	HUMAN Model-based Analysis of Human Errors during Aircraft Cockpit System Design	

D4.1 - Detailed Experimental Design for Cycle 1

Project Number: **Nature:** **Classification:** **Version:** **Parts & Classifications:** **Work Package(s):**	211988 Deliverable Public V1.0 Main document (Public) WP4
Document Timescale: **Start of Document:** **Final version due to:** **Time now:**	Project Start Date: March 1, 2008 T0+15 T0+16 T0+17
Issue Date (dd/mm/yyyy):	04/08/2009
Compiled:	Tina Mioch, TNO
Authors:	Tina Mioch, TNO Jan-Patrick Osterloh, Andreas Lüdtke, OFF Marcus Biella, DLR
Technical Approval:	Andreas Lüdtke, OFF
Issue Authorisation:	Andreas Lüdtke, OFF

DISTRIBUTION LIST		
Copy type[1]	Company and Location	Recipient
C, E	European Commission DG RTD - Directorate H - Transport CDMA 4/138 B-1049 Bruxelles	Stephanie Stoltz-Douchet
T	HUMAN consortium	All HUMAN partners

[1] Copy types:
M = Master copy,
E = Email,
C = Controlled copy (paper),
D = Electronic copy on disk,
T = TeamSite (Sharepoint)

RECORD OF REVISION			
Version	Date (dd/mm/yyyy)	Status Description	Author
V0.1	07/07/2009	Initial version	TiM
V0.2	21/07/2009	Input from OFF	JPO
V0.3	27/07/2009	Update DLR section	MB
V1.0	04/07/2009	Final check	AL

1 Table of Contents

1 Table of Contents 4
2 Abbreviations and Definitions 5
3 Introduction 6
4 Experimental Set-up 6
4.1 Subjects 6
4.1.1 Pilot Characteristics 6
4.1.2 Preparation of Pilots 7
4.2 Schedule 9
4.2.1 Duration of scenarios 9
4.2.2 Number of scenarios 10
4.2.3 Schedule July – October 2009 10
4.2.4 Schedule per week 10
4.2.5 Schedule for the two day session 10
5 Working Hypotheses 11
5.1 General Hypotheses 11
5.1.1 Hypothesis on comparison of actual and predicted behaviour 11
5.1.2 Hypothesis on statistical comparison of actual and predicted errors 11
5.2 Hypotheses about Learned Carelessness 12
5.2.1 High Level Test Conditions 12
5.2.2 Low Level Test Condition 13
5.3 Hypotheses about Cognitive Lockup 13
5.3.1 High-Level Test Conditions 14
5.3.2 Low-Level Test Conditions 14
6 Scenarios 14
6.1 Additional Cockpit Characteristics 15
6.2 Specific Scenario Characteristics 16
6.3 Provoked Operational Errors 16
7 Summary 17

2 Abbreviations and Definitions

Abbreviation	Definition
AFMS	Advanced Flight Management System
AHMI	Airborne Human Machine Interface (Interface for AFMS)
ATC	Air Traffic Control
ATCO	Air Traffic Control Operator
FMA	Flight Mode Annunciation on PFD
GECO	Generic Cockpit Simulator at DLR
PF	Pilot Flying
PFD	Primary Flight Display
PM	Pilot Monitoring
PNF	Pilot Non Flying, see PM
R/T	Radio / Telephony

3 Introduction

This deliverable provides details on the experimental design (including the experimental scenarios and schedule) and the working hypotheses for the first experimental cycle. Previous deliverables (D2.4 and D2.2) described these aspects on a more generic level. Now that the preparation of the physical simulation platform (WP2) and virtual simulation platform (WP3) are finished the final details can be fixed. This deliverable provides these details and can be considered as an update and complement of D2.2 and D2.4.

Section 4 reflects the work that has been done in Task 4.2. This includes a more detailed description of the test schedule of the recruited pilots, which has preliminary been described in D2.2, and how the pilots are prepared for the experiments. Section 5 provides a summary of the working hypotheses derived in Task 4.1. Finally, in Section 6 the final scenarios that have been designed to test the hypotheses are described.

4 Experimental Set-up

The section describes the final experimental design for the first cycle of simulator experiments on the physical simulation platform.

4.1 Subjects

In this section, the experimental subjects are described in detail regarding their

- characteristics,
- experience,
- nationality and
- gender.

4.1.1 Pilot Characteristics

Altogether 53 pilots expressed their interest to act as a subject in HUMAN trials by the end of June 2009:

- Approximately 2/3 of them are first officers.
- Approximately 1/3 of them are experienced pilots with dedicated interest in aircraft design and operations.
- The majority is employed by German airlines.

In addition to that, three further pilots have been identified to take part in the study as PM. Each of them will perform approximately 4-6 simulator sessions. Two of the pilots are commercial airline pilots, one pilot acts as a back-up and is currently employed at a research institute.

4.1.1.1 Experience

Their average age is approximately 32 years (SD: 8 years) and the mean reported flight hours are 6000 (SD: 5000).

Nevertheless, this is obviously only a rough estimate for the experiences of the actual sample of twelve pilots that will perform HUMAN trials in cycle one.

Nearly 50% of the pilots already took part in at least one study at the DLR facilities in Braunschweig. Therefore these pilots have basic knowledge of the GECO in its previous version (A320-esque cockpit environment with flight dynamics of the VFW 614). Experiments for HUMAN will be performed in an A350-esque cockpit environment, still with flight dynamics of the VFW 614.

4.1.1.2 Nationality

To minimise expenses (in terms of planning and travelling) it was decided to treat German pilots as preferential in the first place.

Nevertheless, pilots from other airlines are available as well (e.g. Czech Republic), but they are currently not involved in the experiments.

4.1.1.3 Gender

With a percentage of 10% female pilots in the population of HUMAN-interested pilots their share is higher-than-average (compared to approximately 6% female pilots among commercial pilots).

4.1.2 Preparation of Pilots

All pilots mentioned above have been contacted to check their availability for HUMAN experiments in cycle 1. Given their first response, the conduction of trials is not in danger by a lack of pilots.

Trials will be performed by one test subject from the HUMAN-interested sample plus an additional pilot monitoring (PM). Three PMs have been identified altogether to perform approximately four to six trials each in the HUMAN trials.

Depending on the availability of the PMs, a schedule for twelve test sessions is right now under construction (July 2009).

Month	Week	Day	Date	Pilot Flying
AUGUST	34	MO / TU	17./18.08.	#1
		WE / TH	19./20.08.	#2
	35	MO / TU	24./25.08.	

		WE / TH	26./27.08.	
AUG / SEPT	36	MO / TU	31.08./01.09.	#3
		WE / TH	02./03.09.	#4
SEPTEMBER	37	MO / TU	07./08.09.	
		WE / TH	09./10.09.	
	38	MO / TU	14./15.09.	
		WE / TH	16./17.09.	
	39	MO / TU	21./22.09.	
		WE / TH	23./24.09.	
SEPT / OCT	40	MO / TU	28./29.09.	
		WE / TH	30.09./01.10.	
OCTOBER	41	MO / TU	05./06.10.	
		WE / TH	07./08.10.	
	42	MO / TU	12./13.10.	
		WE / TH	14./15.10.	#5
	43	MO / TU	19./20.10.	#6
		WE / TH	21./22.10.	
	44	MO / TU	26./27.10.	
		WE / TH	28./29.10.	#7
NOVEMBER	45	MO / TU	02./03.11.	#8
		WE / TH	04./05.11.	
	46	MO / TU	09./10.11.	
		WE / TH	11./12.11.	

4.1.2.1 Requirements for the experiments

Only commercial airline pilots are considered as test subjects for HUMAN. Neither private nor military pilots are in the focus of the HUMAN research topics in the first place.

Average knowledge is needed to perform the scenarios for HUMAN. Normal and state-of-the-art procedures will be flown here. Nevertheless, the futuristic approach is in the use of the AHMI. AHMI knowledge (acquired in previous DLR studies) is only available for five pilots of the HUMAN sub-sample. In general, each commercial airline pilot is regarded as capable to perform trials with the DLR's AHMI. As the AHMI can be considered as comparatively complex assistance system (in contrast to Enhanced Vision Systems and Taxi Guidance Systems which don't involve as many operating steps as the AHMI), additional training will be needed.

Pilots will have to accept that clearances will be given in these scenarios via data link; no R/T communication is foreseen in the HUMAN scenarios in cycle 1.

Furthermore, they will have to be aware that – deriving from nowadays operations - AHMI's trajectories / clearances will be given for the complete trajectory in these scenarios - and not partially. This will be ensured by thoroughly briefing.

4.1.2.2 Preparation to fulfil the requirements

It is crucial to ensure that possible errors by the pilots won't be caused by their limited experience with the AHMI but by underlying error producing mechanisms (cf. D1.2 Error Types and Error Producing Mechanisms).

Therefore the following way of briefing the AHMI is in use:

1. Training starts with an AHMI handbook. It is sent to the pilots in advance of the trials and consists of the following chapters:
   - Description of the Navigation Display
   - View Mode Buttons
   - Display Mode Buttons
   - Generating and Negotiation Trajectories
   - Creating and Editing Constraint Lists (main focus of the document)
     - (including examples of the HUMAN [training] scenarios)
2. Training continues with AHMI on site.
   - This will be the ultimate preparation phase in the GECO. Pilots will fly training scenarios based on examples of the HUMAN scenarios. Here for example a manual Rotterdam approach will be flown as well in order to make the pilot used to the VFW 614 flight characteristics.
   - Main focus will be on the (re-)planning functionality of the AHMI in terms of
     - horizontal diversions
     - vertical diversions
     - timed diversions

     both as pilots' and as ATCO's request.
   - While the average period of training will be approx. 180 minutes, it should be highlighted that the practical training will take as long as both the DLR instructor and the pilot himself consider it as necessary.

4.2 Schedule

4.2.1 Duration of scenarios

The average duration of a scenario is approximately 35 minutes. The duration of a scenario shall not exceed fifty minutes even though the eye and head tracking device is now worn on a convenient standard bike helmet.

4.2.2 Number of scenarios

Eight scenarios (cf. D2.4 Annex II) have been created for cycle 1.
This is the highest possible number for scenarios to be conducted in two days including training and in-depth debriefing sessions for each scenario.

4.2.3 Schedule July – October 2009

Pre-trials are conducted in July 2009. According to the amendment, it is expected to conduct twelve trials per two days in the months of August, September and October.

4.2.4 Schedule per week

It is preferred to have slots on "Monday and Tuesday" and "Wednesday and Thursday". Nevertheless, other slots may be done as well, depending on the availability of the pilots.

4.2.5 Schedule for the two day session

Day one will start will a general theoretical briefing, training on both the GECO and AHMI, i.e. not only AHMI functionality will be trained but one Rotterdam scenario will be flown manually as well, to get the pilots used to the VFW 614 flight characteristics. This will last approx. three hours.

A lunch break (60 min.) is planned as well. (Coffee) breaks will take place whenever they are needed, in total additional 60 min. altogether).
Each scenario A-H consists of briefing (5 min.); calibration (10 min.); the test run itself (35 min.) and the subsequent debriefing (30 min.) Therefore the total run time is approx. 80 minutes.

The following order of scenarios is proposed:
Day one:

- Scenario G (Learned Carelessness + trap, but no error) (~80 min.)
- Scenario F (Learned Carelessness build up) (~80 min.)
- Scenario C (Learned Carelessness build up) (~80 min.)
  (+Cognitive Lockup)

Day two will continue with scenarios:

- Scenario D (Learned Carelessness build up) (~80 min.)
- Scenario H (Cognitive Lockup) (~80 min.)
- Scenario A (Baseline) (~80 min.)
- Scenario E (Learned Carelessness build up) (~80 min.)

- Scenario B (Learned Carelessness + trap, error expected) (~80 min.)

5 Working Hypotheses

In this section, the final working hypotheses, as initially described in D2.4, are presented. The working hypotheses guide the design of the scenarios, as the scenarios serve to provoke and test these hypotheses.

Section 5.1 describes general hypothesis for the comparison of actual and predicted activities. In section 5.2and 5.3 we will then describe the working hypothesis for cognitive error production mechanisms relevant for the first experimental cycle (Learned Carelessness and Cognitive Lockup).

5.1 General Hypotheses

This section lists the general hypotheses, mainly regarding the statistical comparison of actual and predicted behaviour.

5.1.1 Hypothesis on comparison of actual and predicted behaviour

Hypothesis: The pilots and the cognitive model will show the same normative behaviour. This includes the execution of procedures, but also decisions and choices made by the pilots and the model.

This means that the behaviour of the model has to be compared to the behaviour of the pilots. The techniques and methods to accomplish this are described in D2.6.
The same behaviour is shown if in decision points, e.g. the same decisions are made, the procedures that are executed are executed by means of the same actions in the same order, and the timing is about the same (for motor and percept actions).
We expect that there are variances in the behaviour of the pilots; this variance will be taken into account when evaluating whether the model and the pilots show the same behaviour. The same variance should be found in the model behaviour.
The behaviour can be extracted by interpreting the eye gaze data, the order of motor actions (e.g. button clicks) and the timing of the motor actions.

5.1.2 Hypothesis on statistical comparison of actual and predicted errors

Hypothesis: The cognitive model will predict a subset of the errors a pilot could perform during the experiments, namely these errors that could be explained by

Learned Carelessness (see 5.2) and Cognitive Lockup (see 5.3). For these errors, the pilots and the cognitive model will show the same number of errors.

This means that we will statistically compare the number of errors, the kinds of errors and if possible the kind of error production mechanisms. The model should predict the same amount of errors, at the same events, with the same variance as the human pilots.

5.2 Hypotheses about Learned Carelessness

We expect that the cognitive model will show the effects of Learned Carelessness after the same number of task repetitions like pilots (10-15 times) when they interact with the AHMI. In addition, we expect that the pilots will show additional effects of Learned Carelessness, when they interact with the old systems of the cockpit, e.g. the autopilot automation.

In order to analyse the effects and hypothesis of Learned Carelessness, we will concentrate our analysis on tasks on the AHMI, because this system is unknown to the subjects, and thus routine could only emerge during the trials. This has the risk that no routine could be established during the experiments, because more than the expected 10-15 repetitions are needed.

In the following the high-level and low-level test conditions for Learned Carelessness are described:

5.2.1 High Level Test Conditions

We distinguish two different types of situations in which Learned Carelessness can arise, causing two different Error Types (Task Omitted, or Application of Bad Rules).

Task Omitted:
(1) T is a (sub-)task which shall be triggered by an event E.
(2) T is a task to check a set of (flight) parameters with regard to prescribed conditions.
(2) Scenarios are flown in which E occurs N times --> pilots perform T without recognizing any deviations from prescribed conditions.
(3) Scenario is flown in which E occurs while pilots experience high workload.

Hypothesis: Pilots will omit task T because they assume (based on past experience during the N repetitions of T) that the prescribed conditions are fulfilled. Thus, they invest all their resources to deal with the high workload situation.

Example: For each ATC uplink, the pilot has to check the horizontal and vertical deviations. On the horizontal display changes are highlighted, but not on the vertical one, thus checking the vertical changes needs more effort. If the pilot has not recognised vertical changes, e.g. on the cruise flight level, it is likely that he will omit this check.

Application of Bad Rules:
(1) T is a task with a decision point (pilot has to decide either to perform action A1 or A2 depending on the context).
(2) Scenarios occur in which T has to be performed N times in contexts in which action A1 is correct.
(3) After these N repetitions of T, a scenario S occurs in which A2 would be correct.

Hypothesis: Pilots will learn a simplified rule for T in which conditions to solve the decision point correctly are no longer present. Thus in S they will not recognize that A2 would be the correct action. Instead they will perform A1 just like in the preceding scenarios.

Example 1:
T = Re-planning. At the beginning of this procedure the PF has to check if the AP is switched to FMS. If not, he has to perform associated actions. This could be a decision point to investigate.
A1 = not pressing any buttons
A2 = pressing the buttons (pushing the knobs) in order to switch the AP to FMS.
N = X. N is a parameter in the cognitive model which can be set to a reasonable value, e.g. 12.

Example 2:
Each activation or mode change that has influence on the AC's flight progress must be verified by the FMA. Not monitoring the FMA is probably well learned and has proceeded into routine behaviour which captures PF and prevents him from observing FMA data. This is the application of bad rules and leads to not monitoring important data.

5.2.2 Low Level Test Condition

Non-performance of an action, cross-check, or check-list.

5.3 Hypotheses about Cognitive Lockup

We expect that the cognitive model will show cognitive lock-up in situations where a particular demanding task needs to be executed, and another task will occur that also needs to be handled.

In the following the high-level and low-level test condition for Cognitive lockup is described, followed by the according low-level test conditions.

5.3.1 High-Level Test Conditions

For Cognitive Lockup (formerly referred to as attention capture), we have defined the following high-level test conditions.

1. Scenario in which pilots have to perform a task T1.
2. T1 is a task that is cognitively very demanding. This increases the 'workload'.
3. Event E1 occurs which requires to interrupt T1 in order to perform T2 immediately (T2 requires immediate action).
4. The pilot will not (immediately) change to task T2, but will finish or continue T1.
5. The reaction time to start T2 is larger than when no other task, or a not very demanding task is executed.

Hypothesis: Pilot will have the tendency to stay with the current task because of a high 'workload' (in this case a task that is very demanding).

5.3.2 Low-Level Test Conditions

The following example implements the high-level test conditions:
The scenario is the same as for learned carelessness. That means that at a particular point in time, the ATC sends a shortcut with a vertical profile that violates the altitude constraint. The model does not notice this violation because it omits the altitude check. Thus, the a/c starts to re-climb to altitude 11000. After a certain while, the model recognizes the climb during regular monitoring of the flight conditions. Correcting the altitude is T2 (of the high-level test conditions).
However, the cognitive model is at the same time executing another task, T1, which is the monitoring of a storm that is potentially critical. This task is very demanding, and all attention is focused on this task. For that reason, the pilot will not switch immediately to T2, but will switch considerably later than in the scenario without the monitoring of the storm.

6 Scenarios

This section described the final scenario design for the first experimental cycle.

There have been eight different scenarios defined, A to H, incorporating different planning, re-planning and system events with direct reference to AHMI/AFMS operation of the experimental flight. The scenarios represent short-distance flights between Zürich (CH), Frankfurt/Main (D) and Rotterdam (NL), respectively, along pre-defined routes covering an en-route section, approach and landing. The traces of the experimental flights form an input with regard to the definition and validation of the pilot model, i.e. the virtual simulation platform (VSP).

The scenarios are refined based on pre-trials that have been performed on the physical simulation platform. The description below reflects an update and refinement of the scenario description in D2.4 (main document and Annex II).

6.1 Additional Cockpit Characteristics

The basic cockpit characteristics are described in the D2.4 document. The layout of the simulator cockpit is derived from the coming Airbus A350 XWB aircraft and incorporates a state-of-the-art glass-cockpit with wide-screen LCD as well as modern input devices like the KCCU (Keyboard Cursor Control Unit) introduced by Airbus on the A380 flight deck. The underlying flight dynamics of the simulator are based on the ATTAS simulation representing a twin-jet VFW 614 built by VFW/Fokker, in use as test aircraft at DLR since 1985.

To support a short and unproblematic training phase of the subject pilots it was decided that flaps, slats, and speed brakes lever as well as gear lever are operated manually, only, even though a future system like the AHMI/AFMS - to be in operation in about 15 to 20 years - most probably will be prepared to control these levers directly. However, this direct system input currently does not represent a standard for airline operation. Thus, the PF is forced to operate the levers himself or the PF may request the PM to do so.

Moreover, after touch down on the runway the PF will have to operate the rudder pedals and brakes manually.

At an early stage of the project it was decided to abandon voice communication between pilot and ATC controller because one of the kernel features of the AHMI/AFMS is regarded the 'gate-to-gate' negotiation of a 4D-trajectory which does not require additional voice communication under normal operational conditions. However, it was decided that ATC commands like "Go direct to …" or "Turn right heading …" should be enabled to get an interaction mix of the pilot with the autopilot/auto-thrust system and the AHMI/AFMS, respectively. Therefore, such messages pop up a window on the AHMI providing the ATC message content and buttons for the pilot to approve the message contents at once. An immediate modification of the active constraint list is not necessary, but an FMS-controlled flight requires the modification of the constraint list.

6.2 Specific Scenario Characteristics

Another kernel feature of the AHMI/AFMS is regarded the time control of the flight, i.e. time constraints for touch-down or any waypoint of the flight plan may be inserted by the pilot or ATC. This feature requires a system setup with an absolute simulation time, which may be used to setup the cockpit view system accordingly, also.

The absolute simulation time will be used to setup daytime and night-time flights, 6 of the scenarios start app. 11:00 (daytime), two of them app. 23:00 (night time).
Within the experiment flights the pilot is requested to find and realise route short-cuts himself. After modifying the constraint list/flight plan the pilot will send the entire remaining trajectory to ATC and, after approval, engage the trajectory for aircraft control. Actually, a sector-based approval of ATC represents the standard operation mode of aircraft within an ATC environment. However, as 'gate-to-gate' operation represents a kernel feature of the AHMI/AFMS the negotiation process between aircraft and ATC covers entire 4D-trajectories, only.

The list of events described by the D2.4 document includes aircraft system messages, i.e. 'fuel pump malfunction' and 'cabin pressure alert'. The cabin pressure alert, in particular, may result in the pilot decision to reduce altitude by manual flight operation, depending on the emergency procedures applied for the particular airline. The PM will have to advise the PF to re-enter automatic flight mode after having reached the corresponding altitude.

Each flight scenario ends with the touch down point (TPT) as part of the constraint list, which must not be "moved" because it represents the runway threshold.

6.3 Provoked Operational Errors

Especially scenarios B and H (cf. D2.4, Annex II) are very likely to produce errors because the pilot has performed successfully more than ten re-planning tasks by now. Within these repetitions checking the vertical display always led to a positive result meaning that the vertical profiles were correct. It is our hypothesis that pilot will omit these checks of the vertical profile due to “learned carelessness”.

This psychological theory of “learned carelessness” states that humans have a tendency to neglect safety precautions if this has immediate advantages, e.g. it saves time because less physical or cognitive resources are necessary. Careless behaviour emerges if safety precautions have been followed several times but would not have been necessary, because no hazards occurred. Then, people tend to omit the safety precautions and the absence of hazardous consequences acts as a negative reinforcer of careless behaviour.

One other specific error type (cf. D1.2) is expected due to the tasks at the end of the scenario. Due to a certain amount of time pressure it is likely that some pilots' attention is still captured by the first task, resulting in neglecting the second one. We expect associated errors in scenarios B and C.

7 Summary

In this document, an update and refinement has been given of the preparation and planning of the first experimental cycle. The test schedule has been described, in addition to how the pilots are prepared for the experiments. A summary of the working hypotheses has been given and an update and refinement of the scenarios that have been set up in a way that the hypotheses can be evaluated and tested. The preparation of the experiments has progressed as planned.