Categorisation of functions in the Guided patient positioning use case with associated colour coding.

4.2.3. Integration of the Outputs of the MTT

Means-end modelling has been used in the modelling of the Guided patient positioning AdCoS to provide insight into the hierarchical structure of the task(s).

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

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

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

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

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

4.3.1. Summary

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

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

Integration to HF-RTP is done by defining structured interfaces between MTTs that can generate task models – input to U-DAT – and MTTs that can generate a structured report – output from U-DAT.

4.3.2. Current status and functionality

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

Within HoliDes the HF-TA and Means-end analysis have been used to create a task analysis for resp. ECG Triggering and Guided Patient Positioning. This information can eventually be used as structured input to U-DAT. Vice versa, U-DAT can create a structured output format that can be used by other tools to create summary reports or further analysis. A first exploration to link the output of HF-TA to U-DAT and U-DAT to HF-Filer has been done.

U-DAT has been applied to the Guided patient positioning and ECG Triggering AdCoS of WP6. Also other AdCoS can make use of this tool.

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

Task Analysis

The input task model covers all needed attributes from a task analysis such as e.g. described in Software for Use. An example task map is depicted in Figure 3. A user role performs a number of ‘essential tasks’ – depicted in orange. Essential tasks are high-level tasks that do not imply a solution direction, but define the intention of the user. Additionally, so-called ‘concrete tasks’ – depicted in blue - can be part of the model, which are based on a solution direction. They define the concrete steps by which the intention can be carried out. Different arrow types model relationships between tasks like “include” (normal arrow) or “optional” (dotted arrow) sub-tasks.

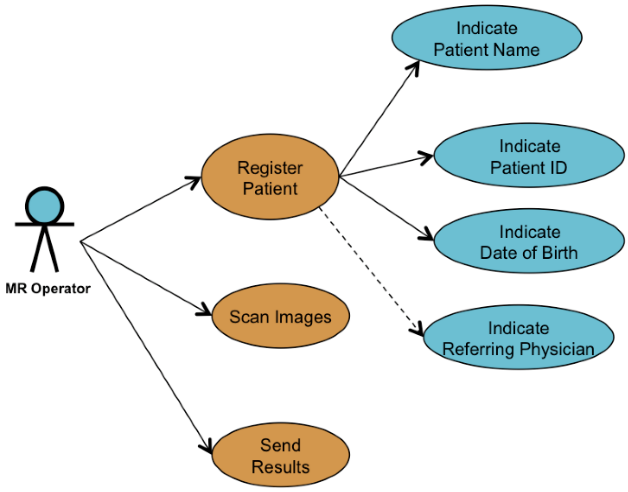


Figure 3: An example of Task Map

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

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

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

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

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

4.3.3. Integration of the Outputs of the MTT

U-DAT has been used to evaluate both the ECG Triggering and Guided Patient Positioning AdCoS from the Health domain, WP6. Figure 4 below shows the corresponding U-DAT for ECG Triggering.

								Criticality		Task Score				
ID #	Context of Use	User Role	From	Relation	Task	Step	User Goal	Hazard	Essential	Succes	Close call	Fail	Comments	Root cause analysis (close call / fail)
Scenario 1: The operator is at the scanner console. **Select the patient data for patient X from the RIS**. Once done, **open the ExamCard database browser and select the ExamCard "Cardiac exam"**. **Call in patient X from the waiting room, check the patient for contra-indications and explain the procedure to the patient**. **Next, ask the patient to change clothes (hospital gown is provided) and clean the skin in order to apply the VCG electrodes. Once the electrodes are applied, check the VCG signal at the display on the operator's console.** You are satisfied with the quality of the VCG signal. **Provide head set and nurse call to the patient, and place the coil on the chest.** Back at the operator's console, you decide to use the manual VCG calibration. **Selecting this option and perform the necessary steps.** **When you are satisfied, start the scan.**														
Task_1	Zone 3 / Console	MR Operator	MR Operator	include	Register a patient		Enter New Exam details to start exam of (the next) patient							
Task_2	Zone 3 / Console	MR Operator	Task_1	include	Select Exam Card		Select the right procedure for the scheduled examination							
Task_3	Zone 3 / Console	MR Operator	MR Operator	include	Check VCG Signal		Verify that the VCG signal that is detected, is good enough to start the examination							
Task_4	Zone 3 / Console	MR Operator	MR Operator	include	Choose Calibration		Ensure that VCG data is calibrated in order to get reliable R-top detection throughout the exam.							
Step_1	Zone 3 / Console	MR Operator	Task_4	specialization		Manual Calibration	Ensure that VCG data is calibrated in order to get reliable R-top detection throughout the exam.							
Step_2	Zone 3 / Console	MR Operator	Step_1	include		Instruct patient	Optimize patient-related circumstances under which VCG calibration is performed							
Step_3	Zone 3 / Console	MR Operator	Step_2	include		Check calibration result	Verify that the VCG calibration is completed successful, operator wants to intervene and repeat the process							
Task_5	Zone 3 / Console	MR Operator	MR Operator	include	Set Heart Rate		At ExamCard level, operator specifies the patient's heart rate (as well as other properties such as laterality, anatomic region etc)							

Figure 4: U-DAT ECG Triggering

The first 7 columns represent the task analysis for ECG triggering. The task model is similar to Figure 3. With the information contained in columns 1-7 it is possible to recreate the task model. Vice versa, a task model from another task modelling tool should be able to provide the same information to create the model in U-DAT. Cooperating with the HFC team an equivalent task model has been created to explore how HF-TA can be used eventually as an input tool to U-DAT. Figure 5 below shows the equivalent task model for ECG triggering using the HF-TA MTT.

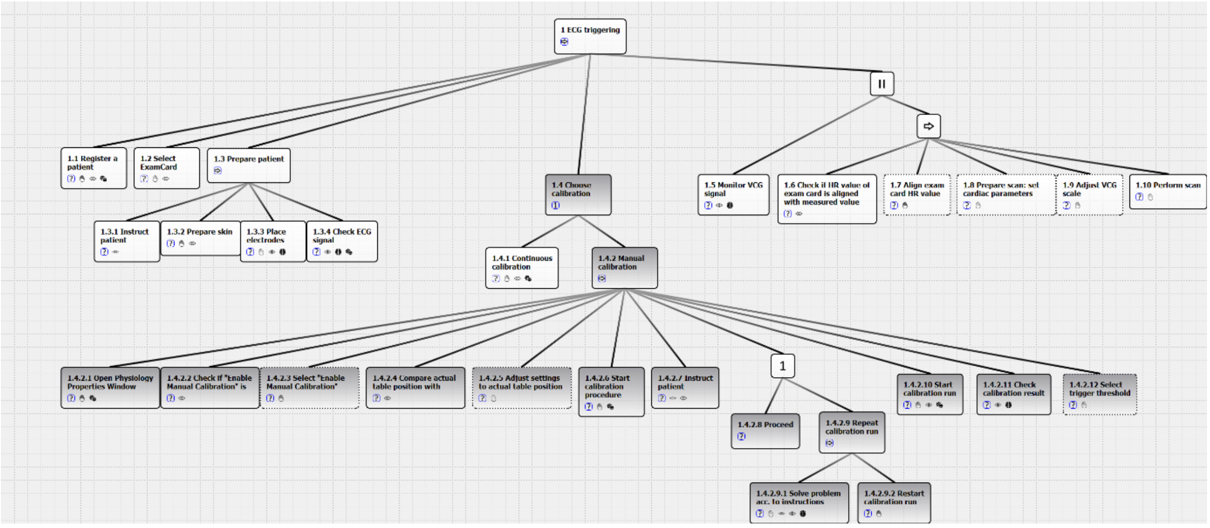


Figure 5: HF-TA based task model for ECG Triggering

On the output side PHI has cooperated with ANYWI to explore how the output from U-DAT can eventually be used as input to HF Filer. We used the data multiple U-DATs to define how this information can be used and presented using HF Filer. Figure 6 below shows a possible output in HF-Filer using U-DAT input data.

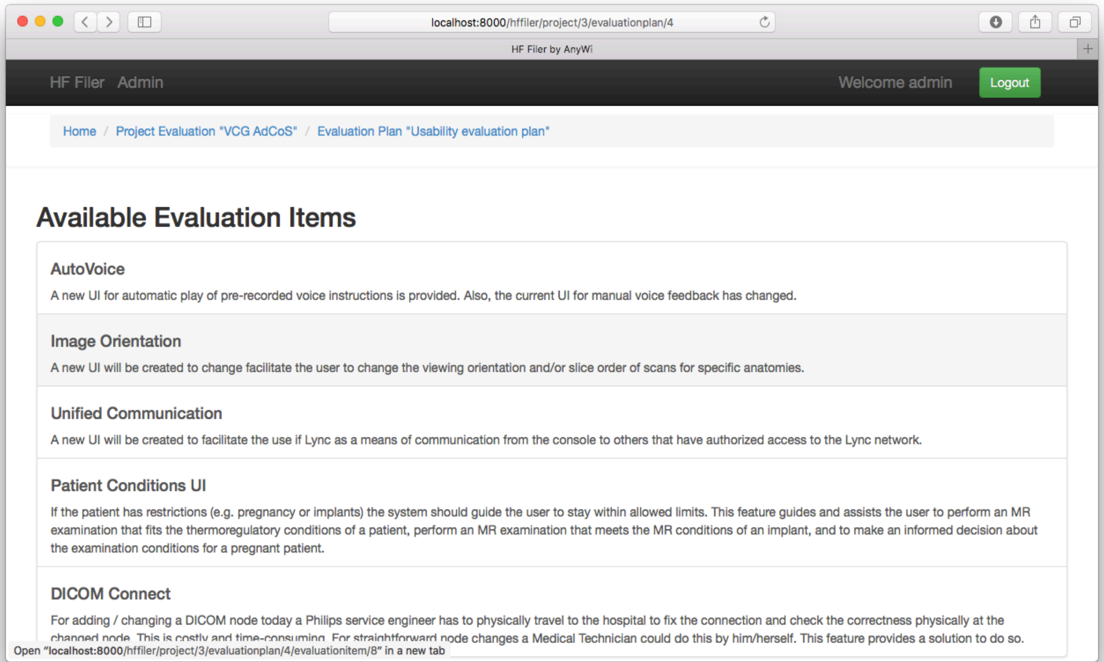


Figure 6: U-DAT based data represented in HF-Filer

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

4.4.1. Summary

As we described in previous deliverable, in our solution, we aim at implicit gestures performed in aeronautic cockpit. In particular, we study pilot's implicit gestures connected with controlling of selected important cockpit elements: Yoke, Touch screen, Navigation control panel, Throttle lever, Electronic flight bag. We define three levels of interaction with a particular element: Full interaction, Touch-and-Go and Unfinished.

For the purpose of implicit gesture recognition we proposed method for detection of transitions between phases of implicit gestures described in previous deliverable. This section contains results of evaluation of this method on collected dataset of implicit gestures.

4.4.2. Current Status and Functionality

At first, it was necessary to improve dataset of implicit gestures collected in Honeywell cockpit simulator, because old version contains insufficient quantity of implicit gestures for correct evaluation. Statistics of current dataset is showed in Table 1. The current version of dataset was used in our experiments. This dataset contains implicit gestures connected with 5 chosen cockpit elements shown in Figure 1. We deal with 12 implicit gestures connected to the 5 cockpit elements.

Table 4: Comparison of statistics for different versions of implicit gesture dataset

	1st version	Current version
Number of videos	30	59
Number of people	10	19
Number of implicit gestures	360	708

We experimented with several settings of the method for implicit gesture recognition. The size of temporal window of frames was set from $n+1 = \{5, 10, 15, \ldots, 45\}$, without skipping of frames, according to the best result. Each probability map from the Pose Machine was subsampled into 47×65

pixels. The set of positive training vectors of frames was increased by randomly time-shifting the annotated moment of gesture phase transition by a normal distribution with standard deviation 1.33 frames. This increased robustness and the size of the training set. Negative vectors of frames were generated from other parts of videos. The resulting random forest consists of 50 trees, where each tree was trained on 1000 positive and 4000 negative randomly chosen training vectors of frames.

Figure 2 illustrates the detection performance of the implicit gesture recognition system for one testing video sequence. The graphs report the response of the detector in time (horizontal axis is time/frame). All of the figures show the response on the "interaction with navigation control panel" implicit gesture, between the first and the second phase. For comparison, the first graph shows the full interaction detection, the second one touch-and-go interaction, and the third one shows the unfinished interaction detection.

These graphs present typical results of our system. The highest responses manifest not only in the location of the given level of interaction (full, touch-and-go, unfinished), but also on the other ones. However, these high peaks are always related to the same cockpit element (e.g. navigation control panel). This behaviour is also confirmed for all implicit gestures in figure 3, where figures 3a (transition between 1^{st} and 2^{nd} phase) and 3c (transition between 2^{nd} and 3^{rd} phase) show results of experiments, where transition moments of implicit gestures related to the same cockpit element were counted as negatives. But figures 3b (transition between 1^{st} and 2^{nd} phase) and 3d (transition between 2^{nd} and 3^{rd} phase) show results of experiments, where transition moments of implicit gestures related to the same cockpit element were not counted as negatives, they were ignored.

However, this behaviour is not malignant, because the borderline between the levels of interaction is unclear. A positive result is that the unfinished gesture can be also reliably detected, without the hand even actually reaching the region of interest (bottom image in Figure 2).

The responses of one gesture detector are constantly low for other parts of the videos, including the other gestures. The recognition of gestures is thus very stable and reliable.

Figure 3 also show results achieved for various frame delay tolerance in recognition of transition between phases. The best results are achieved for 8 frames (22.86 ms) delay for transition between first and second phase of implicit gestures and 24 frames (68.57 ms) delay for transition between second and third phase of implicit gestures. These results also show very similar detection accuracies for smaller delays. So gestures can be detected very accurately in time – start of the second phase at single frame precision (auc: 0.909) and end of the second phase with 22.86 ms (auc: 0.886).

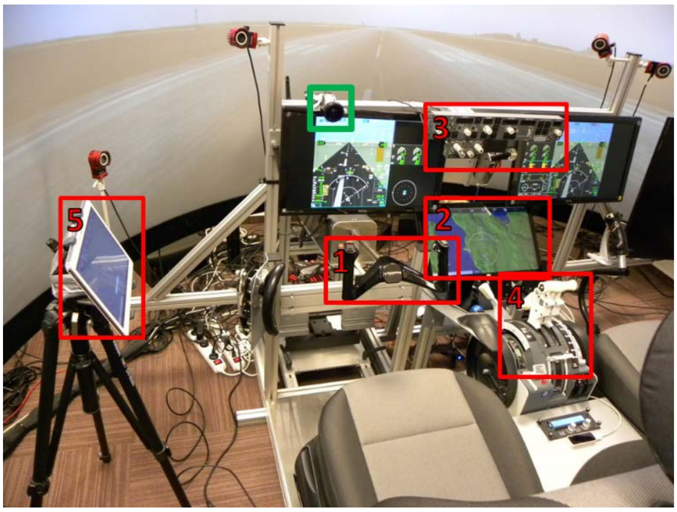


Figure 7: Image of cockpit simulator and its important cockpit elements marked by red rectangles: 1: control wheel, 2: touch screen, 3: navigation control panel, 4: throttle lever, 5: electronic flight bag. Position of RGB camera is marked by green rectangle.

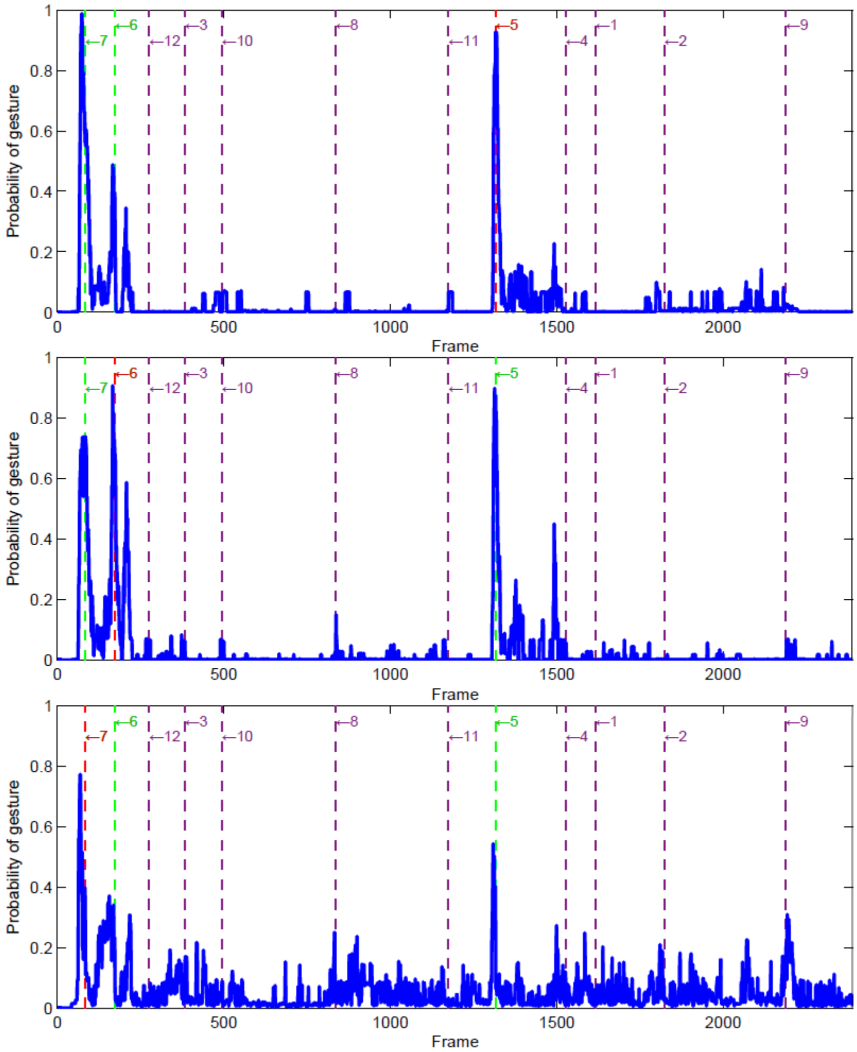


Figure 8: Results of transition recognition between first and second phase of implicit gestures connected with navigation control panel. red dashed line: searched implicit gesture, green dashed lines: implicit gestures connected with the same cockpit element. Labels of implicit gestures: 1: full interaction with throttle lever, 2/3/4: full/touch-and-go/unfinished int. with touch screen, 5/6/7: full/touch-and-go/unfinished int. with navigation control panel, 8/9: full/unfinished int. with throttle lever, 10/11/12: full/touch-and-go/unfinished int. with EFB.

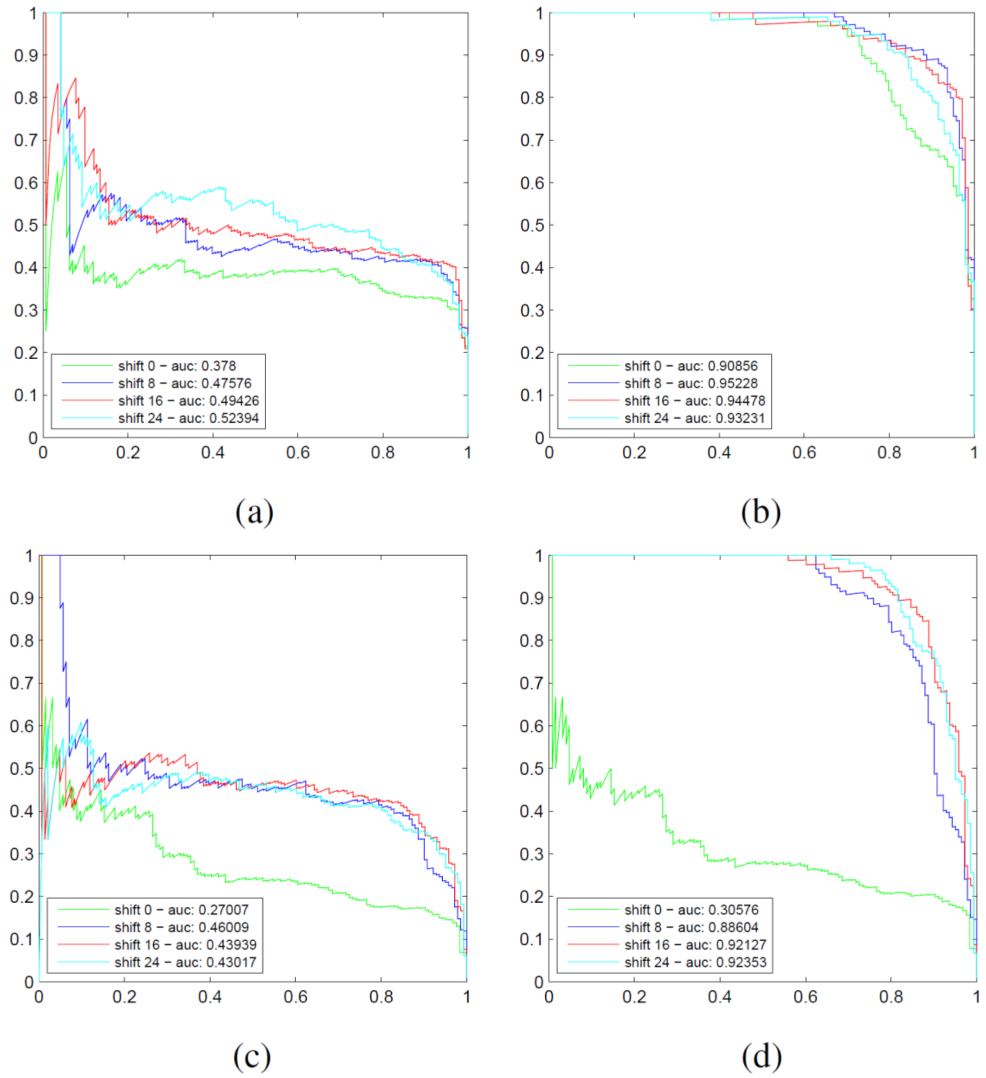


Figure 9: Precision recall curves for transition recognition between phases of implicit gestures for various frame shifts of the moment of transition between phases of implicit gesture. In label for each curve, particular value of area under curve is included. Figures (a) and (b) belong to the transition recognition between first and second phase of implicit gestures. Figures (c) and (d) belong to the transition recognition between second and third phase of implicit gestures. Results shown in figures (a) and (c) belong experiment where transition moments of implicit gestures related to the same cockpit element were not included as negatives, but they were ignored in experiment. Results shown in figures (b) and (d) belong experiment where transition moments of implicit gestures related to the same cockpit element were included as negatives.

4.4.3. Integration of the Outputs of the MTT

Human gestures are one of objective markers for inference of operator psychophysiological state. As part of human communication, the gestures add emotional content to the information that is transmitted. Gestures are either created intentionally though often times in automated way, or unconsciously when they lay psychophysiological state over the expressed information. In this way hidden emotions, attitude, stress or fatigue can be detected [1].

Yet another type of gestures can be considered for human-machine interaction, where the way how human operates the machine reflects the expertise [2], stress [3, 4] or fatigue of the operator [5].

In aircraft cockpit the analysis of gestures supports two safety related aspects. First, gestures provide real time information about the pilot's state and this state information can be combined with other state-related information into a multi-marker classifier. This approach provides more reliable and robust information [6] and is supposed to be the future trend in state classification [7]. Second, gestures provide insight in tasks that the crew is solving and the way how appropriately the crew is solving the tasks. Early recognition of procedural deviation allows for intervention that is cheaper and safer – leaving the crew more time to correct an error.

In WP7 AdCoS for diversion assistance, the information about pilot state is used to trigger the adaption of the system. The diversion assistant evaluates airports in reach from current aircraft position for suitability to unplanned landing. The evaluation of airports considers a number of parameters related to aircraft, route to an airport, and airport. From the parameters a 'difficulty to access and land at an airport' can be derived and combined with actual state of the pilot to bias weights in airport prioritization algorithm. As a result, an airport with easier access/landing will be ranked higher for a fatigue pilot than an airport that is e.g. closer to the aircraft but more difficult to land at, e.g. due to complex morphology or higher traffic.

4.4.4. Showcase – Screenshots and Demonstrator Pics

Figure 10: Aeronautic simulator with the cameras tracking the pose and detecting the implicit gestures.

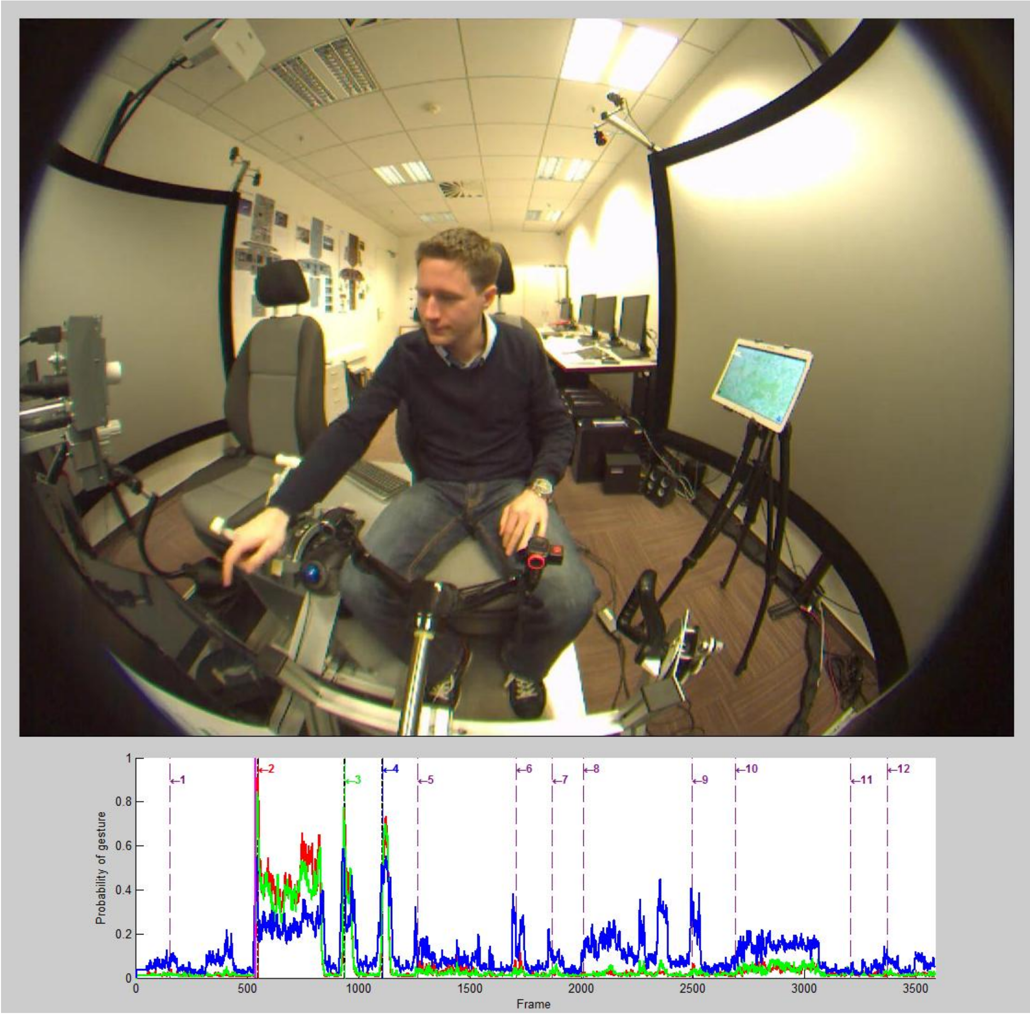


Figure 11: Operator's interface visualizing the detected implicit gestures in time, as related to the input video.

4.5. Detection of operator's head orientation (BUT/HON)

4.5.1. Summary

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

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

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

4.5.2. Current Status and Functionality

We gathered a new video dataset it the Generic Experimental Cockpit (GECO) simulator. The participants were instructed to look several selected displays according to given scenario. At first they were instructed to look at each display for several seconds and then look back at the scenario paper or camera. In the last scenario, they were instructed to go through all the displays without looking back at the paper.

Further, we gathered another video dataset in the GECO simulator with real pilots. We used eye-tracker for this video dataset, so that we captured ground truth position and orientation of the pilot's head together with information about the display at which the pilot was currently looking. At the beginning of each session, the pilots were instructed to look at each display for several seconds, and then followed about 20 minutes long scenario, during which the pilots were conducting various tasks in the cockpit simulator, while their head pose and information about their gaze were still recorded. This dataset will be used to better utilize the connection between the head pose and find the correlation between head pose of the pilot and the position of the display he is watching.

As the head is not usually pointed straight in the direction of the display watched (due to various position of the eyes), this dataset helps to find the relationship between the head pose and the display watched more precisely.

4.5.3. Integration of the Outputs of the MTT

The inputs and outputs of the computer vision module have been identified toward the AdCoS (WP7, Aeronautic Cockpit). Then, the correspondences between the inputs and outputs and the concepts of the HoliDes meta model have been identified. Since the module operates on real-time data from visual sensors, it interacts with the HoliDes meta model indirectly, by processing the external inputs and forwarding the data to the AdCoS's data structures.

For application in diversion assistant AdCoS (WP7, Aeronautics cockpit), the information about the pilot’s head orientation will be used to prevent the attentional tunnelling. Attentional tunnelling is a phenomenon that describes the allocation of attention to a particular channel of information, hypothesis or task goal, for a duration that is longer than optimal, given the expected cost of neglecting events on other channels, failing to consider other hypotheses, or failing to perform other tasks.

The likelihood of getting tunnelled rises in high workload and multitasking environments, which opens space for missing important information or alert with serious consequences: There have been several aircraft accidents for which the attentional tunnelling was identified as the primary cause.

The detection of attentional tunnelling is complex, however a reasonably simple system can be used to mitigate the most prominent effects of the attentional tunnelling – the missed events. This is especially important for EFB applications, such as Diversion assistant, as they can provide useful information, but they can also attract attention of pilots in situation when they should analyse higher priority displays.

The diversion assistant will connect the information about pilot’s head orientation, e.g. pilot looks at the HMI of diversion assistant on EFB, to information about status of other cockpit displays. In case there is a higher priority event shown on any of the higher priority displays, diversion assistant will raise an information window navigating pilot to the display where that event is shown. If pilot does not react, the diversion assistant may even escalate the conflict by switch off its display leaving just the warning message, see Figure 10.

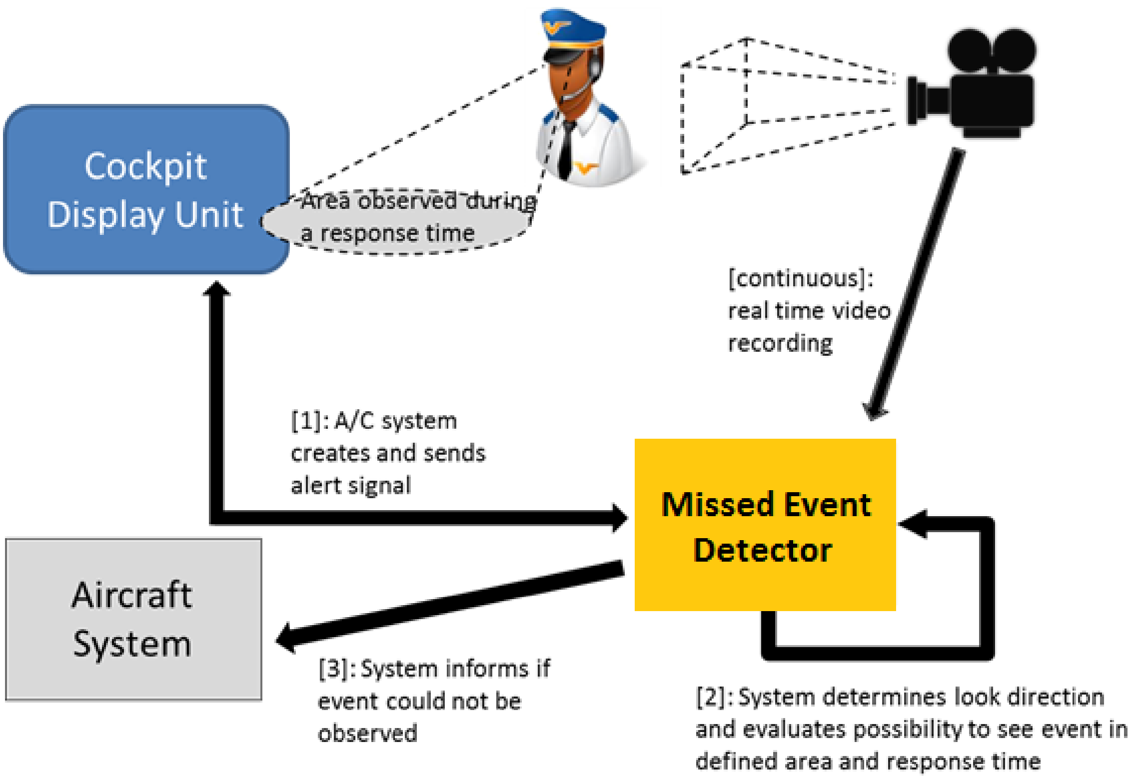


Figure 12: Missed event detector is based on information of pilot's head orientation obtained from the analysis of the cockpit video stream. When the system detects pilot cannot observe an event of high priority it interacts by alerting the pilot.

4.5.4. Showcase – Screenshots and Demonstrator Pics

Figure 13: Aeronautic simulator with the gaze estimation camera (placed on the lower edge of the windshield).

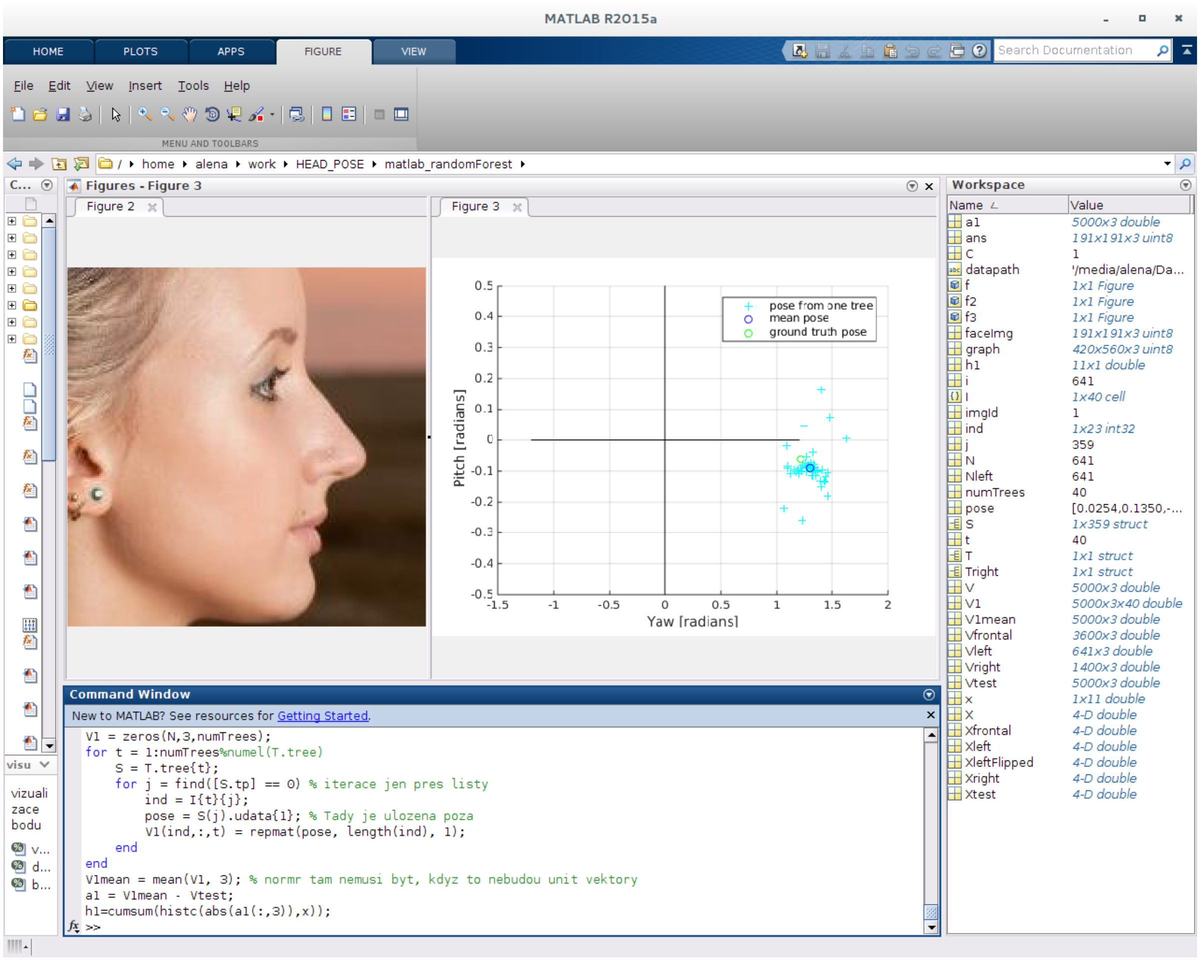


Figure 14: Operator's interface for inspection of the tool's internals – a set of tools and scripts in Matlab.

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

4.6.1. Summary

This MTT serves three purposes: To collect data for the development of the CONFORM-tool (documented in WP3), to understand the specifics of natural lane changes, and to understand users' preferences for automatically driven lane changes. Together, these results are addressing the overtaking-use case and will be part of the IAS demonstrator.

4.6.2. Current Status and Functionality

With the third experiment finished in August 2015, we have finished the data pre-processing and entered the analysis phase. This experiment was specifically directed at the matter of preferences for automatically driven lane changes. Using the human driven lane changes from experiment 2, we clustered different lane change types according to variables such as time headway to a preceding vehicle, driven speed during the lane change, and timing of the lane change itself. Then, the best representative for each cluster was used as an "automation" driving style, to be re-played to the subjects in the driving simulator.

In order to be able to test for a possible preference for a subject's own driving style, each of our 36 subjects in this experiment had also participated in the experiment 2 of this experimental series. Each subject witnessed triplets of different driving styles, re-played in DLR's Dynamic Driving Simulator. After each triplet, the subjects were tasked to indicate which one of the three styles they had liked best, and which one they had like worst.

Altogether, 3 different driving styles were presented, in addition to each subject's own driving style. The results so far indicate that subjects do prefer specific driving styles over their own, yet this pattern is not uniform over all situations. Further analysis will address the question of the root causes for this difference in preference distribution.

4.6.3. Integration of the Outputs of the MTT

The results will be used to train the CONFORM-tool to detect human driving styles from manual driving data. The CONFORM-tool then will be used to execute this function as a module in the IAS-demonstrator. Further, the empirical data from the experiments will be used to model the lane changes the IAS-demonstrator performs for the use case 9.2 of WP 9 (overtaking manoeuvre).

4.7. CPM-GOMS Task Analysis of a lane change for manual and automated driving (DLR)

CPM-GOMS task analysis is applied to model motor, perceptual and cognitive processes necessary for conducting lane changes during driving.

4.7.1. Summary

CPM-GOMS task analysis provides the user with a framework to describe motor, perceptual and cognitive processes of a human interacting with a technical system in detail. Here, we adapt CPM-GOMS to model the processes necessary for a human to accomplish a lane change with an automotive AdCoS on a two-lane highway.

4.7.2. Current Status and Functionality

D5.4 described the general methodology and the data acquisition in detail. Since then, we have made substantial progress in the analysis of the acquired data. In particular, we defined the resources and operators on the motor, perceptual and cognitive level. Resources are entities for actions that the human driver has available classified in the three levels motor, perceptual and cognitive. In contrast, operators are basic action units of a resource. Operators can also have a system unit they act. For instance, the motor resource 'right foot' has (among others) the operator 'press' that can act on the system unit 'throttle'. Example resources and operators are provided in Table 5.

Table 5. Example resources and operators on the three levels motor, perceptual and cognitive with example system units they act on.

level	resource	operator	system unit (example)
motor	left foot	rest	clutch, floor
		press	clutch
	right foot	rest	throttle, brake, floor
		press	throttle, brake
	left hand	rest	indicator lever
		steer	steering wheel
		change mode	indicator lever
		move	to:from:
	right hand	rest	gear stick, wiper lever
		steer	steering wheel
		change mode	gear, wiper
		move	to:from
	head	point at	left mirror, right mirror, wind shield
		move	to:from:
	eyes	gaze at	dashboard, left mirror, …
perceptual	visual	observe	vehicle, platoon, object in environment
		discover	vehicle, platoon, object in environment
		gauge	vehicle, platoon, object in environment
cognitive		decide	
		judge	
		verify	
		detect signal	

automatic procedure

Note: On the motor level in the move operators, the term "to:from:" refers to the fact moving always occurs from a source to a destination location.

Based on the definition of resources and operators, the driver's actions in each lane change (three per driver) were manually coded for the motor and perceptual level. This also includes the attribution of a timing to each operator execution during these lane changes. This gives us the possibility to consider the sequence of actions in a lane change (for an example, see Figure 15) and the average timing of operators (see Figure 16).

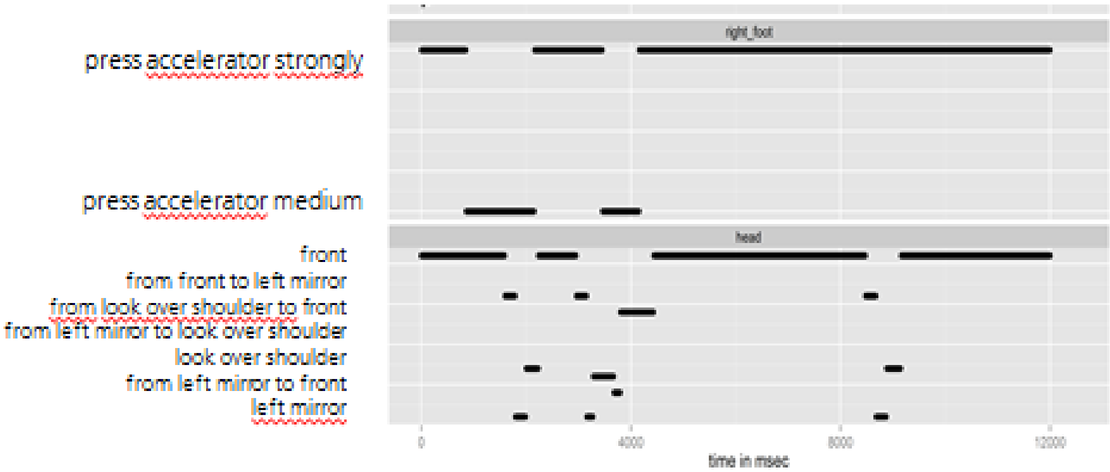


Figure 15. Example of time course of operators in motor resource 'right foot' and 'head' for a lane change from accelerator lane to right lane of one participant.

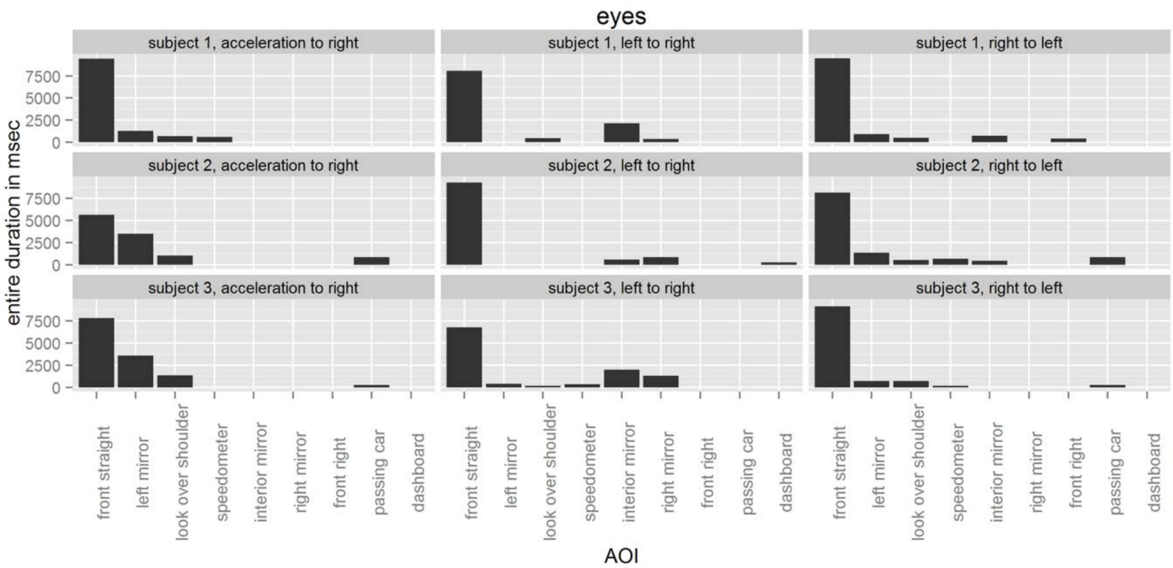


Figure 16. Average timing of the operators of the motor resource eyes for the three lane changes (row) of the three participants (column). Note: AOI = area of interest.

As next steps, we have to accomplish an extensive analysis of the timing of the motor and perceptual operators in order to gain realistic estimates thereof thus aiding current and future modelling activities. Moreover, the coding and timing analysis of operators on the cognitive level needs to be accomplished.

4.7.3. Integration of the Outputs of the MTT

The adapted task analysis for AdCoS can be used at different stages of the AdCoS design process, such as 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. We will apply the method during the overtaking manoeuvre use case of WP9 (9.2) in order to decide on the optimal timing for the handover of control between automated and manual driving. Moreover, the results can be used as basis for cognitive models in WP2.

4.8. Theatre Technique (DLR)

4.8.1. Summary

The Theatre Technique can be used for a low-cost simulation of vehicle automation. It essentially consists of two seat boxes, with interconnected steering wheels and throttle/brake. The driver in seat box two can therefore mimic an automation behaviour, e.g. if the driver in seat box one lets go of the steering wheel.

4.8.2. Current Status and Functionality

The Theatre Technique was used in a two day design session together with IAS on 12-13 November 2015. Aim of the design session was to develop candidate alternatives for a handover-of-control between driver and vehicle in various situations on a two lane highway. This directly addresses the overtaking use case.

As a first step, a number of scenarios were modelled and implemented in the traffic simulation "Virtual Test Drive". These scenarios were then driven by the driver in seat box one (see Figure 17; left side), while a driver in seat box two (right side) provided the service an automation would otherwise. This could consist of either steering, speed control or both.

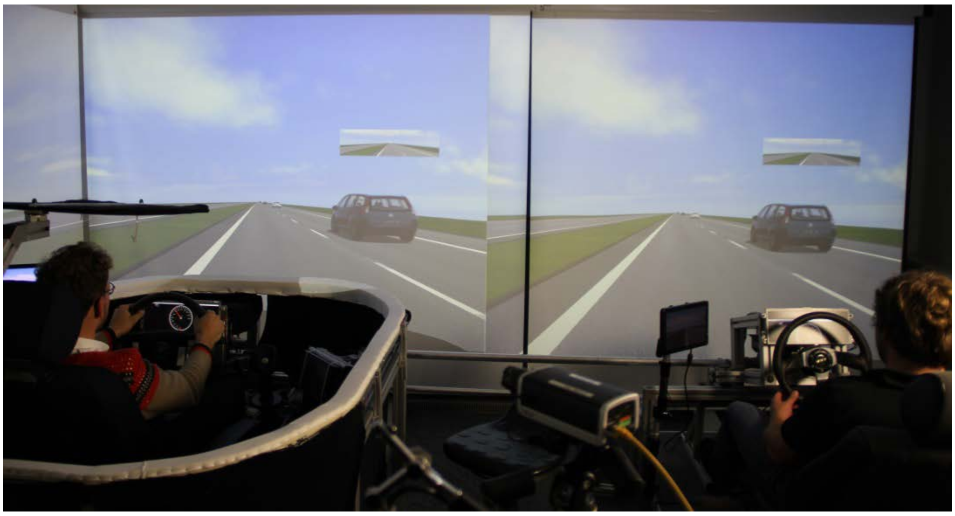


Figure 17. The Theatre Technique as it was used during the Design Session.

The impressions from driving through these scenarios were then directly discussed, documented, and translated in a more abstract representation (see Figure 18). The outcome of the design session has been a novel concept of automation level transitions, involving only a minimum of necessary explicit interaction between vehicle and driver.

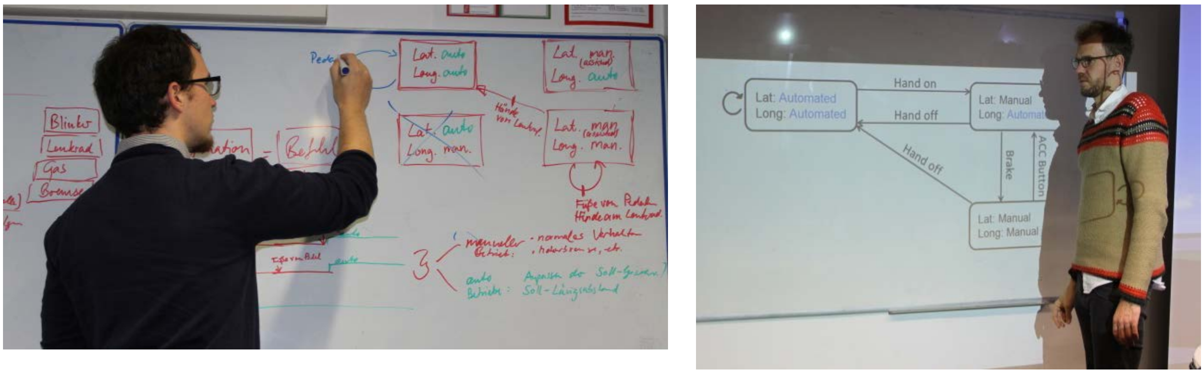


Figure 18. Discussion of the interaction concept.

4.8.3. Integration of the Outputs of the MTT

The Theatre Technique's outcome is a documented novel design concept. This concept then will be implemented, to enable us to test it further in simulated applications. It is planned to then transfer the concept to the IAS demonstrator.

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

The development of the SuRT has already been finished and the final version of this MTT is described in D5.4.

4.10. HF-TA: Human Factors Task Analysis (HFC)

HF-TA is both a method and a tool: as a method, it details the steps of a comprehensive task analysis procedure. As a software tool, it supports this procedure in several ways to make it easier and more efficient.

4.10.1. Summary

The procedure of HF-TA includes several steps, each of which is supported by and represented in the software. Collecting data (from documents, observations, and interviews), exportable visualisation of the collected input, HF expert analyses, and task modelling are included in one platform.

4.10.2. Current Status and Functionality

In the HF-TA tool, the 6 steps of the Human Factors Task Analysis procedure are implemented as tabs in the GUI. The task models resulting from these steps are represented both in table and graphical ("task tree") format. Both representations can be exported (as a *.csv file or image, respectively). Hierarchical, logical and chronological relations between tasks can both be displayed or hidden in the graphical task tree.

- **Theoretical HTA** (hierarchical task analysis): Collecting preliminary knowledge (documentation and training materials) of how a task should be done in the view of the system`s designers. In the tool, the hierarchical structure of subtasks can be entered into a table and viewed as a task tree (graphical representation). The two views are synchronised.
- **Observation**: The theoretical HTA tree is compared to an actual user session by annotating a task path and taking notes on differences to the theoretical view.
- **Card Sorting**: A user`s perception of the task structure can be derived by conducting the Card Sorting usability method directly on-screen within the tool.
- **Interview**: The theoretical HTA is compared with SME's perception of (sub)task demands. A set of predefined (but extendable) questions guide the interviewer and answers can be entered directly for every subtask. For example, it can be of interest which subtask requires which skill.
- **Knowledge Audit**: Also an interview with guidance through standardized questions, but concerning the overall task, the subject matter expert`s bigger picture of it, as well as expert knowledge and strategies.
- **Human Error Template** (HET, for human error identification): The HET checklist is used by an HF expert to classify and evaluate potential errors concerning each subtask.

4.10.3. Integration of the Outputs of the MTT

HF-TA is used in the healthcare domain, in the VCG Triggering and 3D acquisition use cases (PHI). The first feedback from the AdCoS owner was collected, a first round of observations for the VCG Triggering has been done in user sessions, and further feedback cycles are planned after analysing the gathered data. It will be looked into how it could be used together with other MTTs in these use cases (U-DAT) and also in the control room domain. The tool is going to be OSLC-compliant.

4.11. Empirical analysis and validation methods of cognitive processes in the Aeronautic domain (SNV)

Empirical analysis of cognitive processes is being applied to study cognitive workload effects in the aeronautic domain, in the context of DivA use case.

SNV addressed the problem of designing an experiment where it was possible to assess the workload effects on directed attention under different conditions. The main purpose is to use experimental data to learn a classifier able to detect pilots' states, in terms of workload and fatigue. At the moment, behavioural (reaction times) and psychophysiological data (EEG and eye tracking data) from 20 participants performing dual task experiments have been collected.

4.11.1. Summary

A series of experiments has been designed in order to assess the effects of mental workload. Experiments are based on a dual task paradigm in which participants involved in a primary task are asked to perform a secondary interfering task (to test low level visual and high level cognitive attention). Data collected will be used for the construction of a system able to detect pilots' workload.

4.11.2. Current Status and Functionality

The first experiment has been realized. A visual search task has been performed with DMDX experimental software in order to study processes of directed attention. Participants were instructed to identify a red circle (target) among other distracting coloured figures and to press a button corresponding to the portion of the screen where the target appeared.

To induce cognitive distraction a syntactic secondary task has been inserted in the experiment. Specifically, the secondary task was a syntactic transformation task: participants have been asked to transform phrases from active to passive and vice versa. We were interested in observing the behaviour in syntactic transformation of materials with different levels of complexity. Phrases were presented in a headset and participants had to speak aloud in a microphone. This equipment resembled that used by a pilot to talk with the control tower.

Figure 19: Experiment with facelab, Enobio and RTMaps

Reaction times in tasks execution, eye tracking and EEG data have been synchronized by means of RTMaps platform.

The results have highlighted a significant effect of the complexity of materials of the secondary task. Particularly, participants were significantly slow to transform phrases in the condition of ambiguity, i.e., in case of phrases that are designed to intentionally raise doubtfulness or uncertainty as regards the syntactic transformation.

4.11.3. Integration of the Outputs of the MTT

The construction of a classifier able to detect the pilot workload leverages the data collected and analysed by SNV. SNV developed the experimental protocol for the data collection and performed a statistical analysis of such data to detect if participants' behaviour significantly changes in different cognitive tasks. This work satisfies the requirement about the classification of physiological outputs. The integration of this method will be accomplished by uploading the documentation with the description of the method onto the HF-RTP.

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

SNV purpose is to address the need of investigating human performance to assess communication and load processing.

4.12.1. Summary

SNV purpose to assess communication between operators in the energy control room scenario has been pursued by means of an ad-hoc realized questionnaire. Information collected by means of the questionnaire have been used in the designing phase of the AdCoS.

4.12.2. Current Status and Functionality

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

SNV contribution in the framework of WP8 started by analysing common adaptive features to be applied to cognitive and communication processes. Communication processes started to be investigated by collecting the interactions (state of the art) between control room's operators and customers calling for an emergency by means of an ad hoc questionnaire. The questionnaire has been created to assess the lexical and the pragmatic aspects of the interaction. The preliminary results evidenced the need of a more common, formalized grid to interact with costumers to speed up the process with a minor effort. The same kind of baselines will also be collected in a cross-cultural perspective.

4.12.3. Integration of the Outputs of the MTT

Responses to the questionnaire have been used to elaborate a lexicon of the interactions in the energy control room. The lexicon has been used in the design of the communication process and the HMI of the AdCoS. The integration of this method will be accomplished by uploading the documentation with the description of the method onto the HF-RTP.

4.13. Focus group (SNV)

SNV purpose is to contribute to the validation of the AdCoS and to improve the effectiveness of the system collecting relevant information from final users.

4.13.1. Summary

In the context of WP8 SNV developed the protocol for the focus group that has been realized to collect feedbacks about the AdCoS developed by REL for IRN to improve communication flow between the headquarter and the team on the field.

4.13.2. Current Status and Functionality

SNV has developed the protocol to set up and conduct efficiently the focus group. The protocol has been shared and discussed with partners making them able to manage the interview by themselves.

The first step of the procedure was the identification of the moderator, someone very expert in the topic of the interview that has been instructed to interact with participants. Moderator has been provided with the main questions to present to participants to gain relevant feedbacks.

Also important was the identification of an observer, participating in the focus group to the aim to take notes about verbal and non-verbal language used by participants during interactions.

Another fundamental step was the definition of the materials, pictures or textual descriptions, to be presented to participants in order to create the context of the interview. For example, during the focus group realized by REL and IRN a preview of the application realized to facilitate communication between operators and technicians in energy control room.

SNV has defined instructions about the setting of the interview. The whole setting contributes to the success of the focus group. For example, it is important to have a room where participants can seat in circle looking each other and also to set video and audio recorders to avoid the loss of information.

4.13.3. Integration of the Outputs of the MTT

The MTT has been used by REL and IRN for the validation of the application realized to manage the communication flow between operators and technicians. The main aim of this phase is to improve the system making them able to reduce the number of phone calls and the time spent by the operators on the phone with technicians. The same methodology is going to be applied also in the health domain. The integration of this method will be accomplished by uploading the documentation with the description of the method onto the HF-RTP.

4.14. Empirical analysis and validation methods of cognitive and communicative processes in the Automotive domain (SNV)

SNV purpose is to address the need of investigating human performance to assess distraction processes. SNV addressed the problem of designing an experiment where it was possible to assess the distraction driving in a real scenario.

4.14.1. Summary

Empirical analysis of cognitive processes has been applied to study distraction effects in the automotive domain. The data obtained will be implemented in the AdCoS in order to prevent visual and cognitive distraction in safety critical manoeuvres such as lane change.

4.14.2. Current Status and Functionality

In the following the empirical analysis methodology applied to the automotive domain will be described by means of the example of the experiment on drivers visual distraction realized by SNV in collaboration with CRF (De Simone, Presta, Collina & Tango, 2015).

Thirty participants volunteered in the experiment. Subjects have been driving on a test site including a highway part and an extra-urban part. The distraction has been induced by an interfering task consisting in a visual search performed on a secondary touch screen (Surrogate Reference Task (SURT) developed by DLR). The distraction task resembles the interaction of the driver with a real touch screen device mounted on the car dashboard.

Figure 20: CRF vehicle equipment

By means of a proper vehicle equipment configuration, we have been able to collect data about the driving behavior (like the steering angle, vehicle speed, position of the vehicle in the lane, etc.), the driver's head orientation, and the execution of the secondary task (e.g., response times, correct answers, etc.).

We analysed the data in order to compare the driving behavior with and without the SURT in the different road scenarios. The analysis results showed that the distraction task significantly affected the driving performance in both scenarios. We exploited such results to derive the model of the driver's visual distraction based on vehicle dynamics.

4.14.3. Integration of the Outputs of the MTT

The construction of a classifier able to detect the driver cognitive state leverages the data collected and analysed by SNV. SNV developed the experimental protocol for the data collection on the road and performed a statistical analysis of such data to detect if the driver behaviour significantly changes in the different cognitive states. The quality of the real data collected and analysed will contribute to the satisfaction of the requirements about the CR of the final classifier. The integration of this method will be accomplished by uploading the documentation with the description of the method onto the HF-RTP.

4.15. Cognitive Distraction Classifier (TWT)

The Cognitive Distraction Classifier (CDC) is a tool which aims to detect and interpret the level of the drivers' cognitive distraction, based on the driver's behavioural and physical data.

4.15.1. Summary

Driving requires cognitive resources in order to perceive the environment, operate a car, and make decisions. Also during automatic driving the driver should be in control and be able to overtake the system if necessary. Therefore when the driver is cognitively distracted from his/ her primary task of driving, such as due to conversations in the car, there may be a competition for cognitive resources, which could negatively affect the primary task performance and thus safety in traffic.

TWT's CDC aims to detect the driver's level of cognitive distraction based on behavioural (i.e., driving), video (i.e., face-tracking), and audio (i.e. voice) data. The output in the form of the level of distraction and its reliability can be used in AdCoS' to adapt behaviour of the system accordingly. For example, the system could warn the driver with an auditory signal to make him/her alert.

4.15.2. Current Status and Functionality

Previously, we executed an experiment with the goal to investigate the effect of cognitive distraction on facial gestures and driving behaviour. To this end, seven participants carried our three conditions in a driving simulator. In one condition they had the single task of driving, while in the other two conditions they had to perform a mathematical task of different difficulty degree simultaneously.

Figure 21: Driving simulator setup.

We have extensively explored these data offline. The software "Intraface" [8] was used to extract facial recognition marks from two videos that were recorded from the front, and at a 45 degree angle from the side. These recognition marks were integrated to meaningful features. For example, from the marks on the eye-lids, the eye blink rate was derived. The resulting 60 facial features, and 32 driving behaviour features were further analysed and used to develop a classification algorithm using supervised machine learning.

We used two different machine learning algorithms: Naïve Bayes and Boosting (AdaBoost). To classify the data offline, for each participant and for each condition, a part of the data is used to train a classifier, while the remaining data are classified with that trained classifier. As a first step it is common to apply the so-called leave-one-out method (using 90% of the data for training). The advantage being that the more data are used for training, the higher the classification accuracy may be. The disadvantage, however, is that the risk of overfitting is higher: Due to the experimental setup, the rather large training set needs to be put together by means of random selection of data points, increasing the chance of high correlation of neighbouring data (in time). Therefore, as a second step to reflect a more realistic prediction, the data were split into two (equally large) sequentially recorded sets. In addition, we use feature selection to assay whether there are features that help to predict distraction for subsets of drivers or that may be informative for the analysis of distraction of all drivers, i.e., on a more universal level.

Results so far show that the facial features (compared to the driving features) are more important in successful classification of the level of cognitive distraction. The classification analysis is work in progress. An automatic distinction between the different conditions (i.e., level of cognitive distraction) based on machine learning appears to be possible, albeit the classification accuracy varies widely between the three classes at the present stage. To further develop the classification model, audio features will be integrated. We have recorded a first experiment in which we investigate the effect of cognitive distraction on audio (i.e., voice) features. The first audio features have been derived, and current exploration of the data is in progress.

Currently we are building a framework that allows real-time analysis and classification of recorded data, which is crucial for the integration in AdCoS'.

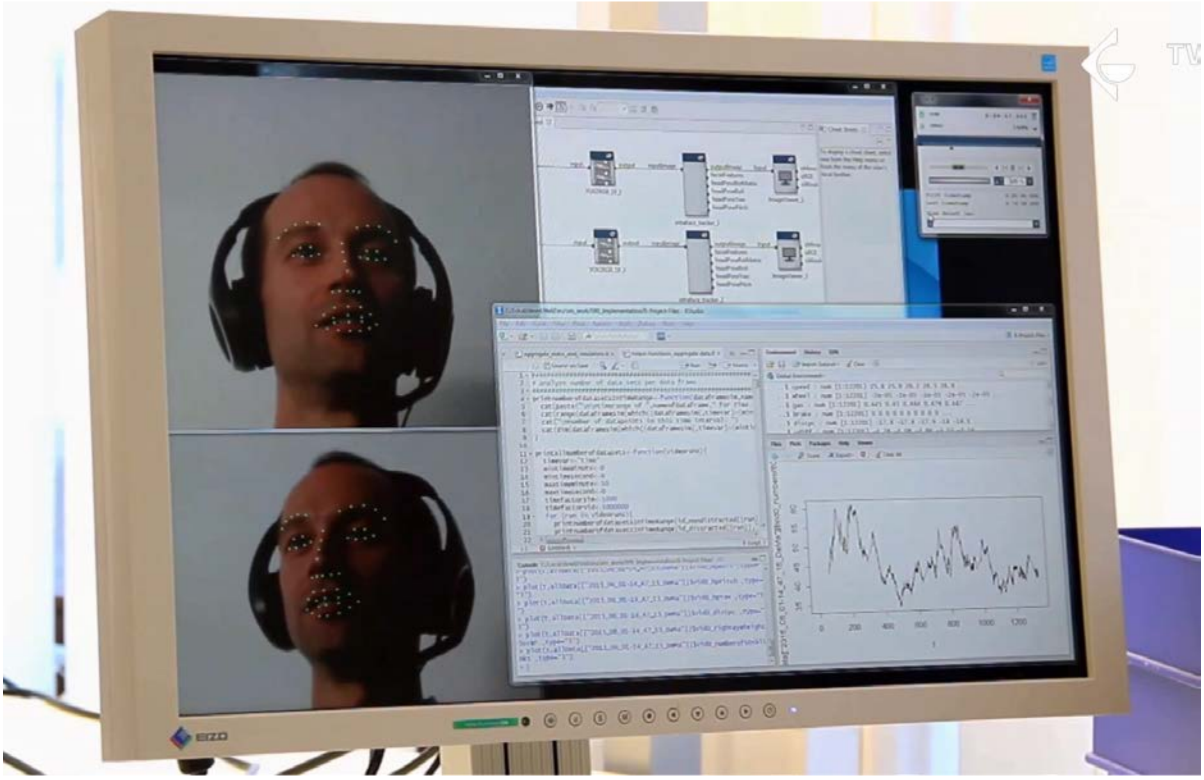


Figure 22: Facial recognition software and data analysis.

4.15.3. Integration of the Outputs of the MTT

In collaboration with TAKATA, we have been preparing joint experiments which are planned for early 2016. In this experiment, around 40 participants will take part in several conditions in a further developed driving simulator, in which either the level of cognitive distraction is manipulated, or the way the system's behaviour adapts.

Together with IBEO we are planning to integrate our offline CDC in their car employing a distracting conversation task. The data that will be collected in that experiment will help us to further develop our classification model based on multimodal features in a natural setting.

5. Update of Requirements and Metrics

This section contains updated requirements, stripped down to the basic requirements and accompanied with evaluation metrics. The document is structured per tasks/MTTs in the Work Package 5; for each task, the list of relevant and important requirements is given, including the requirement ID and name, its description and method of evaluation + evaluation metric.

5.1. HF Filer (AWI)

Requirement: **Reporting on experiment results**

ID: WP6_UMC_HEA_REQ04

WP7_HON_AER_REQ88

WP8_ADS_CTR_REQ28

WP9_CRF_AUT_REQ31

Ver: 0.1

Description: Store a report on experiment data in plain text format according to an itemised experiment plan. The experiment results must be aggregated into a single result in text format for each item in the experiment plan.

Validation & Verification

Method: Qualitative to evaluate if the tool stores reports on experimental data per item in the experiment report.

Metric: User satisfaction

Success: ?

Comment: N/A

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

Requirement: **Compare recorded operator behaviour with official procedures**

ID: WP6_HEA_EBA_REQ_01_v0.2

Ver: 0.2

Description: Match recordings of operator behaviour to procedures to allow comparison between official definitions of procedures and actual operator behaviour.

Validation & Verification

Method: Qualitative evaluation to determine if the technique allows to compare a sequence of user actions (also simulated) with a official definition.

Metric: User satisfaction

Success: ?

Comment: N/A

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

Requirement: **Evaluation of agent action**

ID: WP7_HON_RTP_REQ78

Ver: 1.0

Description: Create a tool/methodology that is able to classify an action of agent (human, machine) being either appropriate or erroneous. It is assumed that the tool has a task/procedure model with all supported alternate actions for a given situation.

Validation & Verification

Method: Quantitative, by success rate of individual implicit gestures recognition

Metric: Success rate, False positive rate, False negative rate

Success: 95% Success rate

Comment: N/A

Requirement: **Evaluation of agent action**

ID: WP7_HON_RTP_REQ78

Ver: 1.0

Description: Create a tool/methodology that is able to classify an action of agent (human, machine) being either appropriate or erroneous. It is assumed that the tool has a task/procedure model with all supported alternate actions for a given situation.

Validation & Verification

Method: Quantitative, by success rate of detection of defined behaviour

Metric: Success rate, False positive rate, False negative rate

Success: 75% Success rate or more

Comment: Moved from the "Detection of operators' head orientation" MTT to better suit the technology.

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

Requirement: Eye-tracker strategy

ID: WP7_HON_RTP_REQ82

Ver: 1.0

Description: Compare benefits and disadvantages of using either head-mounted or cockpit mounted eye-tracker in highly unstable environment (cockpit, car). Define best practices/constraint when either of the two is more relevant.

Validation & Verification

Method: Weighted costs/benefits of different eye-tracker strategies

Metric: Weighted costs/benefits

Success: one technology wins by 10% or more

Comment: N/A

Requirement: Eye-tracker operability

ID: WP7_HON_RTP_REQ83

Ver: 1.0

Description: Investigate strategies of using eye-tracker when the subject needs to - turn head in wide range of angles, - may wear sunglasses or headsets, - undergoes sudden changes in illumination, - may need to change seat, - needs to be monitored for a long period of time.

Validation & Verification

Method: Expert judgement

Metric: outcome of expert judgement

Success: mode of operation verified and feasible

Comment: N/A

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

Requirement: Learning of individual driving behaviour

ID: WP9_DLR_AUT_REQ01

Ver: 1.0

Description: After several manual driven overtaking manoeuvres the driver model has learnt the natural driving behaviour of the driver.

Validation & Verification

Method: Calculation of a confidence value (range between zero and one) based on the observations

Metric: Confidence value

Success: > .9

Comment: N/A

Requirement: Online learning

ID: WP9_DLR_AUT_REQ02

Ver: 1.0

Description: The driver model shall improve stepwise over several overtaking manoeuvres its current knowledge of the driver by considering inputs by the driver (steering angle, brake pedal position, throttle position) while driving manually. The driver model then updates its manoeuvre preferences

Validation & Verification

Method: Expert judgement

Metric: outcome of expert judgement

Success: yes

Comment: N/A

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

Requirement: Handover of control dependent on cognitive load of driver

ID: WP5_DLR_AUT_REQ1

Ver: 1.0 (new)

Description: Handover-of-control between car and driver (and vice versa) should only occur when the driver is capable of handling the information, i.e. handover of control leading to human errors and human behavior leading to automation mode unawareness should be avoided.

Validation & Verification

Method: Expert judgement

Metric: Outcome of expert judgement

Success: yes

Comment: CPM-GOMS provides means to analytically determine phases of cognitive load during lane changes. Thus the results of expert judgement can be enhanced by the results of the CPM-GOMS analysis.

Requirement: Handover of control only when driver possesses sufficient degree of situation awareness

ID: WP5_DLR_AUT_REQ2

Ver: 1.0 (new)

Description: Handover-of-control from car to driver should only occur when the driver has enough situation awareness to take over; thus car/HMI has to provide driver with the information.

Validation & Verification

Method: expert judgement

Metric: outcome of expert judgement

Success: yes

Comment: CPM-GOMS provides means to analytically determine which information the driver takes into account at which time during lane changes. Thus the results of expert judgement can be enhanced by the results of the CPM-GOMS analysis.

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

Requirement: Handover of control dependent on cognitive load of driver

ID: WP5_DLR_AUT_REQ1

Ver: 1.0 (new)

Description: Handover-of-control between car and driver (and vice versa) should only occur when the driver is capable of handling the information, i.e. handover of control leading to human errors and human behavior leading to automation mode unawareness should be avoided.

Validation & Verification

Method: Expert judgement

Metric: Outcome of expert judgement

Success: yes

Comment: Theatre technique is a method to test early in the development cycle whether or not certain handover of control strategies violate certain requirements. Thus, experts can use theatre technique to test and judge certain scenarios.

Requirement: Handover of control only when driver possesses sufficient degree of situation awareness

ID: WP5_DLR_AUT_REQ2

Ver: 1.0 (new)

Description: Handover-of-control from car to driver should only occur when the driver has enough situation awareness to take over; thus car/HMI has to provide driver with the information.

Validation & Verification

Method: expert judgement

Metric: outcome of expert judgement

Success: yes

Comment: Theatre technique is a method to test early in the development cycle whether or not certain handover of control strategies violate certain requirements. Thus, experts can use theatre technique to test and judge certain scenarios.

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

Requirement: Data for distraction model is derived in validated distraction paradigm

ID: WP5_DLR_AUT_REQ3

Ver: 1.0 (new)

Description: The data that is used to train the classifier for the distraction model of the AdCoS was collected using a validated and well-established secondary task for inducing distraction

Validation & Verification

Method: comparison with norms

Metric: Distraction task is described in a norm

Success: yes

Comment: SuRT task is described in ISO/TS 14198:2012(en) (https://www.iso.org/obp/ui#iso:std:iso:ts:14198:ed-1:v1:en)

Requirement: **Data for distraction model is derived in real traffic.**

ID: WP5_DLR_AUT_REQ4

Ver: 1.0 (new)

Description: The data that is used to train the classifier for the distraction model of the AdCoS is collected in real traffic and not in the simulator

Validation & Verification

Method: setup in real car

Metric: yes/no

Success: yes

Comment: The SuRT task can be used in a real car (and has been done by CRF/SNV).

5.9. HF-TA

Requirement: **Task Modelling**

ID: WP6_HEA_ixR_UC01_3D_acquisition_REQ6

Ver: 0.1

Description: The tooling should be able to model the work flow of the tasks with indication of their dependencies and order of execution. It should also capture details of each step of the interaction, like interfaces, the controls and displays involved.

Validation & Verification

Method: Check: Ability of the tool to model the abovementioned details for all important steps under specified circumstances (scenario).

Metric: Successfully modelled sequence of interaction in a given scenario.

Success: yes / no

Comment: N/A

Requirement: **Task analysis data gathering**

ID: WP6_HEA_ixR_UC01_3D_acquisition_REQ7

Ver: 0.1

Description: The tooling supports task analysis procedure and collection of data from document analysis, observations, expert interviews, capturing of users` mental models, as well as error analysis, and combines these in one platform such that they can be further analysed together.

Validation & Verification

Method: Success is measured by the ability of the tool to gather data in the specified procedures of task analysis in one or more use cases.

Metric: completeness of checklist

Success: complete (yes / no)

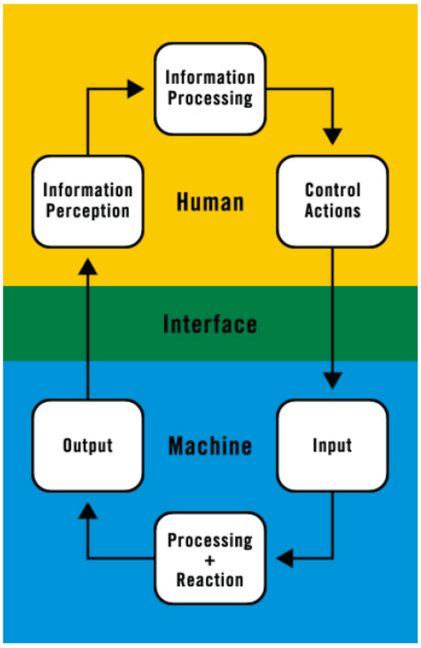
Comment: N/A

Requirement: Perception, Cognition, and Action (PCA) Requirements of Tasks

ID: WP6_HEA_ixR_UC01_3D_acquisition_REQ8

Version: 0.1

Description: The tooling should be able to capture what the human operator needs to perceive for a task (sensory input like vision and hearing), whether the task involves higher cognition, and how s/he has to perform the task (manually, speech).



Device User Interface in Operational Context (adapted from Redmill and Rajan, 1997).

Validation & Verification

Method: Ability of the tool to capture the abovementioned task requirements.

Metric: checklist

Success: completeness of checklist

Comment: N/A

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

Requirement: Cooperation between operators and operational teams

ID: WP8_IRN_CR_REQ007

Ver: 3.0

Description: The AdCoS shall improve the cooperation of the operator with the operational teams.

Validation & Verification

Method: Qualitative. Responses to the questionnaire have been used to elaborate a lexicon of the interactions in the energy control room. The lexicon has been used in the design of the communication process and the HMI of the AdCoS.

Metric: AdCoS' designer judgement

Success: Yes

Comment: N/A

5.11. Empirical analysis and validation methods of cognitive processes in Automotive domain (SNV)

Requirement: Classification of driver's cognitive state

ID: WP9_CRF_AUT_REQ3

Ver: 1.0

Description: The classifier of the driver cognitive state shall be able to do that with a CR > (80÷85)%.

Validation & Verification

Method: CR evaluation in the test phase of the classifier

Metric: CR

Success: to be evaluated

Comment: The construction of a classifier able to detect the driver cognitive state leverages the data collected and analysed by SNV. SNV developed the experimental protocol for the data collection on the road and performed a statistical analysis of such data to detect if the driver behaviour significantly changes in the different cognitive states. The quality of the real data collected and analysed will contribute to the satisfaction of the requirements about the CR of the final classifier.

5.12. Cognitive Distraction Classifier – CDC

Requirement: Algorithm for Analyses of Distraction

ID: WP9_TWT_AUT_REQ04_v0.1

Ver: 1.0

Description: Cognitive distraction level classifier algorithm, based on drivers' behavioural-, video-, and audio data.

Validation & Verification

Method: significance testing of classification accuracy (in experiments in which several levels of distraction are induced)

Metric: p-value, classification accuracy

Success: classification accuracy > chance level & p <= 0.05

Comment: N/A

Requirement: Output of Estimated Distraction Level Computed by Algorithm for Analyses of Distraction

ID: WP9_TWT_AUT_REQ14_v0.1

Ver: 2.0

Description: The estimated output level can be used as an input to realise other driving functionalities such as adaptive feedback (e.g., an auditory warning tone when the driver's level of distraction exceeds a certain threshold).

Validation & Verification

Method: true-false evaluation

Metric: true (when comprehensible output is given) or false (when no output is given, or the output is incomprehensible or out of range)

Success: output = true

Comment: updated requirement

5.13. Driver Distraction Classifier

Requirement: Classification of driver's cognitive state

ID: WP9_CRF_AUT_REQ03

Ver: 2.0

Description: The AdCoS should constantly monitor the driver and classify his/her mental state (i.e. distraction) by using real-time data.

Validation & Verification

Method: Classification rate on validation set

Metric: CR% (Classification Rate Percentage)

Success: CR% > 80%

Comment: N/A

6. References

[1] S. K. Card, A. Newell and T. P. Moran, The Psychology of Human-Compute Interaction, Hillsdale, N.J.: Lawrence Erlbaum Associates, 1983.

[2] E. Hollnagel, Human Reliability Analysis: Context and Control., London: Academic Press London, 1993.

[3] 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.

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

[5] A. Schieben, M. Heesen, J. Schindler, J. Kelsch and 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.

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

[7] G.-B. Huang, Q.-Y. Zhu and C.-K. Siew, “Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks.,” *Extreme Learning Machine: Theory and Applications,* no. 70, pp. 489-501, 2006.

[8] M. A. Regan, C. Hallett and C. P. Gordon, “Driver distraction and driver inattention: Definition, relationship and taxonomy,” *Accident Analysis & Prevention,* vol. 43, no. 5, 2011.

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

[10] F. Tango and M. Botta, “Real-Time Detection System of Driver Distraction using Machine Learning,” *Intelligent Transportation Systems,* vol. 14, no. 2, 2013.

[11] 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.

[12] C. Xu, S. Li, G. Liu, Z. Zhang, E. Miluzzo, Y. Chen, J. Li and 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.

[13] T. Horberry, J. Anderson, M. A. Regan, T. J. Triggs and J. Brown, “Driver distraction: the effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance.,” *Accident Analysis Prevention,* vol. 38, no. 1, pp. 185-191, 2006.

[14] M. A. Regan and 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.

[15] L. M. Trick, J. T. Enns, J. Mills and J. Vavrik, “Paying attention behind the wheel: a framework for studying the role of attention in driving.,” *Theoretical Issues in Ergonomics Science,* vol. 5, no. 5, pp. 385-424, 2004.

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

[17] Q. Ji and X. Yang, “Real-time eye, gaze, and face pose tracking for monitoring driver vigilance,” *Real-time imaging,* pp. 357-377, 2002.

[18] H. Saito, T. Ishiwaka, M. Sakata and S. Okabayashi, “Applications of driver's line of sight to automobiles - what can driver's eye tell,” in *Vehical and information systems conference*, 1994.

[19] Y. Ebisawa, “Unconstrained pupil detection technique using two light sources and the image difference method,” *Visualization and*

intelligent design in engineering, pp. 79-89, 1989.

[20] P. Jimenez, L. M. Bergasa, J. Nuevo, N. Hernandez and I. G. Daza, "Gaze fixation system for the evaluation of driver distractions induced by IVIS," *Online,* 2011.

[21] M.-H. Sigari, M. Fathy and M. Soryani, "A driver face monitoring system for fatigue and distraction detection," *International journal of vehicular technology,* 2013.

[22] M. F. Khan and F. Aadil, "Efficient car alarming system for fatigue detection during driving," *International journal of innovation, management and technology,* pp. 480-486, 2012.

[23] E. Hollnagel, Cognitive Reliability and Error Analysis Method. CREAM, Oxford, UK: Elsevier, 1998.