Test report preliminary testing with system pilots cognitive task analysis

Anna Martins, Matthias Wies (DLR), Sara Silvagni (Deep Blue), Tanja Bos (NLR), Bruno Berberian, Nicolas Maille (ONERA), Barry Kirwan (EUROCONTROL), Alia Lemkaddem, Olivier Grossenbacher (CSEM), Rebecca Charles, Jim Nixon (CRANFIELD), Jaime Diaz Pineda (CATIE)

Short abstract: Future Sky Safety is a Joint Research Programme (JRP) on Safety, initiated by EREA, the association of European Research Establishments in Aeronautics. The Programme contains two streams of activities: 1) coordination of the safety research programmes of the EREA institutes and 2) collaborative research projects on European safety priorities.

The concept of Human Performance Envelope (HPE) was introduced and defined in Deliverable D6.1 "Concept for Human Performance Envelope" of Project P6 "Human Performance Envelope" of Future Sky Safety (FSS). The validation plan of the real time simulations, where the different HPE factors were manipulated in order to provoke a slow degradation of the pilots' performance, is described in FSS Deliverable D6.2. The main goal of this report is to describe the results of the real time simulations. The interactions and connections of the results will subsequently be described in the next deliverable (D6.4).

Programme Manager	Michel Piers, NLR
Operations Manager	Lennaert Speijker, NLR
Project Manager (P6)	Marcus Biella, DLR
Grant Agreement No.	640597
Document Identification	D6.3
Status	Approved
Version	2.0
Classification	Public

This page is intentionally left blank

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Contributing partners

Company	Name
DLR	Ana Martins, Matthias Wies, Marcus Biella
DBL	Sara Silvagni, Carlo Valbonesi
NLR	Frederik Mohrmann, Tanja Bos
DLH	Carsten Schmidt-Moll
ONERA	Bruno Berberian, Nicolas Maille
ECTL	Barry Kirwan
CSEM	Alia Lemkaddem, Olivier Grossenbacher
CRANFIELD	Rebecca Charles, Jim Nixon
IPB/ENSC	Jean-Marc Andre, Jaime Diaz Pineda
TAV	Sylvain Hourlier

Document Change Log

Version	Issue Date	Remarks
1.0	19-12-2016	First formal release
2.0	20-12-2016	Second formal release

Approval status

Prepared by: *(name)*	**Company**	**Role**	**Date**
Ana Martins	DLR	Project (P6) Member	05-12-2016
Matthias Wies	DLR	Project (P6) Member	05-12-2016
Checked by: *(name)*	**Company**	**Role**	**Date**
Bernd Korn	DLR	Quality assurance	15-12-2016
Approved by: *(name)*	**Company**	**Role**	**Date**
Marcus Biella	DLR	Project Manager (P6)	19-12-2016
Lennaert Speijker	NLR	Operations Manager	20-12-2016

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

ACRONYMS

Acronym	Definition
AC	Alternating Current
ACARE	Advisory Council for Aeronautics Research in Europe
ATC	Air Traffic Control
BPM	beats per minute
BR	Breath Rate
CAT	Category
EASp	European Aviation Safety plan
EC	European Commission
ECG	Electrocardiogram
EEG	Electroencephalogram
ELEC	Electric
EMG	Electromyography
EPR	Engine Pressure Ratio
EREA	association of European Research Establishments in Aeronautics
EU	European Union
FAF	Final Approach Fix
FSS	Future Sky Safety
GS	Glideslope
GSR	Galvanic Skin Response
HF	High Frequency
HMI	Human-Machine Interface
HPE	Human Performance Envelope
HR	Heart Rate
HRV	Heart Rate Variability
IAF	Initial Approach Fix
ILS	Instrument Landing System
ISA	Instantaneous Self-Assessment
LF	Low Frequency

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

LOC	Localiser
MATB	Multi Attribute Task Battery
NDB	Non-directional beacon
PPG	Photoplethysmography
RMSE	Root Mean Squared Error
RMSSD	Root Mean Square of Successive Differences
RWY	Runway
SA	Situation Awareness
SACL	Stress Arousal Checklist
SD	Standard Deviation
SESAR	Single European Sky ATM Research
SMI	SensoMotoric Instruments
SRIA	Strategic Research and Innovation Agenda
ST	Stress
TOD	Top Of Descent
VFR	Visual Flight Rules
VLF	Very Low Frequency
WL	Workload
WP	Work-Package

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

EXECUTIVE SUMMARY

Problem Area

The metaphor underpinning the Human Performance Envelope (HPE) concept suggests that, when studying performance degradation and recovery, we need to consider a range of interdependent factors (e.g., workload, fatigue, etc.) as a whole, instead of considering one/two single factors in isolation. If these factors, working alone or in combination, are studied borrowing the envelope metaphor, it can be possible to determine the starting point in which significant performance degradation could affect safety.

Description of Work

In this deliverable we describe potential interactions between the HPE components through a real-time simulation where we collected behavioural, psycho-physiological, performance-based and subjective data. The triangulation of the measurements allows us to determine:

- Points where human performance deteriorates;
- Behavioural and/or physiological markers, which are critical in signalling performance degradation;
- How to increase the envelope improving performance and safety;
- How to develop effective recovery measures through innovative HPE based solutions.

This deliverable presents how the HPE can be measured in pilot-in-the-loop simulations in high fidelity cockpit simulators with Air Traffic Control support. The most promising psycho-physiological, performance-based, and subjective measures that can be triangulated to characterize the HPE and the associated degradation and recovery points are outlined, as well as an experimental paradigm for HPE measurement.

Results & Conclusions

The results of the simulator experiments with airline pilots show that measuring mental workload and stress is possible, especially with physiological measures. Instead, measuring situation awareness with physiological measures is more challenging. Furthermore, the results highlight that certain HPE factors combined degrade performance significantly more than a single HPE factor. This is shown by the performance measures, subjective data and behavioural measures. Overall, the simulator experiments pushed the pilots to their limits and the boundary of the HPE. Thus, safe performance was sometimes touched or even exceeded providing valuable insights for necessary HMI design solutions.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Applicability

This document illustrates the results of the P6 real-time simulation and presents the most promising measures that can be used to monitor the status of the HPE and to support the performance recovery. Also, the document serves as a basis for the final stage of P6, where the intention is to run a further simulation in 2017, with a strong focus on external validity and HMI design solutions for difficult HPE configurations. This current deliverable concludes on the best sensors and measures to support this final simulation. It thus applies to the entire P6 team, in particular to the partners involved in the final Work Package (WP6.4).

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

This page is intentionally left blank

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

TABLE OF CONTENTS

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

LIST OF FIGURES

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

LIST OF TABLES

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

1 INTRODUCTION

1.1. The Programme and the Project

The EC Flight Path 2050 vision aims to achieve the highest levels of safety to ensure that passengers and freight as well as the air transport system and its infrastructure are protected. Trends in safety performance over the last decade indicate that the ACARE Vision 2020 safety goal of an 80% reduction of the accident rate is not being achieved. A stronger focus on safety is required.

Future Sky Safety, established under coordination of EREA, is a Transport Research Programme built on European safety priorities that brings together 33 European partners to develop new tools and new approaches to aeronautics safety. The Programme links the EASp (European Aviation Safety plan) main pillars (operational issues, systemic issues, human performance and emerging issues) to the Flight Path 2050 safety challenges through four Themes:

- **Theme 1** (new solutions for today's accidents) aims for breakthrough research to address the current main accident categories in commercial air transport with the purpose of enabling a direct, specific, significant risk reduction in the medium term.
- **Theme 2** (strengthening the capability to manage risk) conducts research on processes and technologies to enable the aviation system actors to achieve near-total control over the safety risk in the air transport system.
- **Theme 3** (building ultra-resilient systems, organizations and operators) conducts research on the improvement of Systems, Organisations and the Human Operator with the specific aim to improve safety performance under unanticipated circumstances.
- **Theme 4** (building ultra-resilient vehicles) aims at reducing the effect of external hazards on vehicle integrity as well as reducing the number of fatalities in case of accidents.

Together, these Themes and the institutionally funded safety research intend to cover the safety priorities of Flight Path 2050 as well as the ACARE Strategic Research and Innovation Agenda (SRIA) (in particular the Challenges brought forward by ACARE Working Group 4 "Safety and Security").

The Programme will also help coordinate the research and innovation agendas of several countries and institutions, as well as create synergies with other EU initiatives in the field (e.g. SESAR, Clean Sky 2). Future Sky Safety is set up with expected seven years duration, divided into two phases of which the first one of 4 years has been formally approved.

Future Sky Safety P6 addresses Theme 3 (Building ultra-resilient systems and operators) focussed on strengthening the resilience to deal with current and new risks of the humans and the organizations operating the air transport system.

P6 builds on a concept previously proposed in the Air Traffic Management (ATM) domain, extending it to the Human Operators in the cockpit. The aim of the project is to define and apply the Human Performance

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Envelope for cockpit operations and design, and determining methods to recover crew's performance to the centre of the envelope, and consequently to augment this envelope.

The Human Performance Envelope is to some extent a new paradigm in Human Factors. Rather than focusing on one or two individual factors (e.g. fatigue, situation awareness, etc.), it considers a range of common factors in accidents and maps how they work alone or in combination to lead to a performance decrement that could affect safety. The safe region on the envelope is bordered by markers, which can be measured and signalled, allowing the pilots to detect and recover, or enabling external agencies to prompt recovery, or allowing automation to kick in and take over. The Human Performance Envelope will deal with the most crucial people in the accident chain, giving them back-up when they most need it, assuring performance when things get difficult. It will increase safety by focusing on the sharp end of accidents, and consign the term 'Pilot error' to the waste paper bin. The impact will primarily be through improved design and operational practices and is thus expected in the short to medium term.

1.2. Research objectives

FSS Project P6's main goal is to define and apply the concept of the Human Performance Envelope in the terms of cockpit operations and design. Based on the current knowledge about cognitive demands in the cockpit, the project will determine methods to restore the crew's performance to the centre of the envelope, and consequently to augment this envelope, through innovative HMI design, new automation concepts and new flight crew monitoring solutions (with impact on procedures or training).

In particular, by the end of the Project P6 the following results are expected:

- New Guidelines for HMI development, taking into account one dedicated concept of automation.
- General Guidelines for Augmenting the Envelope.
- Demonstrator (i.e. prototype with limited functionalities in an example scenario) of HPE monitoring and regulation solutions implemented in full mission simulators.

This study aims to describe potential interactions between the HPE components through a real-time simulation where we collect behavioural, psycho-physiological, performance-based and subjective data. The triangulation of the measurements allows to determine:

- Points where human performance deteriorates;
- Behavioural and/or physiological markers, which are critical in signalling performance degradation;
- How to increase the envelope improving performance and safety;
- How to develop effective recovery measures through innovative HPE based solutions.

This aim is to present how the HPE can be measured in pilot-in-the-loop simulations in high fidelity cockpit simulators with Air Traffic Control support. The most promising psycho-physiological, performance-based, and subjective measures that can be triangulated to characterize the HPE and the associated degradation and recovery points are outlined, as well as an experimental paradigm for HPE measurement.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

1.3. Approach

The HPE is a new paradigm in Human Performance and Human Factors. It suggests that we are good at designing systems – including pilot-cockpit systems – that protect against single factor problems such as stress, workload and situation awareness, but maybe not against combinations of those factors. The HPE concept also suggests that to protect against such combinations of factors requires a new method of analysis, which could lead to insights to deliver new training, procedural or HMI improvements, and a safer and more resilient system. P6 is in particular focused on potential new HMI solutions to protect against complex multi-factor situations, and this deliverable attempts to answer a single question: *Which measures are most useful in helping to identify HMI improvements to safeguard against HPE degradation?*

1.4. Structure of the document

This document is divided in three main sections:

- First, we present the results of the **pre-tests ran by ONERA and Cranfield** to test the sensors (Sections 2 and 3). The main objectives of these two experiments were to test the reliability of the sensors proposed for the real time simulations and to progress on the study of the respective influence of workload and stress on performances and their potential interaction.
- This is followed by the **results of the real time simulations** (limited to Scenario 1) performed at flight simulator AVES at DLR, presented in Section 4. In particular, the results presented in the document include:
    - The experiment design, to provide a recap of runs performance / subject (Section 4.1);
    - What subjective measure were tried and which one works (Section 4.2);
    - HPE scaling concept based on HPE curve, a tool tested in the simulations to detect performance degradation and pilot awareness and to find points at which performance decrement occurs (Section 4.4);
    - Physiological Parameters Analysis (Section 4.5);
    - Eye tracking data (Section 4.6);
    - Behavioural analysis (Section 4.7).
- Finally, the final section is dedicated to **future analysis** to be performed on Scenario 2. The methodologies to be applied to this analysis are presented in this document. In particular, the focus is on:
    - Competence analysis methodology, to assess the performance of the Pilot Monitoring (Section 5.1);
    - Cognitive Task Analysis, to investigate pilots' behaviour, communication, application of procedures and decision-making (Section 5.2).

The validation plan, including a description of the scenarios as well as the sensors, metrics and tools used have already been described in detail in D6.2 "Test plan for preliminary systems/pilots cognitive task analysis". Therefore these will only be briefly described here.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

2. ONERA PRE-TESTS

In D6.1, we performed a literature review to identify potential sensors for the collection of the physiological data for each HPE factor. These were then tested by ONERA and Cranfield. In common both experiments used the Multi-Attribute Task Battery (MATB II) and the CSEM Smart Harness (see section 6.1.1 of D6.2 for a more detailed description of the harness) to collect physiological test data. In addition, both experiments used the NASA-TLX to assess workload levels (see section 6.2.1.2 of D6.2 for more information about the tool).

2.1. Objectives

The main objectives of this experiment were to test the reliability of the measures/tools proposed for the real time simulations and to progress on the study of the respective influence of workload and stress on performances and their potential interaction. The choice was made to use the Multi-Attribute Task Battery (MATB II), as an example of a simplified environment with tasks similar to those usually performed by pilots. MATB II is described in detail in section 5.1 of D6.2. See also D6.2, in particular sections 5.2.2, for a description of the sensors selected.

2.2. Summary of the Experiment

Performance, Physiological and Subjective measures (Workload levels through NASA-TLX) were collected. Four types of physiological measures were obtained: cardiovascular (ECG), electrodermal activity (galvanic skin response or GSR), Oculometric measures (pupil diameter and number of blinks) and EEG measures. Subjects were 12 cadets of the French Air Force Academy.

Two experimental factors were controlled: cognitive workload and stress level. The former was manipulated as seen in Table 1.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Table 1: Settings of the MATB-II for the two levels of task load

	Task Load	Level 1 (TL1)	Level 2 (TL2)
Tracking	Target Movement	Medium	High
	Joystick Response	Medium	High
System Monitoring	# Green Light Events	2	4
	# Red Light Events	2	4
	# Scale Events	4	8
Communication Events	Other	1	1
	Own	3	3
Resource Management	# failures	4	8
	# max concomitant failures	2	4
	Mean duration of failures	30s	95.6s

Stress was manipulated as follows:

- NS: The subject was alone in the room; no video recording, and no other sounds other than the communication messages.
- S: presence of other people in the room; video recorder was used; cockpit noise was played; subjects were told their performance would be compared with those of the other subjects and sent to the instructors of the French Air Force Academy.

The task load (TL1 vs TL2) condition was counter-balanced as described in the following Table 2.

Table 2: Distribution of task loads per subject

Subject	No Stress		Stress	
S1	TL1	TL2	TL1	TL
S2	TL1	TL2	TL2	TL
S3	TL2	TL1	TL1	TL
S4	TL2	TL1	TL2	TL
S5	TL1	TL2	TL1	TL
S6	TL1	TL2	TL2	TL
S7	TL2	TL1	TL1	TL

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

S8	TL2	TL1	TL2	TL
S9	TL1	TL2	TL1	TL
S10	TL1	TL2	TL2	TL
S11	TL2	TL1	TL1	TL
S12	TL2	TL1	TL2	TL

2.3. Hypotheses

These main hypotheses are summarized in Table 3.

Table 3: Hypotheses about the evolution of the dependent measures

	Task Load increase	Stress increase
1. Performance Measures		
a. RMSD-C	↗	↘ then ↗
b. SYSMON RT	↗	↘ then ↗
c. SYSMON Missed	↗	↘ then ↗
d. Tank out of range	↗	↘ then ↗
e. COMM: false	↗ or →	↘ then ↗
2. ECG		
a. Rythm	↗	↗
b. RMSSD	↘	
3. EDA		
a. Tonic (SCL)	↗	↗
b. Phasic (SCRs)	↗	
4. EEG		
a. PSD Theta	↗	↗ then ↘ for frontal electrodes
b. PSD Alpha	↘	
c. PSD Beta	↗	
d. PSD Delta		
5. Eye Tracking		
a. Mean pupil diameter	↗	↗
b. Blink rate	↘	↗
6. Subjective Workload		
a. NASA TLX (Mean)	↗	

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

2.4. Comparison between the CSEM ECG measures and the biopac ECG measures

The CSEM vest was available only for the three last subjects and was used in addition to the ECG biopac measure. A comparison of the RR intervals calculated by the CSEM algorithm on the data recorded with the CSEM smart vest and the one calculated with the biopac algorithm on the data recorded with the biopac system appeared to be highly similar, as shown in Figure 1.

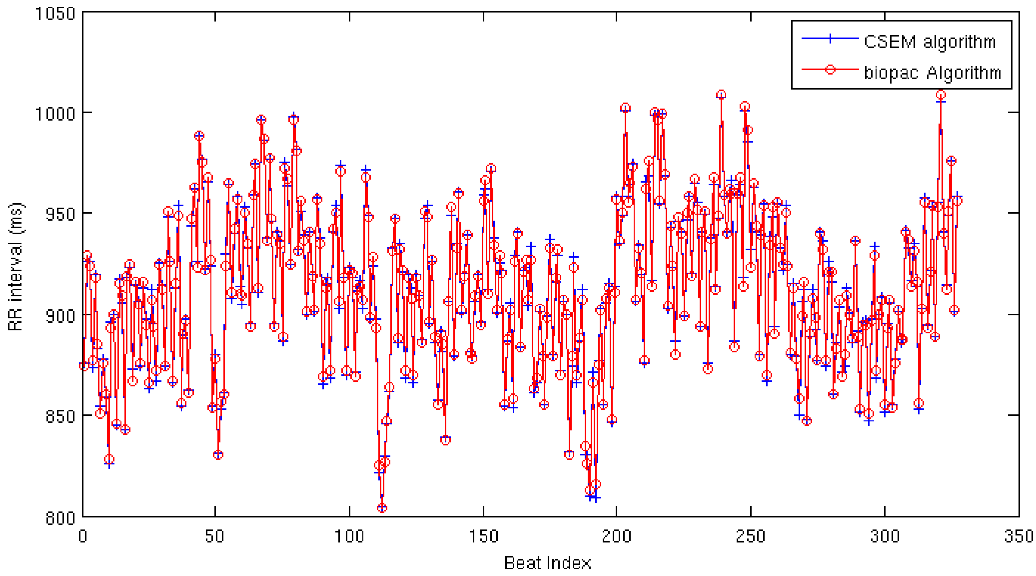


Figure 1: Comparison between the RR intervals calculated by the CSEM smart vest and the biopac system

2.5. Results

Not surprisingly, the estimated cognitive load increased with the level of difficulty (*F(1,8) = 38.567, p<.05)* (see Figure 2). More precisely, the modulation of the difficulty impacted the dimension of the cognitive load (see Figure 3). Mental (*F(1,8) = 12.890, p<.05*), physical (*F(1,8) = 7.426, p<.05*) and temporal (*F(1,13) = 28.148, p<.05*) demands were perceived larger when difficulty increased. At the same time, the frustration induced by the task increased with difficulty (*F(1,8) = 19.629, p<.05*). Finally, perceived effort increased with difficulty (*F(1,8) = 6.347, p<.05*) whereas performance was perceived as significantly decreasing by difficulty (*F(1,8) = 27.422, p<.05*).

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

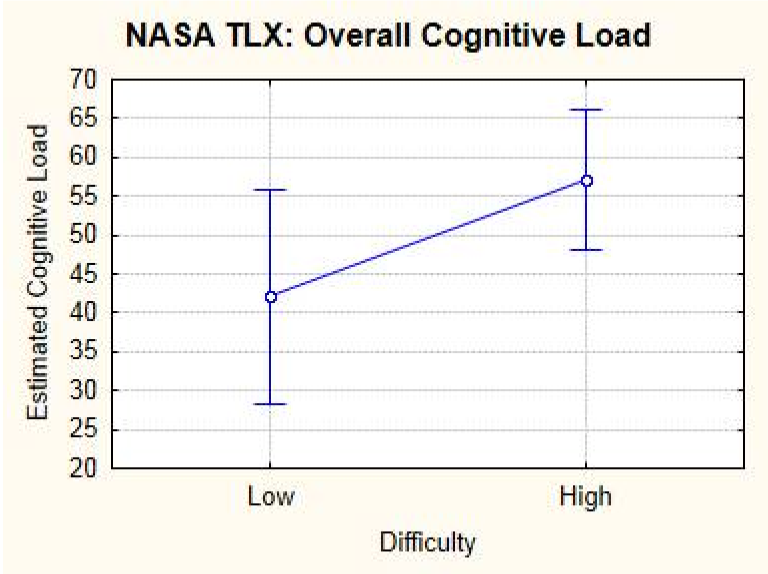
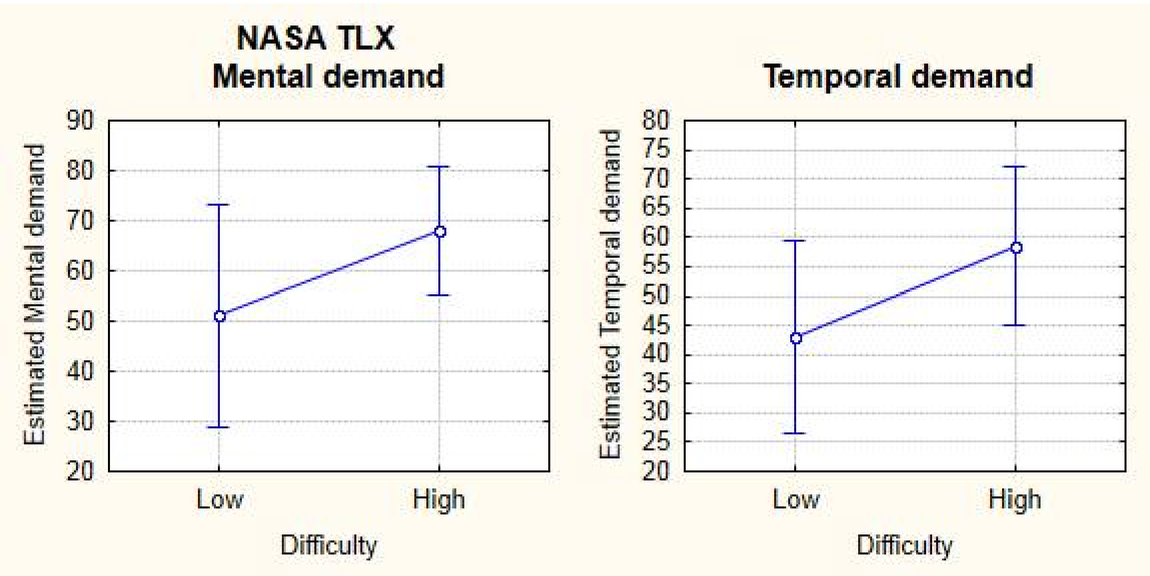


Figure 2: Effect of difficulty on perceived workload



Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

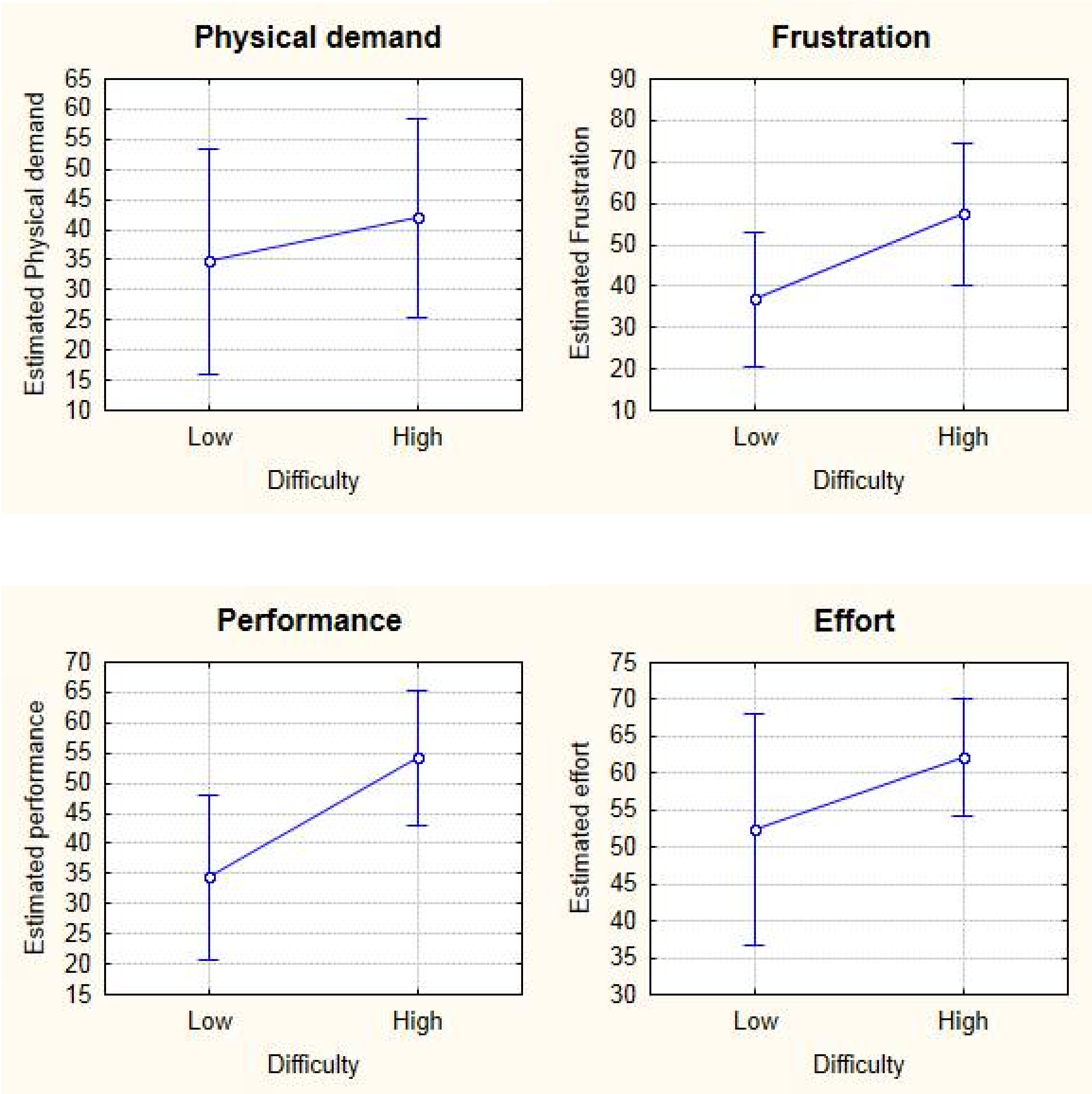


Figure 3: Effect of difficulty regarding the different dimension of the perceived workload

Difficulty also dramatically impacted the level of performance in the MATB subtasks (see Figure 4). Particularly, the increase in difficulty lead to an increase in error regarding fuel management task ($F(1,8) = 50.493$, $p<.05$), an increase in distance to the target in the tracking task ($F(1,8) = 58.566$, $p<.05$) and an increase in misses in the alarm detection task ($F(1,8) = 28$, $p<.05$). In contrast, the level of difficulty has no impact on the reaction time regarding the alarm detection task ($F(1,8) = 2.246$, *n.s*), nor on the mismanagement of communications (no errors recorded in both conditions).

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

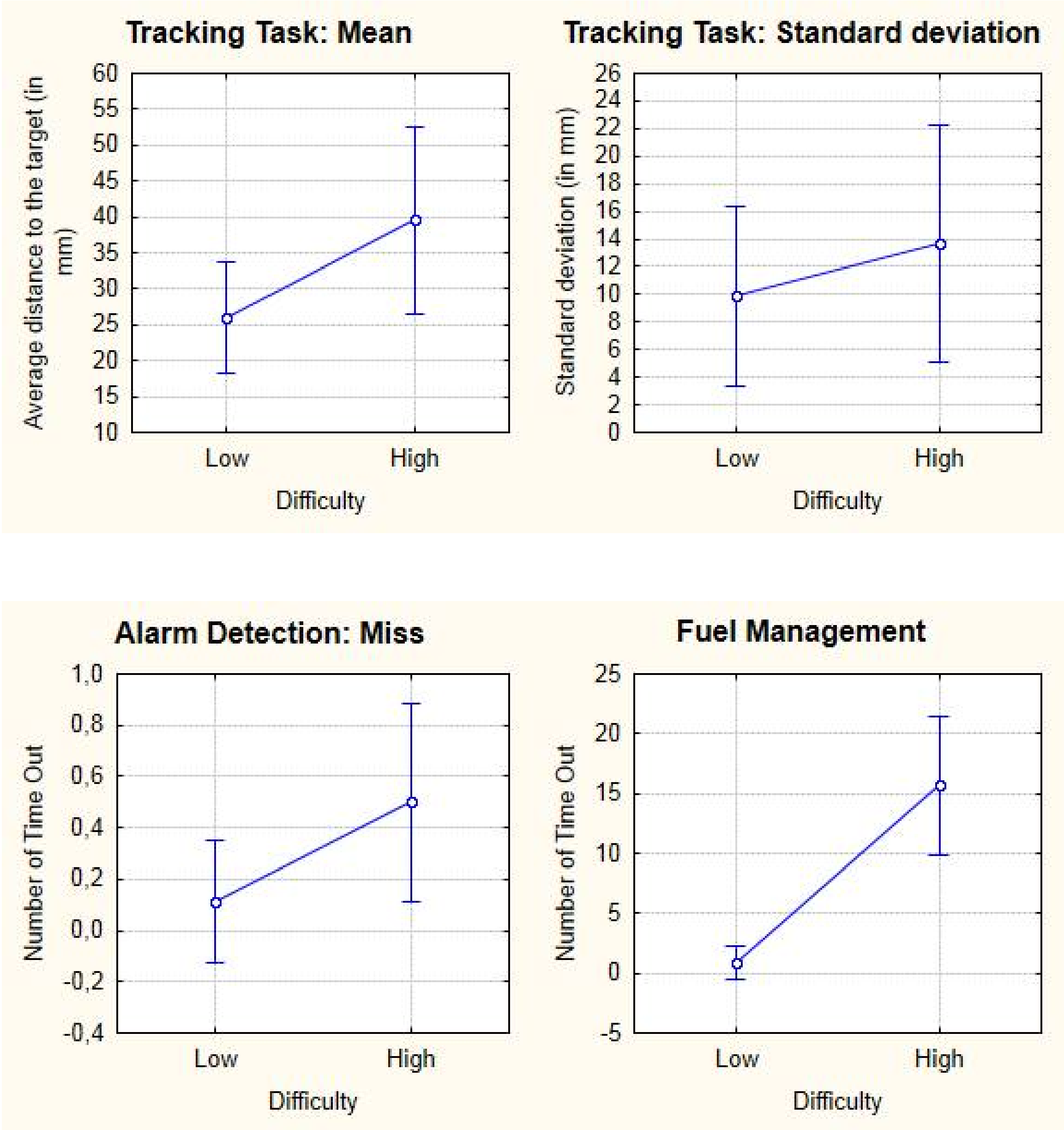


Figure 4: Effect of difficulty regarding task performance

Regarding physiological measures, the level of difficulty impacted pupil dilatation (see Figure 5) with an increase in pupil diameter with the increase in difficulty ($F(1,8) = 14.983$, $p<.05$). The other physiological measures seemed not to have been modulated by the level of difficulty: heart rate ($F(1,8) = 1.184$, *n.s*), galvanic skin response ($F(1,13) = 2.105$, *n.s*) or number of blinks ($F(1,8) = 1.184$, *n.s*).

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

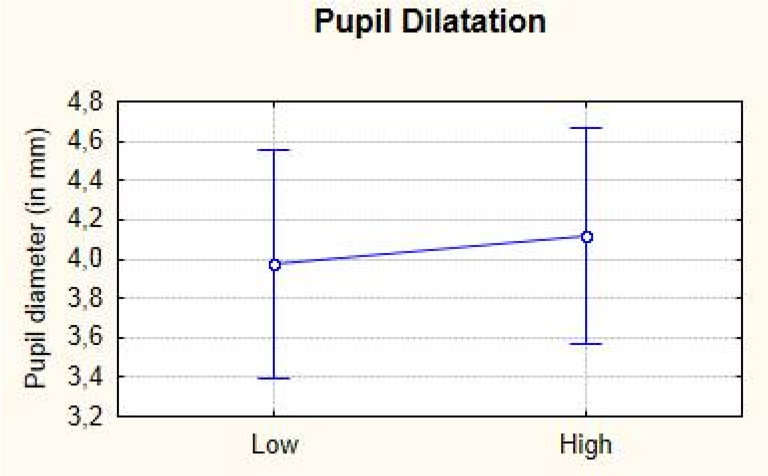
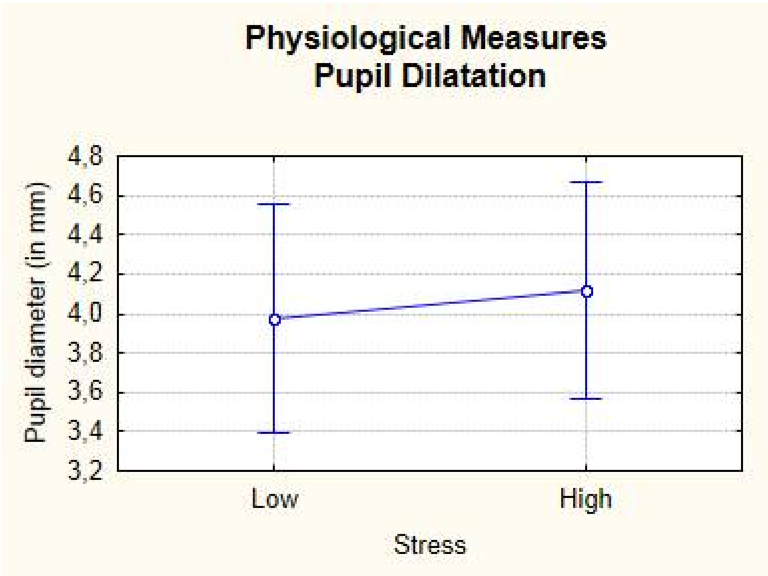


Figure 5: Effect of difficulty regarding pupil dilatation

Physiological measures confirmed the impact of the experimental manipulation on the level of stress (see Figure 6). Indeed, we observed an increase in the ECG frequency ($F(1,8) = 18.466, p<.05$), in the galvanic skin conductance ($F(1,8) = 10.514, p<.05$) and in pupil dilatation ($F(1,8) = 467.001, p<.05$).



Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

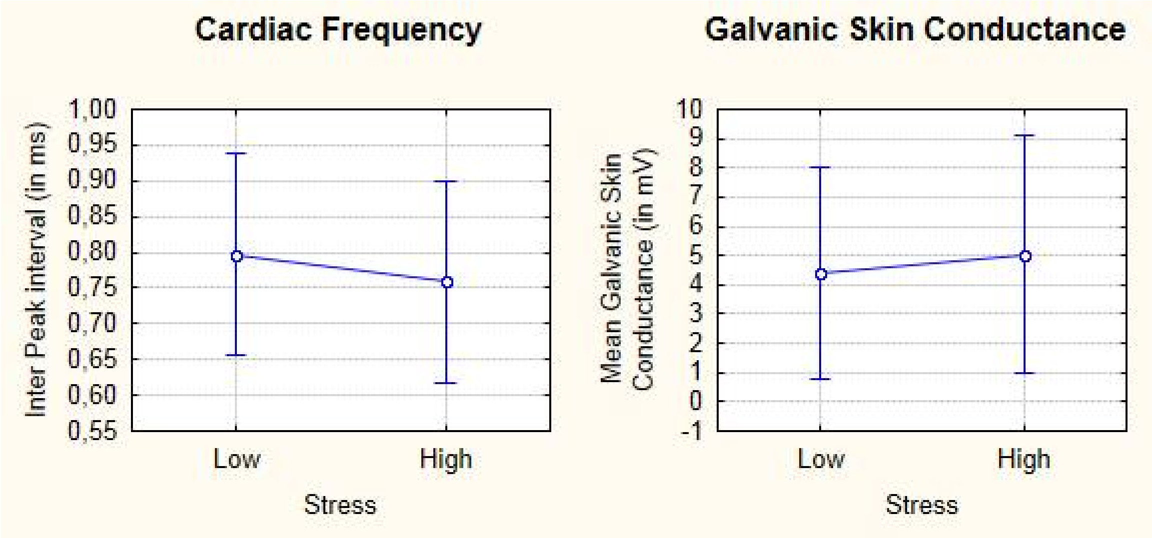


Figure 6: Effect of Stress in pupil dilatation, heart rate and electro-dermal activity

Interestingly, we also observe an increase in performance with the increase in stress. Particularly, stress significantly reduces the reaction time of the participant (see Figure 7). At the same time, participants acknowledged more effort in the stress condition (*F(1,8) = 5.802, p<.05)*. Finally, regarding EEG signal, whereas the effect of difficulty is more diffused, we observed a clear increase in frontal activity (F1, F2 & Fz) whatever the oscillation band considered (see Table 4). In Table 4 the + means a significant increase in electrical activity and – a significant decrease. An empty cell means no significant effect of factor on this oscillation band for this channel. The data suggested that stress seemed to stimulate frontal activity and enhance both engagement in the task and performance.

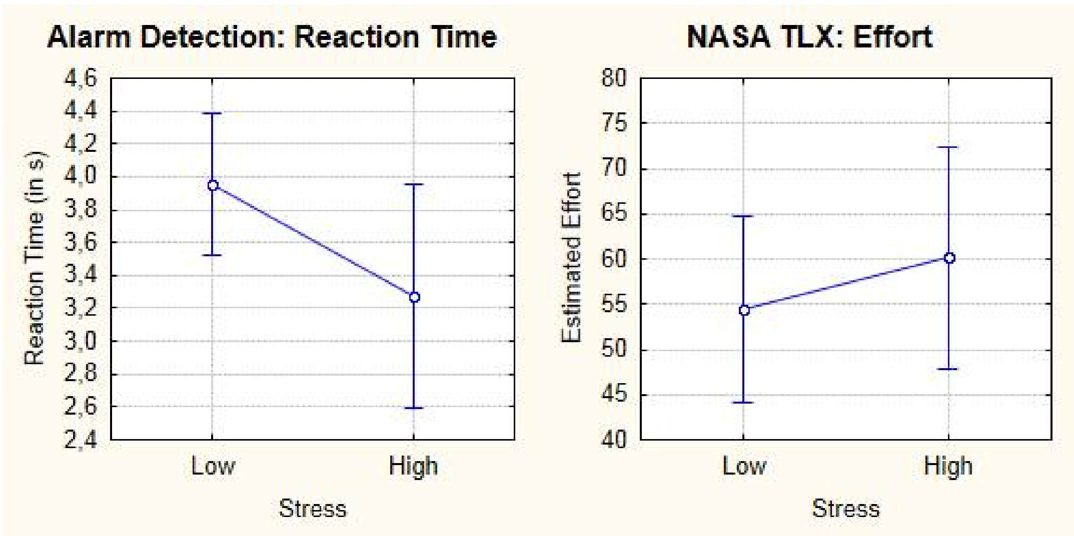


Figure 7: Effect of Stress regarding both task performance (reaction time in alarm detection) and perceived effort.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Table 4: Statistical analysis of EEG signal by oscillation band

		Stress (S)	Difficulty (D)
FP1	Delta	+	+
	Theta	+	
	Alpha	+	
	Beta	+	+
FP2	Delta	+	
	Theta	+	+
	Alpha	+	
	Beta	+	
Fz	Delta	+	
	Theta		
	Alpha	+	+
	Beta	+	+
F3	Delta	+	-
	Theta		
	Alpha		
	Beta	+	-
Cz	Delta	+	(+)
	Theta	+	
	Alpha		
	Beta	-	-
Pz	Delta		
	Theta		+
	Alpha	(+)	
	Beta		
P4	Delta		
	Theta	+	
	Alpha	+	+
	Beta		

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

2.6. Conclusion

In this study, we showed that both the manipulation of stress and difficulty were relevant. The manipulation of the task load through MATB was proved to impact both performance measure and perceived workload whereas the manipulation of task context (social pressure, noisy environment) impacted stress, particularly regarding physiological measures like ECG, GSR and oculometric measure. Interestingly, our results also confirm the relevance of both the measure proposed and the captors used to grab the modification induced by these two factors. Particularly, we noted an increase in electrodermal activity, in pupil dilatation and in heart rate frequency with increase in stress level whereas pupil dilation also increases in case of increase in difficulty.

Interestingly, we also observed a positive effect of stress regarding performance measure since the participant performs the task faster with stress. This increase in performance comes with an increase in participant engagement in the task. Indeed, participants reported more effort in the stress condition. Finally, such increase in engagement was confirmed by an increase in electrical activity in the frontal area. In other words, under social pressure, the participants seemed to engage more resources to improve their performance. This result confirms the bidirectional relation between stress and task performance. If stress could have an inhibitory effect in extreme condition, the stress induced in our simulated condition did not reach such level. In the simulated level of stress, the stress appears as beneficial for performance.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

3. CRANFIELD PRE-TESTS

3.1. Overview

The purpose of this experiment is to develop a systematic understanding of how elements of human performance in aviation can be inferred from physiological data. Specifically, we precisely define a number of tasks which relate to flying tasks performed by an airline pilot. The outputs of this study can inform the analysis of the data from the simulation activity at DLR and further our understanding of relationships between physiological parameters and performance of pilots when engaged in different tasks.

The study is framed by the idea of a performance envelope. A pilot has a region where task performance is acceptable and a region where task performance is not acceptable. This unacceptable performance may be due to a number of factors including task-load, mental workload, or divided attention.

The purpose of the study is to determine patterns of physiological response that can indicate a likely trend towards poorer or better performance depending on the task completed by the participant.

In this study participants were required to perform a number of tasks demanding different elements of cognition while aspects of their physiology were being measured. Physiological markers were recorded using a smart-harness. The smart harness contains three sensors which are worn next to the participant's skin to record heart rate variability, respiration rate and oxygen saturation.

The study has three overall hypotheses:

1. There will be association between the physiological data and task-load.
2. There will be differences in patterns of physiological data depending on task type.
3. There will be differences in patterns of physiological response data depending on the task-load gradient.

3.2. Method

3.2.1. Participants

Forty four male participants took part in this study. Due to missing or incomplete data, five participants were excluded from the final analysis, resulting in 39 participants with a mean age of 34.1 years (SD 10.94). This number was not informed by a power calculation since no readily available data is available to reliably compute effect sizes. A sample size of forty participants was deemed to give the best balance between the duration and complexity of the experimental task and the required number of data points to use parametric statistics subject to the appropriate assumptions being met.

All participants had normal or corrected to normal vision and none reported consuming alcohol since waking prior to taking part in the experiment. No participants were excluded for health reasons. Four participants also stated that they had some flying experience. However these participants were not professional pilots.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

3.2.2. Tasks

The Multi Attribute Task Battery 2 (MATB II) was used to deliver tasks to participants (Figure 8). The MATB II can model a variety of tasks which are central to aviation tasks but do not require a qualified pilot on which to assess performance. MATB II use in experimental studies is well documented in the literature and the MATB II has demonstrated content validity, construct validity and face validity.

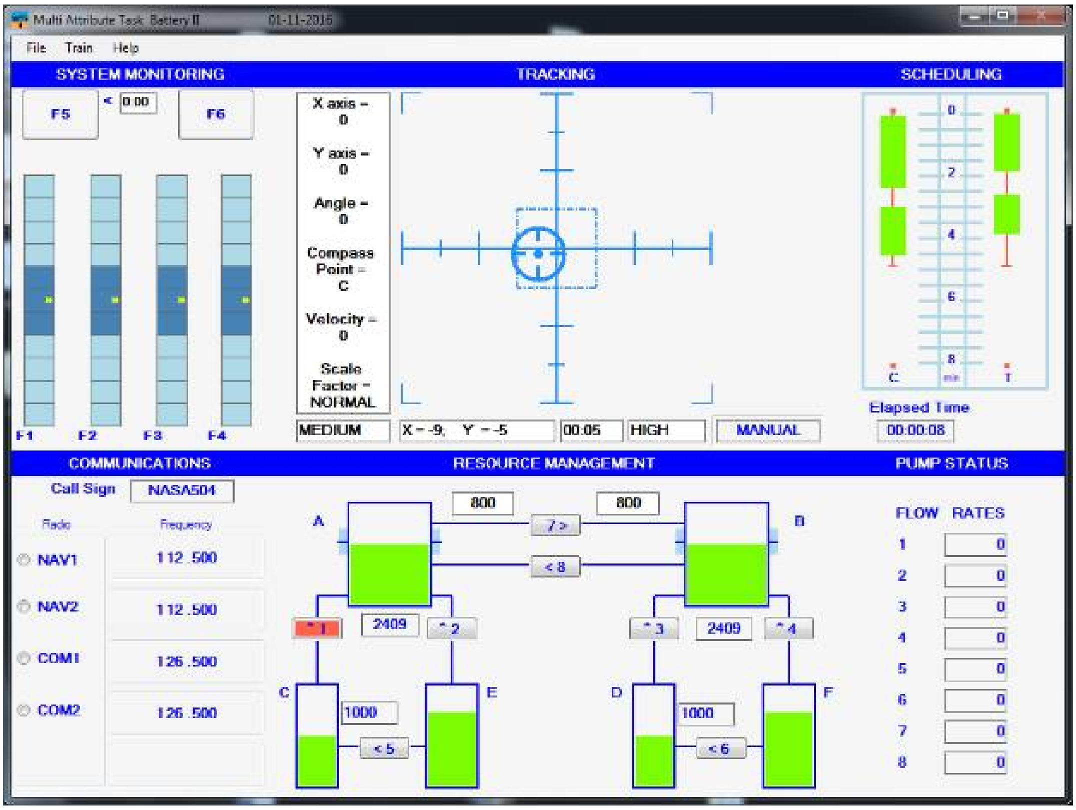
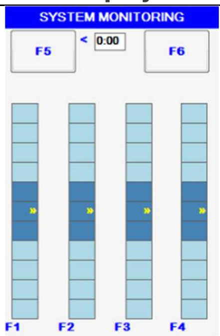
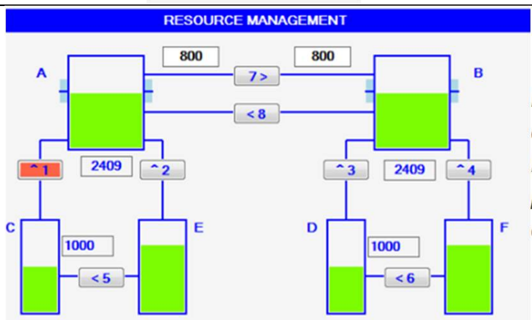
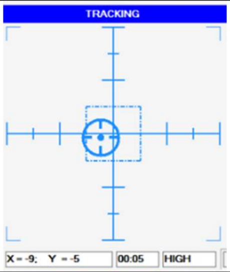


Figure 8: MATB II Interface.

Four tasks from the MATB II were used in this study: the resource management task, the tracking task, the system monitoring task, and the communication task (Table 5). The system monitoring and resource management tasks were combined creating three experimental tasks. These tasks represent key elements of cognition required in the aircraft cockpit.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Table 5: MATB II interface elements and task descriptions

Task name	Display	Task
System monitoring		*Participants detect changes to the colour of the buttons (F5 and F6) and out of range movement of the scales (F1-F4)*
Resource management		*Participants switch pumps on and off to maintain flow. Participants detect failed pumps (red pump has failed) and adjust their plan.*
Tracking		*Participants maintain moving target (circle) within inner square.*
Communications		*Participants listen to auditory messages and dial in correct frequency when an ownship (NASA504 in this example) announcement is made.*

Two task-load levels were used: low and high. The task-load was manipulated by adjusting the number of events in each task. The event frequencies are shown in Table 6.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Table 6: MATB II task-load stimulus

Task	Stimulus	Number of Events in 5 minutes	
		Low Task-load	High Task-load
Systems monitoring and resource management tasks (resmon)	Pump Failure	6 x 5 seconds	10 x 15 seconds
	Gauge Alert	6	30
	Green Light Off	4	10
	Red Light On	4	10
Communications task (comms)	Communications	8	20
Tracking task (tracking)	Tracking	Default Low	Default High

3.3. Design

A mixed design was used. The within-subjects factors are task type (tracking, comms and resmon tasks) and task-load (high or low). The between-subjects factor is gradient (low to high, or high to low task-load order presentation) which was randomised. Task order was also randomised. Each participant completed two five minute sessions (one high, one low task-load) of each of the three tasks, and two five minute baseline measurements pre and post task. Performance measures were captured for each task.

During each block, physiological data from the participants was acquired by the smart-harness. Each five-minute block was followed by a two-minute period during which subjective workload measurements were taken using the Bedford Workload scale and the NASA-TLX.

3.4. Dependent Variables

3.4.1. Biographical data recorded from participants

Sample data was recorded from participants to include age, gender, and flying experience. Participants who reported taking regularly prescribed medicine which affected heart rhythm or who wear implantable cardiac devices were screened out of the study.

3.4.2. Measurement of physiological parameters

Measurement of the physiological data was made using a proprietary smart harness designed and manufactured by CSEM SA. Data representing heart rate variability and respiration rate was collected electro-physiologically. Data representing oxygen saturation was collected optically and GSR was recorded

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

using finger plesmography. In addition, eye blinks were measured post-trail using video footage from a webcam recording at a resolution of 480 x 600 pixels.

The vest and LTMS-S system has been developed by CSEM SA. It consists of three sensors that measure heart rate (using a two lead ECG), oxygen saturation, core body temperature, breathing rate and physical activity. The system has been clinically validated using an experimental protocol including rest and movement (walking / running) as well as two, 24-hour recordings for circadian measurement, comparing the sensors to reference measurements (Chetelat et al., 2015). Each trial consisted of 10 and six healthy males. This validation exercise concluded that the quality of the measurement is comparable to clinical medical devices.

3.4.3. Measurement of subjective workload

Following each experimental trial, subjective workload was measured using the NASA TLX. Spare capacity was measured using the Bedford Workload scale. Both of these scales are well represented in the human factors literature and have been shown as sensitive to changes in task demand.

3.5. Procedure

An example timeline from a morning experimental session is shown in

Figure 9. Participants were firstly given a voucher for participation in the study. Participants were the briefed and asked for informed consent. When informed consent was given participants were asked to randomly select their participant number, and select a piece of paper indicating the task-load condition they would be completing (low to high or high to low). Each piece of paper was discarded after it was selected. Three cards were also presented to the participants stating the tasks on the back. They were asked to select the cards one at a time which determined the task order randomly.

Participants then provided brief biographical details and a stress and arousal checklist on a PC Participants were then asked to change into the smart harness and trained to stable performance on the MATB II for twenty five minutes. After this participants completed a five minute baseline measurement sitting quietly looking at the MATB II screen. Participants then completed a total of six, five-minute blocks of activity on the MATB II. During each block, physiological data from the participants was acquired by the smart-harness. Each five-minute block was followed by a two-minute period during which subjective workload measurements were taken using the Bedford workload scale and the NASA TLX. Participants were then instructed to change and given a full debrief.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

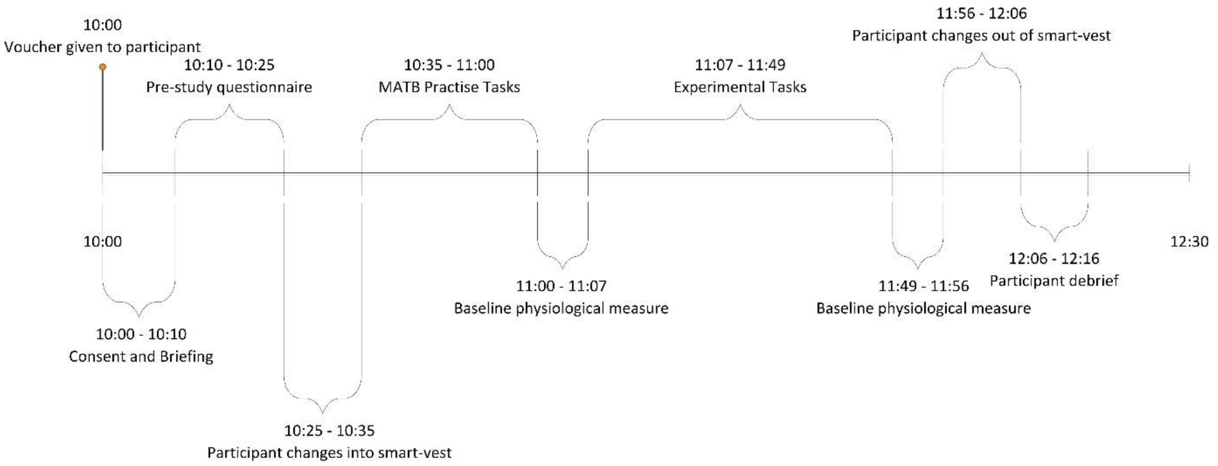


Figure 9: Example of am session timeline for experiment

3.6. Data Analysis

The physiological data were recorded in real time. The harness was worn from before the MATB II training to the end of the experimental procedure as one continuous recording. The required recordings were eight five minute segments for each person. These were distinguished by markers throughout the recording made by the participant by ‘tapping’ one of the sensors. These markers were cross checked with the time recorded for the start of each activity by the researcher. Each recording was then split into eight individual recordings using bespoke software provided by CSEM SA. These recordings were of five minutes duration. The quality of each of these recordings was checked by CSEM SA to ensure that the sensors were collecting accurate information. At this stage six participants were excluded from the analysis due to poor quality recordings. The first 50 seconds and last ten seconds of each recording were then removed resulting in eight four minute segments per participant. This was in accordance with guidance provided by the European Heart Journal (Force, 1996). Mean heart-rate (HR), breathing rate (BR), blood oxygen saturation (SpO_2) was then calculated along with the frequency measures of Standard deviation of the N-N interval (SDNN), Very low, low and high frequency spectral density (VLF, LF, HF) variables for each four minute segment for each participant ensuring all zero measures were removed. The data cleansing process can be seen in Figure 10.

Future Sky Safety has received funding from the EU’s Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

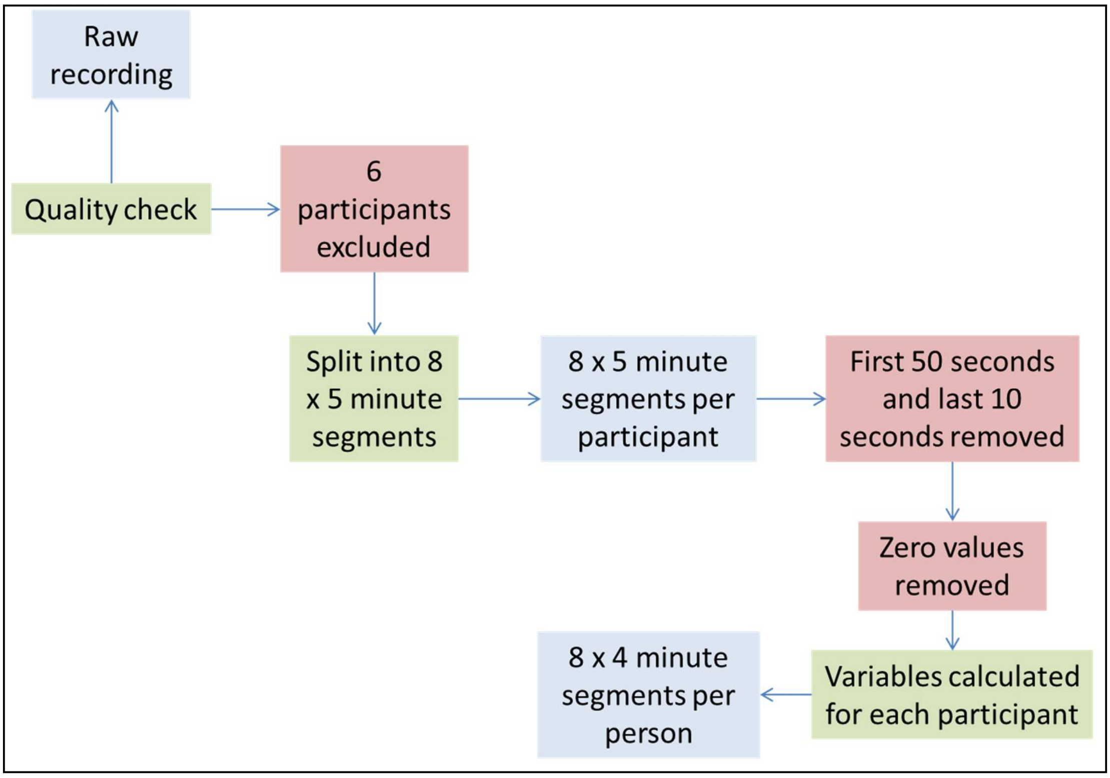


Figure 10: Data cleansing process

3.7. Results

3.7.1. Workload Analysis

3.7.1.1. NASA-TLX

The NASA-TLX is analysed using 2 × 3 rANOVA (repeated measures analysis of variance). Task-load (low vs. high) and task (communications vs. tracking vs. monitoring) comprise the fixed factors. For both analysis N = 39. No data was missing or excluded from the analysis. Exploration of the data indicated significant departure from normality across the NASA-TLX subscales. A natural logarithmic transformation plus a linear constant to avoid zeros ($x' = \ln x + k$) stabilised variance and improved the departures from normality. This transformation was selected empirically by computing $\sqrt{x + k}$, $\log_{10} x + k$, $\frac{1}{x}$, and $\ln x + k$. The natural logarithmic transformation had the greatest effect on reducing significant departures from normality as assessed by the Kolmongorov-Smirnoff test using the standard normal distribution as the reference distribution. Marginal means are transformed back ($x = e^{x'} - k$) to retain the meaning in the scales. All significant interactions are ordinal. As such, pairwise comparisons have been conducted since the ordinal interaction indicates that the effect is the same at each level of the other factor. Pairwise comparison significance values are Bonferroni adjusted to control the familywise error rate. The new value for α is 0.0028 (0.05/18 tests).

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Factor plots for all scales are shown in Figure 11. Overall significant differences were found between low and high task-loads and between different task types. All scales showed significant differences between high and low task-loads. These differences are dependent on the scales under examination. It is illustrative to group the mental demand, effort and temporal demand scales together as the first group for consideration. The frustration, performance and physical demand scales comprise the second group, following a similar pattern of significance and magnitude. The mental demand, effort and temporal demand scales show increasing magnitude of mental workload across the communications, tracking and monitoring tasks. All differences in workload between tasks were significant for the mental demand and temporal demand scales. Differences between mean effort were significant only between the communications task and the tracking task, and the communications task and the monitoring task. For the frustration, performance and physical demand scales the patterns of means across tasks is similar. For each scale, the mean workload reported for the tracking task in the high condition is the highest reported by participants. Overall only the communications task is significantly different from the tracking and monitoring tasks across these three scales.

Mental demand

A significant effect of task load ($F_{1,38}$ =87.60, $p<0.01$, $\eta_p^2=0.70$) and task type ($F_{2,76}$ =48.01, $p<0.01$, $\eta_p^2=0.56$). The Greenhouse-Geisser correction is applied to the interaction term df since a significant departure from sphericity is present in the data ($W^2 = 0.55$, $p<0.01$).A significant, ordinal interaction between task-load and task-type was present ($F_{1.38,52.55}$ =5.12, $p<0.05$, $\eta_p^2=0.70$). Higher task-load led to significantly higher reported mental demand. Monitoring elicited the highest mental demand followed by tracking and finally the comms task. Pairwise comparisons indicate significant differences between all task types (communications vs. tracking, mean transformed difference = -0.44, $p<0.0028$; communications vs. monitoring, mean transformed difference = -0.71, $p<0.0028$; tracking vs. monitoring, mean transformed difference = -0.27, $p<0.0028$).

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

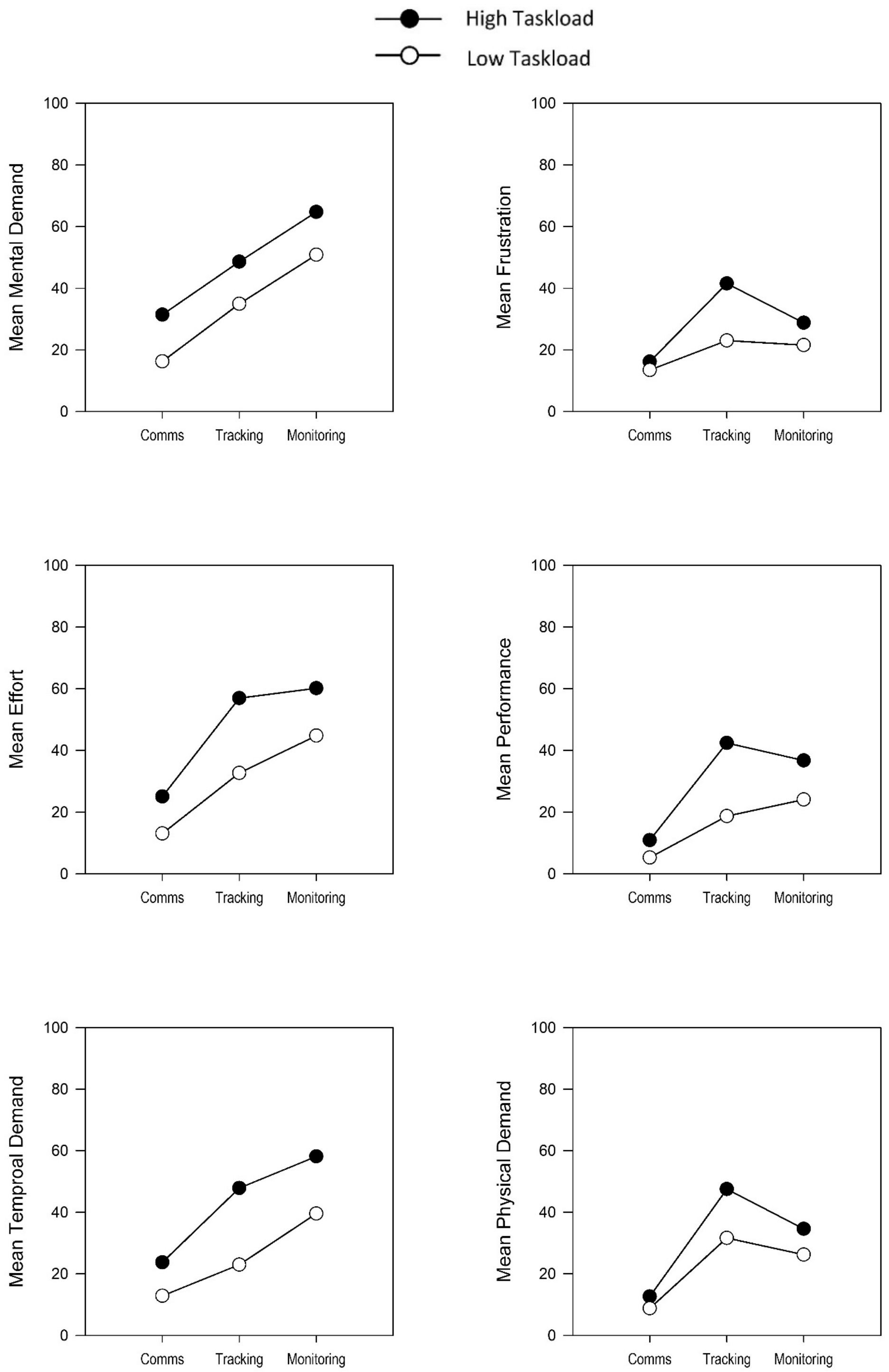


Figure 11: Factor plots for task-load × task-type for each subscale of the NASA TLX

Effort

No significant departures from sphericity were identified in the task type or interaction terms using Mauchly's test of sphericity. No corrections to df will be made. A significant effect of task load ($F_{1,38}$ =76.06, p<0.01, η_p^2=0.67) and task type ($F_{2,76}$ =71.32, p<0.01, η_p^2=0.65). A significant, ordinal interaction

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

between task-load and task-type was present ($F_{2,76}$ =5.12, $p<0.05$, $\eta_p^2=0.08$). Higher task-load led to significantly higher reported mental demand. Monitoring elicited the highest effort followed by tracking and finally the comms task. Pairwise comparisons indicate significant differences between the communications task, and the tracking and monitoring tasks (communications vs. tracking, mean transformed difference = -0.63, p<0.0028; communications vs. monitoring, mean transformed difference = -0.78, p<0.0028) No significant difference was found between the tracking vs. monitoring tasks (mean transformed difference = -0.15, p>0.0028).

Temporal Demand

No significant departures from sphericity were identified in the task type or interaction terms using Mauchly's test of sphericity. No corrections to df will be made. A significant effect of task load ($F_{1,38}$ =80.47, $p<0.01$, $\eta_p^2=0.68$) and task type ($F_{2,76}$ =40.76, $p<0.01$, $\eta_p^2=0.52$). A significant, ordinal interaction between task-load and task-type was present ($F_{2,76}$ =4.35, $p<0.05$, $\eta_p^2=0.10$). Higher task-load led to significantly higher reported mental demand. Monitoring elicited the highest temporal demand followed by tracking and finally the comms task. Pairwise comparisons indicate significant differences between all tasks (communications vs. tracking, mean transformed difference = -0.45, p<0.0028; communications vs. monitoring, mean transformed difference = -0.74, p<0.0028; tracking vs. monitoring, mean transformed difference = -0.29, p<0.0028).

Frustration

No significant departures from sphericity were identified in the task type or interaction terms using Mauchly's test of sphericity. No corrections to df will be made. A significant effect of task load ($F_{1,38}$ =42.27, $p<0.01$, $\eta_p^2=0.53$) and task type ($F_{2,76}$ =18.98, $p<0.01$, $\eta_p^2=0.33$). A significant, ordinal interaction between task-load and task-type was present ($F_{2,76}$ =7.37, $p<0.02$, $\eta_p^2=0.16$). Higher task-load led to significantly higher reported frustration. The tracking task elicited the highest frustration followed by monitoring and finally the comms task. Pairwise comparisons indicate significant differences between the communications task, and the tracking and monitoring tasks (communications vs. tracking, mean transformed difference = -0.51, p<0.0028; communications vs. monitoring, mean transformed difference = -0.35, p<0.0028). No significant difference was found between the tracking vs. monitoring tasks (mean transformed difference = -0.17, p>0.0028).

Performance

No significant departures from sphericity were identified in the task type or interaction terms using Mauchly's test of sphericity. No corrections to df will be made. A significant effect of task load ($F_{1,38}$ =124.44, $p<0.01$, $\eta_p^2=0.76$) and task type ($F_{2,76}$ =48.6, $p<0.01$, $\eta_p^2=0.56$). A significant, ordinal interaction

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

between task-load and task-type was present ($F_{2,76}$ =7.50, p<0.02, η_p^2=0.17). Higher task-load led to significantly higher reported frustration. The tracking task elicited the highest frustration in the high task-load condition. Pairwise comparisons indicate significant differences between the communications task, and the tracking and monitoring tasks (communications vs. tracking, mean transformed difference = -0.78, p<0.0028; communications vs. monitoring, mean transformed difference = -0.80, p<0.0028). No significant difference was found between the tracking vs. monitoring tasks (mean transformed difference = -0.03, p>0.0028).

Physical Demand

No significant departures from sphericity were identified in the task type or interaction terms using Mauchly's test of sphericity. No corrections to df will be made. A significant effect of task load ($F_{1,38}$ =59.68, p<0.01, η_p^2=0.61) and task type ($F_{2,76}$ =48.27, p<0.01, η_p^2=0.56). No significant interaction was found ($F_{2,76}$ =2.5, p>0.05). Higher task-load led to significantly higher reported physical demand. The tracking task elicited the highest physical demand in both the low and high task-load conditions followed by monitoring and finally the comms task. Pairwise comparisons indicate significant differences between the communications task, and the tracking and monitoring tasks (communications vs. tracking, mean transformed difference = -0.86, p<0.0028; communications vs. monitoring, mean transformed difference = -0.68, p<0.0028). No significant difference was found between the tracking vs. monitoring tasks (mean transformed difference = -0.20, p>0.0028).

3.7.1.2. Bedford workload scale

The modal Bedford rating for the low task condition is 1 on the 10-point scale corresponding to the "Was it a piece of cake" qualitative descriptor (see Table 7). The modal workload rating in the high task-load condition is 3 corresponding to the "There was enough time to easily attend to additional tasks". No significant contingency was found between the distribution of scores in the low and high task load conditions although the comparison approached significance (χ^2 (6) = 12.39, p=0.054). However, we must also point out that low expected frequencies are present and so the analysis remains inconclusive.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Table 7: Bedford workload scale

First level qualitative descriptor	Second level qualitative descriptor	Low Task-load	High Task-load
Workload satisfactory.	Was it a piece of cake?	26	15
	Was there more spare time than would ever be needed to attend to additional tasks?	3	5
	There was enough time to easily attend to additional tasks.	10	18
Workload satisfactory without reduction.	Was there ample time to attend to additional tasks?	0	1

3.7.2. Performance Analysis

Different measures which can be used to characterise performance associated with the MATB II tasks are available from the software. In some cases post-processing of the data is required to generate meaningful performance data.

3.7.2.1. Resource Monitoring Task

For this task, we have characterised success as minimal deviation from 2500 units of flow for tanks A and B. The MATB II software output is the tank volumes for Tank A and Tank B. Firstly, we computed the mean squared deviation from 2500 for each tank:

$$Tank\ Deviation_{A,B} = \frac{\sum(Tank_{A,B} - 2500)^2}{n}$$

The resulting values characterise an average, squared deviation from the optimum flow rate in each tank. Participants may adopt a strategy whereby they focus on one tank at the expense of the other. In order to mitigate this effect, a mean of the mean squared deviations from each tank were computed. The square root of this value then provides a single, positive number which represents the absolute average deviation from 2500 units of flow, characterising performance on this task:

$$Overall\ Deviation = \sqrt{\frac{Tank\ Deviation_A + Tank\ Deviation_B}{2}}$$

Exploratory analysis of the flow deviation data revealed significant positive skew and departure from normality. This can be common with this kind of data whereby a small number of participants are simply

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

worse at the task. The more extreme values in the tail cannot be classified as outliers. A log_{10} transformation was applied to minimise skew and stabilise variance. Graphs show untransformed data and standard errors have been computed using the untransformed data (Figure 12). A paired sample t-test indicates a significant difference between mean transformed pooled deviation from flow (t_{38}=3.7, $p<0.01$, η_p^2= 0.27); pooled deviation from flow was highest in the high task-load condition.

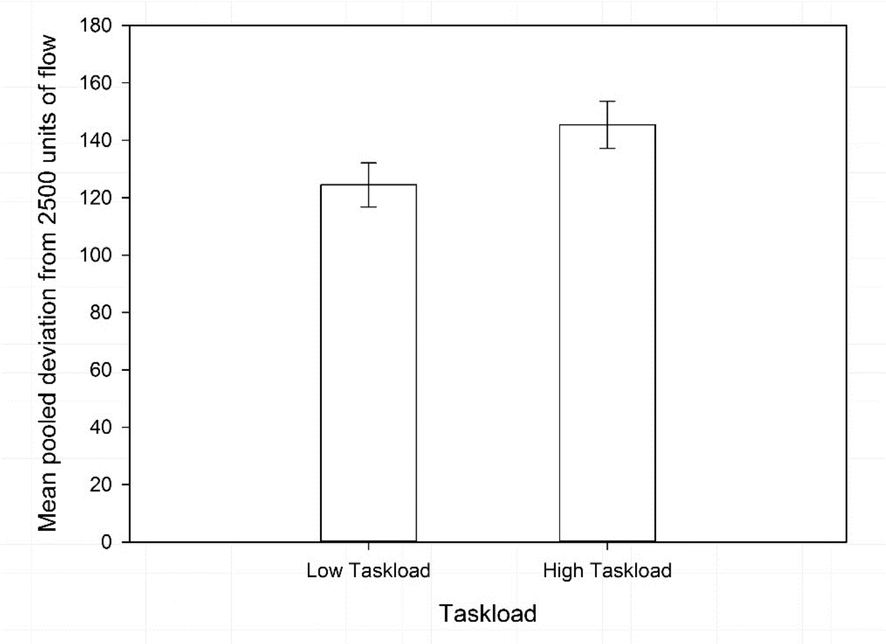


Figure 12: Mean pooled deviation from 2500 units of flow for the high and low task-load conditions

3.7.2.2. Tracking Task

The tracking task comprises the root-mean squared error (RMSE) characterising the amount of deviation in pixels from the target. For each participant, a mean of this RMSE value is taken across the trial to give a single, positive number which represents the average deviation from the target, characterising performance on this task. In addition to the mean RMSE, the standard deviation of the RMSE across the sample is included in the analysis. The mean RMSE gives an indication of the average deviance from the target. The standard deviation of the RMSE gives an indication of the variability of participant response in the tracking task. For example, a RMSE standard deviation of zero would indicate minimal variability around the mean RMSE for a participant.

Exploratory analysis of the tracking data revealed significant positive skew. This can be common with this kind of data whereby a small number of participants are simply worse at the task. The more extreme values in the tail cannot be classified as outliers. A log_{10} transformation was applied to minimise skew and stabilise variance. Graphs show untransformed data and standard errors have been computed using the

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

untransformed data (Figure 13). A paired sample t-test indicates a significant difference between mean transformed tracking error (t_{38}=37.19, p<0.001, η_p^2= 0.97); tracking error was highest in the high task-load condition. A paired sample t-test indicated significant differences in the standard deviation of the RMSEs between the two conditions (t_{38}=11.09, p<0.001, η_p^2= 0.76); variability of tracking error was highest in the high task-load condition.

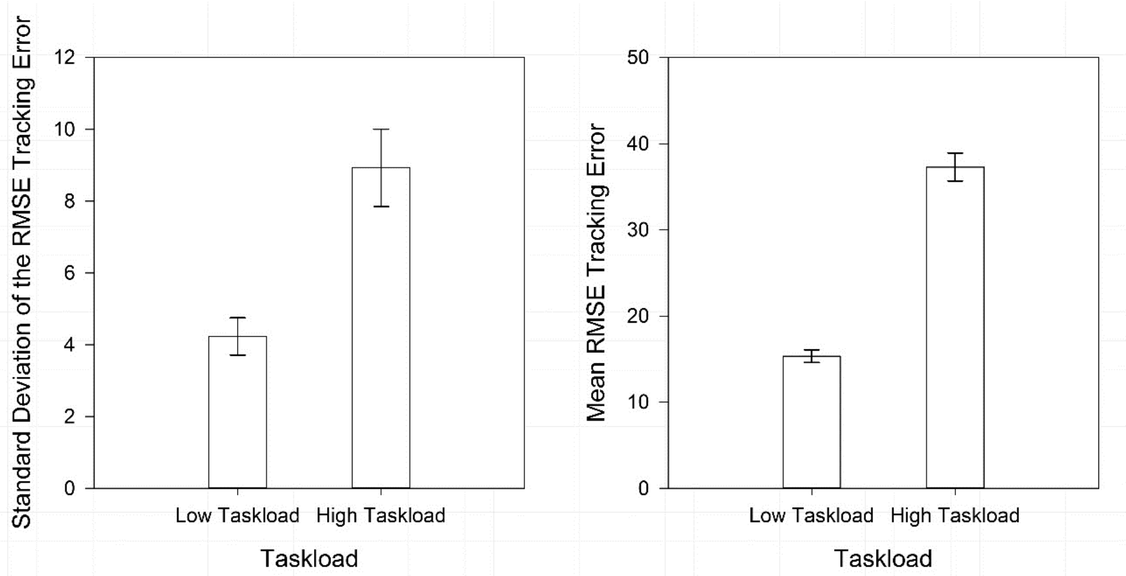


Figure 13: Mean RMSE tracking error and standard deviation of mean RMSE tracking error for high and low task-load conditions

3.7.2.3. *System Monitoring Task*

Two measures have been used to characterise performance on this task. Firstly, reaction time to the target (correct response to absence of a green lamp or presence of a red lamp; correct response to extreme scale deflection). Secondly, the number of time-outs characterising when a demanded response has not taken place. The reaction time data is ratio data and is treated as such in the analysis. The missed response data is discrete, count data and or analysis, these values are summed and tested against an expected uniform distribution to test for task-load effects.

Exploratory analysis of the reaction time data revealed significant positive skew and departure from normality across all conditions and tasks (response to lamps and to scales). This can be common with this kind of data whereby a small number of participants are simply worse at the task. The more extreme values in the tail cannot be classified as outliers. A $\log_{10}$ transformation was applied to minimise skew and stabilise variance. Data from one participant in the scale task was lost. Graphs show untransformed data and standard errors have been computed using the untransformed data. No significant difference between the transformed reaction time data between task-loads was found for either the lamp task (t_{38}= 0.80, p>0.05) or the scale task (t_{37}=0.62, p>0.05).

The total number of time-outs for the scale and lamp subtasks are shown in Table 8. A χ^2 goodness-of-fit test was conducted. An expected distribution specifying the same number of timeout responses in each category was used for comparison. Analysis indicates no significant difference in the distribution of timeouts for the lamp subtask between the high and low task-load conditions (χ^2= 1.58, df=1, p>0.05). A significant difference in the distribution of timeouts for the scale subtask was found between the high and low task-load conditions (χ^2= 61.32, df=1, p<0.001).

Table 8: Total number of time-outs for the scale and lamp subtasks

Timeouts for the Lamp Subtask		Timeouts for the Scale Subtask	
Low task-load	High task-load	Low task-load	High task-load
55	69	68	195

3.7.2.4. Communications task

In the communications task, participants are required to select correct radio and frequency in response to Ownship instruction. The number of incorrect radio and frequency instructions are recorded. For analysis, these values are summed and tested against an expected uniform distribution to test for task-load effects.

The total number of incorrect responses to the radio and frequency commands are shown in Table 9. A χ^2 goodness-of-fit test was conducted. An expected distribution specifying the same number of incorrect responses in each category was used for comparison. Analysis indicates no significant difference in the distribution of radio selection errors between the high and low task-load conditions (χ^2= 1.69, df=1, p>0.05). A significant difference in the distribution of frequency selection errors was found between the high and low task-load conditions (χ^2= 4.9, df=1, p<0.05).

Table 9: Number of incorrect radio and frequency selection responses.

Incorrect Radio Selected		Incorrect Frequency Selected	
Low task-load	High task-load	Low task-load	High task-load
11	18	13	27

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

3.7.3. Physiological Data Analysis

A structured approach to measurement of the physiological data has been taken. For each physiological measure, a 2 × 2 × 3 mixed ANOVA was conducted. The three factors are task-load (low task-load vs high task-load), gradient (low task-load to high task-load vs high task-load to low task-load) and task (communication vs tracking vs system monitoring). Corrections for deviations in sphericity in the task-type condition used the Greenhouse-Geisser corrected df. Greenhouse-Geisser ε is reported to characterise the departure where significant. Gradient is a between-subjects factor. All other factors are within-subjects.

3.7.3.1. Frequency Domain Measures

Three frequency domain measures were taken during the trial: very-low frequency, low frequency and high frequency.

Very low frequency

Overall, the main effect of task-type on very low frequency HRV was significant ($p < 0.01$, $\eta_p^2 = 0.21$). Pairwise comparisons indicate significant differences in the very-low frequency component between the communications task and both other tasks (system monitoring and tracking task). No other significant effects or interactions were found.

Significant departure from sphericity was indicated for the task type effect ($\chi^2(2) = 7.94$, $p<0.02$, $\varepsilon = 0.82$) The main effect of task type was significant ($F_{1.2, 34.2} = 7.2$, $p <0.01$, $\eta_p^2 = 0.21$). Bonferroni corrected pairwise comparisons showed significant differences between the communications and tracking task (mean difference = 769.5ms, SE = 303.2, p<0.05) and the communication and system monitoring task (mean difference = 885.0ms, SE = 295.8, p=0.02). No significant pairwise difference between the tracking and system monitoring task was found.

The main effect of task-load did not reach significance ($F_{1, 33} = 0.19$, $p = 0.66$, $\eta_p^2 = 0.01$). No main effect of gradient was found ($F_{1, 33} = 0.01$, $p = 0.98$, $\eta_p^2 < 0.01$). No significant interactions were found between task-load and task-type ($F_{2, 64} = 0.04$, $p = 0.96$, $\eta_p^2 = 0.01$), task-load and gradient ($F_{1, 33} = 0.02$, $p = 0.88$, $\eta_p^2 = 0.01$) or task type and gradient ($F_{1, 33} = 0.23$, $p = 0.80$, $\eta_p^2 = 0.07$).

Low frequency

Overall, the main effect of task-type had a significant effect on low frequency HRV ($p = 0.01$, $\eta_p^2 = 0.21$). Pairwise comparisons indicate significant differences in the low frequency component between the communications task and both other tasks (system monitoring and tracking task). No other significant effects or interactions were found.

Significant departure from sphericity was indicated for the task type effect ($\chi^2(2) = 9.12$, $p<0.02$, $\varepsilon = 0.80$) and the task type × task-load interaction ($\chi^2(2) = 24.78$, $p<0.01$, $\varepsilon = 0.65$) The main effect of task type was significant ($F_{1.6, 50.1} = 8.56$, $p = 0.01$, $\eta_p^2 = 0.21$). Bonferroni corrected pairwise comparisons showed significant differences between the communications and tracking task (mean difference = 833.1ms, SE =

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

140.3, p<0.01) and the communication and system monitoring task (mean difference = 706.5ms, SE = 130.0, p<0.01). No significant pairwise difference between the tracking and system monitoring task was found.

The main effect of task-load did not reach significance ($F_{1, 32} = 1.02$, $p = 0.32$, $\eta_p^2 = 0.03$). No main effect of gradient was found ($F_{1, 32} = 0.61$, $p = 0.44$, $\eta_p^2 < 0.01$). No significant interactions were found between task-load and task-type ($F_{1.3, 41.3} = 0.04$, $p = 0.90$, $\eta_p^2 = 0.01$), task-load and gradient ($F_{1, 32} = 0.08$, $p = 0.78$, $\eta_p^2 = 0.03$) or task type and gradient ($F_{2, 33} = 0.50$, $p = 0.60$, $\eta_p^2 = 0.02$).

High Frequency

Overall, the main effect of task-type had a significant effect on HF frequency component ($p = 0.02$, $\eta_p^2 = 0.12$). Pairwise comparisons indicate significant differences in the high frequency component between the communications task and both other tasks (system monitoring and tracking task). No other significant effects or interactions were found. The main effect of gradient approached significance ($p = 0.1$, $\eta_p^2 = 0.08$). No other significant effects or interactions were found.

Significant departure from sphericity was indicated for the task type effect ($\chi^2(2) = 10.75$, $p<0.01$, $\varepsilon = 0.77$). The main effect of task type was significant ($F_{1.5, 49.5} = 4.31$, $p = 0.02$, $\eta_p^2 <0.12$). Bonferroni corrected pairwise comparisons showed significant differences between the communications and tracking task (mean difference = 125.9ms, SE = 41.2, p=0.02) and the communication and system monitoring task (mean difference = 111.7ms, SE = 130.0, p=0.01). No significant pairwise difference between the tracking and system monitoring task was found.

The main effect of task-load did not reach significance ($F_{1, 32} = 2.45$, $p = 0.13$, $\eta_p^2 = 0.07$). The main effect of gradient approached significance ($F_{1, 32} = 2.93$, $p = 0.1$, $\eta_p^2 = 0.08$). No significant interactions were found between task-load and task-type ($F_{2, 64} = 0.72$, $p = 0.49$, $\eta_p^2 = 0.02$), task-load and gradient ($F_{1, 32} = 0.04$, $p = 0.85$, $\eta_p^2 = 0.01$) or task type and gradient ($F_{2, 33} = 0.40$, $p = 0.67$, $\eta_p^2 = 0.01$).

Low frequency/high frequency ratio

Overall, the main effect of task-type had a significant effect on LF/HF frequency component ($p = 0.03$, $\eta_p^2 <0.13$). Pairwise comparisons indicate significant differences in the high frequency component between the communications task and the system monitoring task. No other significant effects or interactions were found.

The main effect of task type was significant ($F_{2, 52} = 4.0$, $p = 0.03$, $\eta_p^2 <0.13$). Bonferroni corrected pairwise comparisons showed significant differences between the communications and the system monitoring tasks (mean difference = 0.73ms, SE = 0.26, p=0.03) No significant pairwise difference between the tracking and system monitoring task or the tracking and communications task were found.

The main effect of task-load was not significant ($F_{1, 26} = 0.11$, $p = 0.74$, $\eta_p^2 <0.01$). The main effect of gradient was not significant ($F_{1, 32} = 0.04$, $p = 0.83$, $\eta_p^2 <0.01$). No significant interactions were found

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

between task-load and task-type ($F_{2, 52}$ = 0.59, p = 0.56, η_p^2 = 0.02), task-load and gradient ($F_{1, 32}$ = 1.7, p = 0.2, η_p^2 = 0.06) or task type and gradient ($F_{1, 32}$ = 0.1, p = 0.90, η_p^2 < 0.01).

Task Type effects in the frequency domain

A consistent effect in the frequency domain is the discrimination of the communications task and the other two tasks, regardless of the task-load effect. This is shown graphically in Figure 14.

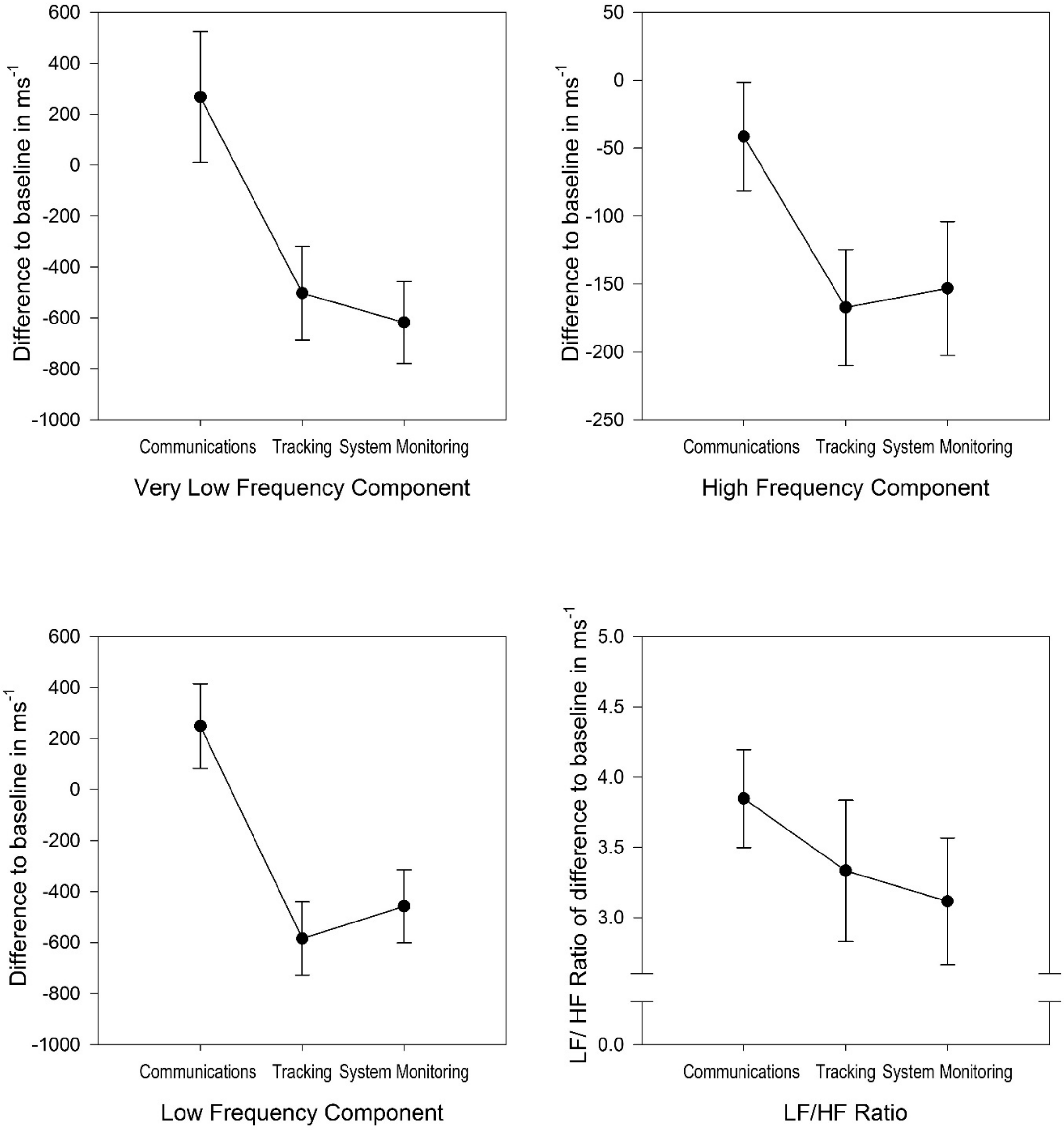


Figure 14: Estimated marginal means for the task-type effect for very low, low, high and low/ high ratio frequency domains. Error bars show one SE

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

3.7.3.2. *Mean breath rate*

Overall, the main effect of task-load had a significant effect on mean breath rate ($p = 0.03$, $\eta_p^2 = 0.14$). The main effect of gradient approached significance ($p = 0.1$, $\eta_p^2 = 0.08$) and the interaction between task-load and task-type approached significance ($p = 0.07$, $\eta_p^2 = 0.08$). No other significant effects or interactions were found.

Significant departure from sphericity was indicated for the task type effect ($\chi^2(2) = 23.09$, $p<0.01$, $\varepsilon = 0.66$) and the task type × task-load interaction ($\chi^2(2) = 15.30$, $p<0.01$, $\varepsilon = 0.73$). The main effect of task type was not significant ($F_{1.3, 43.6} = 0.28$, $p = 0.66$, $\eta_p^2 <0.01$). The main effect of task-load was significant ($F_{1, 33} = 5.45$, $p = 0.03$, $\eta_p^2 = 0.14$). The main effect of gradient approached significance ($F_{1, 33} = 4.34$, $p = 0.05$, $\eta_p^2 = 0.12$). The interaction between task-load and task-type approached significance ($F_{1.4, 47.8} = 2.79$, $p = 0.07$, $\eta_p^2 = 0.08$) No significant interactions were found between, task-load and gradient ($F_{1, 33} = 0.46$, $p = 0.50$, $\eta_p^2 = 0.01$) or task type and gradient ($F_{2, 33} = 0.21$, $p = 0.67$, $\eta_p^2 < 0.01$).

3.7.3.3. *Mean Heart Rate*

Overall, the main effect of task-load had a significant effect on mean heart rate ($p = 0.03$, $\eta_p^2 = 0.15$). No other significant effects or interactions were found.

The main effect of task type was not significant ($F_{2,60} = 0.11$, $p = 0.90$, $\eta_p^2 <0.01$). The main effect of task-load was significant ($F_{1, 30} = 5.18$, $p = 0.03$, $\eta_p^2 = 0.15$). The main effect of gradient was not significant ($F_{1, 30} <0.01$, $p = 0.96$, $\eta_p^2 <0.01$). No significant interactions between task-load and task-type ($F_{2, 60} = 1.24$, $p = 0.30$, $\eta_p^2 = 0.04$), task-load and gradient ($F_{1, 30} = 0.04$, $p = 0.84$, $\eta_p^2 = 0.01$) or task type and gradient ($F_{2, 30} = 0.54$, $p = 0.59$, $\eta_p^2 < 0.02$) were found.

3.7.3.4. *Blood oxygen level (Sp02)*

No significant effects or interactions were found.

Since these data are percentage data an arcsine transformation has been applied in order to conduct the ANOVA procedure. The percentage data have been transformed to probabilities through division by 100 percent. These probabilities have then been transformed:

$$Sp0_{2(Transformed)} = arcsine\sqrt{p}$$

All ANOVA outputs use the transformed data.

No significant main effects for task type ($F_{2,54} = 0.60$, $p = 0.55$, $\eta_p^2 = 0.02$), task-load ($F_{1, 27} = 0.85$, $p = 0.37$, $\eta_p^2 = 0.03$) or gradient ($F_{1, 28} =0.15$, $p = 0.70$, $\eta_p^2 < 0.01$) were found. No significant interactions between task-load and task-type ($F_{2, 54} = 1.5$, $p = 0.23$, $\eta_p^2 <0.01$), task-load and gradient ($F_{1, 28} = 0.05$, $p = 0.87$, $\eta_p^2 = 0.02$) or task type and gradient ($F_{2, 28} = 0.33$, $p = 0.72$, $\eta_p^2 = 0.02$) were found.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

3.7.3.5. SDNN

Overall, the main effect of task type had a significant effect on mean SDNN ($p = 0.01$, $\eta_p^2 = 0.21$). No other significant effects or interactions were found.

Significant departure from sphericity was indicated for the task type × task-load interaction ($\chi^2(2) = 10.13$, $p<0.01$, $\varepsilon = 0.76$). The main effect of task type was significant ($F_{2,56} = 7.52$, $p = 0.01$, $\eta_p^2 = 0.21$), estimated marginal means are shown in Figure 15. Bonferroni corrected pairwise comparisons showed significant differences between the communications and tracking task (mean difference = 13.5ms, SE = 3.1, p=0.01) and the communication and system monitoring task (mean difference = 12.9 , SE = 2.9, p<0.01). No significant pairwise difference between the tracking and system monitoring task was found.

The main effect of task-load was not significant ($F_{1, 28} = 0.79$, $p = 0.38$, $\eta_p^2 = 0.03$). The main effect of gradient was not significant ($F_{1, 28} = 1.74$, $p = 0.22$, $\eta_p^2 = 0.06$). No significant interactions between task-load and task-type ($F_{1.5, 42.6} = 0.22$, $p = 0.74$, $\eta_p^2 < 0.01$), task-load and gradient ($F_{1, 28} = 0.01$, $p = 0.97$, $\eta_p^2 < 0.01$) or task type and gradient ($F_{2, 28} = 0.35$, $p = 0.71$, $\eta_p^2 = 0.01$) were found.

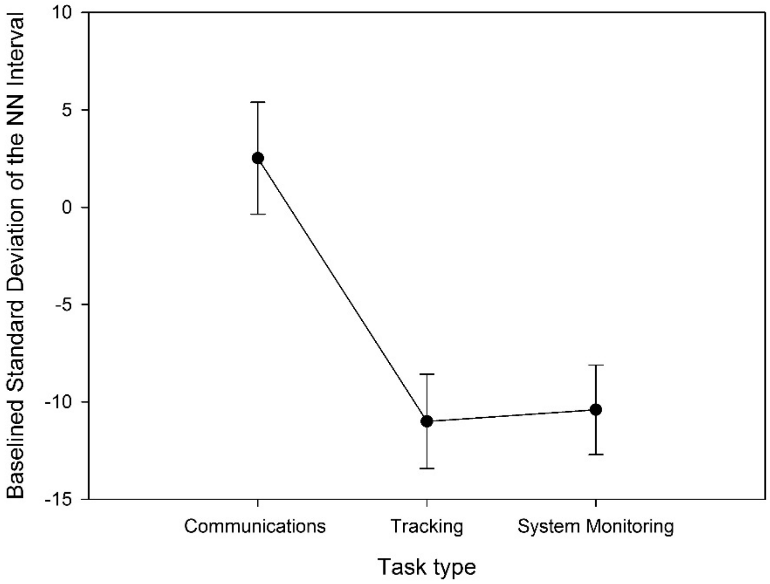


Figure 15: Estimated marginal means for mean SDNN across all task-types

3.7.3.6. Blink rate

Overall, the main effect of task type had a significant effect on mean blink rate ($p < 0.01$, $\eta_p^2 = 0.75$). A significant interaction was found between task-type and task-load. Pairwise testing indicates that task-load is discriminated in the visual tasks only: system monitoring and tracking tasks and not in the communications task.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Significant departure from sphericity was indicated for the task type main effect ($\chi^2(2) = 7.95$, $p<0.02$, $\varepsilon = 0.84$). The main effect of task type was significant ($F_{1.7,61.8} = 109.2$, $p < 0.01$, $\eta_p^2 = 0.75$). The main effect of task-load was not significant ($F_{1, 37} = 1.50$, $p = 0.23$, $\eta_p^2 = 0.04$).

A significant interactions between task-load and task-type ($F_{2, 74} = 7.78$, $p <0.01$, $\eta_p^2 <0.01$), task-load and gradient ($F_{1, 28} = 0.01$, $p = 0.97$, $\eta_p^2 <0.01$) or task type and gradient ($F_{2, 28} = 0.35$, $p = 0.71$, $\eta_p^2 = 0.17$) were found. Estimated marginal means are shown in Figure 16. Pairwise comparisons show a significant difference in the number of blinks between the high and low task-load conditions for the tracking condition ($t_{37} = 2.8$, $p=0.01$, $\eta_p^2 = 0.20$, two-tailed test), and for the system monitoring condition ($t_{37} = 2.3$, $p=0.03$, $\eta_p^2 = 0.13$ two-tailed test). No significant difference was found between task-loads in the communications task ($t_{37} = 1.1$, $p=0.26$, $\eta_p^2 = 0.03$ two-tailed test).

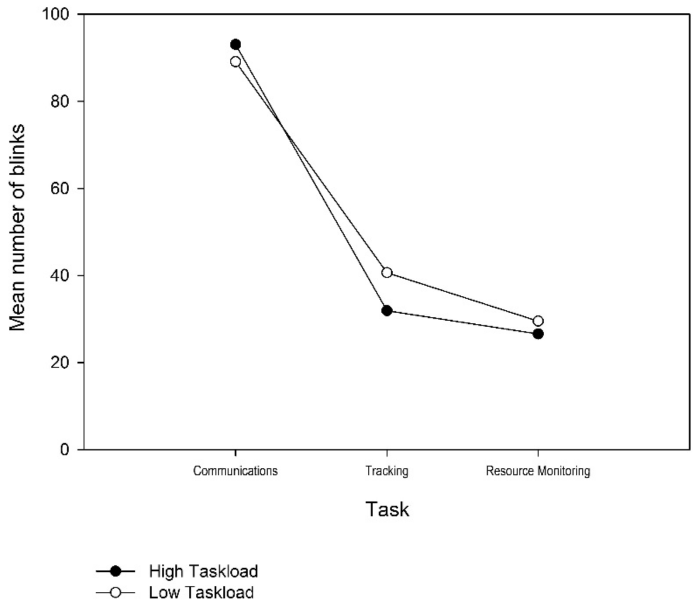


Figure 16: Estimated marginal means showing blink-rate for the task-type × task-load interaction

3.8. Summary of Results

Hypothesis 1: There will be association between the physiological data and task-load.

Mean heart rate and mean breathing rate successfully discriminated between task-loads in all tasks. Blinking rate discriminated task load in the visual tasks but not in the communication task.

Hypothesis 2: There will be differences in patterns of physiological data depending on task type.

Task-type tended to be differentiated between the communications task and the other two visual tasks. Although the task loading for each of the tasks was high, it may be that higher capacity in the echoic

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

sensory store can explain these results. Even at high task-loads, auditory tasks may elicit less workload given the higher capacity in the echoic sensory store.

Hypothesis 3: There will be differences in patterns of physiological response data depending on the task-load gradient.

No systematic pattern of differences arose between participants who were presented with a high to low task load gradient as opposed to a low to high task-load gradient. It may be that a slower rate of rise or fall is required for this effect to be discriminated by the physiological data.

3.9. Conclusion

In this study we have measured physiological variables in response to different tasks from 39 participants. Participants completed tasks presented on the NASA MATB II. Tasks included communications, system monitoring and a tracking task using a joystick. Participants completed two conditions: a high task-load condition and a low task-load condition. Gradient was also manipulated between participants: either high to low task-load, or low to high task-load.

Results confirm the validity of the task schedules. Subjective mental workload was rated as significantly higher in the high task-load condition. Performance measured on a variety of variables was significantly better in the low task-load condition.

The smart-vest physiological measurement technology from CSEM SA was used to capture a variety of physiological measures. Mean heart rate and mean breathing rate discriminate the task-load for each task. Mean heart and breathing rate are significantly higher in the high task-load conditions. The frequency domain measures to include low, very low frequency domains and the LF/ HF ratio discriminate between the communications task and both other tasks (system monitoring and tracking). Differences in the high frequency band between these groups of tasks approached significance but did not reach it. The frequency domain measures did not discriminate the low and high task-load conditions suggesting that these measures are less sensitive to workload changes in response to changes in task-load. This pattern of differences observed in the frequency domain measures may be explained by the higher capacity of the echoic sensory store. A high task-load in the communications task may still be managed more easily since the higher capacity auditory modality is engaged.

Finally, the mean number of eye-blinks was significantly different in response to task-load in the visual tasks (system monitoring and tracking) when compared to the communications task.

These results broadly support the use of physiological data collected in this non-invasive way for the definition and understanding of the human performance envelope in aviation.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

4. FIRST REAL TIME SIMULATION – RESULTS AND DISCUSSION

4.1. Summary of the Experimental design

The main objectives of the experiment carried out were to determine the validation of the HPE concept and to investigate the impact on the pilots' performance of the three factors selected (workload, stress and SA). The hypothesis was based on the verification that three factors combined require lower levels of factors (medium) than one factor alone (high levels of workload or stress or SA) to degrade the performance.

4.1.1. Scenario 1

To test this hypothesis, in Scenario 1 an approach phase with an A320-200 was simulated at the DLR research simulator AVES, a motion simulator with six degrees-of freedom. Eight different experimental conditions were tested in the simulator, ending in eight runs of 15/20 minutes in which pilots had to manage a course of events expected to produce variation of factors during the simulated flights (see Table 10). The events associated with the increase of workload were turbulence (medium or high), approach and runway change. Low fuel situation, delay vectors and loud noise were used to produce an increase in stress. Finally, the reduction of SA was generated by the events Low visibility, Localiser interference and Wind shift. The first run was taken as the baseline, with only basic tasks to perform (Flying an Instrument Landing System (ILS) approach with manual control) and nominal levels of workload (WL), stress (ST), and situation awareness (SA). Runs 2 and 3 included respectively medium workload and high workload, while runs 5 and 6 included, in turn, high stress and reduced SA. Run 7 and 8 were a combination of medium respectively high workload, stress and reduced SA. Run 1, an ILS approach without any event, was taken as the baseline.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Table 10: List of events that took place in each run

Run	Events
1. Baseline	None
2. Medium workload	Medium turbulence throughout whole scenario
3. High workload	High turbulence throughout whole scenario
4. Very high workload	High turbulence throughout whole scenario Approach and RWY change during initial approach (between IAF and FAF)
5. High stress	Low fuel situation throughout whole scenario, Delay vectors during initial approach (between IAF and FAF) Loud noise during final approach (between FAF and landing)
6. High reduced situation awareness	Low visibility throughout whole scenario Localiser interference during final approach (between FAF and landing) Wind shift during final approach (between FAF and landing)
7. Medium workload,	Medium turbulence throughout whole scenario
Medium stress,	Low fuel situation throughout whole scenario Delay vectors during initial approach (between IAF and FAF)
Medium reduced situation awareness	Low visibility throughout whole scenario Localiser interference during final approach (between FAF and landing)
8. High workload,	High turbulence throughout whole scenario
High stress,	Low fuel situation throughout whole scenario Delay vectors during initial approach (between IAF and FAF) Loud noise during final approach (between FAF and landing)
High reduced situation awareness	Low visibility throughout whole scenario Localiser interference during final approach (between FAF and landing) Wind shift during final approach (between FAF and landing)

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

4.1.2. Scenario 2

The following events took place which increased workload and stress, and reduced SA:

- Approaching Bremen with standard fuel for 50 min remaining flight time
- Preparation of CAT1 approach RWY 27
- Go-around during ILS approach RWY 27 due to slow preceding (VFR) traffic
- During downwind ELEC AC BUS 1 Failure
- Constantly increasing workload due to the procedure and the time needed
- Decision making and handling of complexity under low fuel conditions
- CAT2 ILS approach RWY 09
- Possible engine flame-out due to amount of fuel

Data collected in Scenario 1 were used to validate the data collected in Scenario 2. In other words, physiological, performance and subjective measures collected in, for example, high workload (as seen in run 3) was used to identify those same instances of high workload in Scenario 2. Thus, in this scenario the HPE factors were not treated as independent, but dependent variables.

4.2. Subjects

Ten A320 first officers from a major European airline participated in the experiment. All were male, aged between 28 and 36 years (M = 31 yrs, SD= 3.28). On average they had 3125 hours (SD= 1557) in the A320, with an average of 678.7 hours (SD= 43) in the last month. They had a total flight experience ranging from 2250 to 7000 hours (M = 4045 hours, SD= 1569). The subjects were asked to fly manually and land the aircraft, following ATC instructions and working with the Pilot Monitoring (a 'confederate' captain from the same airline).

The complete set of data was available for Run 1 and Run 8, while nine pilots out of ten performed Run 3 (See

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Table 11). Run 2 was only performed by 3 pilots, so data related to that run were not analysed at all. Altogether, a total of 29 runs were analysed with respect to Performance data, Performance Curve and Debriefings. Physiological data for pilot 4, 6 and 7 were not recorded due to technical difficulties with the smart vest. Reliable behavioural markers are not available for pilot 6 Run 4 as almost all his observed behaviours were due to a flu condition (reported during the debriefing).

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Table 11: Summary of the conditions performed during the simulation

		Scenario 1								Scenario 2	*Total*
		Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Run 7	Run 8		
Pilots	1	✓	✓	✓	✓	✓	X	✓	✓	✓	*8*
	2	✓	✓	✓	✓	✓	✓	X	✓	✓	*8*
	3	✓	X	✓	✓	✓	X	X	✓	✓	*5*
	4	✓	X	✓	✓	✓	✓	✓	✓	✓	*8*
	5	✓	X	✓	✓	✓	✓	✓	✓	✓	*8*
	6	✓	X	X	✓	✓	✓	✓	✓	✓	*7*
	7	✓	X	✓	✓	✓	✓	✓	✓	✓	*8*
	8	✓	X	✓	✓	✓	✓	✓	✓	✓	*8*
	9	✓	X	✓	✓	X	✓	X	✓	✓	*5*
	10	✓	✓	✓	✓	X	✓	✓	✓	✓	*8*
Total		*10*	*3*	*9*	*10*	*8*	*8*	*7*	*10*	*10*	

4.3. Subjective data

Five subjective assessment tools were selected to assess workload (ISA and NASA-TLX), fatigue (Samn-Perrelli Fatigue scale), situation awareness (10D-SART), and stress (Stress Arousal Checklist). Also, at the end of each run of Scenario 1 an interviewer conducted a structured debriefing session in which pilots were invited to provide a graphical representation of their performance through the Performance Curve, to discuss their overall impression of the simulation and to explain their behaviour and choices during the execution of the tasks.

4.3.1. ISA

The Instantaneous Self-Assessment (ISA) measurement was used to assess the pilot's workload in real time. The pilot was asked to rate his workload on a scale from 1 to 5 every 2 minutes during all runs of scenario 1.

Figure 17 shows the rated ISA workload of all pilots and all runs of scenario 1. The results reveal that the workload of run 1 was significantly lower compared to all other runs ($p < .001$). The highest workload was achieved in run 4 with 3.13, which is the very high workload run. However, there was not significant

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

difference to run 3 as the high workload run. Additionally, there was no significant difference in the workload between run 3 (high workload) and run 7 (medium workload, medium stress, medium reduced situation awareness). However, there was a significant difference in workload between run 3 (high workload) and run 5 (high stress) and run 4 (very high workload) and run 5 and 6 (highly reduced situation awareness). The workload in run 5 was significantly lower compared to run 3 ($p < .05$). Furthermore, the workload of run 5 and run 6 was significantly lower compared to run 4 ($p < .05$). There was no significant difference between the workload of run 3 and run 6.

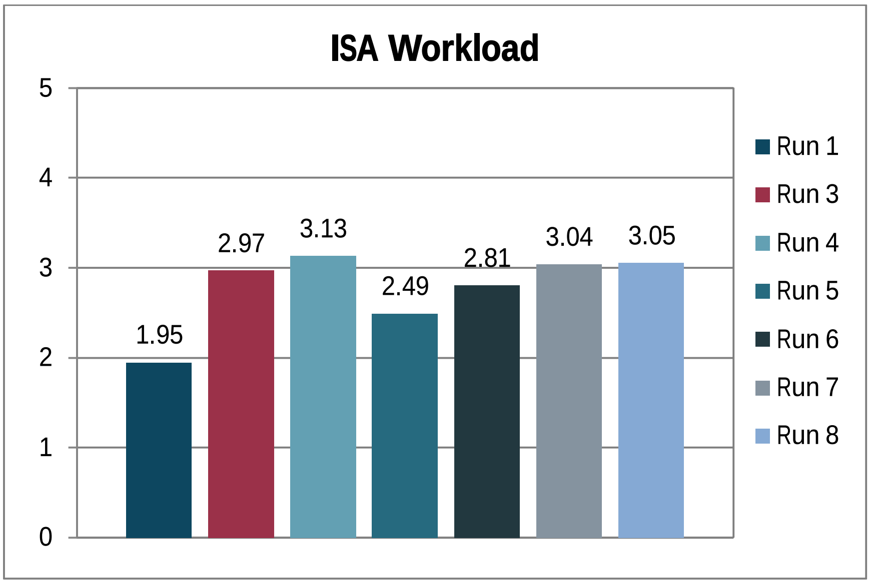


Figure 17: ISA workload rating of all pilots

4.3.2. NASA-TLX

The NASA-TLX was used to subjectively evaluate the workload of each run of scenario 1 and scenario 2. The weighting procedure of the NASA-TLX was done as part of the initial briefing. After each run the pilots were required to rate their workload on a scale from 1 to 20 along six factors: mental demand, physical demand, temporal demand, performance, effort, and frustration level.

The overall workload score for each pilot per run was computed by multiplying each rating by the weight given to that factor. Figure 18 shows the results of all runs of scenario 1 and scenario 2. As with the ISA results, the workload of run 1 was significantly lower compared to all other runs and compared to scenario 2 ($p < .005$). The highest workload of scenario 1 was achieved in run 8. However, there was no significant difference between run 3 (high workload) and run 4 (very high workload) as well as between run 3 (high workload) and run 7 (medium workload, medium stress, medium reduces situation awareness). Nevertheless, the workload of run 5 (high stress) was significantly lower compared to run 4 ($p < .05$). There was no significant difference between the NASA-TLX workload of run 3, run 5 and run 6.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

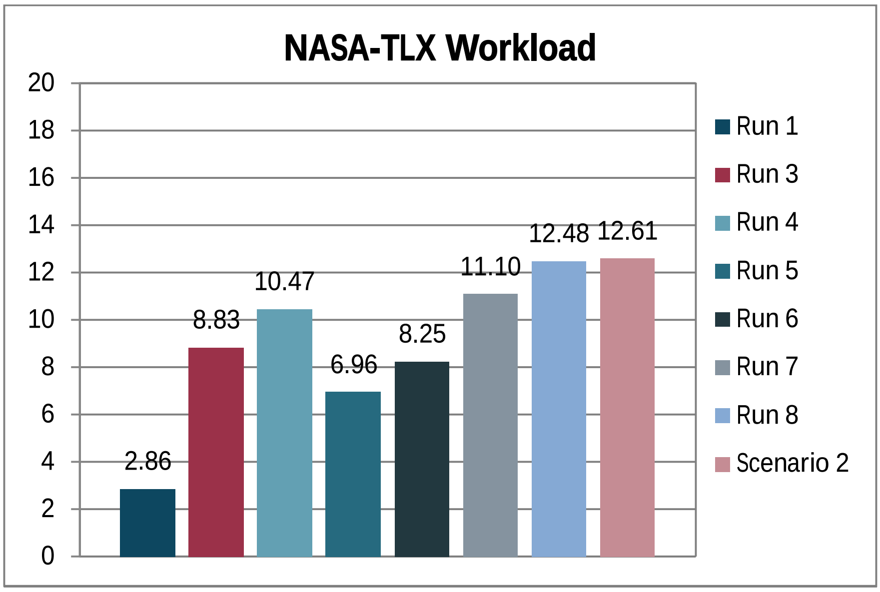


Figure 18: NASA-TLX workload rating of all pilots

4.3.3. Samn-Perelli Fatigue scale

Subjects were asked to rate their level of fatigue using the 7-point Samn-Perelli scale before and after each simulator session. The seven levels of the scale are recalled below:

1. Fully alert, wide awake;
2. Very lively, responsive, but not at peak;
3. Okay, somewhat fresh;
4. A little tired, less than fresh;
5. Moderately tired, let done;
6. Extremely tired, very difficult to concentrate;
7. Completely exhausted, unable to function effectively.

As shown by the following table (

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Table 12), some data are missing, because either some runs have not been made by all the pilots, either the Samn-Perelli data have not been recorded (No data cells). Too few data are available for run 2 and so this run was not included in the analysis.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Table 12: Ratings by each pilot using the 7-point Samn-Perelli scale

Pilot	Scenario 1 Run 1		Scenario 1 Run 2		Scenario 1 Run 3		Scenario 1 Run 4		Scenario 1 Run 5		Scenario 1 Run 6		Scenario 1 Run 7		Scenario 1 Run 8	
	Before	After	Before	After	Before	After	Before	After	Before	After	Before	After	Before	After	Before	After
1	No data	No data	2	2	No data	No data	2	1	2	2	2	2	1	2		
2	2	2	No data	No data	2	3	No data	No data	No data	No data	No data	No data			No data	No data
3	2	2			2	2	2	1	2	2					1	1
4	1	1			1	2	3	1	2	2	1	2	2	3	2	2
5	1	1			1	1	3	2	1	1	1	1	2	3	1	1
6	2	2					2	2	2	1	1	2	2	2	3	2
7	3	2			1	1	2	1	3	1	1	2	1	1	2	1
8	1	3			1	2	2	1	2	3	1	1	1	2	3	1
9	2	1			1	1	2	1			1	1			2	1
10	1	1	2	2	1	1	1	1			2	1	1	1	2	2

Considering ratings both before and after each run, it can be noticed that the fatigue level is always low and should not have an effect on the pilot performances (see Figure 19 below).

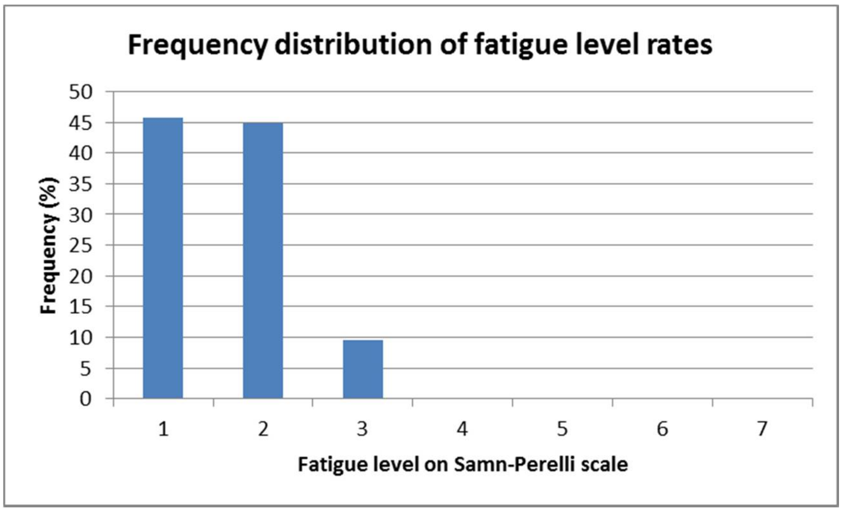


Figure 19: Frequency distribution of fatigue level rates

Nevertheless, further analyses were done. The next figure shows (Figure 20) that if we consider both the rating before and after each run, there is no significant difference between the runs.

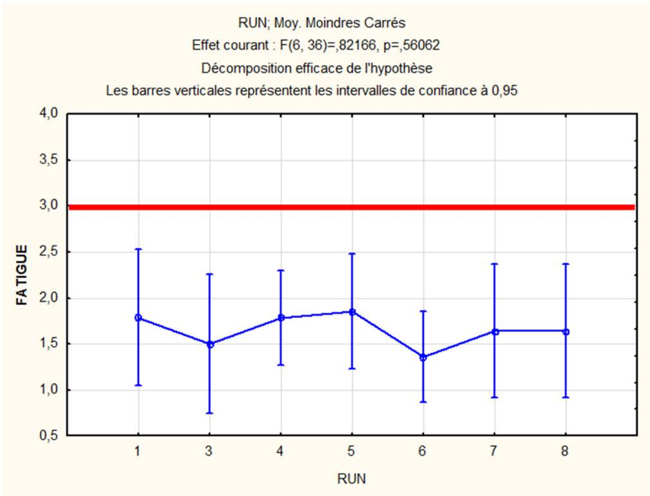


Figure 20: Comparison of fatigue levels before and after each run

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Moreover, the comparison between the fatigue level before each simulation session and after that session does not reveal any significant difference. So each run does not increase significantly the fatigue level of the pilot and so the fatigue level should not decrease performances during the run (see Figure 21).

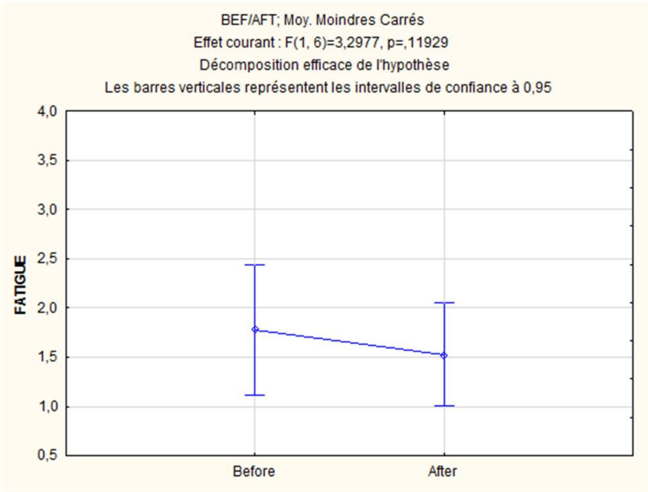


Figure 21: Comparison of fatigue levels before and after each simulation session

As pilots could have 5 simulation sessions in a single half day, the cumulative effect has been controlled. The next figure (Figure 22) displays the evolution of the fatigue level before the 5 consecutive simulation sessions for all pilots who had such a sequence. It can be notice that there is no cumulative effect.

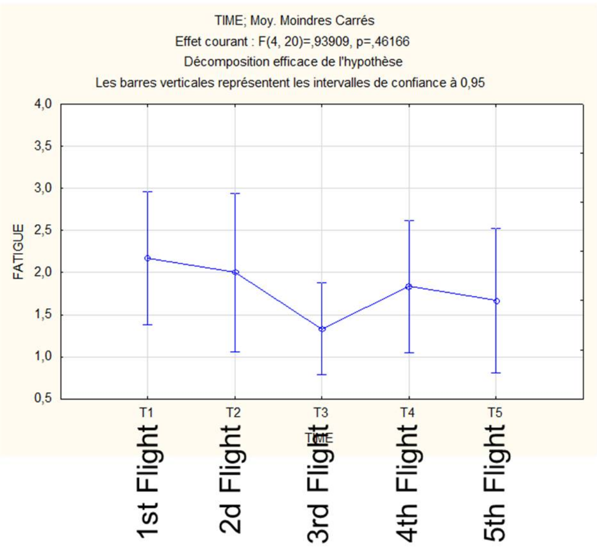


Figure 22: Evolution of fatigue level throughout simulation sessions

To conclude, the fatigue level, measured through Samn-Perelli scores is globally at a low level before and after each simulation session. Moreover, there is no significant difference of the fatigue level for the

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

different conditions (runs) and there is no cumulative effect generated by the sequence of 5 simulation sessions in a single half day.

Therefore, for this study, we can assume that fatigue is not a factor that will influence performances between the experimental conditions of scenario 1.

4.3.4. 10D-SART

The 10D-SART was used to subjectively measure situation awareness. Subjects rated each of the ten dimensions on a 7-point rating scale: familiarity of the situation, focusing of attention, information quantity, information quality, instability of the situation, concentration of attention, complexity of the situation, variability of the situation, arousal, and spare mental capacity.

A composite SART score was calculated using the following formula: SA = U – (D – S), where: U = summed understanding, D = summed attentional demand, and S = summed attentional supply. It applies that the lower the score the lower the situation awareness of the pilots. Figure 23 shows the results. A statistical analysis revealed that the situation awareness in run 1 was significantly better compared to all other runs ($p < 0.05$). Additionally, the situation awareness was significantly lower in run 8 (high workload, high stress, highly reduced situation awareness) compared to run 7 (medium workload, medium stress, medium reduced situation awareness). The lowest situation awareness of the runs of scenario 1 was achieved in run 8.

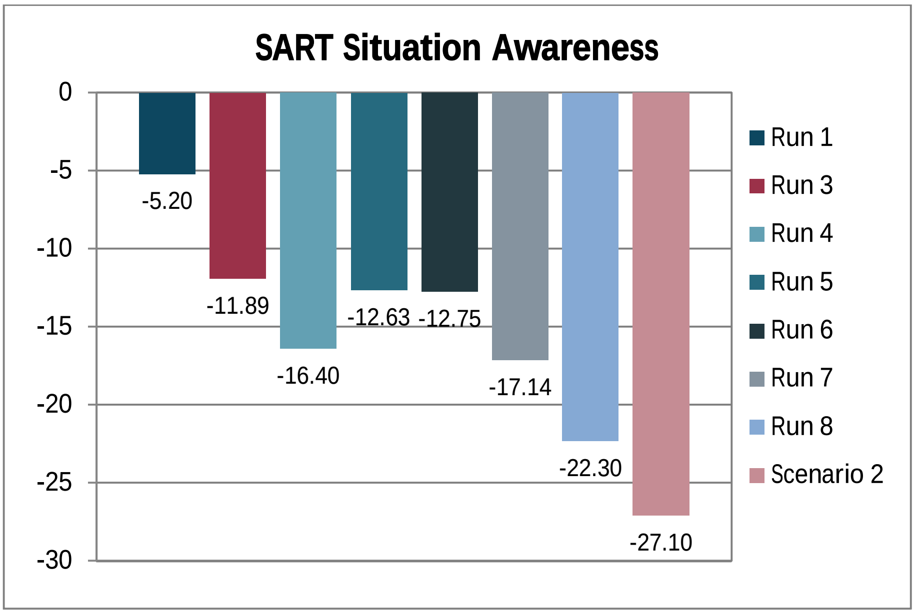


Figure 23: SART situation awareness rating of all pilots

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

4.3.5. Stress Arousal Checklist

The SACL was used to subjectively measure stress. The pilots were asked to rate their level of stress during the run on the basis of two 4 point scales (stress and arousal). The scale points are: definitely feel, feel, unsure and do not feel.

The SACL scale is dichotomized and the items were scored in the direction of the scale concept as 0 or 1. The different item scores were then summed. Figure 24 shows the results. The stress of run 1 was significantly lower compared to all other runs ($p < .05$). Even though the stress was higher in run 8 (high workload, high stress, highly reduced situation awareness) compared to run 7 (medium workload, medium stress, medium reduced situation awareness), no significant difference could be found. The highest stress was achieved in run 8 of scenario 1.

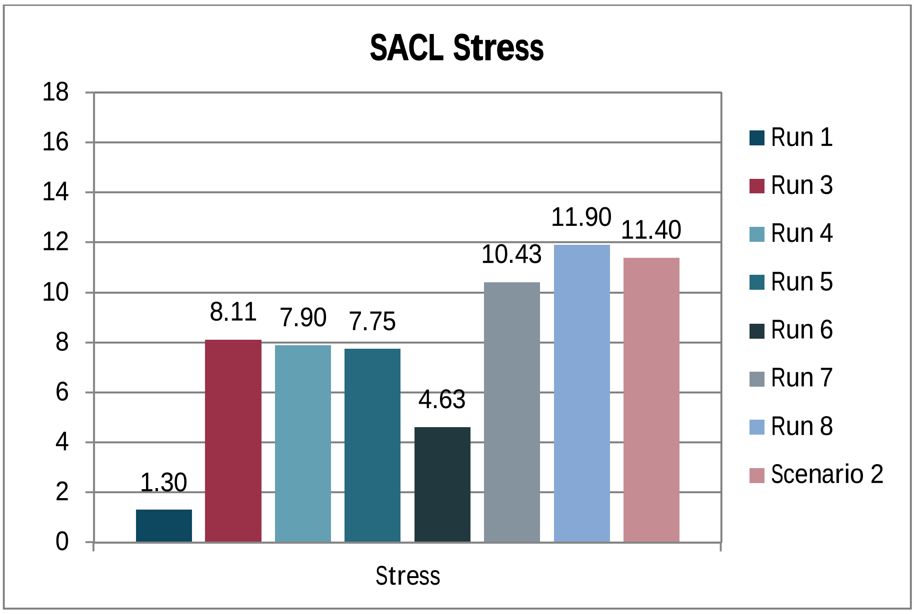


Figure 24: SACL stress rating of all pilots

4.3.6. Debriefings

Each Scenario 1 run was followed by a structured debriefing session conducted by an interviewer, in which pilots were invited to discuss around the following points:

- **Performance Self-assessment** – through a graphical representation called Performance Curve (see Figure 25), the pilots were asked to indicate how they felt and how they self-assessed their own performance during the run. Self-assessments could be expressed as single points onto the curve or as a line from one point to another. The graphical representation was an excellent tool to invite the pilots to discuss about their own performance perception, and

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

was used by facilitators to investigate why pilots positioned themselves onto a specific point and why they moved from one point to another (if happened).

- **Overall evaluation of the simulation** – to get feedback on how realistic was the simulation, if the event or sequence of events were surprising, if they behaved as they would have done in real life, and if they felt in control of the situation.
- **Performance overview** – to explore it they might have done something differently, if they had the feeling that the way they performed had an impact on the safety of the flight, and if they felt affected somehow by the previous run(s).
- **Performance / Possible Insights for design** – this session was used to understand if specific information, new instruments or re-designed HMI, and different ways to interact with the others in the cockpit may be beneficial to improve pilots' performance.
- **Behavioural markers** – the concept of behavioural markers was illustrated to pilots and discussed with them, to understand if, according to them, there are specific indicators in the cockpit that point out that their colleague's performance or their own one is declining. Internal and external markers were discussed, with the goal of collecting new markers to be included in the list started during the Workshop in Rome (cfr. D6.2 – Chapter 3.2) and analysing the ones collected by the facilitator during the simulation live observation.

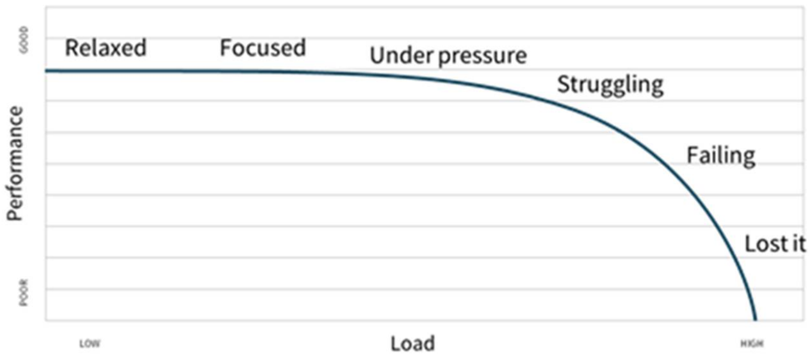


Figure 25: Performance Curve used during the debriefings to collect pilots' performance self-assessment

The benefits of debriefing conduction were twofold. From one side, it gave the opportunity to pilots to discuss about their experience and capitalise on it, learning from the simulation. From the other side, it was a chance for the simulation team to have a better understanding of what happened during each run, asking pilots to provide explanations on specific behaviours or actions undertaken, and to collect new data that might be used in the next project activities. From the debriefings, it is possible to extract information that could help in data interpretation.

The Performance Curve has proven to be an excellent tool to have an estimation of pilots' feelings and self-assessed performance. Together with "quantification-like" anchors for pilots' performance, the curve

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

was really good to travel through the sequence of events of the simulation and to support the discussion with pilots about the scenarios.

Outcomes from each debriefing are reported in Appendix A, while the analysis of Performance Self-assessment through the Performance Curve can be found in Section 4.4.

4.3.7. Discussion

Both results of NASA-TLX and ISA show that the runs in which workload was manipulated (e.g. Run 3, Run 4, Run 7 and Run 8) were the ones with the higher values, perceived as more mentally demanding than the others. This proves that we were able to manipulate this factor in the simulation. However, the lack of significant difference between run 3 and run 6 means that we were not able to completely isolate situation awareness and workload from each other. Thus, in the runs characterised by reduced situation awareness, a certain – undesired - component of workload was affected at the same time. Looking at SART analysis too, we can see how, a part from Run 7 and Run 8 (in which SA was manipulated), lack of SA was found in Run 4 as well (high workload condition). This probably means that in real-time realistic simulation is really difficult to create conditions able to influence workload without affecting situation awareness at the same time, and vice versa, or that is really difficult for pilots to discriminate the difference in the factors affected in each run. SACL results seem to confirm this trend.

On the other side, we can see that the cumulative effect of factors tends to be worse than the single factor effect in all the subjective assessment tools. Lack of significant difference from a statistical point of view may be due to the low number of subjects, and despite of that results are encouraging to prove the HPE concept. Also, results from Samn-Perelli scale prove that we were able to avoid undesired fatigue effects on simulation, thus limit the results of all the analysis to the effect of the three selected factors.

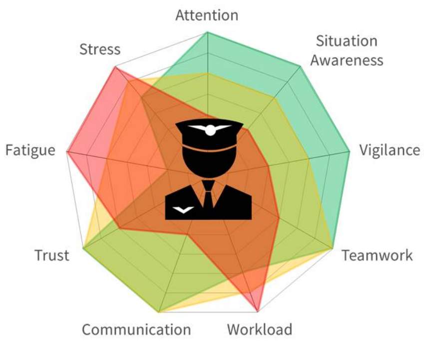
4.4. HPE curves / scaling concept

The Performance Curve was developed to provide evidences of the expected performance decrement related to the factor(s) manipulation, and to explore potential areas for recovery. As reported in D6.2:

"Initially the pilot is relaxed but alert, in the zone, the person who as a passenger we all want behind the controls. As we begin to 'push' individual elements (whether alone or in synchrony) the pilot will bring more cognitive resources into play to maintain performance. As we push further, there may be small decrements in performance – not yet safety significant – but noticeable to them or another pilot, and perhaps in the system performance measures we are going to use. The pilot is under pressure, but at the moment it's nothing they can't handle. As we push further, however, we would expect to see more significant performance decrements, perhaps errors of judgement or decision-making, missing key signals, less ideal flight performance, etc. If we can push them even further they will be aware they are failing. What may happen next is conjecture linked to stress/strain theory. If you stretch a piece of wire by adding

more weight, you obtain the curve as above. It is smooth as the wire stretches. However, at a certain point, the wire's properties become elastic, and for a short period as you add more weight, the function is no longer smooth but erratic. And then it snaps. [...] The point is, behaviour can become erratic, and the pilots may do something unexpected."

The performance curve translates the idea of the different levels of performance (Good, Acceptable, Degraded) that might be obtained through the simultaneous variation of different factors (see Figure 26).



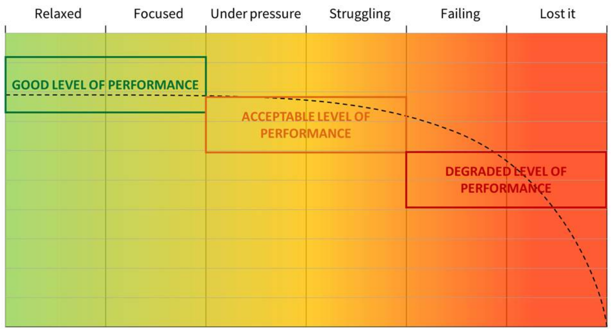


Figure 26: HPE Concept and HPE curve

The questions to be answered through the HPE curve are thus the following:

- Is the performance curve effectively able to track the performance degradation?
- Is the self-assessed performance different in one affected factor runs (e.g. Run 3-4-5-6) with respect to the multiple affected factors runs (e.g. Run 7-8)?
- Can we use the curve to scale the HPE concept?

4.4.1. Results

Position of pilots onto the curve visibly changed among runs with respect to three parameters:

- Overall position of the pilots, corresponding to the **Performance Centre** (median of the values touched during the run);
- The **range of the performance** (Max value – Min value), that gives the idea of the performance shift during the run;
- **Number and direction of movements**, which communicate what happened during the run and if there was a performance recovery or not.

All the runs were analysed and compared on the basis of these parameters. To compare the data collected through the performance curves, the different positions onto the curve were translated into discrete values (likewise a Likert scale). As several pilots also mentioned intermediate status, we decided to

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

consider the following association of values: Relaxed=1; Between Relaxed and Focused=2; Focused=3; Between Focused and Under Pressure=4; Under Pressure=5; Between Under Pressure and Struggling=6; Struggling=7; Between Struggling and Failing=8; Failing=9; Between Failing and Lost it=10; Lost it=11. Finally, the translated positions were integrated with the comments and explanations collected in the debriefing phase.

The results obtained per each run of Scenario 1 are the following:

- **Run 1.** Almost all pilots, apart from one, expressed their performance as one static point onto the curve – "*between focused and relaxed*", "*I was relaxed all the time*". For Pilot 1, Run 1 was the first run on the simulation, so he reported that his mental workload was higher at the beginning as he needed to familiarise with the simulator controls (side stick, thrust level and EPR) that were different from a real aircraft. Once familiarised with the simulator, his performance moved backwards "*from Focused to Relaxed*". The **Overall Run 1 Value** (calculated by the median performance centre of all pilots) is **1.5**.
- **Run 2**. As only three subjects performed this run, resulting data were not further analysed.
- **Run 3**. Performance assessment showed higher variability among pilots in this run than in Run 1. The majority of pilots (5 out of 9) expressed their performance as a static point onto the curve, and, among them, three reported they stayed in a "comfortable" position ("*I stayed a bit more than relaxed, in a condition of high arousal" – "it was basically just the approach, so I stayed more or less focused all the time*"), while the other two reported a bit more critical level of performance ("*Under Pressure, struggled with stability on vertical flight path*" – "*Under pressure, but I was not really confident with the behaviour of the simulator*[1]"). About the other four pilots, three of them moved from an acceptable level of performance to a critical one - shifting from the left to the right - due to the turbulence ("*Turbulences were really strong. At the very beginning of the approach I was always focused as it was a basic flight through turbulences. But at some point it increased very much and the more we get close to the ground the more I struggled because the margins started to reduce and I felt less in control of the situation*"), while one moved back and forth onto the curve, passing from focused at the beginning to failing when he instructed the go-around, to focused again and then again to under pressure ("*I was feeling like failing at the moment of go-around as I was target-fixated and wanted to land the aircraft, afterward focused. But then again almost under pressure, because I was not on a standard pattern when I re-tried the approach*"); for this pilot then, the go-around helped in recovery the performance, while for the other pilot who performed a go-around in Run 3 (Pilot 3) the performance was not affected by that procedure. The **Overall Run 3 Value is 3.5**.

[1] It was the first run of the simulation for this subject.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

- **Run 4**. Higher number of movements can be seen on Run 4 Performance Curve. The majority of pilots (7 out of 10) moved from an acceptable level of performance to a critical one (shifting from the left to the right). For all of them, this shift was due to the complexity of the combination of NDB approach and cross-wind ("*The lowest point was when I tried to prepare Non-Precision and I had to think what to do for NP flight, flying manually… a lot of stuff at the same time*" - "*Demanding. The most difficult part was the alignment with RWY track and then starting the descent, not easy to follow the NDB needle with heavy crosswind").* The other three pilots expressed their performance as a static point onto the curve, even if with different performance severity. Of these three, only one positioned himself in the "green" area of the performance due to his past experience with the type of approach[2] ("*Quite close to the focused phase, a bit in between with under pressure*"), while the other two assessed their own performance as critical ("*lot closer to struggling from the final approach onwards. I had problem in staying aligned with the runway and keeping the proper speed"* – "*Between Struggling and Failing, the scenario was definitively demanding and with many unexpected things.").* In this run, four pilots (Pilot 2, Pilot 3, Pilot 8 and Pilot 9) performed a Go-Around, but for all of them this didn't have any impact nor on their performance self-assessment neither in recovery the situation. The **Overall Run 4 Value** is **4**.
- **Run 5.** The majority of the pilots who performed Run 5 (6 out of 8) expressed their performance as a static point in the acceptable area of the performance curve, with only two pilots struggling in certain phases of the run. In particular, 6 pilots out of 8 assessing their own performance as acceptable; almost all of them reported to be Relaxed or Focused all the time, while the one who positioned himself between focused and under pressure had Run 5 as first run of the simulation. The other two pilots reported a first shift from left to right (until struggling), followed by a movement backwards towards an acceptable level of performance. For one of these two the situation became critical when he realised about the low fuel situation, struggled while deciding what to do next, and improved – and the performance recovered - when he realised the Go-Around was not needed. The other pilot suffered the lack of fuel as well, plus the combination with the loud noise. The situation improved when the noise ended. No one performed a Go-Around in this run. The **Overall Run 5 Value** is **3**.
- **Run 6.** The self-assessment was quite scattered in Run 6. The majority of pilots (4 out of 8) referred to be static onto one position of the curve, three of them mentioned to be focused and one between focused and under pressure during the entire run. Then, one pilot out of 8 moved from relaxed to between focused and under pressure ("*Warning about the localiser being bad, increased workload a bit, but all in an area where it was easy to fix*"), one pilot moved from right to left (from struggling to focused) because of the lack of confidence with the simulator, and two

[2] Pilot 9 mentioned he did the NDB approach several times

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

pilots moved back and forth, starting in the relaxed/focused area, than shifting to the right towards struggling, and finally moving back to the relaxed/focused area again. During the debriefings, they explained that those movements depended on the sequence of events they encountered in this run (localiser interference, wind shift) and once passed those events the situation naturally improved ("*As soon as we were on the glide and on the localiser, stabilised, I went back to focused*"). In this run, only one pilot (Pilot 9) performed a Go-Around, without any impact on his self-assessed performance. The **Overall Run 6 Value is 3**.

- **Run 7.** Unfortunately Run 7 suffers a lack of data that make difficult to derive definitive conclusions on its results. However, despite only 7 pilots out of 10 performed this run, the HPE curve showed a natural shift towards the right end. In fact, only 1 pilot out of 7 stated that his performance remained stable in the good performance area of the curve (between Focused and Under Pressure for the whole run) despite the Go-Around. Then, for two pilots the performance changed but remained acceptable ("*I was under pressure but I didn't have the feeling I degraded my performance, the performance was still there despite the load of work*"), while for other two it passed from a good level to an almost degraded one, with one pilot that showed a severely degraded performance - Lost it - during the landing phase (*"I felt like being a "bit behind", I was losing the overview of the whole situation. I was not able to carry on secondary tasks, like calculating better wind impact in my mind. I was quite close to losing it").* Finally, 2 pilots out of 7 moved backward on the curve, one from the acceptable level of performance to a good level of performance (Pilot 7), the other moving a bit from almost struggling to Under pressure (Pilot 5). Pilot 5 reported that he improved his performance after the Go-Around ("*After the go around the weather changed, the wind was better so it was easier*"), while for the other the situation changed when the localiser interference stopped. A part from Pilot 5, Go-Around was performed by other three pilots (Pilot 1, Pilot 4 and Pilot 6) without any impact on their self-perceived performance. Despite the lack of data with respect to the other runs, the **Overall Run 7 Value** is quite high, calculated as **5**.
- **Run 8.** Lot of movements onto the curve can be observed in this run. Only 2 pilots out of 10 remained stable during the whole run, one between Focused and Under pressure (borderline between good and acceptable level of performance) and the other Struggling almost all the time. One pilot (Pilot 5) moved back from an acceptable level of performance to a good one, basically when the weather improved. Six pilots out of 10 felt their performance shifted from the left to the right. Among them, one moved towards the right, but remained in the area of a good performance ("*Everything was almost as I expected so when we declared the emergency we get a direct radar for the final. I would say that was not more than Focused*"), two seem to remain in the acceptable level of performance, uncomfortable or struggling but still in control of the situation, and the other three pilots lost the control of the situation, ending in the degraded area of human performance ("*The low point was very short final, the last 10-15 seconds, was a speed drop, then got too high. Too little time to do a proper instrument check" – "I did not realise the wind shift – many things happened at the same time! Fuel was my biggest concern").* Finally, 1

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

pilot out of 10 mentioned he moved several times back and forth onto the curve, in the area between Focused and Under Pressure/Struggling ("*I started with radar vector for direct approach, so I was probably under pressure; then I moved back to focused because I had a plan, but in the approach itself the localiser went off and I went between under pressure and struggling, and we made the missed approach. However, as everything went quite fast we realised that we had enough fuel for a second approach and even for a third one, so after the completion of the go around and the new approach we started to get back in between under pressure and focused*"). A total of three Go-Arounds were performed in this run, and only one helped the pilot to recover his performance. The **Overall Run 8 Value is 4.75**.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

All the performance values in terms of Performance centres, range, movements and directions are reported in Table 13.

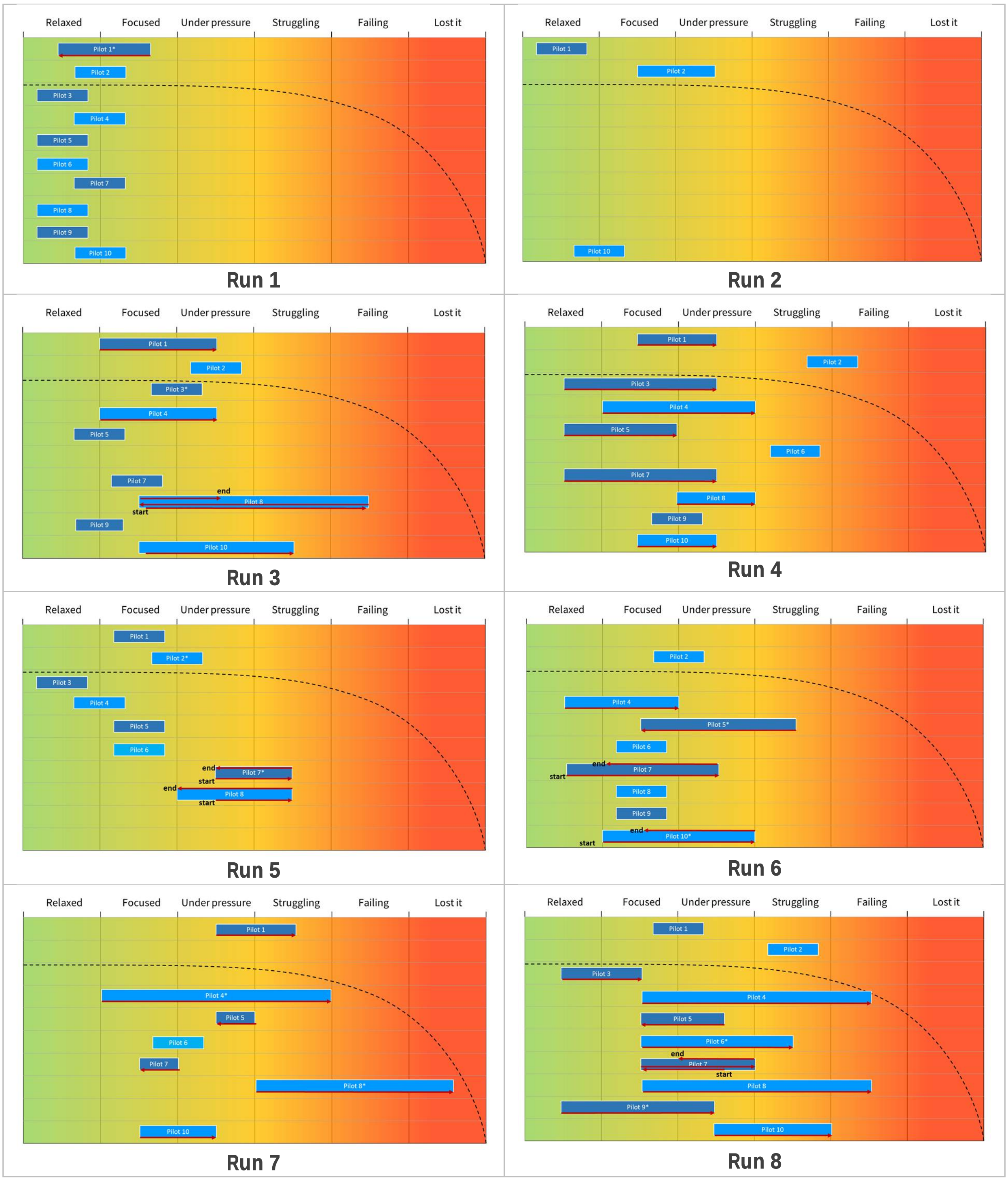


Figure 27: Summary of all the Performance curves for Scenario 1

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Table 13: Summary of all the Performance values for Scenario 1

Subject	Performance centre	Performance Range	Worse Performance	N° of movements	Main direction	Performance recovery
RUN 1						
S1	**2**	2	3	1	Backward ←	Yes
S2	**2**	0	2	0	n.a.	n.a.
S3	**1**	0	1	0	n.a.	n.a.
S4	**2**	0	2	0	n.a.	n.a.
S5	**1**	0	1	0	n.a.	n.a.
S6	**1**	0	1	0	n.a.	n.a.
S7	**2**	0	2	0	n.a.	n.a.
S8	**1**	0	1	0	n.a.	n.a.
S9	**1**	0	1	0	n.a.	n.a.
S10	**2**	0	2	0	n.a.	n.a.
RUN 3						
S1	**3,5**	3	5	1	Forward →	No
S2	**5**	0	5	0	n.a.	n.a.
S3	**4**	0	4	0	n.a.	n.a.
S4	**3,5**	3	5	1	Forward →	No
S5	**2**	0	2	0	n.a.	n.a.
S6	**n.a.**	n.a.	n.a.	n.a.	n.a.	n.a.
S7	**3**	0	3	0	n.a.	n.a.
S8	**5**	6	9	3	Forward →	Yes
S9	**2**	0	2	0	n.a.	n.a.
S10	**5**	4	7	1	Forward →	No
RUN 4						
S1	**4**	2	5	1	Forward →	No
S2	**8**	0	8	0	n.a.	n.a.
S3	**3**	4	5	1	Forward →	No
S4	**4**	4	6	1	Forward →	No
S5	**2,5**	3	4	1	Forward →	No
S6	**7**	0	7	0	n.a.	n.a.
S7	**3**	4	5	1	Forward →	No
S8	**5**	2	6	1	Forward →	No
S9	**4**	0	4	0	n.a.	n.a.
S10	**4**	2	5	1	Forward →	No
RUN 5						
S1	**3**	0	3	0	n.a.	n.a.
S2	**4**	0	4	0	n.a.	n.a.
S3	**1**	0	1	0	n.a.	n.a.
S4	**2**	0	2	0	n.a.	n.a.
S5	**3**	0	3	0	n.a.	n.a.
S6	**3**	0	3	0	n.a.	n.a.
S7	**6**	2	7	2	Back/Forth←→	Yes
S8	**5,5**	3	7	2	Backward ←	Yes
S9	**n.a.**	n.a.	n.a.	n.a.	n.a.	n.a.
S10	**n.a.**	n.a.	n.a.	n.a.	n.a.	n.a.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Subject	Performance centre	Performance Range	Worse Performance	N° of movements	Main direction	Performance recovery
RUN 6						
S1	**n.a.**	n.a.	n.a.	n.a.	n.a.	n.a.
S2	**4**	0	4	0	n.a.	n.a.
S3	**n.a.**	n.a.	n.a.	n.a.	n.a.	n.a.
S4	**2,5**	3	4	1	Forward →	No
S5	**5**	4	7	1	Backward ←	Yes
S6	**3**	0	3	0	n.a.	n.a.
S7	**3**	4	5	2	Forward →	Yes
S8	**3**	0	3	0	n.a.	n.a.
S9	**3**	0	3	0	n.a.	n.a.
S10	**4**	4	6	2	Forward →	Yes
RUN 7						
S1	**6**	2	7	1	Forward →	No
S2	**n.a.**	n.a.	n.a.	n.a.	n.a.	n.a.
S3	**n.a.**	n.a.	n.a.	n.a.	n.a.	n.a.
S4	**5**	6	8	1	Forward →	No
S5	**5,5**	1	6	1	Backward ←	Yes
S6	**4**	0	4	0	n.a.	n.a.
S7	**3,5**	1	4	1	Backward ←	Yes
S8	**8,5**	5	11	1	Forward →	No
S9	**n.a.**	n.a.	n.a.	n.a.	n.a.	n.a.
S10	**4**	2	5	1	Forward →	No
RUN 8						
S1	**4**	0	4	0	n.a.	n.a.
S2	**7**	0	7	0	n.a.	n.a.
S3	**2**	2	3	1	Forward →	No
S4	**6**	6	9	1	Forward →	No
S5	**4**	2	5	1	Backward ←	Yes
S6	**5**	4	7	1	Forward →	No
S7	**4,5**	3	6	3	Backward ←	Yes
S8	**6**	6	9	1	Forward →	No
S9	**3**	4	5	1	Forward →	No
S10	**6,5**	3	8	1	Forward →	No

4.4.2. Conclusion

The position of pilots on the Performance curve indicated how they felt with respect to their performance during the execution of the different runs. Despite they had never used such a tool, pilots proved to be particularly accurate in positioning themselves onto the curve to describe their feeling and their performance during each run. A summary of the overall self-assessed performance, on average, per each pilot on each run is illustrated in the following figure (Figure 28).

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

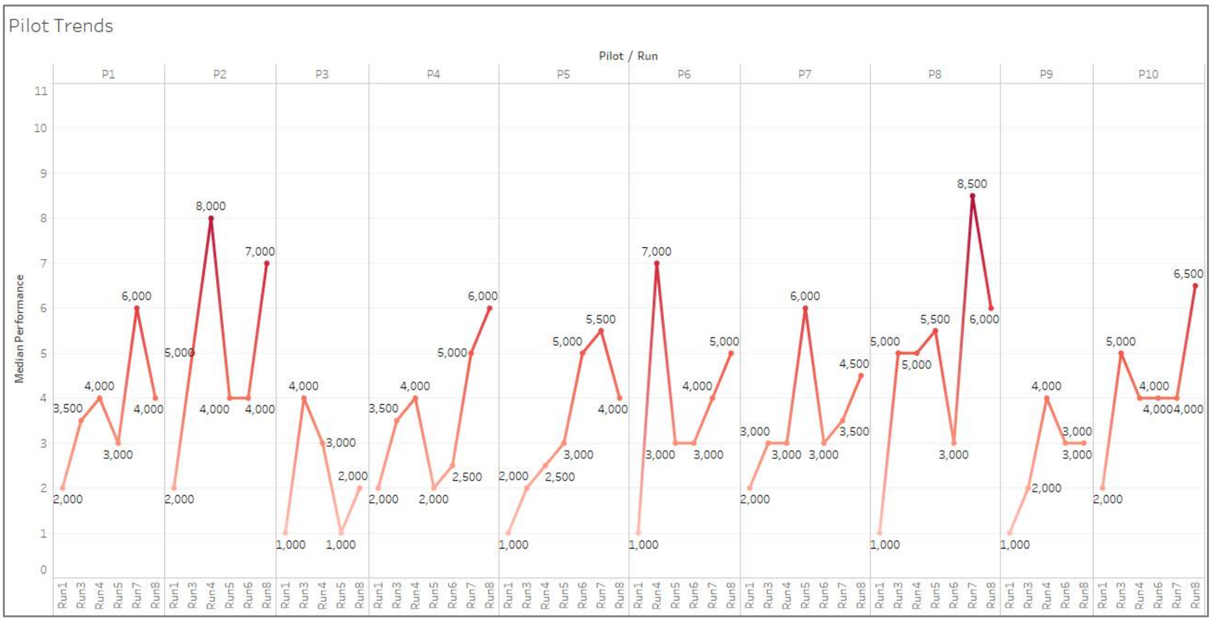


Figure 28: Pilot self-assessed performance trends per each run

If we compare the Overall values of all the runs (based on the central value of the overall performance on that run - see Figure 29) we can observe how the self-assessed performance is, on average, worse in Run 7 and Run 8 than in the other runs (with 5= Under Pressure). Despite the lower number of subjects that performed Run 7, the median values expressed by pilots for this run are higher than the single factor runs. Run 7 (medium level of variation for all the three factors) was thus perceived by pilots as more demanding than the high workload or high stress condition.

A statistical comparison between the Overall centre values of performance of Run 1, Run 3 and Run 7 (Mann Whitney test) shows that Run 1 significantly differ from Run 3 (statistic=5.0, p value=0.0004) and Run 7 (statistic=0.0, p value=0.0002), but also that Run 7 is statistically different from Run 3 (statistic=13.5, p value=0.0301).

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Figure 29: Overall run values comparison

Run 7 and Run 8 also shows an higher average variability (i.e. difference between minimum value and maximum value expressed onto the curve, for all pilots – see Figure 30) than the other runs, meaning that pilots performance moved more in Run 7 and Run 8 compared to the others. This difference is even more visible in Figure 31, where the average best performance position is compared to the average worse performance position per each run. While in the baseline the pilots stayed static, in the two workload runs (Run 3 and Run 4) they shifted onto the right, still remaining in the area that we defined "Acceptable level of performance". From the figure it can be noticed that in Run 7 not only the average worse performance is higher, but best performance as well compared to the others. This means that since the very beginning of the run the pilots felt they were challenged by the simulated events, and their performance was poorer since the beginning, until almost exceeding the limits of the acceptable performance (as it actually happened for two pilots).

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

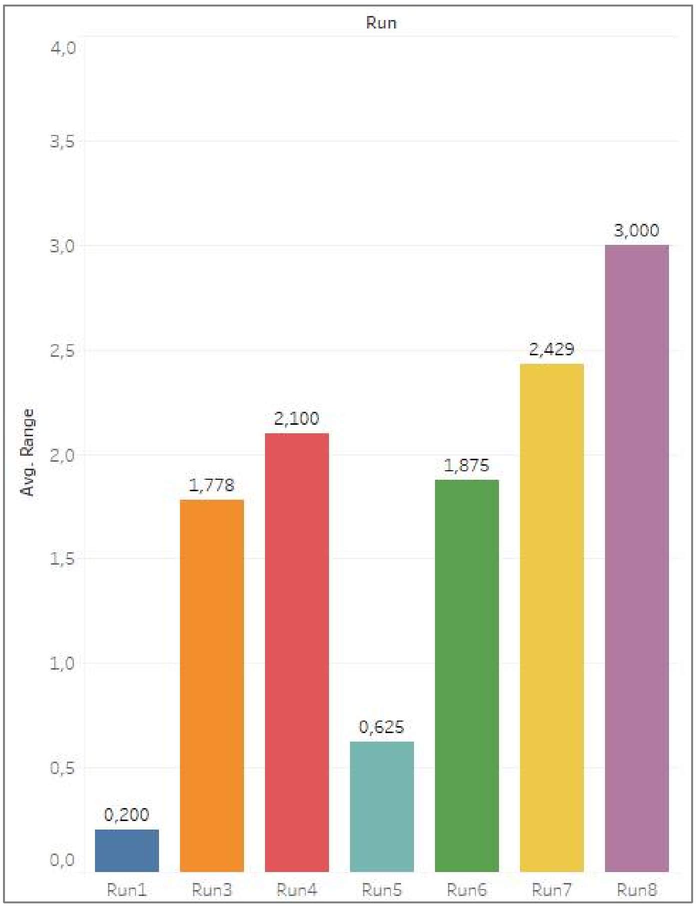


Figure 30: Average Variability per Run

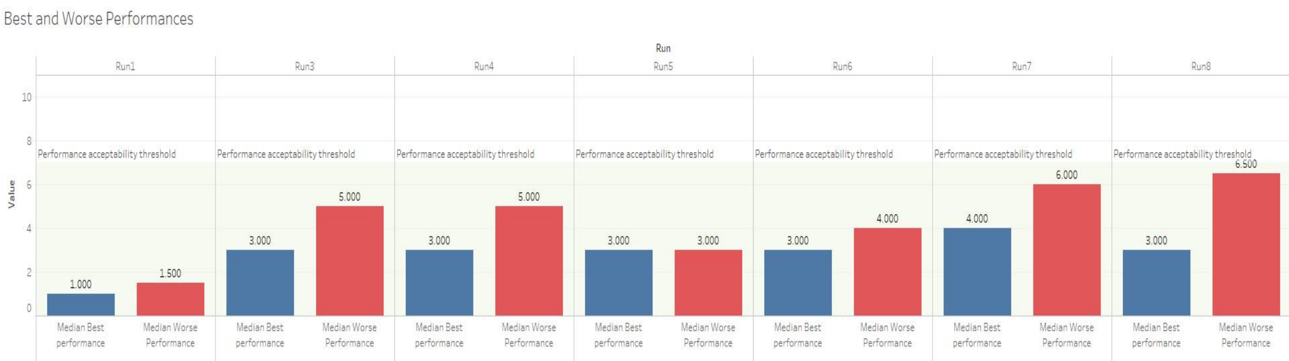


Figure 31: Comparison between Average best performance and average worse performance per Run

Finally, we analysed the distribution of frequency for the baseline, Run 3, Run 4, Run 7 and Run 8. The frequency was calculated by counting the number of times pilots passed on that specific point of the curve. If we look at the distribution of frequency in the single factor conditions (Run 3 and Run 4) compared to the multiple factors conditions (Run 7 and Run 8) we can see how the multiple factors conditions induced a shift in the self-assessed performance towards the right end of the curve, with higher number of pilots that passed the under pressure condition and felt struggling during the execution

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

of the runs. By limiting this comparison to Run 1, Run 3 and Run 7 (the key ones to prove the HPE concept) the shift towards the right end of the curve is even more evident (see Figure 33).

	Performance Value										
Run	Relaxed	Rel/Foc	Focused	Foc/Press	Under Pressure	Press/Str..	Struggling	Strug/Fail	Failing	Fail/Lost	Lost it
R1	6	5	1								
R3		4	6	7	7	3	3	2	1		
R4	3	4	6	8	6	2	1	1			
R7		1	3	4	4	4	3	2	1	1	1
R8	2	2	7	9	9	5	5	3	2		

Figure 32: Distribution of frequency for Run 1 – Run 3 – Run 4 – Run 7 – Run 8

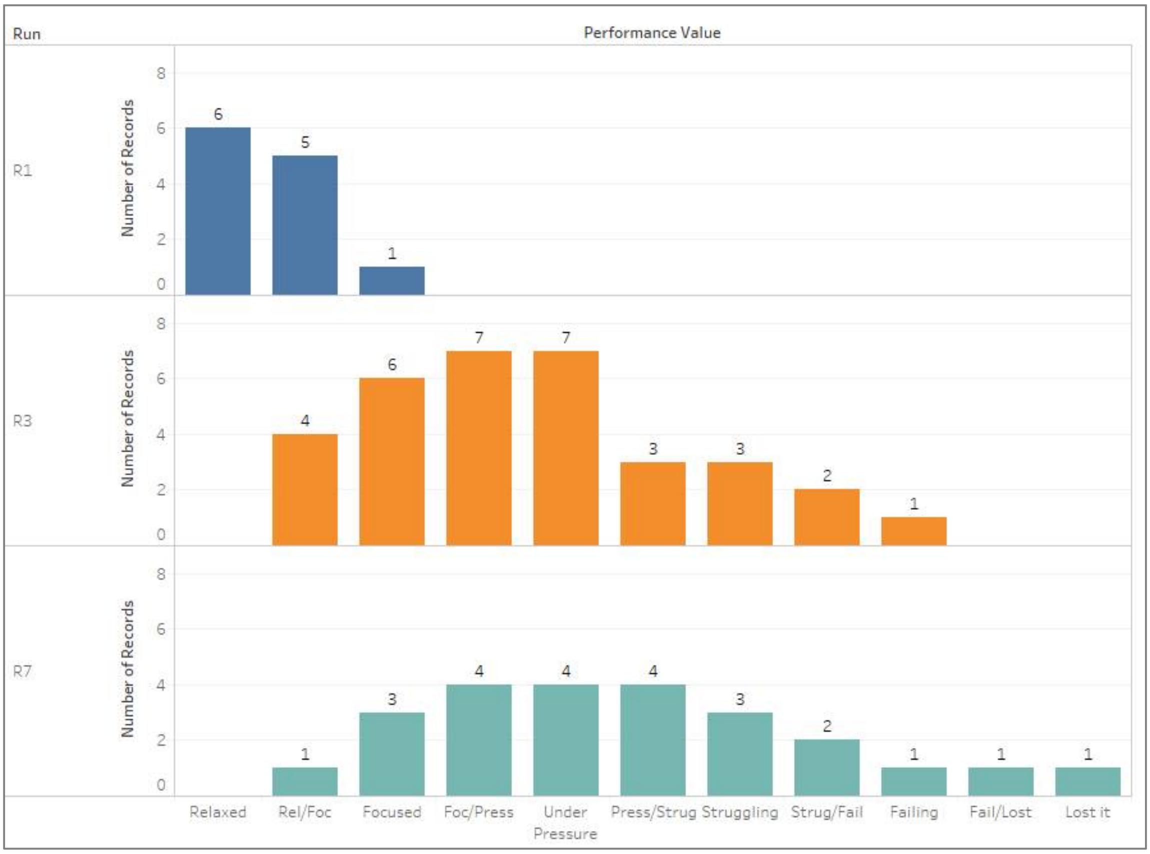


Figure 33: Distribution of frequency for Run 1 – Run 3 – Run 7

From all these analysis, it seems that **performance in Run 7 (medium WL - medium Stress – medium SA) was perceived as more degraded than performance in Run 3 (high WL condition).** Again, unfortunately definitive conclusions cannot be derived due to the low number of subjects for Run 7, but the trend is

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

definitively interesting and new data should be collected to confirm it. Also, the self-assessed performance has to be compared with data collected on the actual performance (i.e. simulator data from manual flying), to check if an improvement or a decrease in the perceived performance corresponds to a variation into simulator data and to see how good pilots were in evaluating the way they acted. Triangulation of data between HPE curve position and actual performance can also help in understanding if the HPE curve can be used in isolation as well to discriminate different levels of performance among runs.

4.5. Physiological Parameters

The CSEM system (Chételat et al., 2011, 2015) for monitoring the physiological parameters (smart vest) is based on the use of three cooperative sensors, as illustrated in Figure 34. The different electrodes (A, I and S) shown in Figure 34 allows the measurements of the following signals:

- Two electrocardiograms (ECG) leads;
- A transthoracic bio-impedance;
- Skin temperature;
- Accelerometer;
- Multi-channel photoplethysmograhy (PPG).

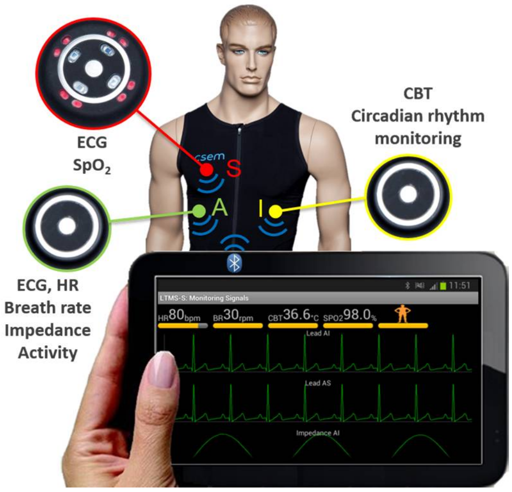


Figure 34: CSEM smart vest, connected to a tablet to indicate different physiological parameters in real time

From these signals several physiological measures can be estimated:

- **R-R intervals**, the time elapsing between two consecutive R waves in an ECG. This measure is often given in milliseconds;

- **Heart rate (HR)**, the number of contractions (beats) of the heart per minute (bpm). In other words, the speed of the heartbeats;
- **Heart rate variability (HRV)**, is the variation of the NN intervals (normal R-R intervals). This variability can be estimated by utilizing different methods. There are three major methods used in this study to extract the corresponding features:
  - Time-domain methods:
    - **SDNN**, standard deviation of the NN intervals and often described as the total variability;
    - **RMSSD**, the square root of the mean squared difference of successive NNs.
  - Frequency-domain methods; by analysing different bands of frequency, interesting physiological measures can be extracted:
    - **HF**, High frequency (0.15 – 0.4 Hz). Describes the parasympathetic modulation;
    - **LF**, Low frequency (0.04 – 0.15 Hz). Describes the sympathetic modulation;
    - **VLF**, very low frequency (0.0033 – 0.04 Hz);
    - **LF/HF**, “Autonomic” balance. The balance between the sympathetic and parasympathetic activity.
  - Geometric methods; these methods are used to analyse the NN intervals by converting them into a geometric pattern:
    - **Mode,** the range of the most frequently detected values of the NNs;
    - **Amplitude mode**, the percentage of NNs with the most frequent duration.
- **Breath rate (BR)** or respiration rate, counted as number of breaths per minute;
- **Activity classification** includes: lying, sitting/standing, or walking;
- **Perfusion index**, describe the process of supplying blood to the tissues, and is obtained in percentage.

Given the data collected and the quality of the signals, the most promising results were obtained with the ECG derived measures. Therefore, **heart rate** and **heart rate variability** features were the measures selected to show the variation of each one of the factors presented in this study. This has also been confirmed in a review study by Straussberger & Schaefer (2004).

4.5.1. Results

This section begins with a single pilot analysis (section 4.5.1.1), where pilot 2 was selected since he/she performed almost all the runs and in addition the physiological response showed to be very promising. The second part of this section (4.5.1.2) shows the results that were derived from the group analysis that was conducted using two different approaches. The first analysis was taking into consideration the entire run, whereas the second approach only phase 2 in each run was exploited.

Future Sky Safety has received funding from the EU’s Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

4.5.1.1. Single pilot analysis

Figure 35, Figure 36 and Figure 37 show the changes of HR values during the Run 3, Run 4 and Run 8. The different colour codes highlight the different flight phases. Phase 1 begins at the start of the run and ends at the *TOD glideslope*. Phase 2 bridges between the *TOD glideslope* and the *decision altitude.* Phase 3 on the other hand side spans from the *decision altitude* to the end of the run. The duration of phase 3 varies very much between one run and another since it will all depend on whether the pilot decides to do a *go around* or not.

Figure 35 illustrates how the HR of pilot 2 increases during the Run 3. It started out with a HR of 87 bpm and reaches 104 bpm by the end on the entire run. However, the maximal HR response was obtained during phase 3 and the pilot hit a maximum HR value of 113 bpm.

In Run 4, the amount of workload was increased from high to very high workload. This is clearly expressed in Figure 36 where the pilot found him/herself in a situation where the decision of a *go around* was taken. Phase 3 started out with the *go around* and the pilot reached a peak HR of 132 bpm, while the peak during phase 2 was at 118 bpm.

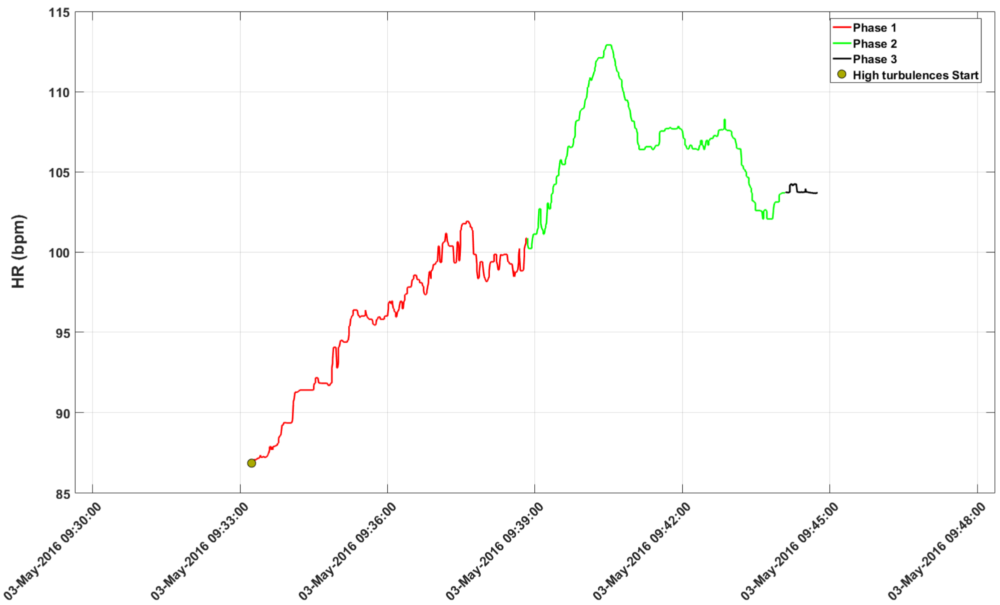


Figure 35: The changes of HR of pilot 2 during Run 3 (high workload). The 3 different flight phases are highlighted in different colours

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

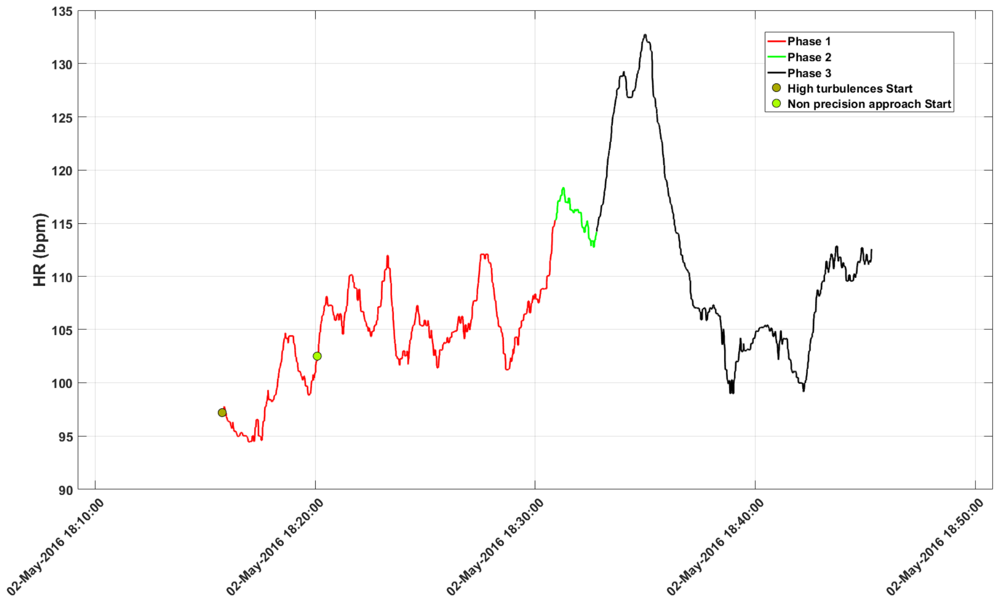


Figure 36: The evolution of the HR values of pilot 2 during Run 4 (very high workload)

While Run 3 and Run 4 consisted of an increased workload in different degrees, Run 8 incorporates also high stress and reduced SA in addition to the high workload. Figure 37 demonstrates the HR values during Run 8 for pilot 2. Besides of having the different phases highlighted in different colours as in Figure 35 and Figure 36, the scenario events (e.g. loud noise, wind shift etc.) are tagged as well. The physiological response was less evident compared to Run 4 were the pilot had taken a decision to do a go around. However, the HR reached a value of 116 bpm during phase 2 which is nearly the same as in the case of Run 4 (118 bpm).

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

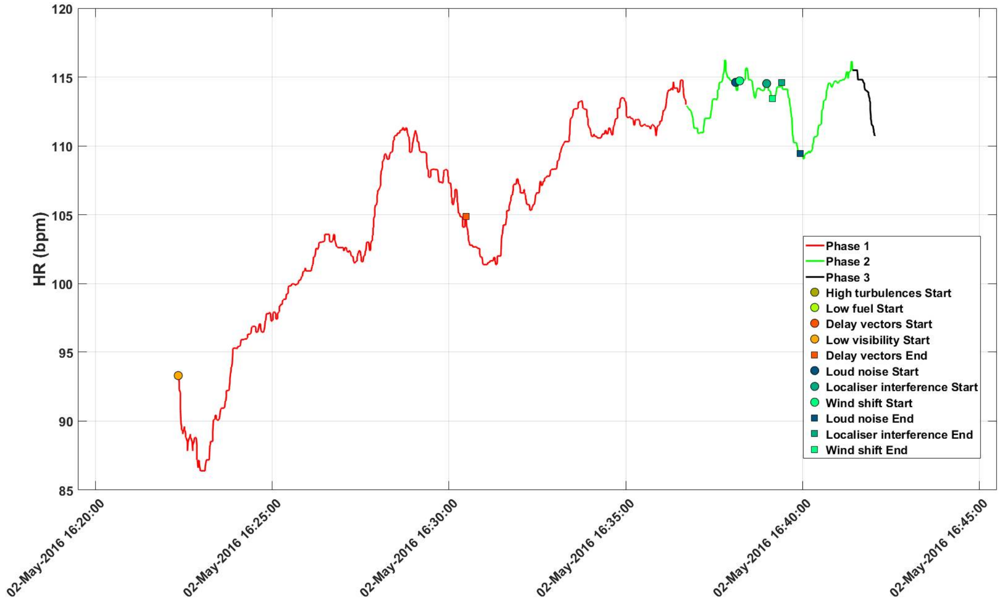


Figure 37: The HR evolution of pilot 2 during Run 8 (high workload, high stress, high reduced SA)

Figure 38 summarizes the HR distributions during all the runs for pilot 2. A nice increase in HR is observed when the workload is increased (Run 2 to Run 4). The high stress run (Run 5) shows a very wide distribution of HR values which indicates that the stress added in this run did not have an impact on the HR of this pilot. A Kruskal-Wallis test showed that all the test between each pair of runs were significant ($p>0.05$) besides Run 2 vs Run 3 and Run 6, Run 6 vs Run 3 and lastly Run 8 vs Run 4.

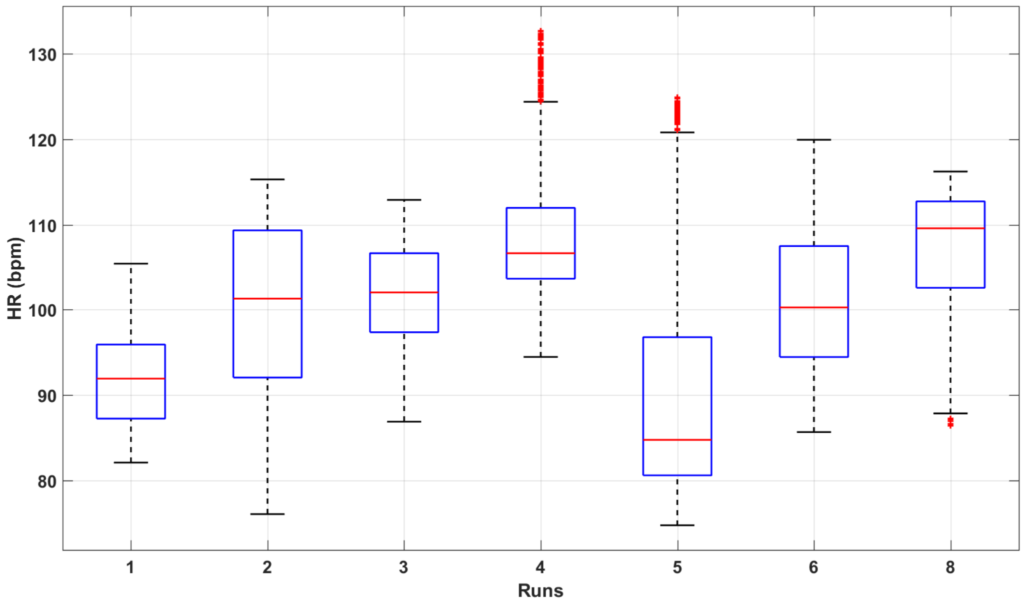


Figure 38: HR distributions of pilot 2 of all the runs

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

The distributions of the HRV features (SDNN, HF, LF and VLF) are shown in Figure 39. The SDNN (Figure 39 A) shows the same trend as the HR but inversely (a decrease of SDNN with added workload). Run 2, Run 5 and Run 6 showed no significant difference when compared to Run 1. The medium workload (Run 2), high stress (Run 5) and reduced situation awareness (Run 6) seem to have no impact on the SDNN (at least not significantly). On the other hand side, the analysis of the HRV features that were extracted from the frequency domain showed interesting results when it came to Run 5 (high stress), in particular HF. The Run 6 remains difficult to interpret due to the large distribution, there is always a trend of decrease but it is not significant.

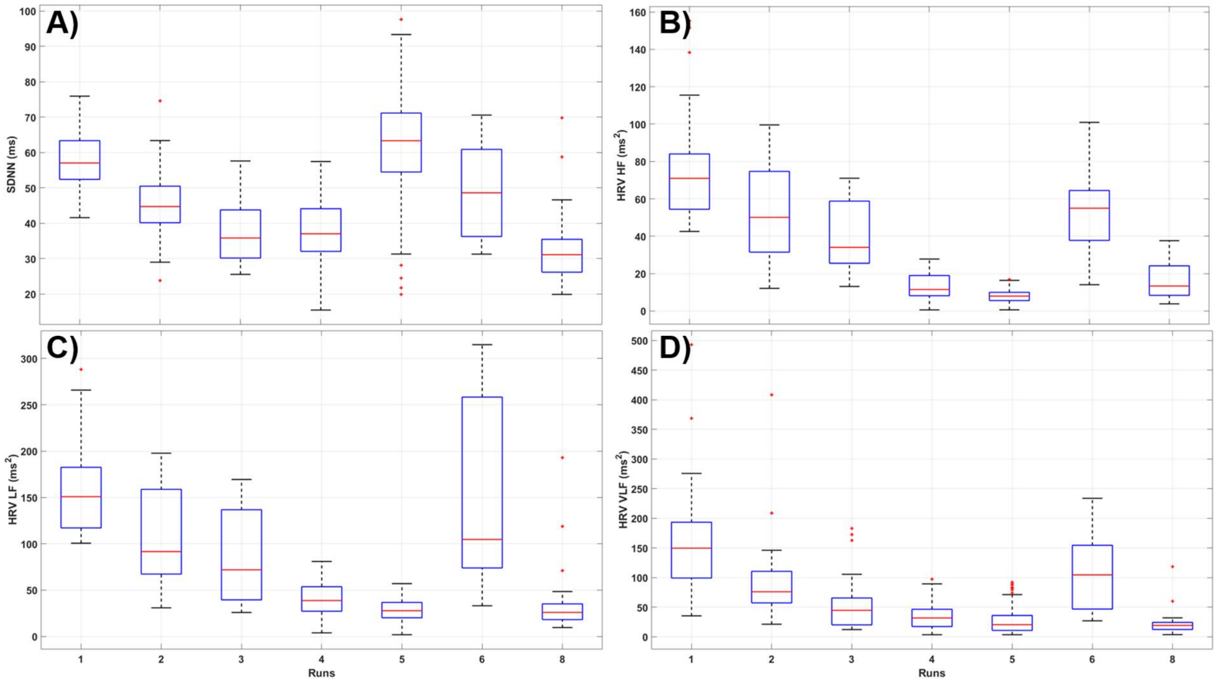


Figure 39: Distributions of HRV features (SDNN, HF, LF and VLF) of pilot 2 during 7 different runs

4.5.1.2. Group analysis

In terms of group analysis, the number of pilots used in this analysis is shown in

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Table 14. Run 2 was excluded from the group analysis since the number of pilots was too low. The analysis was performed on the features in their original form and in a normalised form. The initial hypothesis was that, the normalization would decrease the inter-subject variability and the physiological response would be more prominent. All the runs were normalised by the baseline (Run 1) with the assumption that the pilot was not affected by the tasks given to him/her during Run 1.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Table 14: The total number of pilots in every run

	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Run 7	Run 8
Nbr of Pilots	7	2	7	6	3	5	3	5

Analysis of the entire Runs

Figure 40 shows the distribution of the HR values for all the pilots. The same trend is observed in this Figure as in Figure 38. The HR was increased with an increase of workload, while the stress factor (Run 5) decreased. Run 7 is hard to interpret but this is not surprising since the number of pilots used is too low. In terms of statistical difference, Run 7 was not significant when compared to Run 1, when the HR values were not normalised. While the normalised HR values showed a no significant difference between Run 6 and Run 8.

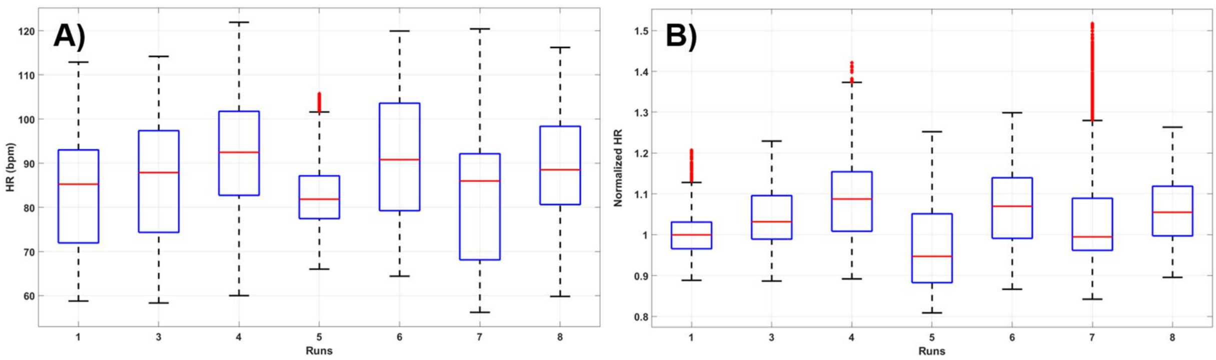


Figure 40: The changes of HR for all the pilots during each run

The SDNN (Figure 41) shows a decrease when workload is applied alone. Therefore, Run 3, 4 and 8 were significant (p>0.05) when compared to the baseline (Run 1). The same runs, plus Run 6 were also statistically significant when compared to Run 5. Equivalent results were obtained with normalised SDNN (Figure 41 B), were Run 7 was also significant compared to Run 5.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

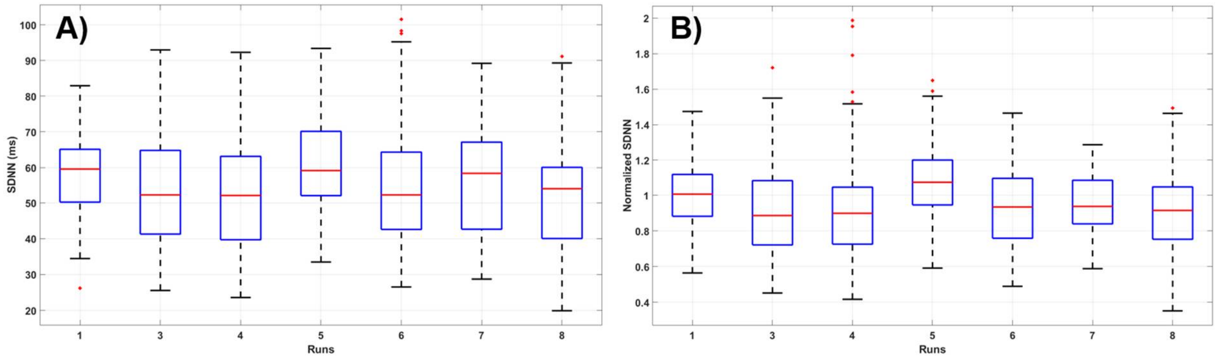


Figure 41: The changes in SDNN of all pilots during all runs when SDNN is not normalised and normalised

As in the single pilot analysis (section 4.5.1.1), the HRV features derived from the frequency domain showed to be sensible to Run 5 (high stress) as well. Figure 42 A) show the distribution of the HF feature, the only run that was not significantly different to the baseline was Run 6 (High reduced SA). Run 3 was not significant different to runs with increased workload or other added factors (Run 4, Run 7 and Run 8). Run 4 was also not significant to Run 7 and Run 8 and in addition Run 7 and Run 8 were not significant different between them. The comparison between Run 6 and Run 7 are weakly significant (p=0.048). Same results were obtained when HF was normalised (Figure 42 B) with even stronger p-values.

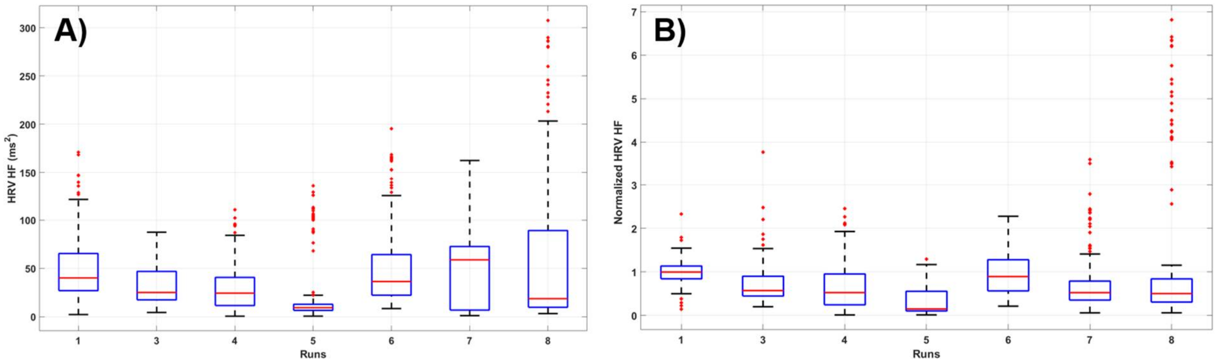


Figure 42: The distribution of HF values of all the pilots during each run. A) Not normalised and B) normalised

When considering the LF feature (Figure 43), the pair of runs which were not significant when compared to each other were Run 3 vs Run 6 and Run 7, Run 4 vs Run 7 and Run 8. The runs which had a weak significant difference are; Run 4 vs Run 8 (p=0.039), Run 6 vs Run 7 (p=0.042) and Run 7 vs Run 8 (p = 0.043). These three last combination showed to be not significant different between them when the features were normalised (Figure 43 B). Otherwise the statistical differences were the same as in the non-normalised case.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

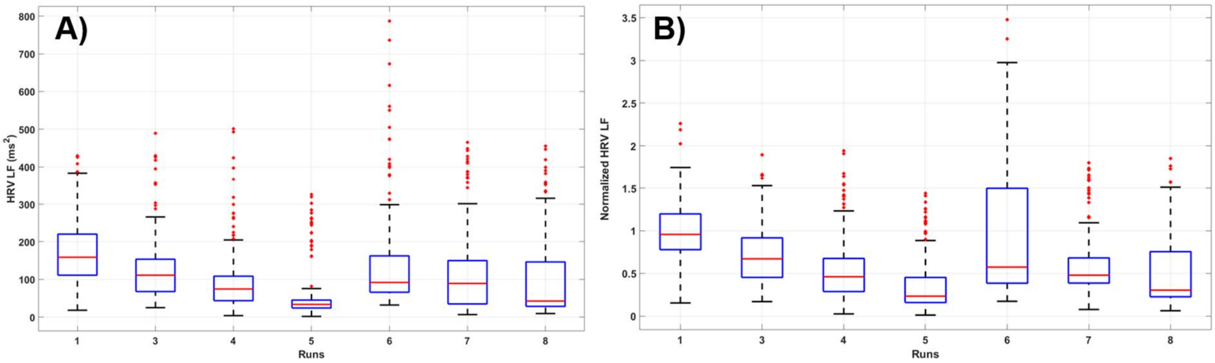


Figure 43: The distribution of the changes in LF of all pilots during each run. A) Not normalised and B) normalised.

Figure 44 shows the same tendency as the remaining previous features which were derived from the frequency domain. Stress (Run 5) is the factor which shows the highest significant difference when compared to the baseline (Run 1). Run 7 was not significant different when compared to Run 3, Run 4 and Run 6. Run 8 was also not significant different to Run 4 and Run 5. The last pair of runs that were not significant is Run 3 and Run 6. The same results were obtained when the VLF was normalised (Figure 44 B).

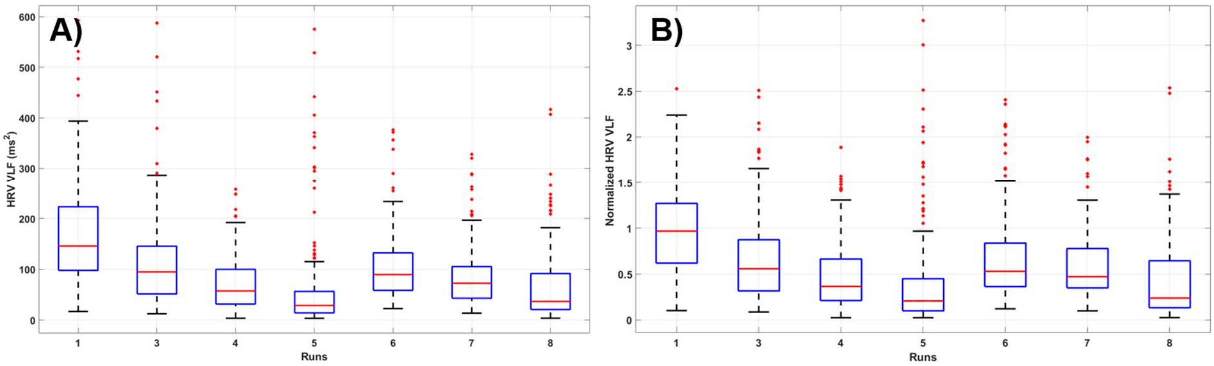


Figure 44: The distribution of VLF value of all pilots during each run. A) Not normalised and B) normalised

Analysis of phase 2 in each Run

While the analysis in the previous subchapter was based on the entire run, in this section the analysis was concentrated on phase 2. As in the previous subchapter, the results are presented when the features are not normalised and when they are. The normalization procedure is the same as previously, which means using Run 1 as baseline. Additional normalization methods were explored in this analysis, for instance using phase 1 to normalize phase 2. However, these results are not shown here since the outcome did not bring any added value to the analysis.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

The first physiological feature analysed in this section is the HR, Figure 45 shows the distribution of the HR values. All pair of runs were significant different from each other beside Run 3 vs Run 8 in the non-normalised case. Run 3 vs Run 7 were weakly significant when the HR was normalised (Figure 45 B).

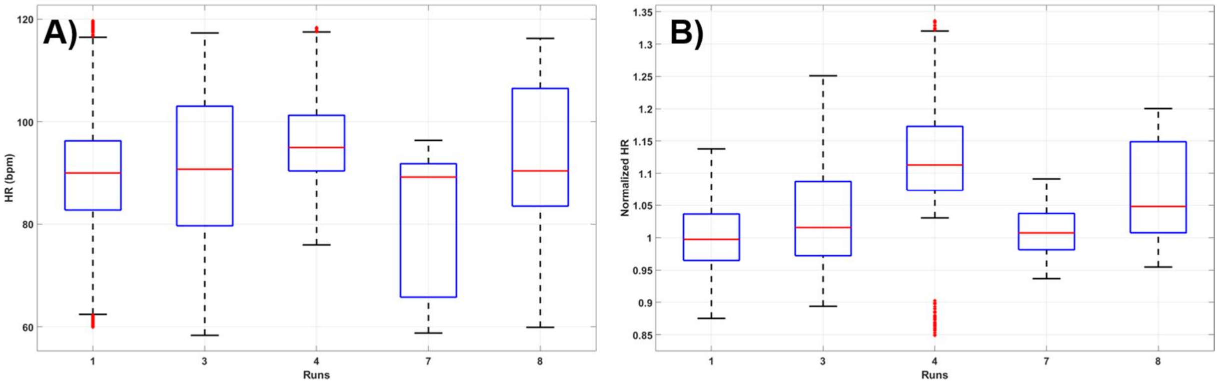


Figure 45: The distribution of HR values of all pilots during phase 2 of each run. A) Not normalised and B) normalised

The HRV feature that was derived from the time domain, SDNN (Figure 46), showed a significant difference when the workload alone was added to the baseline. Therefore the only pair of runs that were significant different are Run 1 vs Run 3 and Run 4. Run 7, which contained a mix of the factors, was significant different to Run 4, but only when the SDNN was not normalised (Figure 46 A).

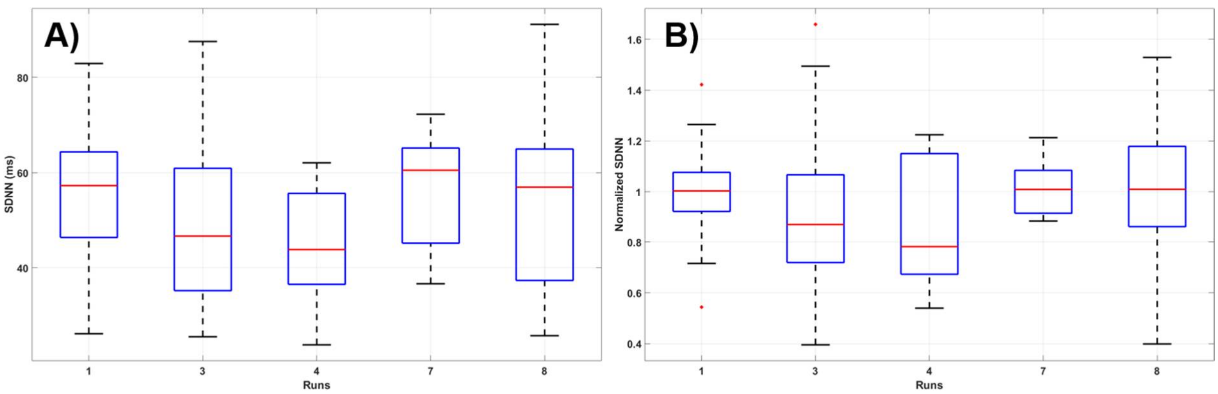


Figure 46: The distribution of the SDNN for all the workload runs during phase 2. A) Not normalised and B) normalised

The HF showed almost no significant difference between the runs when we analyse phase 2 (only Run1 vs Run 4 was significant, however weak p = 0.034). In the normalised case, all the runs were significant with respect to Run 1 (Run 7 and Run 8 with weak p-values).

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

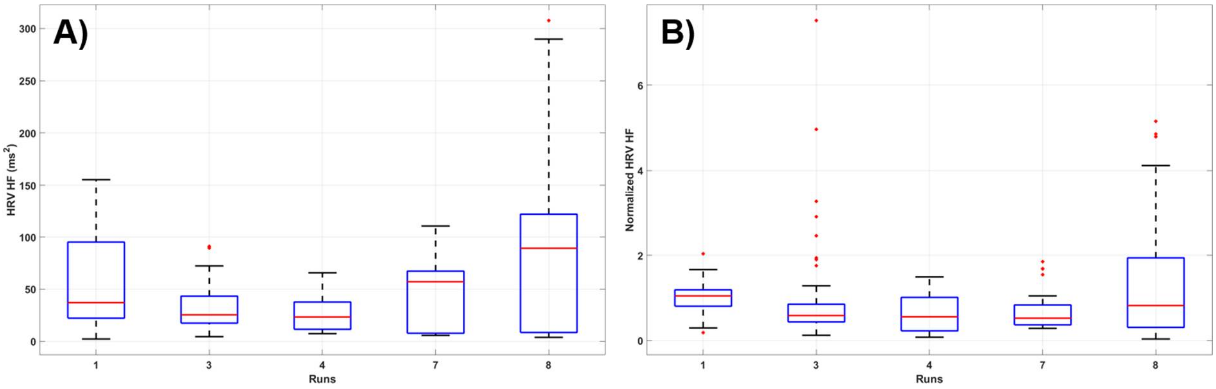


Figure 47: The distribution of the HF for all workload runs during phase 2. A) Not normalised and B) normalised

In the same way as HF (Figure 47), the LF showed to only be significant when Run 1 is compared to Run 4. This significance was strengthen when the LF was normalised and in addition Run 8 and Run 3 showed also be to significant different (weak though, $p = 0.002$ and $p = 0.035$).

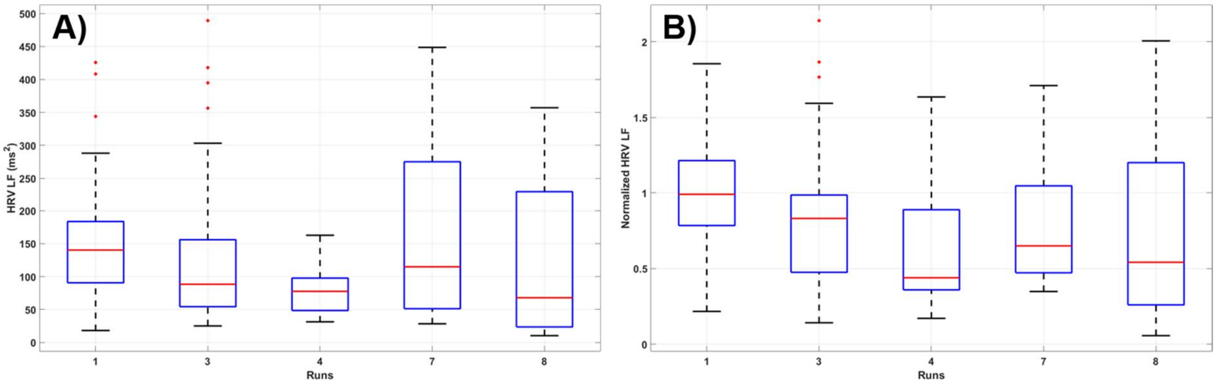


Figure 48: The distribution of the LF for all workload runs during phase 2. A) Not normalised and B) normalised.

In the case of the VLF feature (Figure 49) all the runs were significant different with respect to baseline (Run 1). There was even a significant difference between Run 3 and Run 4, but not when the VLF was normalised (Figure 49 B). The remaining results are the same for the normalised and none normalised analysis, but with a weaker significant p-values when the normalised analysis was conducted (in particular Run 7).

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

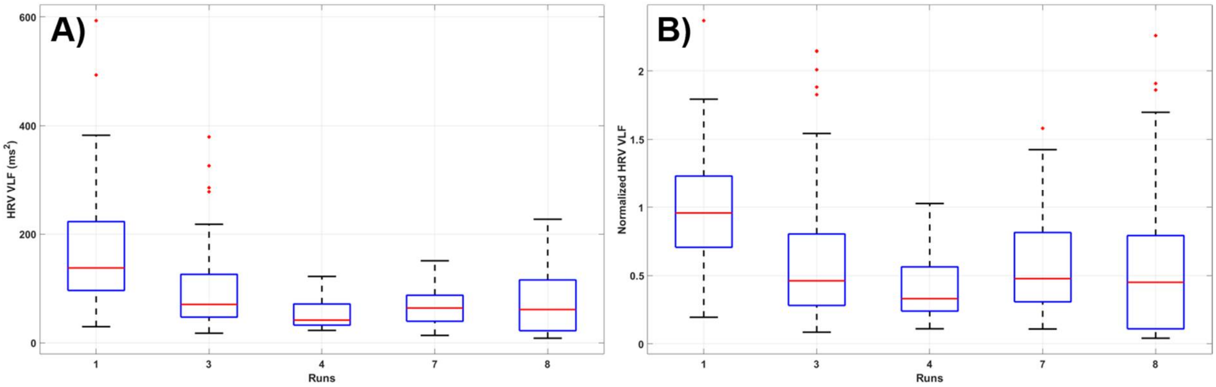


Figure 49: The distributions of VLF for all workload Runs during phase 2. A) Not normalised and B) normalised

4.5.2. Discussion

The results were presented by starting with a single pilot analysis. As expected Cinaz et al. (2013), there was a clear increase of HR during the duration of the different runs (besides Run 5) when looking at the single pilot analysis. However, this was not as evident when the group analysis was conducted, but this changed as soon as the HR was normalised. This indicates the importance of normalizing the HR when conducting the group analysis and in particular when the analysis was constrained to phase 2.

As in previous studies (Cinaz et al., 2013; Pendleton et al., 2016), the SDNN has showed a clear pattern of a decrease with an increase in workload. However, there was no significant difference between the runs with only workload and the once with added factors. Therefore, the SDNN obtained in this analysis was not sensitive enough to distinguish between the runs containing only workload and runs with mixed factors. In addition, since the phase 2 analysis were based on recordings of very short duration, it was very hard to draw any conclusion with respect to all the features based on HRV (i.e. SDNN, HF, LF and VLF).

As the previous physiological features and as the literature have already shown (Cinaz et al., 2013; Luque-Casado et al., 2016; Pendleton et al., 2016; Taelman et al., 2009), the HRV features demonstrated also a significant difference between the runs. However, the factor that exhibited the highest response was the run with high stress and this has also been highlighted in the work of Taelman et al. (2009). The spectral analysis of the HRV consisted of first the HF, which measures fast variations in the frequency domain and reflects the parasympathetic (vagal) modulation. HF reached the highest significance when Run 5 was compared to the baseline in both the single and group analysis. LF feature showed the same tendency as HF in the group analysis. However, in the single pilot analysis the LF response was almost the same when Run 5 was compared to the baseline and when Run 8 was compared to the baseline. This is not surprising, since LF expresses slower variations of the frequency domain that reflects both parasympathetic and sympathetic modulation in comparison to HF were only the parasympathetic modulation is taken into account. The VLF feature has showed the same tendency as LF, however the interpretation of the physiological meaning remains controversial in the literature (Malik et al., 1996). The run which showed no clear pattern in this case is Run 6 (reduced SA) and it is therefore difficult to interpret it. There was always a tendency of a decrease, but it was not significant in the single pilot analysis and the same for HF in the group analysis. On the other side no literature was found so far describing whether degradation of SA has an effect on the physiological response. Run 8, which contains the mixed factors, showed to be highly significant when compared to Run 1 in the single pilot analysis and also did Run 5 (high stress alone). However, in the group analysis the distribution of Run 8 start to increase in width, but remained always significant with respect to Run 1. This is not surprising since the data in Run 5 consisted of only 3 pilots whereas Run 8 contained data from 5 different pilots. Overall, the outcomes of this work is consistent with previous studies, however more data is needed to confirm this and make sure that the results are not biased in one or another direction.

4.5.3. Conclusion

The outcomes of this experiment have shown that physiological measures such as HR, SDNN, HF, LF and VLF can be sensitive to an increase in workload and/or stress. The runs with SA degradation on the other side showed very often no clear pattern and remain therefore difficult to interpret. HR and SDNN were particularly sensitive to the increase in workload, while the HRV features derived from the spectral analysis (HF, LF and VLF) showed a significant response to the increase of stress as well. The single pilot versus group analysis showed the importance of normalizing HR values when conducting the group analysis (in particular the phase 2 analysis). This is not surprising since an “absolute” HR value is much more subject dependent than HRV that express a type of “variability” which results in HR being more sensitive to inter-subject variability. The outcomes of this work is consistent with previous studies, however we are conscious that more data is needed to confirm this and make sure that the results are not biased in one or another direction.

4.6. Eye-tracking data

4.6.1. Data Acquisition

Eye tracking glasses

For the recording of the eye movements the Eye Tracking Glasses of SensoMotoric Instruments (SMI) were used. SMI’s eye tracking technology provides binocular tracking up to a 120 Hz sampling rate. Combined with a high definition scene camera and automatic parallax compensation this ensures accurate data over all distances. The SMI BeGaze analysis software supports aggregation of eye tracking data over multiple participants and allows qualitative visualization as well as quantitative analysis of eye tracking data. Data and visuals such as heat maps or Key Eye Tracking Metrics can be exported for further analysis.

Future Sky Safety has received funding from the EU’s Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Definition of areas of interest

A set of 22 areas of interest were defined in the cockpit. Four of them (numbers 2, 3, 4 and 5) match simulator's windows while the other ones are dedicated to instrument panels (see Figure 50). The eye-tracking system continuously evaluates from the pupil measurements and the head position which area of interest the operator is looking at. A default error value is recorded when the pilot does not watch any of these area of interest.

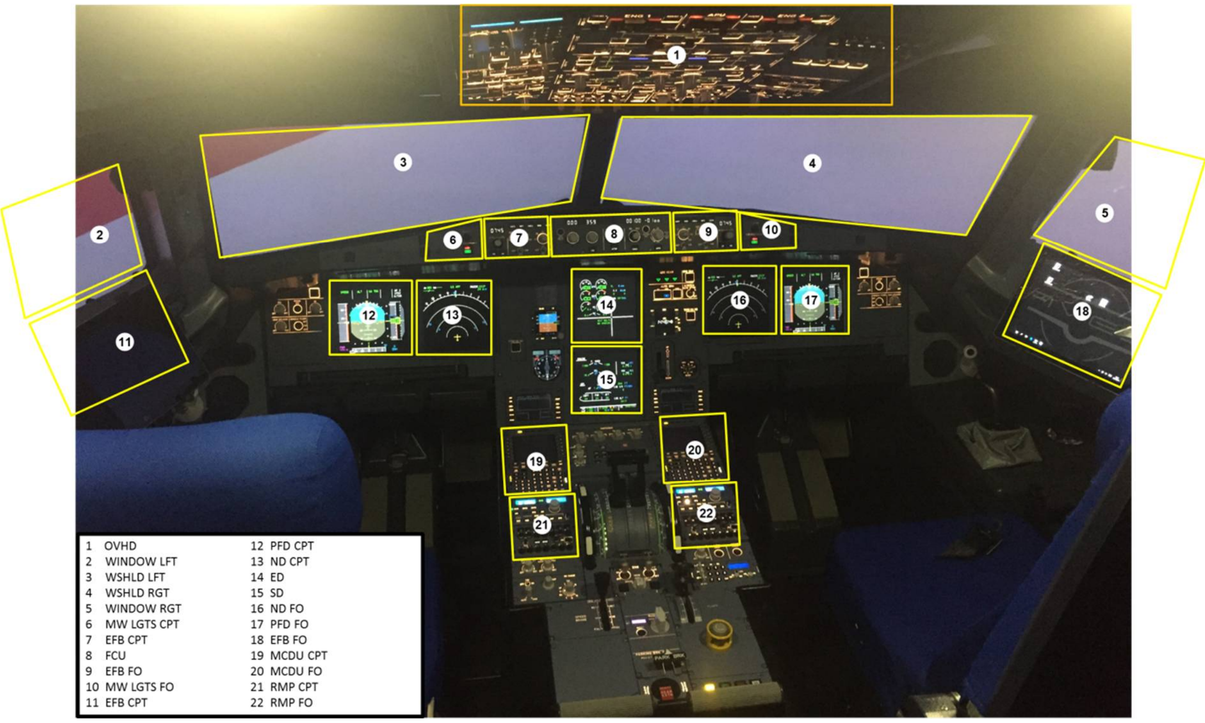


Figure 50: Areas of interest defined in the simulator

4.6.2. Data pre-processing

4.6.2.1. Data extraction

From the raw data, which is a continuous stream of pupil size estimates and other parameters, specific pupil responses need to be retrieved for dedicated parts of the flight or for specific events. The eye-tracking data and the simulator data are synchronized through a common time stamp. So the extraction of eye data for analysis requires first the processing of simulator data in order to identify the target time stamps.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

In a first time, it was decided to identify in each flight:

- The beginning of the simulation scenario
- The top of descent (TOD)
- The crossing of 1000ft above runway elevation
- The crossing of 750ft above runway elevation (height decision for some approaches)
- The crossing of 200ft above runway elevation (height decision for other approaches)
- The touch down
- If relevant, the beginning of the go-around manoeuvre

Figure 51 shows simulation data for one flight and the identified specific times. The beginning of the scenario and the touch down were extracted from videos, while the top of descent was adjusted according to localiser and glide slope deviations. Other markers were calculated from the altitude parameter.

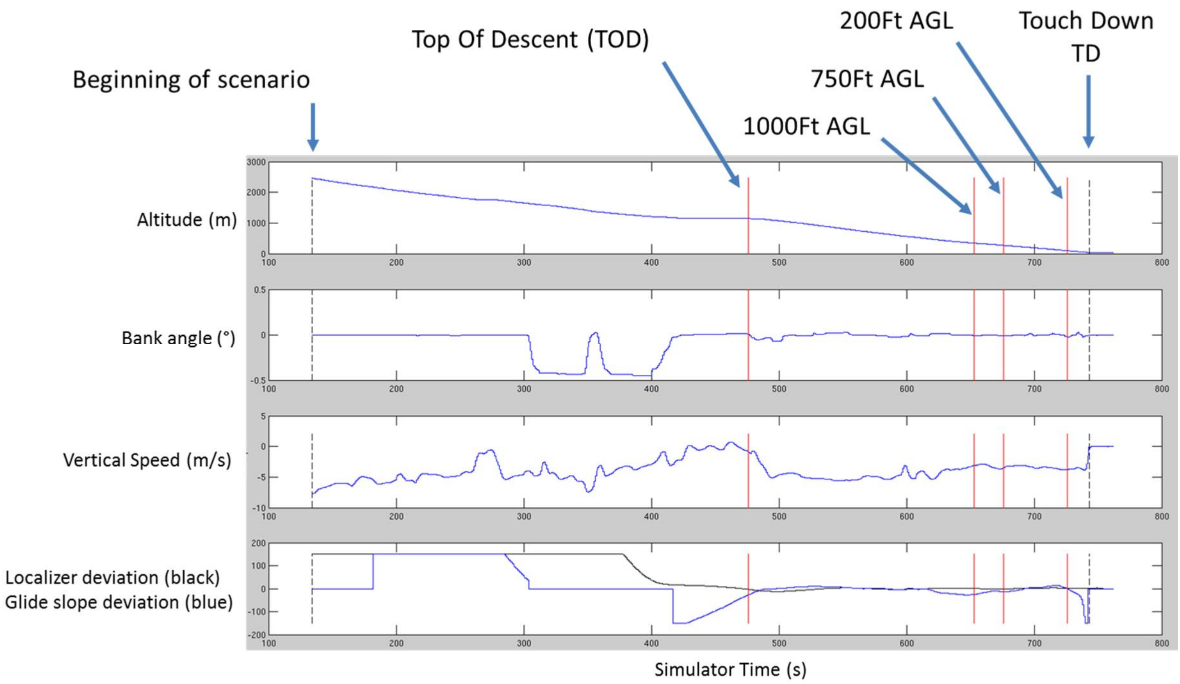


Figure 51: Simulator flight parameters (altitude, airplane bank, vertical speed, localiser, and glide slope deviations) and identified markers

Then, the variables of interest are left and right pupil sizes, and the identification of which area of interest the subject is looking at. They can now be extracted for the part of the flight we want to analyse.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

4.6.2.2. *Cleaning data*

Regardless of the eye tracking device used, gaze-loss samples are unavoidable (i.e. from eye blinks). In our case, missing values can be recorded as -1 or 0 values. Also the signal has to be cleaned from these data before performing statistical analysis. The basic cleaning of the data has been done in three steps (see also Figure 52):

1. Remove bad data
2. Interpolate missing sample
3. Filter with a 9-point average filter

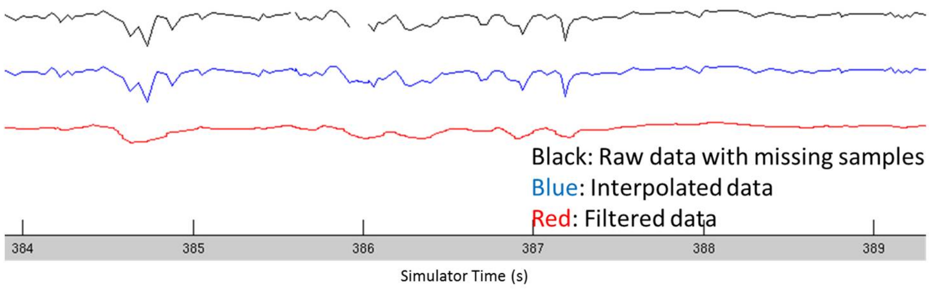


Figure 52: pupil radius data at the tree steps of the cleaning process

4.6.3. Description of the analysis methodology

The analysis focused on run 1, 3, 4 and 8. Run 1 constitutes a baseline (low workload, low stress, high situation awareness).

4.6.3.1. *Lighting conditions*

For all scenarios, simulator's light conditions were equivalent (same lighting in the simulator, same lighting of the database). Nevertheless, the lighting level of window screens that display the simulated outside world was different from the lighting level of the other display screens (see Figure 53). So it has been decided to keep only data when the pilot was looking at inside displays.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Figure 53: View of the cockpit during simulation

Then, for each pilot and each run, the percentage of time (from the beginning of the scenario to the 200ft marker) spent to look outside (areas of interest 2 to 5), at inside panels (other areas of interests) or, at other location have been calculated. Figure 54 displays all the results.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

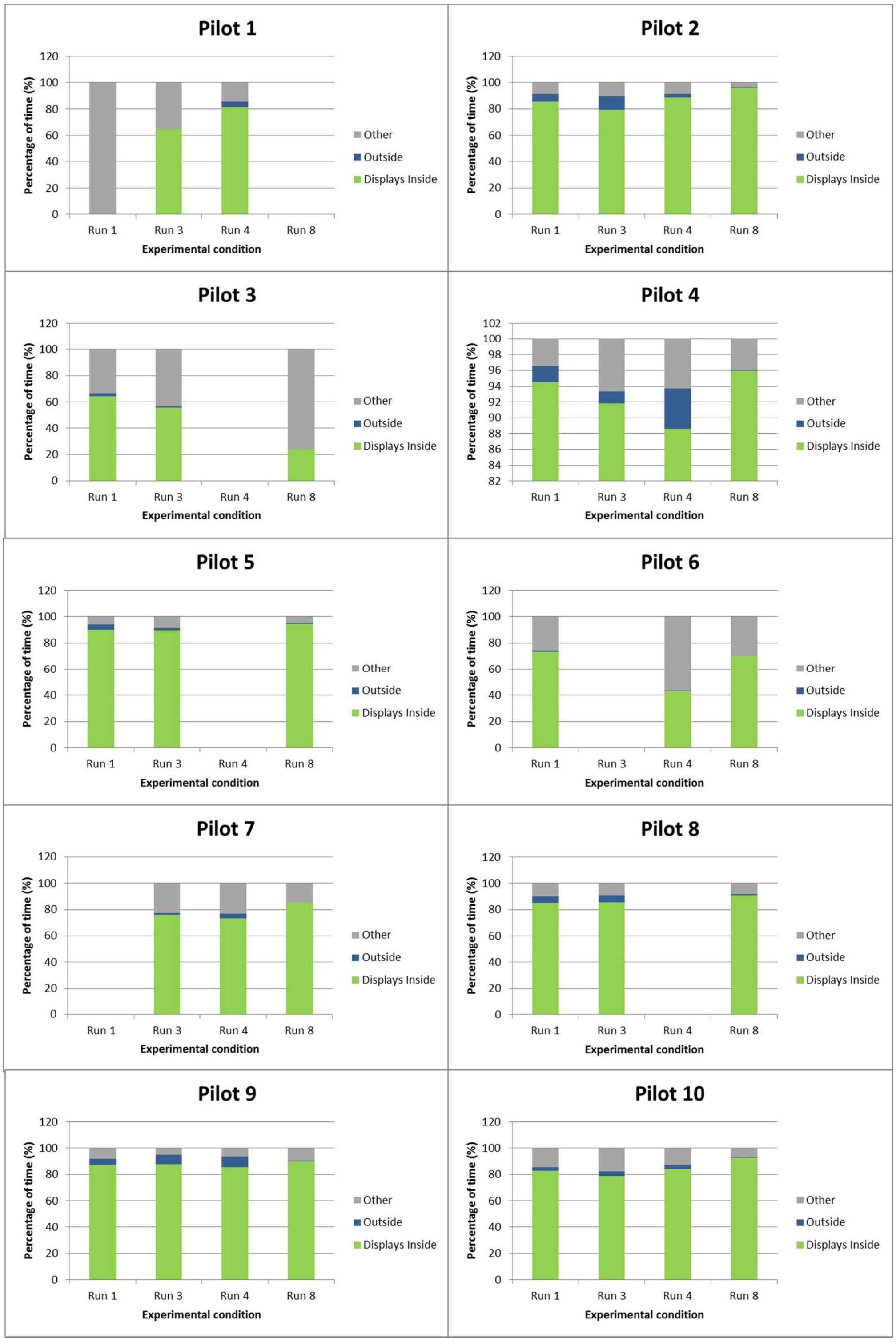


Figure 54: Percentage of time spent to look outside (blue), at inside displays (green) or, at other location (grey) for each pilot and each run

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

These results indicate that some data are missing (i.e. run 8 for pilot 1). Either the run had not been made by the pilot, either the simulation data were not available, either the eye-tracking data were corrupted or not available. Moreover, it can be seen that some data are not coherent with a piloting activity. As an example, in run 8, pilot 3 watched 22% of its time at inside display and 78% at other location. Such a distribution is not coherent and reveals calibration problems (probably the pilot touched the glasses after the calibration process). So it was chosen to remove from the data set all the questionable data.

Data from pilots 2, 5, 8, 9 ad 10 are kept for the analysis (pilot 7 is discarded because the baseline - run 1- is missing). For these data, the pupil radius will be analysed only when the pilot is looking at inside panels (over 80% of the time).

4.6.3.2. Comparisons between experimental conditions

The aim is to compare the pupil radius between experimental conditions (Run 1 versus Run 3 versus Run 4 versus Run 8). The comparison can be done pilot by pilot. For this analysis, it is required to have equal sample sizes for each experimental condition. Then a repeated measure ANOVA can be conducted. Thus for each run the sequence of measures will be resampled (see Figure 55).

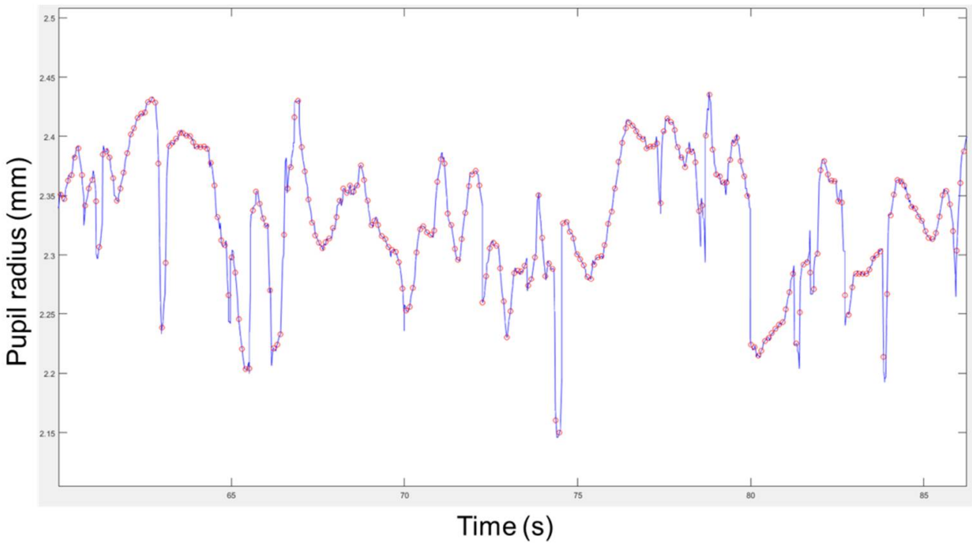


Figure 55: Example of resampling (red circles) of filtered radius measures (blue line)

Then, the comparison between experimental conditions can be done with all the pilots together. Nevertheless, has we have only few pilots for each condition, it the set of all the measures are kept for this analysis. Moreover, it is required to normalize the pupil radius for each pilot in order to be able to compare data from several pilots. Data from Run 1 (baseline condition) have been used for the normalization process.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

4.6.3.3. Overview of the analyses made

Comparisons have been made for several parts of the flight:

- From the beginning of the scenario to the 200Ft marker
- From the beginning of the scenario to the Top of Descent
- From the Top of Descent to the 1000Ft marker
- From the 1000Ft marker to the 200Ft marker

As we wanted to have comparable activities between the runs, flights with an interrupted landing before the 200Ft marker have been removed from the first and the last analyses. So the data taken into account for the analyses are summarized by the following table (Table 15):

Table 15: Runs used for analysis of eye-tracking data

	Run 1	Run 3	Run 4	Run 8
Pilot 2			Go Around before 200Ft marker	
Pilot 5			No simulator data	Go Around before 200Ft marker
Pilot 8			No eye tracking data	
Pilot 9				
Pilot 10				

For each part of the flight, comparisons of the pupil radius between conditions have been made by pilot. Measures from both eyes have been taken into account in these analyses.

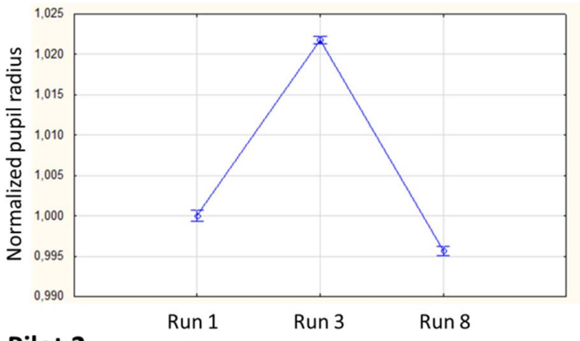
4.6.4. Results

The following figures display results of ANOVA analyses in which the mean is represented by a circle and the bars represents the confidence interval (p=0.95). Tables display results of a post-hoc Newman-Keuls test where red values indicate conditions which are significantly different.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

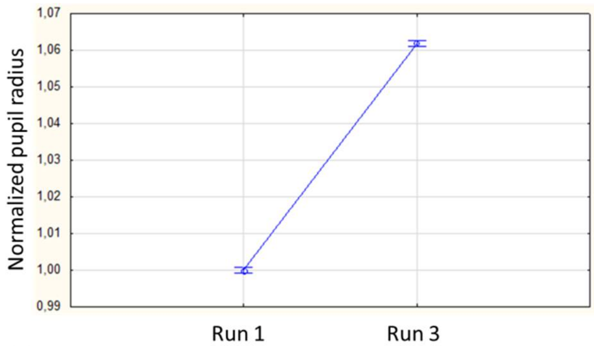
4.6.4.1. Overall descent (from the beginning of the scenario to the 200ft marker)

The following results (Figure 56) show that for all pilots, results are significantly different for all the runs compared.



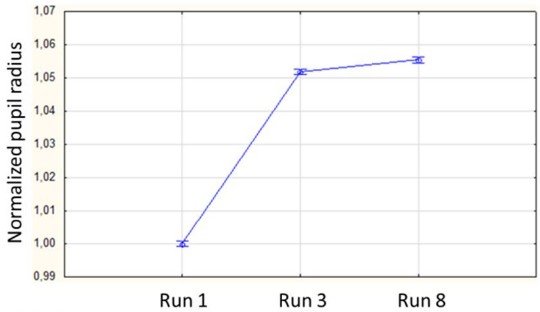
Pilot 2

Cellule N°	Test de Newman-Keuls ; variable VD_1 (results_new.s Probabilités Approximatives des Tests Post Hoc Erreur : MC Intra = ,00205, dl = 47998, RUN	{1} 1,0000	{2} 1,0218	{3} ,99565
1	R1_P2		0,000009	0,000009
2	R3_P2	0,000009		0,000022
3	R8_P2	0,000009	0,000022	



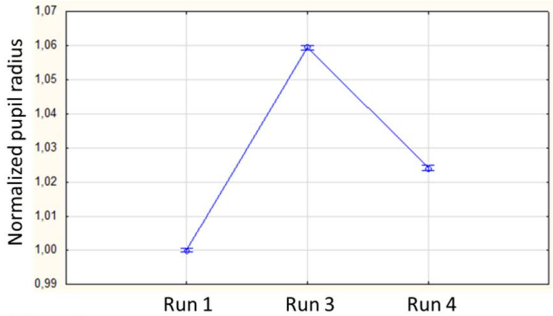
Pilot 5

Cellule N°	Test de Newman-Keuls ; variable VD_1 (results_new Probabilités Approximatives des Tests Post Hoc Erreur : MC Intra = ,00433, dl = 23999, RUN	{1} 1,0000	{2} 1,0619
1	R1_P5		0,000009
2	R3_P5	0,000009	



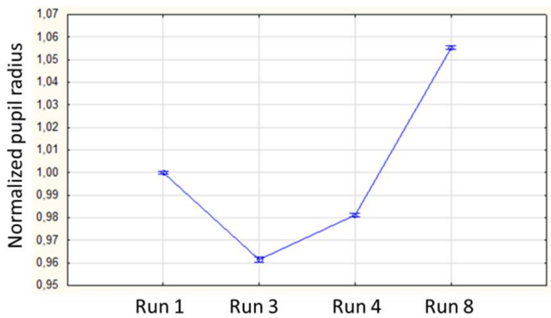
Pilot 8

Cellule N°	Test de Newman-Keuls ; variable VD_1 (results_new.s Probabilités Approximatives des Tests Post Hoc Erreur : MC Intra = ,00463, dl = 47998, RUN	{1} 1,0000	{2} 1,0518	{3} 1,0555
1	R1_P8		0,000009	0,000022
2	R3_P8	0,000009		0,000009
3	R8_P8	0,000022	0,000009	



Pilot 9

Cellule N°	Test de Newman-Keuls ; variable VD_1 (results_ne Probabilités Approximatives des Tests Post Hoc Erreur : MC Intra = ,00256, dl = 47998, RUN	{1} 1,0000	{2} 1,0594	{3} 1,0241
1	R1_P9		0,000022	0,000009
2	R3_P9	0,000022		0,000009
3	R4_P9	0,000009	0,000009	



Pilot 10

Cellule N°	Test de Newman-Keuls ; variable VD_1 (results_new.sta) Probabilités Approximatives des Tests Post Hoc Erreur : MC Intra = ,00502, dl = 71997, RUN	{1} 1,0000	{2} ,96144	{3} ,98120
1	R1_P10		0,000022	0,000009
2	R3_P10	0,000022		0,000009
3	R4_P10	0,000009	0,000009	
4	R8_P10	0,000009	0,000008	0,000022

Figure 56: Normalised pupil radius of overall descent

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Nevertheless, there is no clear impact of the workload on the pupil radius. When the workload increases to a high value (Run 3), the pupil radius increases for 4 pilots but decreases for the last one. Moreover, when the workload continues to increase (Run 4), the pupil radius decreases for pilot 9 but increases (compared to run 3) for pilot 10.

For Run 8 that combines stress and workload increase with a degraded situation awareness (compared to Run 1), the pupil diameter increases for two pilots (8 and 10), but decreases for pilot 2.

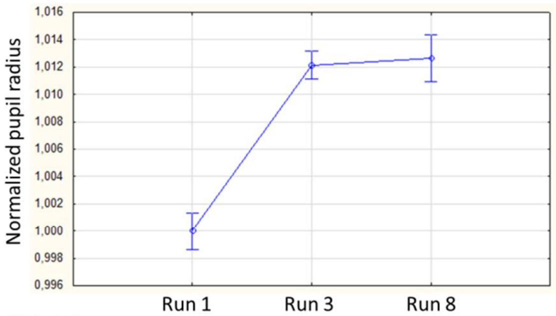
Based on these results, is it not necessary to conduct an analyses combining data of different pilots for the overall descent.

4.6.4.2. From Top of Descent to 1000ft

This part of the flight is shorter and allows less variability for the pilot activity. The aircraft is already aligned on the runway and the pilot maintains the airplane on the final descent path.

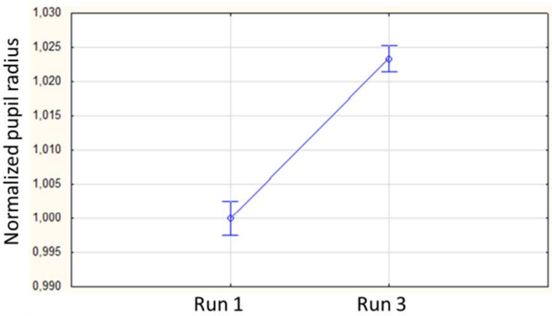
In this case, it can be noted that for all pilots, the pupil radius significantly increases between run 3 and run 1 (see Figure 57). So the workload increase from the baseline to the high level induces an increase of the pupil radius. Nevertheless, the increase of workload from a high level to a very high level (run 4) does not imply an increase of the pupil radius for all pilots (see pilot 9).

For Run 8 that combines stress and workload increase with a degraded situation awareness (compared to Run 1), the pupil diameter increases for the three pilots (2, 8 and 10) and the increase is bigger or equal to the one observed for the high workload condition (run 3 and 4).



Pilot 2

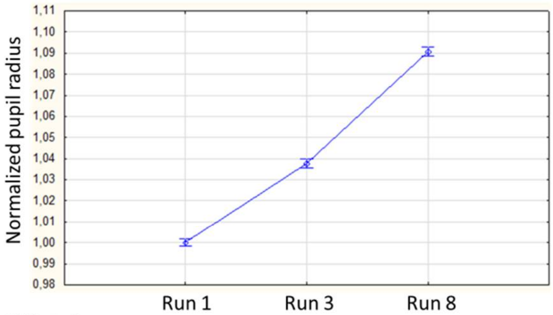
Cellule N°	Test de Newman-Keuls ; variable VD_1 (results_Ech_ Probabilités Approximatives des Tests Post Hoc Erreur : MC Intra = ,00070, dl = 2998,0			
	RUN	{1} 1,0000	{2} 1,0122	{3} 1,0126
1	R1_P2		0,000009	0,000022
2	R3_P2	0,000009		0,612634
3	R8_P2	0,000022	0,612634	



Pilot 5

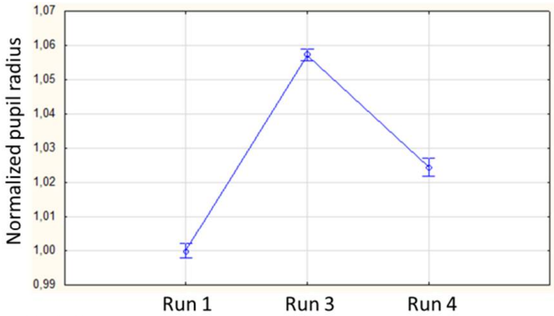
Cellule N°	Test de Newman-Keuls ; variable VD_1 (results_ Probabilités Approximatives des Tests Post Hoc Erreur : MC Intra = ,00142, dl = 1499,0		
	RUN	{1} 1,0000	{2} 1,0234
1	R1_P5		0,000009
2	R3_P5	0,000009	

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.



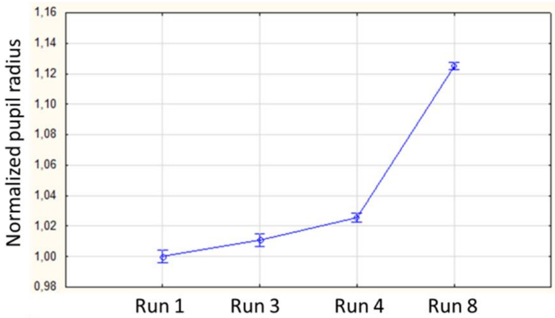
Pilot 8

Cellule N°	Test de Newman-Keuls ; variable VD_1 (results_ Probabilités Approximatives des Tests Post Hoc Erreur : MC Intra = ,00161, dl = 2998,0			
	RUN	{1} 1,0000	{2} 1,0377	{3} 1,0906
1	R1_P8		0,000009	0,000022
2	R3_P8	0,000009		0,000009
3	R8_P8	0,000022	0,000009	



Pilot 9

Cellule N°	Test de Newman-Keuls ; variable VD_1 (results Probabilités Approximatives des Tests Post Ho Erreur : MC Intra = ,00184, dl = 2998,0			
	RUN	{1} 1,0000	{2} 1,0572	{3} 1,0245
1	R1_P9		0,000022	0,000009
2	R3_P9	0,000022		0,000009
3	R4_P9	0,000009	0,000009	



Pilot 10

Cellule N°	Test de Newman-Keuls ; variable VD_1 (results_Ech_15 Probabilités Approximatives des Tests Post Hoc Erreur : MC Intra = ,00480, dl = 4497,0			
	RUN	{1} 1,0000	{2} 1,0108	{3} 1,0256
1	R1_P10		0,000027	0,000022
2	R3_P10	0,000027		0,000009
3	R4_P10	0,000022	0,000009	
4	R8_P10	0,000008	0,000022	0,000009

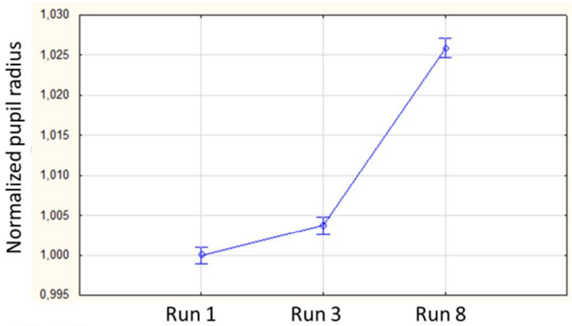
Figure 57: Normalised pupil radius of final approach between TOD and 1000ft

4.6.4.3. From 1000ft to 200ft

In this part of the flight, the aircraft is supposed to be stabilized on the descent track. The pilot activity is very constrained and the duration of this part of the descent is short.

Globally, the workload increases bring an increase of the pupil radius, even if run 3 of pilot 10 and run 4 of pilot 9 are not completely coherent with this result. For run 8, an increase of the pupil radius is observed and this increase is bigger than for run 3 or run 4. See Figure 58.

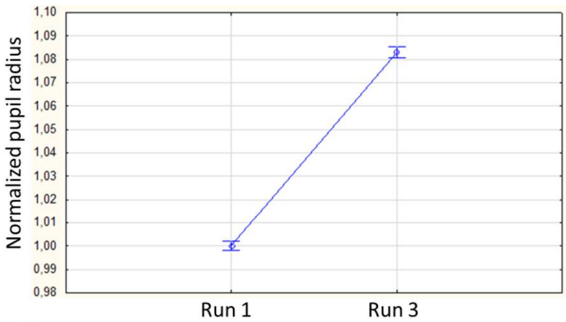
Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.



Pilot 2

Test de Newman-Keuls ; variable VD_1 (results_Ech_
Probabilités Approximatives des Tests Post Hoc
Erreur : MC Intra = ,00048, dl = 2998,0

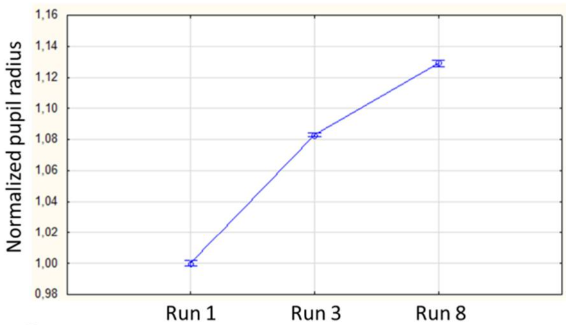
Cellule N°	RUN	{1} 1,0000	{2} 1,0037	{3} 1,0259
1	R1_P2		0,000012	0,000022
2	R3_P2	0,000012		0,000009
3	R8_P2	0,000022	0,000009	



Pilot 5

Test de Newman-Keuls ; variable VD_1 (results_
Probabilités Approximatives des Tests Post Hoc
Erreur : MC Intra = ,00214, dl = 1499,0

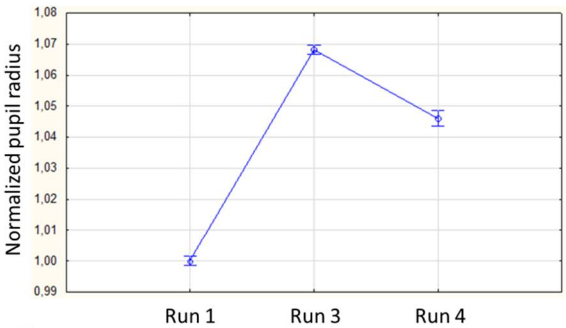
Cellule N°	RUN	{1} 1,0000	{2} 1,0830
1	R1_P5		0,000009
2	R3_P5	0,000009	



Pilot 8

Test de Newman-Keuls ; variable VD_1 (results_E
Probabilités Approximatives des Tests Post Hoc
Erreur : MC Intra = ,00106, dl = 2998,0

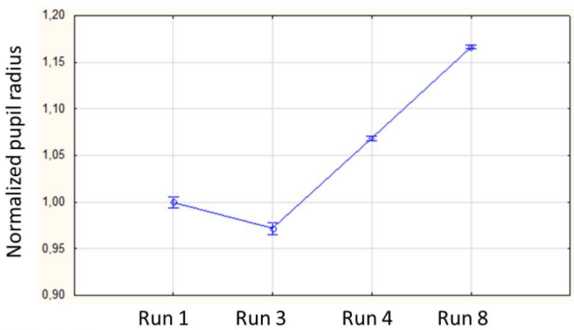
Cellule N°	RUN	{1} 1,0000	{2} 1,0830	{3} 1,1291
1	R1_P8		0,000009	0,000022
2	R3_P8	0,000009		0,000009
3	R8_P8	0,000022	0,000009	



Pilot 9

Test de Newman-Keuls ; variable VD_1 (results_
Probabilités Approximatives des Tests Post Hoc
Erreur : MC Intra = ,00105, dl = 2998,0

Cellule N°	RUN	{1} 1,0000	{2} 1,0681	{3} 1,0461
1	R1_P9		0,000022	0,000009
2	R3_P9	0,000022		0,000009
3	R4_P9	0,000009	0,000009	



Pilot 10

Test de Newman-Keuls ; variable VD_1 (results_Ec
Probabilités Approximatives des Tests Post Hoc
Erreur : MC Intra = ,00792, dl = 4497,0

Cellule N°	RUN	{1} 1,0000	{2} ,97162	{3} 1,0685
1	R1_P10		0,000009	0,00000
2	R3_P10	0,000009		0,00002
3	R4_P10	0,000009	0,000022	
4	R8_P10	0,000022	0,000008	0,00000

Figure 58: Normalised pupil radius of final approach between 1000ft and 200ft

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

4.6.5. Conclusions

The outcomes of this experiment suggest that there is no clear impact of the workload on the pupil radius when the measure is done on the overall descent. Nevertheless, when the study is done on smaller part of the flight where the pilot activity is more comparable between experimental conditions (sections 4.6.4.2 and 4.6.4.3) a workload increase comes with an increase of the pupil radius. It has also to be noticed that the workload level is not uniform during each run (as an example the approach and runway change in run 4 implies an increase of workload but only for a limited period of time during the run) and therefore the pupil radius measure should be tested on even smaller part of the flight, associated to dedicated events. This study would be done to complete the available results. Finally, no group analysis has been conducted because of the small number of usable data. Hence, more data is needed in order to obtain more reliable conclusions with respect to effects of different HPE factors on pupil radius.

4.7. Behavioural Markers

During each run, two external observers identified and collected a set of behaviours that could be potential indicators of performance decline, hereafter called Behavioural Markers. A preliminary list of pilots' markers was collected and consolidated during workshops with operational experts, on the basis of similar studies performed in ATC (Edwards, 2013, 2014). This list was used to set up a grid of aspects and elements to be observed during the simulations, and the list was discussed directly with the subjects during the debriefings to validate it and integrate new markers if emerged during the simulations.

Two kinds of markers intended to be collected during the simulations:

- **External Markers** – observable behaviours that are commonly associated to a performance decrease, collected during simulation live observations and off-line video check;
- **Internal Markers** – subjectively experienced indicators that people use to know that their internal state is reaching the tolerance limits and might negatively impact on the performance, collected during the debriefings.

4.7.1. Results

4.7.1.1. External markers

From the data collected during the workshops with operational experts, four categories of markers were identified as the most commonly associated to a decrement of pilot's performance:

- Changes in pilot's <u>Behaviour</u> (reaction time, movements, interaction with tools etc.);
- Changes in pilot's <u>Performance</u> (intended as lack of accuracy in procedures application);
- Changes in pilot's <u>Communication</u> (variation in tone and contents of non-procedural communication);
- And <u>Physiological</u> changes, observable from outside (such as breathing rate or gaze fixation).

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

The full list of external markers collected during the workshops, divided per type of factor, is reported in Table 16.

Table 16: List of External Behavioural Markers, collected during the workshops in Rome

External Behavioural Markers	
Category	***Marker***
Behavioural changes	Longer reaction times
	Action paralysis, lack of physical / cognitive reaction
	Problems in changing the attention focus
	Repeated actions and reiterated actions & mistakes
	Change in speed and regularity of movements
	Mumbling / Thinking out loud
	Cognitive stiffening, trouble in changing idea/opinion/plan
	Compulsive use of setting and controls
	Amplification of own typical body movements, or onset of unusual body movements
	Use of external supports (paper brain and automation)
	Irritation and emotional reactions
	Movements on the seat / motion restless
	Lack of attention
	Disorientation among HMIs
Physiological changes	Change in breath rate (long breathings, deep breathings, panting)
	Gaze fixation
	Fast eye movements
	Lapse of memory
Performance changes	Miss ATC / colleague's communications / Absence of response
	Loss of accuracy on secondary tasks
	Omissions, skipping or forgetting things
	Deviation from procedures (shortcuts), lack of accuracy in executing procedures
	Change in tasks prioritisation
	Speed in going through procedures
	Centralization of control and reduction of teamwork, rigidity in the roles
	Increase of questions/requests to colleague
	Not resuming an interrupted task
	Misinterpretation of abbreviations
	Lack of alternative solutions / plan B
Communication changes	Change in the tone of voice, volume and pitch
	Change in the amount of communication (too much/too little)

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

	Change in the communication content
	Stop talking
	Complaining
	Language switch, usually from English to mother tongue
	Irritated communication
	Use of the wrong wording
	Communicate using body movements (indicate controls or mime actions)

The four categories were used to drive the experts' attention during the live observation and video check. In this preliminary phase, all the aspects and variations related to Communication, Performance, Physiology and Behaviours were collected, without associating them to a good or degraded performance or estimating the severity of the markers. All the indicators collected are reported onto the timeline of events, and will be used in triangulation with physiological data and performance data to identify potential early indicators or recurrent markers of performance decrements.

With respect to the four categories, the main types of indicators observed during the simulations were:

- Behavioural:
  - Mumbling, Thinking out loud;
  - Seat adjustment, movements on the seat (upward, forward, back);
  - Mouth movements, such as tighten lips, licking lips, tongue between lips etc.;
  - Moves hand/fingers on the side-stick, change in sidestick grasp;
  - Sniff, snort, cough;
  - Gesturing;
  - Smiling, laughing, smirking;
  - Tense facial expression, grimaces;
  - Adjust eyeglasses, microphone, headset;
  - Touch or scratch nose, cheek, head;
  - Shaking head;
  - Looking around;
  - Making sounds, vocal noises or humming.
- Communication:
  - Change in the communication style, from chatty to silent;
  - Sniggering;
  - Stutter, hesitate, flounder;
  - Question, express doubts;
  - Complaining;
  - Joking;
  - Communicate using body movements (indications, nods etc.);

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

    - Slips of memory.
- Physiological:
    - Heavy or deep breathing;
    - Swallow.
- Performance:
    - Talking over colleague;
    - Mistakes in calculations (remaining time, remaining fuel);
    - Forget actions/procedures (checklists, aircraft configuration);
    - Under-specification of actions/plans;
    - Miss workload question;
    - Miss ATC calls/information;
    - Give up control to Pilot Monitoring;
    - Do not react/reply to Pilot Monitoring questions;
    - Repeat commands/questions.

Together with that, during the observations the observers took note of other three key aspects of pilots' performance: if and when they realise about aircraft/environmental conditions (fuel, wind), and if they realise that on their own or guided by Pilot Monitoring; if and when they declare emergency, and again if they did it on their own initiative or guided by Pilot Monitoring; if they perform a Go Around. These aspects were collected (and so far included under the "Performance" category) to support the analysis of performance data, in particular with respect to the right and punctual application of company procedures. The values (positive or negative) of these three aspects and their impact on the evaluation of the overall pilot performance need to be evaluated with an operational expert. Also, the support of an operational expert is desirable to perform an additional video check and complement the analysis of all the other Performance indicators that are difficult to be spotted by a non-operational expert, such as loss of accuracy on secondary tasks, deviation from procedures (shortcuts), lack of accuracy in executing procedures, lack of alternative solutions / plan B.

All the indicators observed and collected during the run executions were reported on a timeline, together with the sequence of events simulated in each run and the phase of flight in which the indicators were observed. Examples of the outcome of this mapping can be found in the two tables below.

The timeline will help during the data triangulation phase, when these markers will be mapped onto physiological and performance data to see if they can be used as early warnings or as indicators of performance decrement.

Table 17: Markers collected for Pilot 9, Run 1

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Run 1	12/05, 12:21:23		Duration: 00:11:27				
Scenario characteristic	Time		Phase of flight	Behavioural marker	Time		Category
	From	To			From	To	
Start	12:21:23	end	Slot 1				
Top of Descent	12:28:07	-	Slot 2				
				Tongue between lips & tightening lips	12:31:00	12:31:02	*Behavioural*
Decision altitude	12:32:09	-	Slot 3				
				Tightening lips	12:32:11	12:32:12	*Behavioural*
Touch down (FRA)	12:32:28	-		Tightening lips	12:32:25	12:32:28	*Behavioural*
				Tightening lips	12:32:36	12:32:38	*Behavioural*
End	12:32:50	-					

Table 18: Markers collected for Pilot 7, Run 8

Run 8	11/05, 09:55:54		Duration: 00:23:21				
Scenario characteristic	Time		Phase of flight	Behavioural marker	Time		Category
	From	To			From	To	
Start	09:55:54	end	Slot 1				
High turbulences	09:55:54	end					
Low fuel	09:55:54	end					
Low visibility	09:55:54	end					
Delay vectors	09:55:54	10:00:47					
				Realises about low fuel	09:56:28	-	*Performance*
				Discussion about fuel (PM proposes PF to ask for direct to LETKI, to avoid being in emergency)	09:59:00	-	*Performance*
				Tongue between lips	10:03:39	10:03:40	*Behavioural*
Top of Descent	10:05:48	-	Slot 2				
				Moves on the seat	10:07:22	10:07:28	*Behavioural*
				Adjust grip on sidestick	10:07:30	10:07:32	*Behavioural*
				Adjust grip on sidestick	10:07:36	10:07:38	*Behavioural*
Loud Noise	10:07:57	10:09:38					
Wind shift	10:08:46	10:09:11					
Localiser interference	10:08:46	10:09:52					
				Can't hear PM indication ("stay on green line")	10:09:00	10:09:02	*Performance*
				Tilt head upward	10:09:06	10:09:08	*Behavioural*
				Moves on the seat	10:09:15	10:09:17	*Behavioural*
				Panting	10:09:18	10:09:20	*Physiological*
				Adjust headphone	10:09:26	10:09:27	*Behavioural*
Decision altitude	10:09:33	-	Slot 3				
Go around	10:09:33	-					*Performance*
				Adjust headphone	10:09:41	10:09:41	*Behavioural*
				Sniffs	10:09:41	10:09:41	*Behavioural*
				Deep breath	10:09:53	10:09:57	*Physiological*
				Sniffs	10:10:24	10:10:24	*Behavioural*
				Sniffs	10:11:23	10:11:24	*Behavioural*
				Sniffs	10:11:30	10:11:31	*Behavioural*
				Sniffs	10:11:49	10:11:50	*Behavioural*
				Scratches face	10:12:42	10:12:43	*Behavioural*
				Sniffs	10:13:22	10:13:23	*Behavioural*
				Sniffs	10:13:57	10:13:58	*Behavioural*
				Sniffs	10:14:19	10:14:20	*Behavioural*
				Deep breath	10:14:52	10:14:53	*Physiological*
				Pants answering WL question	10:16:17	10:16:20	*Communication*
				Moves on the seat	10:16:52	10:16:57	*Behavioural*
				Stretching words while speaking ("Okkkkkk....")	10:17:48	10:17:50	*Communication*
Touch down (FRA)	10:18:56	-					
End	10:19:15	-					

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

4.7.1.2. *Internal markers*

From the data collected during the workshops with operational experts, three categories of markers were identified as the most commonly associated to a decrement of pilot's performance:

- Changes in pilot's Behaviour[3] (e.g. mumbling);
- Changes at Physiological level, (such as breathe rate or muscle tension).
- Changes in pilot at Cognitive level (variation in tone and contents of non-procedural communication).

The full list of internal markers collected during the workshops, divided per type of factor, is reported in Table 19.

Table 19: List of Internal Behavioural Markers, collected during the workshops in Rome

Internal Behavioural Markers	
Category	***Marker***
Behavioural changes	Mumbling
	Tightening grip on the side stick
Physiological changes	Clammy palms
	Hot flushes and sweat
	Increase in breath rate
	Increased heart rate
	Muscle tension (especially neck and shoulders)
	Dry mouth
Cognitive changes	Cognitive tunnelling
	Inability or difficulty to think ahead and anticipate
	Trouble in changing idea/opinion/plan
	Disorientation among several instruments and HMIs
	Change in time perception (lose track of time, feeling that time is faster/slower, feeling need to save time)
	Inability to face issues or new requests
	Losing the global picture of the situation

[3] Intended as behavioural observed only by the pilot him/herself.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

	Think in circle, mentally repeating checklists
	Difficulty in keeping concentration on the task
	Inability to prioritise activities

The three categories and the markers were used to drive the experts' attention during the debriefing performed with the pilots. Questions about behavioural markers were asked once per pilot, taking advantage of the longest break available, i.e. generally when the other pilot was performing Scenario 2. It is worth noting that while external markers were observed during the runs as they occurred, internal markers – due to their nature – were only collected during the debriefing. Pilots referred them to their own experience during the run(s) previously performed or to their personal experience on the job.

With respect to the three categories, the main markers collected during the debriefings were:

- Behavioural:
  - Mumbling;
  - Holding breath;
  - Tongue slips in between teeth;
  - *Recovery*: a few deep breaths.
- Physiological:
  - Increase of temperature (*Recovery*: adjust air conditioning);
  - Increased heart rate;
  - Stomach feeling;
  - Dry mouth;
  - Tingling fingers;
  - Feeling restless;
  - Feeling high state of arousal;
  - Sweating.
- Cognitive:
  - Cognitive tunnelling (*Recovery*: look around the cockpit);
  - "Hearing but not listening";
  - Misinterpreting abbreviations;
  - Realising things are missed or forgotten;
  - Monitor colleague's symptoms of distress to calm down himself;
  - Relying completely on Pilot Monitoring and not paying attention to the information Pilot Monitoring treats;
  - Cannot anticipate situations;
  - Cannot shift attention from primary task.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

4.7.2. Discussion

The overall picture that emerges from the analysis of external markers is that individual variability played a big role in the number of markers observed during each run per each pilot, and it seems difficult to use them to discriminate the runs. However, it is interesting to note that for almost all pilots (a part from two) Run 1 was the one with the lowest number of markers observed. Also, all pilots that performed Run 7, a part from one (Pilot 1) showed the highest number of markers in that run, while the number of markers on the other runs is extremely variable (see Table 20). The pattern for Run 1 and Run 7 is confirmed even by comparing the frequency of markers per minute. So, markers in Run 7 are more frequent than in Run 1, regardless the duration of the run.

Table 20: Number of markers collected in each run, per each pilot[4]

		Tot markers	*Behavioural*	*Communication*	*Performance*	*Physiological*
S1	Run 1	**3**	2	1	0	0
	Run 3	**3**	0	0	0	3
	Run 4	**3**	1	1	0	1
	Run 5	**41**	40	0	0	1
	Run 7*	**24**	20	3	0	1
	Run 8	**26**	20	3	2	1
S2	Run 1	**0**	0	0	0	0
	Run 3	**0**	0	0	0	0
	Run 4*	**19**	9	3	2	5
	Run 5	**19**	10	0	3	6
	Run 6	**31**	17	4	2	8
	Run 8	**11**	6	1	0	4
S3	Run 1	**9**	2	1	0	6
	Run 3*	**11**	1	2	1	7
	Run 4*	**25**	5	6	6	8
	Run 5	**14**	2	0	2	10
	Run 8	**18**	6	2	5	5
S4	Run 1	**0**	0	0	0	0
	Run 3	**4**	3	0	0	1
	Run 4	**13**	8	0	0	5
	Run 5	**13**	9	0	1	3
	Run 6	**14**	9	2	1	2
	Run 7*	**32**	22	0	6	4
	Run 8	**23**	15	2	3	3
S5	Run 1	**5**	5	0	0	0
	Run 3	**16**	11	1	3	1

[4] Stars indicate the run in which pilots performed a Go Around

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

		Tot markers	*Behavioural*	*Communication*	*Performance*	*Physiological*
	Run 4	**28**	21	0	0	7
	Run 5	**33**	22	2	2	7
	Run 6	**28**	18	2	4	4
	Run 7*	**56**	40	0	3	13
	Run 8*	**21**	13	0	3	5
S6	Run 1	**2**	0	0	0	2
	Run 4	*n.a. due to flue condition*				
	Run 5	**6**	3	0	1	2
	Run 6	**12**	5	3	1	3
	Run 7*	**54**	43	0	3	8
	Run 8	**16**	13	0	3	0
S7	Run 1	**13**	13	0	0	0
	Run 3	**12**	11	0	0	1
	Run 4	**15**	10	0	0	5
	Run 5	**20**	14	1	1	4
	Run 6	**24**	23	1	0	0
	Run 7	**35**	29	1	4	1
	Run 8*	**27**	18	2	4	3
S8	Run 1	**1**	1	0	0	0
	Run 3*	**13**	8	0	3	2
	Run 4*	**15**	10	2	3	0
	Run 5	**26**	12	3	6	5
	Run 6	**15**	9	0	4	2
	Run 7	**38**	29	0	3	6
	Run 8	**3**	2	0	1	0
S9	Run 1	**4**	4	0	0	0
	Run 3	**21**	21	0	0	0
	Run 4*	**30**	26	0	3	1
	Run 6*	**67**	52	5	7	3
	Run 8*	**60**	51	3	3	3
S10	Run 1	**9**	5	0	0	4
	Run 3	**3**	2	0	0	1
	Run 4	**14**	10	0	0	4
	Run 5	**20**	7	2	2	9
	Run 7	**28**	12	3	6	7
	Run 8	**5**	1	1	2	1

With respect to the type of markers, the most collected ones have been behavioural markers, followed by physiological ones. The low number of Performance markers can be due to a bias in the observers, who were not operational experts. This bias can be overcome by performing an additional video check with an operational expert, mainly focused on the identification of performance (and communication) markers. Also, the high number of markers in the Behaviour category can be explained by the broader scope of that

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

category, which includes several different indicators from facial expression to mumbling to physical movements. With respect to the communication markers, it has to be said that for several pilots it was not possible to investigate that category, as all the non-standard conversation was in German and the observers were non German speakers. In absolute terms, the most frequent ones among all pilots were the ones related to mouth movements, followed by movements on the seat and heavy or deep breathing. It has to be said that this absolute counting may be influenced by the fact that two pilots in particular showed a very high frequency of that behaviour (almost a systematic facial pattern); however changes in facial expression and particular mouth movements were observed in almost all pilots, so it could be worthwhile to investigate this aspect in a more systematic way, even using tools and software for Facial expression recognition.

If we compare the number of markers per phase of flight, we can see that how the higher number of markers are observed in the first slot (from the run start to Top of Descent), which is also the longest one with a mean duration – considering all the runs - of 10 minutes in a range that goes from 4 minutes up to 24 minutes. Slot 3, from Decision Altitude to the end of the run, is the one in which the lowest number of markers have been observed, but the mean duration of that slot is 45 seconds (excluding the runs in which pilots performed a Go Around). The situation changed in case of Go Around, with peak of markers recorded during and after the execution of the procedure. Looking at the patterns in the same run, markers after the Go Around were not only more than in the previous slots, but were more frequent than before. In general, the main trend emerging from the analysis of external behavioural markers is that when the task becomes more difficult pilots tend to accentuate their own typical behaviours (for example biting lips or moving on the seat), behaviours that can be found in normal situations too. This was particularly evident in Pilot 5 and Pilot 9, the two pilots that showed the highest number of markers in absolute terms. If we compare markers collected in Run 3 and Run 7 for Pilot 5 (

Table 21Table 21 and Table 22), we can see how the increase of task complexity in Run 7 generated an increase in number of facial expressions and mouth movements that were already recorded in the other run.

Table 21: Markers collected for Pilot 5, Run 3

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Run 3	10/05, 12:09:26		Duration: 00:13:32				
Scenario characteristic	**Time**		**Phase of flight**	**Behavioural marker**	**Time**		**Category**
	From	To			From	To	
Start	12:09:26	end	Slot 1				
High turbulence	12:09:26	end					
				Movement on the seat (moves body upward)	12:09:36	-	*Behavioural*
				Eye scratching	12:09:50	12:09:52	*Behavioural*
				Movement on the seat (moves body upward)	12:09:53	-	*Behavioural*
				Movement on the seat (moves body upward & forward)	12:10:06	12:10:07	*Behavioural*
				Missed ATC call	12:10:09	-	*Performance*
				Missed ATC call	12:10:20	-	*Performance*
				Missed ATC call	12:11:08	-	*Performance*
				Tightening lips	12:11:36	-	*Behavioural*
				Some heavier breathings	12:13:21	12:13:28	*Physiological*
				Doesn't recall ILS name ("*ehm... ehm... ILS* ")	12:14:05	12:14:07	*Communication*
				Tightening lips	12:14:15	-	*Behavioural*
				Grimace	12:14:33	-	*Behavioural*
				Tightening lips	12:14:58	-	*Behavioural*
Top of Descent	12:16:52	-	Slot 2				
				Tightening lips	12:18:33	-	*Behavioural*
				Tightening lips	12:18:56	-	*Behavioural*
				Tightening lips	12:21:28	-	*Behavioural*
Decision altitude	12:22:03	-	Slot 3				
Touch down (FRA)	12:22:27	-					
End	12:22:58	-					

Table 22: Markers collected for Pilot 5, Run 7

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Run 7	10/05, 09:32:32		Duration: 00:28:27				
Scenario characteristic	Time From	Time To	Phase of flight	Behavioural marker	Time From	Time To	Category
Start	09:32:32	end	Slot 1				
Medium turbulences	09:32:32	end					
Low fuel	09:32:32	end					
Low visibility	09:32:32	end					
Delay vectors	09:32:32	09:42:57					
				Screatches his cheek	09:33:17	-	*Behavioural*
				Deep breath	09:33:59	-	*Physiological*
				Realises about low fuel	09:34:10	-	*Performance*
				Swallows	09:34:28	-	*Physiological*
				Licking lips	09:36:25	-	*Behavioural*
				Tense facial expression (crooked mouth)	09:36:54	-	*Behavioural*
				Licking lips	09:38:05	-	*Behavioural*
				Licking lips and swallow	09:38:32	09:38:35	*Behavioural*
				Moves his head back	09:38:56	-	*Behavioural*
				Swallows	09:39:37	-	*Physiological*
				Tightening lips	09:40:04	-	*Behavioural*
				Licking lips	09:40:36	-	*Behavioural*
				Swallows	09:41:08	-	*Physiological*
				Moves index finger on the side-stick	09:41:08	09:41:18	*Behavioural*
				Licking lips	09:41:27	-	*Behavioural*
				Licking lips	09:42:09	-	*Behavioural*
				Licking lips	09:43:05	-	*Behavioural*
				Smirk	09:43:10	-	*Behavioural*
				Deep breath	09:45:01	09:45:03	*Physiological*
				Licking lips	09:45:20	-	*Behavioural*
				Licking lips	09:46:25	-	*Behavioural*
				Moves index finger sullo stick	09:46:34	-	*Behavioural*
				Deep breath	09:47:30	09:47:34	*Physiological*
				Moving index finger sullo stick	09:47:45	09:47:48	*Behavioural*
				Tongue between lips	09:48:17	-	*Behavioural*
Top of Descent	09:48:30	-	Slot 2				
Localiser interference	09:48:30	09:48:56					
				Licking lips	09:49:02	-	*Behavioural*
				Deep breath	09:49:08	09:49:10	*Physiological*
				Licking lips	09:49:38	-	*Behavioural*
				Licking lips	09:50:12	-	*Behavioural*
				Licking lips	09:50:37	-	*Behavioural*
Decision altitude	09:51:44	-	Slot 3				
				Licking lips	09:51:49	-	*Behavioural*
Go around	09:51:56	-					*Performance*
				Licking lips	09:51:59	-	*Behavioural*
				Licking lips	09:52:46	-	*Behavioural*
				PM declares emergency	09:53:20	-	*Performance*
				Tighting lips	09:54:05	-	*Behavioural*
				Licking lips	09:54:31	-	*Behavioural*
				Tense facial expression (crooked mouth)	09:54:59	-	*Behavioural*
				Licking lips	09:55:01	-	*Behavioural*
				Deep breath	09:55:03	09:55:05	*Physiological*
				Licking lips	09:55:37	-	*Behavioural*
				Moving on seat	09:55:40	-	*Behavioural*
				Licking lips	09:56:03	-	*Behavioural*
				Deep breath	09:56:24	09:56:26	*Physiological*
				Licking lips	09:56:57	-	*Behavioural*
				Swallows	09:57:08	-	*Physiological*
				Biting lips	09:57:50	-	*Behavioural*
				Swallowing	09:58:15	-	*Physiological*
				Deep breath	09:58:23	09:58:26	*Physiological*
				Licking lips	09:58:29	-	*Behavioural*
				Licking lips	09:58:44	-	*Behavioural*
				Deep breath	09:58:50	09:58:52	*Physiological*
				Tense facial expression (crooked mouth)	09:58:56	-	*Behavioural*
				Tense facial expression (crooked mouth)	09:59:18	-	*Behavioural*
				Licking lips	09:59:26	-	*Behavioural*
				Licking lips	10:00:19	-	*Behavioural*
Touch down (HAJ)	10:00:37	-					
End	10:00:59	-					

Of course, several other aspects can play a role in the number and type of markers observed during each run, a part from the individual variability and complexity of the run. These aspects are the run sequence, the outcome in the previous run, the interaction with the confederate pilot and pilot's tiredness. It has also to be considered that the number of subjects is quite low for the analysis of an aspect such affected by individual variability.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

However, the experience reported in the debriefings shows that pilots are good in recognising signals of a performance decrement in their colleagues, and that in real operations they use these signals to apply strategies such as task switching and teamwork to recover from a potentially dangerous situation. Looking at pilots' debriefings, it is possible to observe that most reported markers are the following:

- **Decrease of the amount of communication**: most pilots (4) observed that in difficult situations communication becomes more "sterile" and tends to focus only on the execution of the tasks ("no jokes");
- **Change in the communication style** and/or tone of voice: most pilots (3) reported that communication tends to become more straightforward or the contrary more uncertain straightforward, or the voice itself of the person changes – e.g. higher pitch;
- **"Hearing but not listening"**: some pilots (2) reported that they can perceive themselves (or observe colleague) hearing ATC or colleague's communication but cannot really grasp the meaning when experiencing distress.
- **Forgetting things or skipping steps:** some pilots (2) reported they tend to forget and/or skip things. This require them to ask for repetition or makes them unable to resume a task if interrupted;
- **Focussing on primary task only:** some pilots (2) reported they can feel themselves not able to focus on secondary tasks as usual when being in a difficult situation. This is true also when observing a colleague;
- **Sweating:** some pilots (2) reported they feel sweating when facing a difficult situation;
- **Heart rate:** some pilots (2) reported they feel an accelerated heart rate when task demand increase and there is a threat for performance decrease.

For what concerns recovery measures, some pilots (2) reported conscious breath control as a way to get out of cognitive tunnelling. Another "mitigation" measure (for 2 pilots) is looking around the cockpit and/or moving a bit around, in order to disengage from attention fixation.

As next steps of the analysis, the external markers will be triangulated with physiological and performance data, to see if recurrent patterns can be identified and check if behavioural markers can be used to recognise areas in which performance is degraded, or if they are associated to a particular physiological state.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

4.8. Performance measures

Pilot performance was evaluated on the basis of the ability to manually fly the aircraft along a trajectory or along certain target values. Therefore, various simulator data were recorded to calculate aircraft deviations from the trajectory or target values. Deviations of the following parameters were calculated and used as performance measures:

- Speed
- Heading or track
- Altitude
- Vertical speed
- Localiser
- Glideslope

Additionally, the latitudinal and longitudinal point of touchdown on the runway was recorded and was used as a further performance measure. All these parameters were selected because they reflect the pilots' primary task to fly the aircraft. Deviations of these factors from the trajectory and target values can thus be well used as measurements of performance.

4.8.1. Results

The performance data under analysis are the deviation from localiser (LOC) and glideslope (G/S) of the Instrument Landing System (ILS) during final approach between the top of descent of the ILS and the decision altitude. The deviations are indicated in dots, two dots being a full deviation on the localiser and glideslope deviation scale. The localiser deviation of the ILS indicates the deviation to the left or right of the runway centreline. The glideslope deviation of the ILS indicates the upper and lower deviation of the 3° approach slope. The time span with the occurrence of the localiser failure in run 8 is disregarded and not part of analysis in order not to falsify the results.

With respect to the deviation from localiser, Figure 59 shows that Run 1 had a significantly lower deviation from localiser compared to all other runs ($p < .05$) except Run 5 ($p > .05$). The highest deviation was achieved in Run 8 with 0.18 dots. Additionally, Run 5 had a significantly lower deviation compared to run 6, 7 and 8 ($p < .05$), run 6 had a significantly lower deviation compared to run 8 ($p < .05$), and run 8 had a significantly higher deviation compared to run 7 ($p < .05$).

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

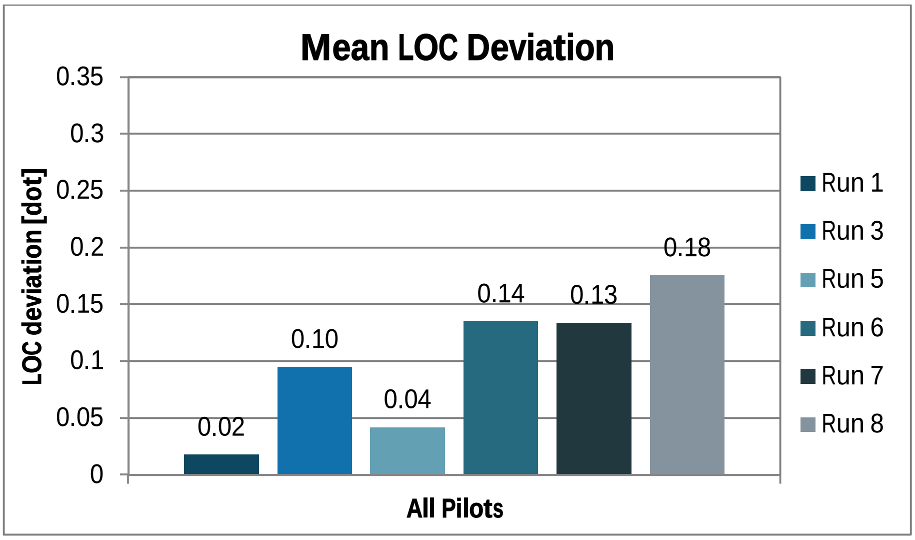


Figure 59: Mean deviation from localiser

Similar results were found with respect to glideslope deviation (Figure 60). Run 1 had a significantly lower deviation from glideslope compared to all other runs ($p < .05$). Additionally, Run 3 had a significantly lower deviation compared to Run 7 ($p < .05$) and Run 8 ($p < .05$), and run 7 and 8 had a significantly higher deviation compared to run 5 ($p < .05$). Run 7 and Run 8 also had the highest deviation from glideslope with 0.29 dots.

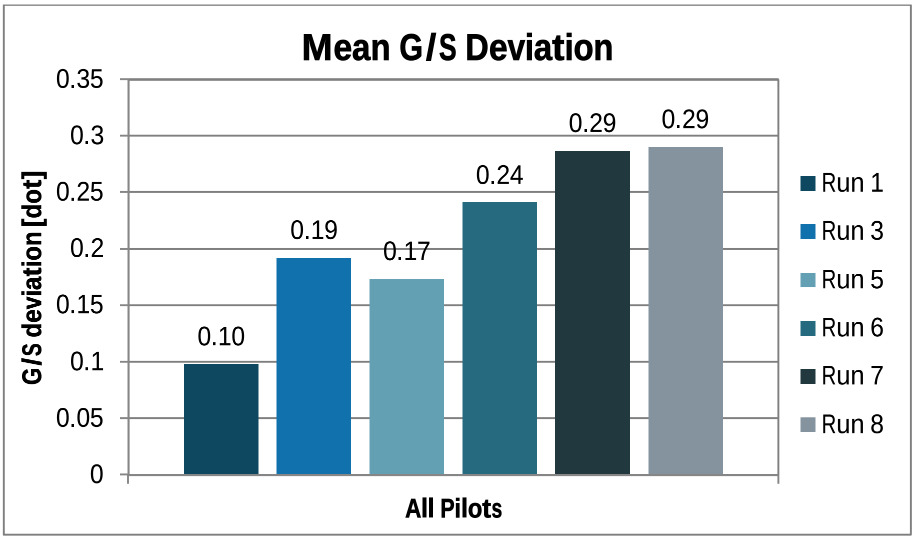


Figure 60: Mean deviation from glideslope

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

4.8.2. Conclusions

Comparing the deviations from localiser and glideslope of all runs with each other needs to be done carefully, as in some runs (run 1, 3, 7, 8) turbulences were enabled compared to some other runs (run 1, 5, 6). Nevertheless, some clear conclusions can be drawn, especially of the runs with turbulences enabled.

A comparison of run 3 and 8 shows that the deviations from localizer and glideslope were much higher in run 8. This highlights that if several HPE factors are degraded at the same time, a lower pilot performance results compared to if only one HPE factor is degraded. Furthermore, a comparison of run 7 and 8 shows that a higher degradation of the same combined HPE factors leads to a lower performance.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

5. FURTHER/FUTURE ANALYSIS

5.1. Competencies analysis – methodology

5.1.1. Introduction

To assess the performance of the Pilot Monitoring, a continuous competency based rating technique as introduced in D6.2, was used. In comparison to the PF, for the PM competencies objective indicators such as inputs to aircraft systems or the efficiency of flight path are lacking. The competencies are of such nature that until now is it not possible to assess them in an automated, objective way. A tool was developed allowing an observer (e.g., an instructor pilot) to play the videos from the PM scenarios and meanwhile provide a rating for the pilot performance on the following three competencies:

- Situation Awareness,
- Problems Solving and Decision Making,
- Application of procedures.

Early in the project, it was recognised that performance measures for the PM scenario were needed, to triangulate the results and validate the HPE.

Current competency rating methods used in the field of simulator training and research have the disadvantage of having in general a low interrater reliability. In most methods, a final score is given to the complete scenario whereas the performance may change in the course of a scenario. The final rating provides limited insight. The observers are provided a scenario description and example behaviours per event and the corresponding rating, to standardise the rating frameworks and thereby increasing the interrater reliability.

Through the tool, the observer is able to monitor the selected competencies while playing a video recording of a scenario, and provide a rating for the PM performance on these competencies, at any given time. A four point scale is provided but the rating is set by means of a slider that allows any ratings in between two discreet values. Each change in performance rating is time stamped and saved. This also is expected to stimulate rating segments of a scenario one by one, which should again increase the reliability.

In the next sections the design and development of the tool, the application of the tool on the pilot monitoring experiment, and results are described.

5.1.2. Design and Development

The development of the tool was performed in three rapid prototyping iterations. The concept and the tool were evaluated in small experiments. This section describes the process of development and intermediate evaluations.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

5.1.2.1. *The tool*

The tool integrates the competency framework, the administration of ratings and the video that needs to be assessed.

The first build prototype (Figure 61) was built around custom made online media player which could present the observer with pre-defined video playlist. Next to the video window interactive competency rating sliders were presented. The sliders were accompanied with a four point coloured rating scale (Exceeds, Meets, Below, Unacceptable), and basic competency descriptions for each of the four scale points. Beneath the video a window was positioned that presented each the competency rating value graphically over time.

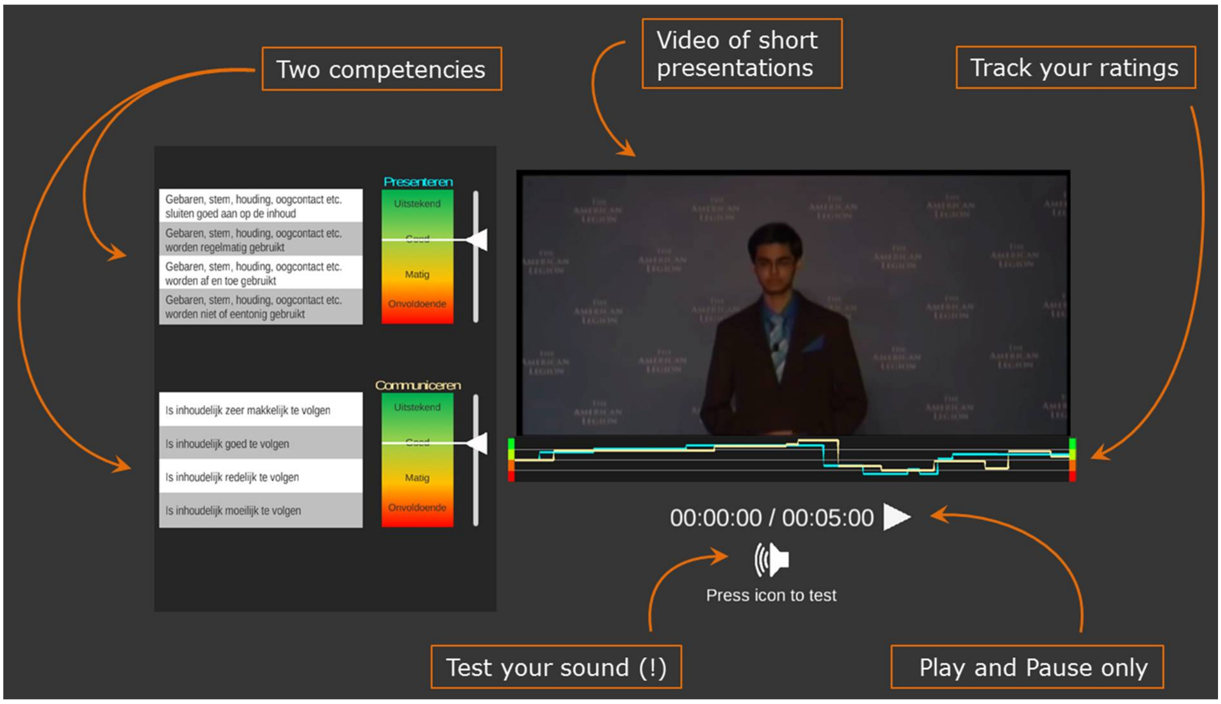


Figure 61: First prototype with two competencies and a video presentation about general topic

The target audience for the first prototypes were colleague researchers from within the department. Due to the mixed academic backgrounds of the prototype target audience, videos were selected that didn't require aviation competency knowledge, containing five student presentations about general topics about 1 minute each. 'Presenting' and 'Communicating' were selected as competencies to make it straight forward the prototype target audience. With the aim to standardize the rating framework the observers' frame of mind was supported with scenario specific performance indicator in the form of an additional table. The table contained performance indicators for each competency and for each of the ratings on the four point scale.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

5.1.2.2. *Intermediate evaluations*

With the goal of obtaining high fidelity competency ratings in mind, it was required that the observers would review and rate all competencies rating as often as possible, while watching the video. With the first A/B test a prototype of the application, the concept and the HMI were evaluated. Additionally, we were interested if a timed notifier would influence the rating frequency and attention distribution between the multiple competencies.

Since it is difficult to get a large sample size with pilot instructors, we changed the subject of the performance measurement for this A-B test, allowing inviting other people to join the A-B test. The performance measurements were done on short presentation. The participants were asked to rate two competencies: presentation skills (non-verbal) and communication (verbal). We expected a general consensus about the interpretation of these competencies, in order to replace the 'expert role'.

The randomized experiment contained one version without notifier (Version A) and one with notifier (Version B). In the version with notifier, the competency rating slider knob had its own built-in timer. Without any observer input for 30 seconds the rating button would gradually colour- and started blinking red after a period of 30 seconds. The notifier could be deactivated by clicking again on the knob to inform the current rating value is still valid or by sliding it to a new competency value. In total 19 participants tested the application, 16 of them provided valid data. The participants were randomly assigned to the application with and without notifier. Nine participants tested the application with notifier and seven participants tested the application without notifier. After the experiment the observers were asked to fill out a questionnaire with questions about the concept in general, the tool's HMI and (if applicable) their experience with the notifier.

The results included:

- In general, the users were satisfied with the application (intuitive, ease of use, design etc.).
- The users prefer a dynamic slider in comparison with static rating buttons.
- The notifier had no influence on the frequency of rating and was experienced as slightly negative.
- Rating two competences simultaneously does not require a high workload, according to the users. However, without training you cannot take all the specific indicators in account.

In a second test, the C-D test, the objective was to determine the effect of the rating standardization with the scenario specific indicators on the interrater reliability (Inter Class Coefficient, ICC). In total 14 participants tested the application. The participants were randomly assigned to the application with and without scenario specific indicators. Seven participants tested the application with specific indicators and five participants tested the application without specific indicators.

Overall interrater reliability of the C-D test (ICC .079 to .215) was lower, compared to the results of the A-B test (.625 to .756). The participants, who received the specific indicators as part of their training, showed a slightly higher reliability for one of the competencies (communication) than the participants

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

who did not received the specific indicators. However, this is not the case for competency 1 (presentation). We did not have strong evidence that the training with specific indicators has influence on the reliability of the ratings. Nevertheless, the specific indicators are included in the final design for application in assessing the PM competencies in the project. The indicators are there to clarify ambiguous competences, which help to increase the overall reliability of the observation scores.

5.1.2.3. *Final Design*

For the Future Sky Safety experiments it was decided to include the notifier. The experiment sessions last for 45 minutes. A notifier might help the participants to rate more frequently and was set at 5 minutes.

Also the scenario specific indicators were made available to the observers. The indicators are there to clarify ambiguous competences, which help to increase the overall reliability of the observation scores.

Figure 62: The rating tool with an experiment unrelated video for instructional demonstration purposes

While the intermediate evaluation tool versions only required the observers to watch one short video, the final tool was fitted with additional software logic to handle more complex features like individual user logins, handle multiple videos, and tracking of the progress. Whereas the intermediate tools were only trailed within NLR's internal computer network, the final tool was required to keep all data safe while instructors were accessing the online tool from other parts of the world. The login system made sure all sensitive content was protected from public access and made it also possible to personalize the rating sessions for each instructor. Each instructor was provided with a personal anonymized login id. Using the login ID the tool could automatically load from the pre-defined database which videos the instructor had

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

to observe and in which order. During an actual rating session the tool automatically logged the progress over time of each flight video session in the database. This made it possible for instructors to close the rating tool and continue at a later moment without losing the previous logged rating data. After finishing each video the next video would automatically loaded into the tool. Each logged rating in the database was coupled with the anonymized login id, making it with the data analysis possible to group ratings for each flight session video per individual observer.

5.1.3. Application of the competency rating tool in the PM Experiment

For the application of the competency rating tool to the data collected in the PM experiment the following data streams are relevant (Figure 63):

1. The scenario 2 experiment was run on the DLR AVES simulator,
2. Ten subjects participated in the experiment,
3. The video camera recorded each subjects performance and behaviour from an instructors point of view,
4. The ten videos were loaded in a predefined order into the online rating tool,
5. Instructors were invited to remotely observe each subject through the online tool on three competencies using the standardized rating framework,
6. All the competency ratings were stored in a secure database.

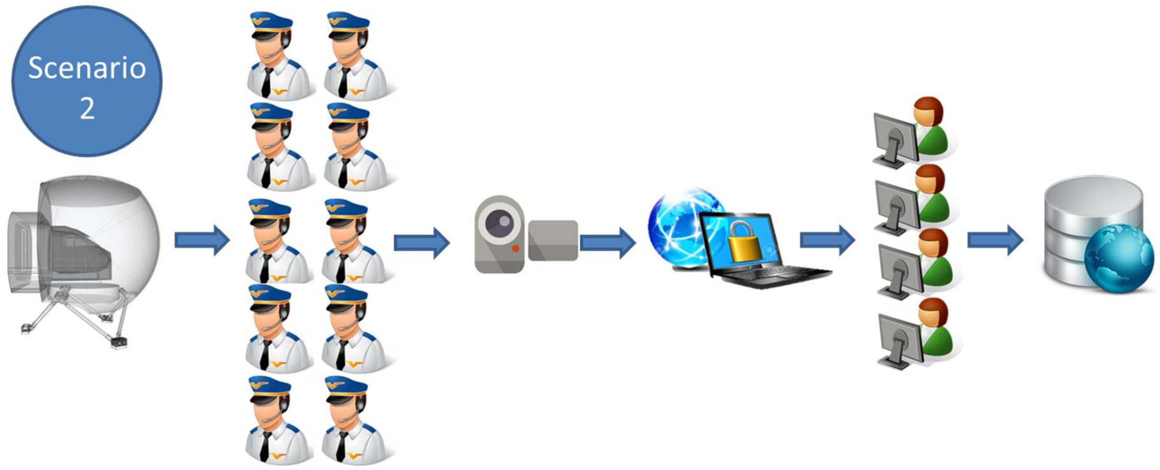


Figure 63: Overview of the rating tool data in and output

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Online hosting

To make the tool accessible for all worldwide invited observers the tool was hosted on an NLR hosted server containing also the website with training material, the scenario 2 video's and the database storing the actual rating data. The server access to this experiment was secured with a login system that required login details to protect unwanted access to the sensitive experiment 2 video data. Once experiment setup was completed, the login details were personally sent to the invited instructors.

Tool instructions

After login in on the secure website the observers were presented with instruction material, with the goal of the experiment, how the tool works, what they could expect during the flight scenario and what they were supposed to do.

Initially the instructors were presented with instruction tool HMI and informed they only had to observe the three competencies from the pilot monitoring in the right seat. Subsequently the instructors were presented with the general and scenario specific indicators (0). Again, this was to set the observers frame of mind by providing everybody with standardized the rating framework. Instructors were informed it that it was of great importance to be familiar with the competencies and rating levels. For each of the three competencies, performance indications were given on the four point rating scale and additionally divided per flight phase and that contained examples on the four point rating scale. It was additionally suggested to open the detailed performance indicators on a second computer screen or print the table so the indicators could be checked whenever needed.

The next instructional step was to present the flight phase details (0) of scenario 2 which starts in the vicinity of Bremen Airport in Germany; starting with a descent phase towards runway 27, a go-around, followed with an workload increasing AC BUS 1 fault, taxing the FO with new LAPA runway calculations for runway 27. This result into the conclusion that runway 27 is unsafe requiring the FO to do another LAPA calculation for runway 09 and subsequently the last phase is the approach to runway 09 itself (Figure 64).

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

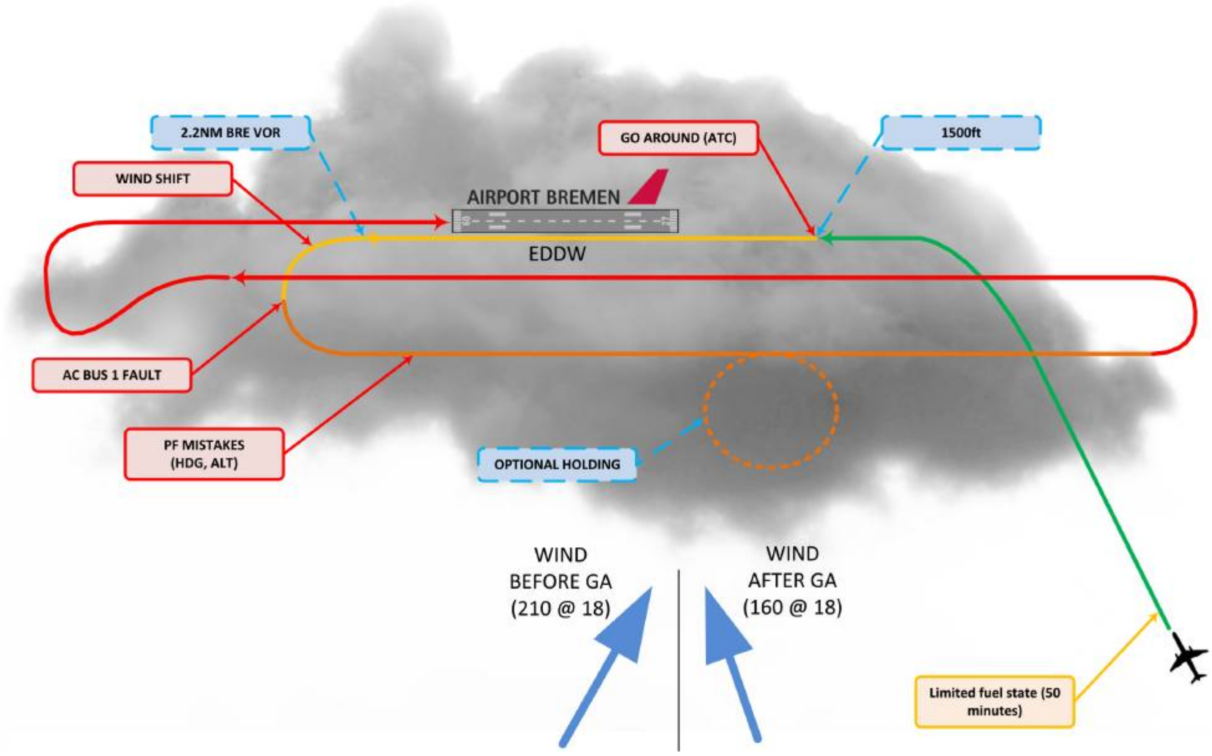


Figure 64: Scenario 2 flight phases and events

As final instructional step before starting the actual experiment the instructor were able to test the rating tool at their own pace using a demonstration tool (Figure 62). The demo tool replicated the actual tool, but contained a different video and did not store any ratings into the database.

Observers

To collect data using the online tool, observers were invited to voluntarily use the online application starting in their spare time. The online data collection was open from the second week in June until the second week of September 2016. The observers were all qualified ATPL airline captains, with an instructor rating or with an academic aerospace background; to make sure they had experience with competency ratings of pilots.

Method

Due to the voluntary nature to participate as observer, it was anticipated that not every observer could observe and rate video recording of each individual pilot, which implicated watching 10 videos with more than six and halve hours of video material, excluding the training of the online tool. The ten, to be rated, pilot video recordings were therefore prioritized into a flight session playlist on the basis of observed events that would provoke a rating change by the observer. Each observer, using the online tool, had the same video playlist order which unfortunately would not cancel out the possible learning effect over time. This method was nevertheless chosen to maximize the possibility of having multiple independent ratings on the prioritized videos, making it possible to later-on further analyse the ratings in-between instructors.

5.1.4. Results

A total of 1116 unique rating were made on the three competencies situational awareness, decision making and application of procedures. Ratings were provided by four independent observers with the online tool. The observers rated in total 21 video's (Table 23). One observer rated all ten pilots and flight sessions with pilots 3, 4, 5, and 10 were rated multiple times making it possible to determine the consistency and intra-class correlation between observers.

Table 23: The number of rating changes per flight session and instructor

Pilot	1	2	3	4	5	6	7	8	9	10
Observer										
1	41.0	41.0	23.0	16.0	10.0	30.0	41.0	34.0	30.0	18.0
2			132.0	148.0	139.0					130.0
3			34.0	34.0	X					28.0
4			58.0	33.0	46.0					50.0

As an example, the ratings for pilot 3 are depicted in Figure 65 to Figure 67. More results are included in Appendix D.

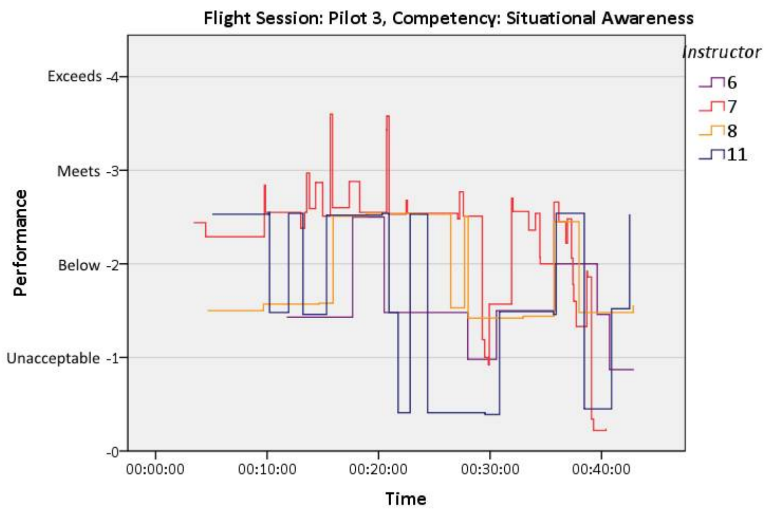


Figure 65: Pilot Monitoring performance ratings on situation awareness for pilot 3, by four observers

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

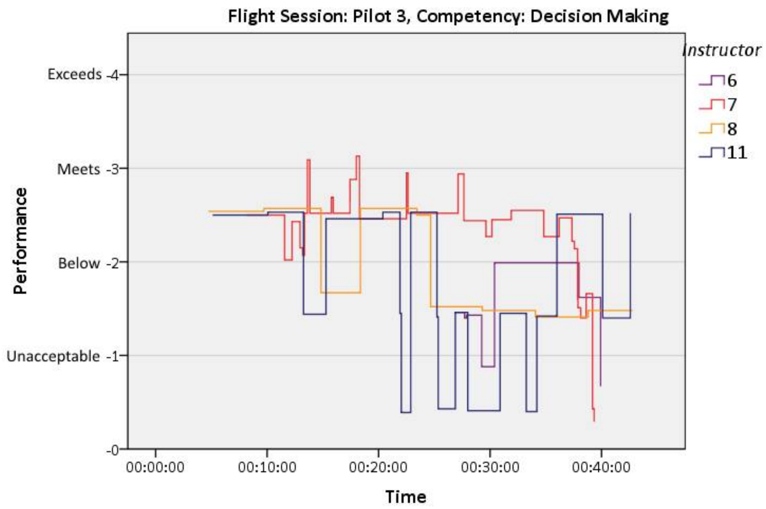


Figure 66: Pilot Monitoring performance ratings on problem solving and decision making for pilot 3, by four observers

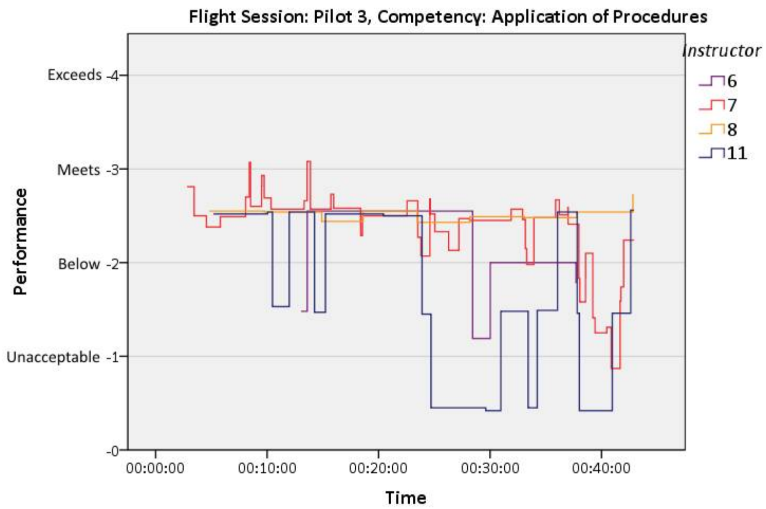


Figure 67: Pilot Monitoring performance ratings on application of procedures for pilot 3, by four observers

Reliability

In order to compute the reliability of the ratings, the Intra-class Correlation Coefficient (ICC) is used, in particular the two-way random single class measures ICC(2,1) and average measures (2,2), together with the other statistical measures: mean, standard deviation, t-tests and confidence intervals. According to the literature, the ICC can be interpreted as follows: 0-0.2 indicates poor agreement; 0.3-0.4 indicates fair agreement; 0.5-0.6 indicates moderate agreement; 0.7-0.8 indicates strong agreement; >0.8 indicates almost perfect agreement. In this study we used 0.5 as the consensus threshold.

Since the timing that the instructors change their rating is not predetermined and instructors change their rating when they feel the performance is changing, there are large variations in the timestamps. This makes it hard to compare the data points and made it necessary to resample the data to come to equal intervals over which the average rating was calculated and used for calculating the ICC.

Several resampling methods were used to evaluate the ICC:

- the rating every 5 seconds,
- the average rating at intervals of 1 minute
- the average rating at intervals of 5 minutes

The ICC is calculated per competency, for each flight session. The averages of each test are depicted in Table 24.

Table 24: Average Intra-class Correlation Coefficient for four raters at different levels of detail for four flight sessions

	Group		Situational Awareness	Decision Making	Application of Procedures	Total average
Average	Score at	*5 sec*	0.19	0.26	0.23	0.23
ICC	Mean per	*1 min*	0.17	0.33	0.25	0.25
	Mean per	*5 min*	0.42	0.39	0.31	0.37
	Total average		0.26	0.33	0.26	0.28

*ICC Two Way Random, Single Measures

The average Intraclass Correlation Coefficient for ‘Situational Awareness’ is ICC (2,1)= .262, for ‘Decision making’ is ICC (2,1)= .311, for ‘Application of procedures’ is ICC (2,1)= .250. All the average ICC scores are below the threshold of 0.5. In conclusion, using the data of a single rater would be inappropriate.

Since we collected data of multiple raters for 4 flight sessions, the ICC for average rating was also calculated (Table 25).

Future Sky Safety has received funding from the EU’s Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Table 25: Average Intra-Class Correlation Coefficient between the average ratings and the four raters at different levels of detail

	Group		Situational Awareness	Decision Making	Application of Procedures	Total average
Average	Score at	*5 sec*	0.45	0.53	0.51	0.5
ICC	Mean per	*1 min*	0.48	0.6	0.48	0.52
	Mean per	*5 min*	0.72	0.65	0.61	0.66
	Total average		0.55	0.59	0.53	0.56

*ICC Two Way Random, Average Measures

The average Intra-class Correlation Coefficient for ‘Situational Awareness’ is ICC (2,2) = .550; for ‘Decision making’ is ICC (2,2) = .593; for ‘Application of procedures’ is ICC(2,2) = .533.

The ICC with average ratings yielded better results. The ICC values above 0.5, the level of agreement is considered of sufficient reliability for determining the performance.

What is also obvious from both Table 24 and Table 25, is that by increasing the sampling interval, and calculating the average rating within the interval, the ICC increases.

5.1.5. Discussion

A tool was developed allowing for rating pilot monitoring performance on a continuous time scale for the competencies: Situational Awareness, Problem Solving and Decision Making and Application of procedures. In total four observers made ratings of four of the flight sessions of scenario 2, one of them for all 10 flight sessions. The ratings allow for the comparison with other data from the experiment in order to provide a measure of performance for the pilot monitoring.

From the results of four raters over four flight sessions, the calculated ICCs yielded the following results:

- The ICC increases slightly by increasing the sampling window to e.g. five minutes.
- The interrater reliability of the continuous ratings appeared to be too low to use the ratings from a single instructor, independent of the sampling method.
- The average of the ratings made by three to four instructors is sufficiently reliable. For Pilot 3 the ICC (2,2) value was relatively high. Therefor the data of pilot 3 will be used, and in specific the average scores (between the 4 raters) as data for performance in the comparison with other data (such as mental representations of CATIE).
- The number of competencies (3) may be too high to simultaneously monitor. In previous experiments with two competencies to rate the ICC values were higher.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

5.2. Cognitive task analysis – methodology

5.2.1. Test runs

A panel of 10 Pilots has been chosen to perform the scenario 2. During the runs, an expert panel, integrating human factor experts and experts of flight simulations have observed and monitored the pilots devoted to the "Pilot Monitoring (PM)" responsibility/role (Figure 68). The analysis was focused on their behaviour, communication, application of procedures and decision-making.

Figure 68: Control room during the test of scenario 2

After the simulation, a cognitive walkthrough has been performed. The videos of the 2 webcams were synchronized using the software "Noldus observer XT" to support temporal events of the runs (coming back). The first camera was used to recover an overall vision of the cockpit (panorama) and the second one was implemented to observe the use of the EFB (Figure 69).

Figure 69: Panorama and EFB view

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

During the interviews, a guideline has been used to structure the exchange and to recover the information within the same pattern for each pilot. The scenario has been segmented in seven phases (during which decisions making were expected from the Pilot Monitoring – PM) and each phase was divided in "cues" (inputs needed), "what has to be understood" (state of the situation), "what are the options" (what the PM can do), "what has to be done" (to control the situation). For instance, Figure 70 shows the guideline of the phase 5.

Pilot X	**PHASE 5- Third LAPA Calculation (RW09)**	
Description	Weather like above3. LAPA calculation analyzes RWY09 but will only be possible with given Abnormal	
Trigger event	"Could not land on RW 27..." well - "Cannot land on runway 9 as is"	
BASELINE	OBSERVED	QUESTIONS FOR COGNITIVE WALKTHROUGH
Cues ?? - on EFB, the LAPA HMI provides info (highlighted when there is a problem) - have to do an "automatic landing", but... - ...in the OM-B, "warning no auto roll out possible"		- Why did you do a third LAPA calculations ? (too much wind crosswind, no automatic rollout possible) - How does that change your vision of what you have to do ? - Is this a situation very peculiar ?
What has to be understood - LAPA calculation analyzes RWY09 but will only be possible with : • Given Abnormal • Weather change • "Emergency only" in EFB - in the OM-B, PM must realize that only a manual rollout is possible (no automatic roll out, so the pilot will have to manually leave the runway after landing).		
What are the options - Declare "emergency" - Land elsewhere (not played in this scenario)		- Do you have a lot of practice in this kind of procedure ?
What has to be done - Declare "Emergency" in EFB Redo the LAPA with: • Given Abnormal • Weather change • "Emergency only" in EFB - The PM must refer to the OM-B to discover that only a manual rollout is possible and prepare for that.		- Does it still manageable ?

Figure 70: Cognitive walkthrough guideline for phase 5 "Third LAPA Calculation RWY09"

The interviews of the Cognitive Walkthrough were started by an explanation of the project objectives and the confidentiality of data. After that, the scenario was explained through the 7 phases and the pilot was requested to focus on the events/actions phase by phase.

Pilots were requested about their situation awareness to understand how the decision where taken. Experts in flight simulations provided their support to understand different situations, interpreting in real time what information was needed or forgotten. Guidelines and questionnaires of the Cognitive walkthrough are presented in the Appendix E.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

5.2.2. Data processing

The videos taken into the cockpit were synchronized with the Noldus Observer and a quick identification of the nodes was performed to facilitate the analysis. Thus, an exhaustive research of relevant inputs, events, decision and actions was carried out (Figure 71) and the remaining time of flight was calculated (in time and kilograms of fuel). Figure 72 shows an example of the events for the pilot 1; all other data are presented in the Appendix F.

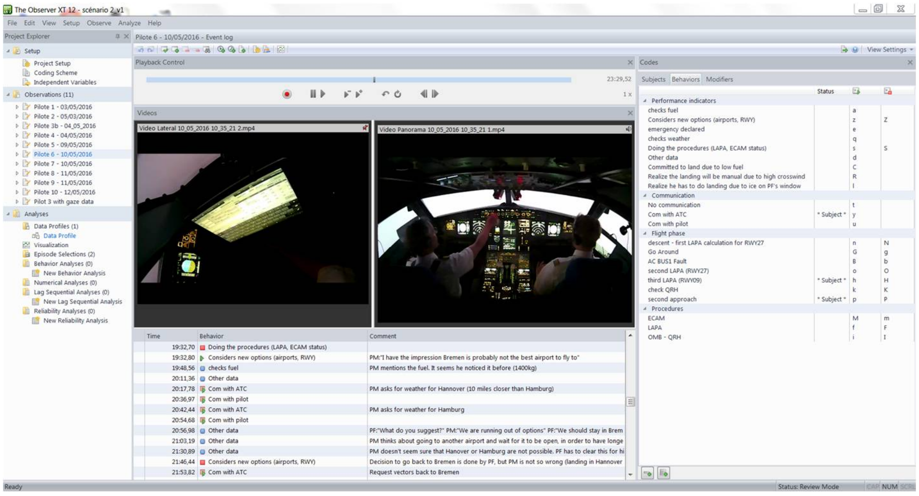


Figure 71: Characterization of information with Noldus Observer XT

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Pilot 1	Video time	Scenario time (VT-0'00'')	FOB (kg)	Time remaining (regarding FOB)	stickers for representation
Start of the recording	00:00	NA	NA	NA	NA
Actual start	00:00	00:00	2000	00:50:00	(50'00''/2000kg) Actual start
Start LAPA 27	02:34	02:34	1897	47:26	(47'26''/1897kg) Start LAPA 27
Finish LAPA 27	04:17	04:17	1828	45:43	(45'43''/1828kg) Finish LAPA 27
Go Around	05:27	05:27	1782	44:33	(44'33''/1782kg) Go Around
Request vectors for another approach	07:52	07:52	1685	42:08	(42'08''/1685kg) Request vectors for another approach
Climb 4000 (doesn't notice)	07:56	07:56	1682	42:04	(42'04''/1682kg) Climb 4000 (doesn't notice)
Low on fuel by PM	08:03	08:03	1678	41:57	(41'57''/1678kg) Low on fuel by PM
AC Bus Failure	08:38	08:38	1654	41:22	(41'22''/1654kg) AC Bus Failure
PF begins ECAM procedures	09:26	09:26	1622	40:34	(40'34''/1622kg) PF begins ECAM procedures
1.4T of fuel left (mentioned by PM)	13:13	13:13	1471	36:47	(36'47''/1471kg) 1.4T of fuel left (mentioned by PM)
PM continues ECAM procedures	13:34	13:34	1457	36:26	(36'26''/1457kg) PM continues ECAM procedures
New weather, given by ATC	15:00	15:00	1400	35:00	(35'00''/1400kg) New weather, given by ATC
Consider new airport	15:35	15:35	1376	34:25	(34'25''/1376kg) Consider new airport
Ends ECAM procedures	16:59	16:59	1320	33:01	(33'01''/1320kg) Ends ECAM procedures
Overall LAPA calculation (no specific runway)	17:38	17:38	1294	32:22	(32'22''/1294kg) Overall LAPA calculation (no specific runway)
Other airports denied	18:38	18:38	1254	31:22	(31'22''/1254kg) Other airports denied
ATC gives vectors back to Bremen	18:43	18:43	1251	31:17	(31'17''/1251kg) ATC gives vectors back to Bremen
PM declares emergency	19:18	19:18	1228	30:42	(30'42''/1228kg) PM declares emergency
End of overall LAPA	21:08	21:08	1154	28:52	(28'52''/1154kg) End of overall LAPA
PM decides to go on RWY09	21:22	21:22	1145	28:38	(28'38''/1145kg) PM decides to go on RWY09
PM does a new LAPA, for RWY09 this time	22:20	22:20	1106	27:40	(27'40''/1106kg) PM does a new LAPA, for RWY09 this time
End of LAPA, PM knows he can land with emergency	23:45	23:45	1050	26:15	(26'15''/1050kg) End of LAPA, PM knows he can land with emergency
No automatic landing possible	24:54	24:54	1004	25:06	(25'06''/1004kg) No automatic landing possible
Enters RWY09 in FMS	26:58	26:58	921	23:02	(23'02''/921kg) Enters RWY09 in FMS
PM asks for fire brigade	27:27	27:27	902	22:33	(22'33''/902kg) PM asks for fire brigade
Briefing for RWY09	28:52	28:52	845	21:08	(21'08''/845kg) Briefing for RWY09
PM knows he has to disengage AP at 80''	29:28	29:28	821	20:32	(20'32''/821kg) PM knows he has to disengage AP at 80''
PF asks to go check OM-B	30:10	30:10	793	19:50	(19'50''/793kg) PF asks to go check OM-B
PM found nothing important in OM-B	31:56	31:56	722	18:04	(18'04''/722kg) PM found nothing important in OM-B
PF asks to go check QRH	32:18	32:18	708	17:42	(17'42''/708kg) PF asks to go check QRH
PM finds nothing in QRH	33:33	33:33	658	16:27	(16'27''/658kg) PM finds nothing in QRH
Clear to land	34:12	34:12	632	15:48	(15'48''/632kg) Clear to land
PM checks weather	34:35	34:35	616	15:25	(15'25''/616kg) PM checks weather
Committed to land	35:29	35:29	580	14:31	(14'31''/580kg) Committed to land
Ice on window	36:09	36:09	554	13:51	(13'51''/554kg) Ice on window
PF switch	36:45	36:45	530	13:15	(13'15''/530kg) PF switch
AP/AT disengaged	37:00	37:00	520	13:00	(13'00''/520kg) AP/AT disengaged
Touchdown	37:13	37:13	511	12:47	(12'47''/511kg) Touchdown

Figure 72: Time events for Pilot 1

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

5.2.3. Data structure

The information was analysed to construct the mental representation of the pilot and to define its impact over 3 parameters: workload; airport and runways selection; and limitations for landing.

To facilitate the comparison between pilots and the “expected behaviour”, the information was structured as follow (Figure 73):

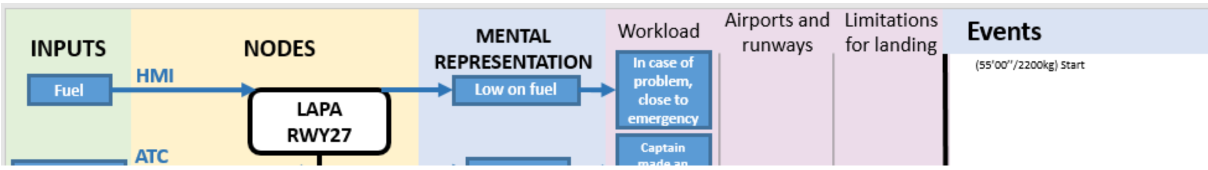


Figure 73: Structure of scenario 2 and mental representation of the pilot

- Inputs: Relevant information coming from the HMI, the Air Traffic Controller (ATC) or the captain (PF – Pilot Flying). The inputs are the cues that should create/change the mental representation.
- Nodes: or “scenario phases”. Actions and moments during the scenario where the pilot must perform procedures or take decisions.
- Mental Representation: how the input and the situation were understood (meaning, impact, consequences, etc.)
- Workload, Airports and runways, Limitations for landing: Impact of the mental representation on each parameter. They will provide an approach about the impact of misinterpretations.
- Events: Timeline with the events observed during the run (Figure 72).

Legends and colour code:

The colours of the boxes are an indication of the performance of the pilot to recover the information (good source and time) and the level of understanding of the input and awareness:

➢ In the **INPUTS** column:

- **green** Means that the PM perceived the input better than expected (searching, knowledge, briefings, etc.)
- **blue** Means that the PM perceived the input without any help.
- **cyan** Means that the PM perceived the input in an acceptable manner but not with the best performance.
- **red** Means that the PM missed the cue, and that the captain or the ATC gave him the correct information

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Every input has been linked with a legend that define the source of the information (PF: Pilot Flying; ATC: Air Traffic Controller; LAPA: Landing Calculation; OMB and QRH: Operational Manuals; HMI: Airplane instruments)

- In the **MENTAL REPRESENTATION** column:
  - **green** Means that the PM understood the situation better than expected (anticipating decisions)
  - **blue** Means that the PM understood the input and situation without any help.
  - **cyan** Means that the PM understood the input and situation with some help.
  - **red** Means that the PM didn't understand the input correctly, and that he didn't have an acceptable representation of the situation. The box means that the pilot didn't perceive an input.

Grey boxes in the NODES column indicate decisions or procedures that were not performed (i.e. consider new airport or OMB procedure (Figure 74).

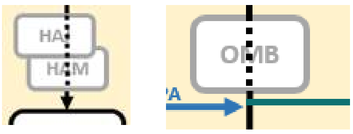


Figure 74: Nodes column: Procedure not performed

5.2.4. Independent data analysis

An "expected" escenario has been constructed to compare the variations of mental representation between pilots (Figure 75).

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

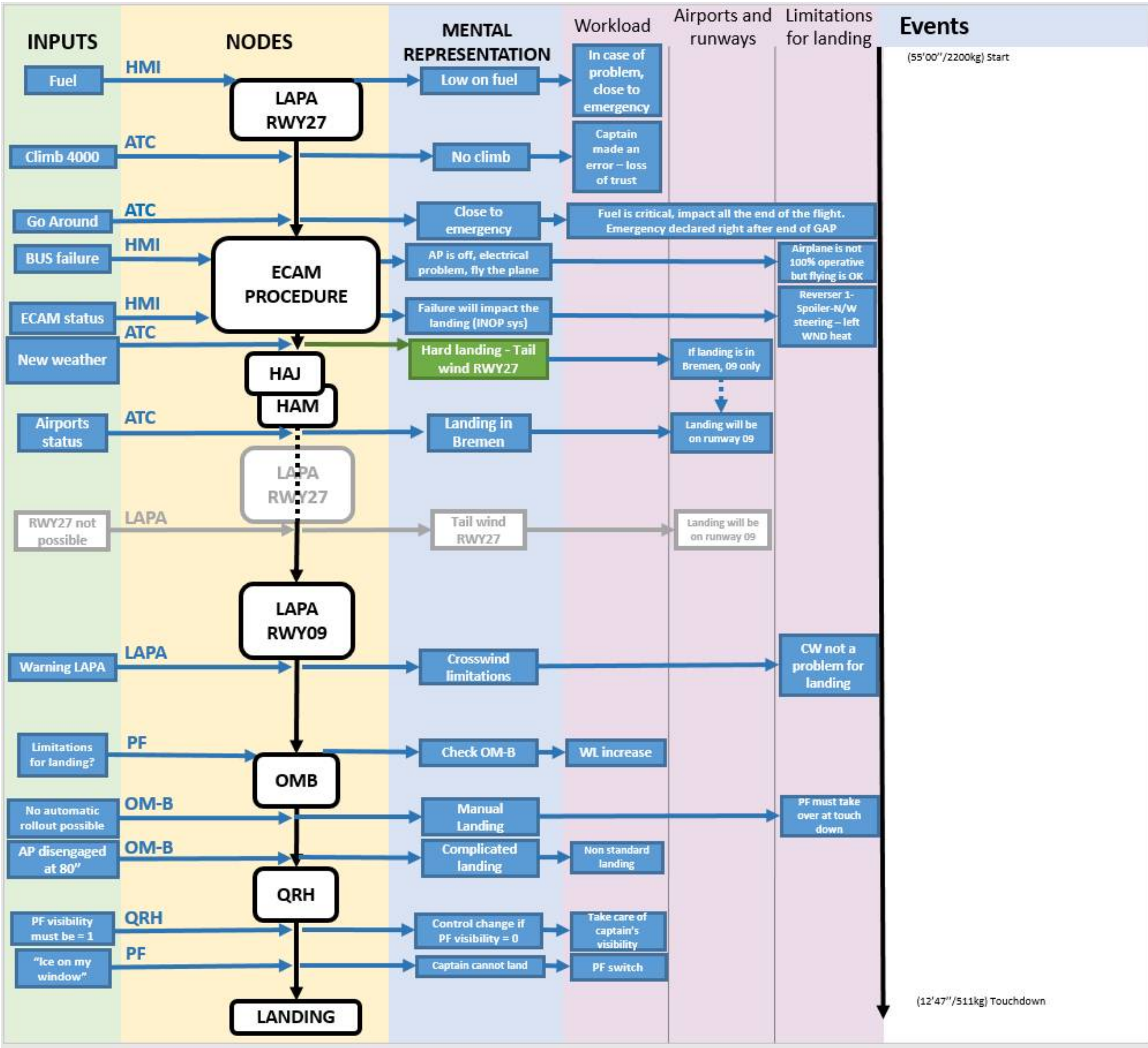


Figure 75: Expected scenario and mental representation

In this graphic, the pilot should understand the New Weather provided by the ATC 

as a limitation to land in runway 27 (RWY27) 

. Moreover, this situation awareness should allow him to avoid the calculations for landing (LAPA) in RWY27 (Figure 76).

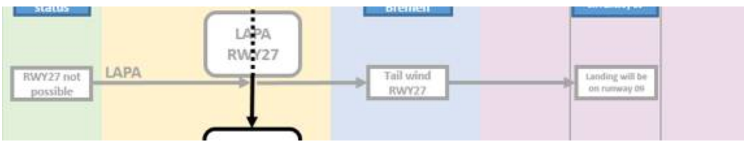


Figure 76: Interpretations of wind shift should avoid the landing calculations of RWY27

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

6. CONCLUSION

This deliverable will be followed by D6.4 “Recommendations recovery measures and HMI implementation”, which will include the methods and principles developed to recover performance if outside of the human performance envelope and the first guideline for HMI development taking into account one dedicated concept of automation.

There is a significant literature on how to measure factors such as workload, situation awareness, stress, fatigue, etc. (see D6.1). But most studies in the formal human performance literature only measured single factors, and many studies were in low-fidelity contexts, e.g. with lab studies and students, rather than a full-scope simulator and certified line pilots. Furthermore, many studies reviewed did not really ‘push’ the pilots out of their comfort zone, in realistic but challenging and unexpected scenarios.

The first step was to collect the key measures and sensors and try them out in a lab setting; if they didn’t work in a neatly controlled environment, they would definitely not ‘fly’ in a full-scope simulator. In order to do this, two small-scale studies were performed by ONERA and Cranfield University, testing a battery of measures. This led to the identification of sensors and measures which could be applied in the full-scope simulator study.

The second step was to develop realistic and challenging scenarios. For this step there were two sub-goals. The first, very focused on the HPE, was a set of escalating variants of a scenario (Scenario 1), going from (Run 1) very routine flight with no problems, to (Run 8) a very difficult situation with multiple factors affecting the pilots. This set of runs, although having high realism, was really focused on ‘internal validity’, enabling us to explore how the factors affected performance and how they interacted with each other, and to see how good each individual sensor/measure performed, alone, or in conjunction (triangulated) with other measures. The second scenario (Scenario 2) was more concerned with external validity, focusing on how pilots used the HMI in a stepwise series of challenges during a single run of around 40 minutes duration. So, broadly speaking, Scenario 1 asked whether the HPE is a useful construct that we can manipulate and measure effectively, and Scenario 2 asked whether these types of simulations grant us insights enabling us to improve the HMI and cockpit safety.

The third step was to run the scenarios with a full scope state-of-the-art A320 simulator and ten commercial pilots (all male). Due to a compressed timescale and occasional simulator problems (as is normal), not all pilots completed all eight runs of Scenario 1, though they all completed the main runs and Scenario 2. Inevitably some data losses occurred, but it was possible to evaluate all the sensors and measures, and the pilots reported unanimously that the scenarios (both Scenario 1 & Scenario 2) were realistic and challenging.

The fourth step was the analysis of a very large and complex dataset, requiring a multi-disciplinary team with different analytic skill sets. For the focus on the HPE, three factors were manipulated: workload, stress, and situation awareness. Fatigue was also measured to make sure that pilots were not tired, as otherwise fatigue may have ‘polluted’ the results. The analysis used a battery of measures:

- Physiological measures
- Behavioural measures
- Subjective/cognitive measures
- Performance measures

The best results were for measuring mental workload and stress – measuring situation awareness was more challenging. Additionally, the results indicated that certain combinations of factors did indeed combine to degrade performance significantly. One of the behavioural measures in particular showed that the 'edges' of safe performance were sometimes touched or even exceeded. Thus, the HPE is a useful concept and can be manipulated and measured in a realistic operational context.

Furthermore, the results have shown that physiological measures such as HR, SDNN, HF, LF and VLF can be sensitive to an increase in workload and/or stress. The run with situation awareness degradation on the other side showed very often no clear pattern and remains therefore difficult to interpret. HR and SDNN were particularly sensitive to the increase in workload, while the HRV features derived from the spectral analysis (HF, LF and VLF) showed a significant response to the increase of stress as well. The single pilot versus group analysis showed the importance of normalizing HR values when conducting the group analysis (in particular the phase 2 analysis). This is not surprising since an "absolute" HR value is much more subject dependent than HRV that express a type of "variability" which results in HR being more sensitive to inter-subject variability.

Scenario 2 analysis will be provided in D6.4a. It will show that the conducted simulation can indeed yield new insights into cockpit HMI design.

The wider applicability of this work and this current deliverable is that the HPE concept has to an extent been validated, and should be considered not only by aviation, but in other contexts where multiple factors can impact on safe human performance, and where there is a desire to protect safety in non-nominal situations. An additional outcome for aviation concerns training of pilots for non-nominal scenarios and emergencies. Several pilots commented that these scenarios were harder than those they undertake for recurrent training. So there may be insights for training, especially as a number of accidents relate to having to recover in relatively sudden high stress/workload situations.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

7. REFERENCES

Chételat, O. et al. (2011). A Highly Integrated Wearable Multi-parameter Monitoring System for Athletes. *15th Nordic-Baltic Conference on Biomedical Engineering and Medical Physics (NBC).* Aalborg, Denmark, pp.148-1451.

Chételat, O. et al. (2015). Clinical validation of LTMS-S: A wearable system for vital signs monitoring*. 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC),* Milan, 2015, pp. 3125-3128.

Cinaz, B. et al. (2013). Monitoring of mental workload levels during an everyday life office-work scenario. *Personal and Ubiquitous Computing*, *17*(2), pp. 229-239

Edwards, T. (2013). Human Performance in Air Traffic Control. Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy.

Edwards, T., Sharples, S., Kirwan, B., Wilson, J. R., and Balfe, N. (2014). Identifying Markers of Performance Decline in Air Traffic Controllers. In *Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics AHFE 2014*

Luque-Casado, A. et al. (2016). Heart rate variability and cognitive processing: The autonomicresponse to task demands. *Biological Psychology, 113*, pp. 83–90.

Martins, A., et al. (2016). Test Plan for Preliminary Systems/Pilots; Cognitive Task Analysis; *Future Sky Safety P6 Human Performance Envelope Report D6.2*.

Pendleton, D. M., et al. (2016). Mental engagement during cognitive and psychomotor tasks: Effects of task type, processing demands, and practice. *International Journal of Psychophysiology, 109*, pp.124-131

Silvagni, S., et al. (2015). Concept for Human Performance Envelope; *Future Sky Safety P6 Human Performance Envelope Report D6.1*.

Straussberger, S., and Schaefer, D. (2004). A psychophysiological investigation of the concept of monotony in ATC: effects of traffic repetitiveness and traffic density. In *Proceedings of 1st international Conference on Research in Air Transportation (ICRAT),* University of Zilina, Slovakia, 199-208.

Taelman, J. et al. (2009). Influence of Mental Stress on Heart Rate and Heart Rate Variability. *4th European Conference of the International Federation for Medical and Biological Engineering (ECIFMBE)*, 2009, Antwerp, Belgium, pp.1366-1369.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Appendix A. DEBRIEFING TRANSCRIPTIONS

Appendix A.1 Pilot 1

Run 1 – Run 3 (02/05/2016)

CURVE

Where were you during the whole run? When did you move from one point to another?

Run 1: Mental workload was high during this run because the commands into the simulator were different from a real airplane. Performance goes backwards (from Focused to Relaxed) because of the familiarization with the simulator: I had to familiarise with side stick, thrust level and EPR (see the answer to the question *"Was the simulation realistic for you?"*)

Run 3: Performance goes from Focused to Under Pressure where I reached the maximum peak of workload.

GENERAL

Was the simulation realistic for you?

Simulator was completely different because the steering of the simulator was not like in a real plane:

- Side stick: plane didn't react to the command as expected
- Thrust level: it was tough to move
- EPR (Exhaust Pressure Ratio) as power setting: Lufthansa uses another measurement (N1 indication) and this made more difficult to make some estimates. The pilot didn't have the exact value as in a normal flight and he just tried on the basis of his experience

Do you feel you reacted as you would have done if it had been real?

Yes.

What is your general impression about the scenario?

Both runs were totally realistic (daily business).

Note: in both runs getting familiar with the simulator took the pilot 50% of his concentration.

Where you very much surprised by the events?

Run 3: Turbulences took part of the concentration as in real situation. The situation became more demanding but nothing unknown: *"Nothing that would put it to the limit"* "*a kind of hard but nothing that is impossible to handle*".

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Was there a particular moment where you felt less in control?

No.

PERFORMANCE

In hindsight, what might you have done differently?

Nothing.

Did you feel your performance was affected by the previous run?

Positive effect: the pilot became more familiar with the simulator.

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

Did you feel in a hurry to get any particular information from your colleague or from the cockpit? If that was the case, which one?

No.

How was your interaction with:

- **The captain?**

It was good. Although the Captain wasn't aware about the new procedures and terminology (he retired 2 years ago).

- **The ATC?**

He wanted the ATC communicating the weather information instead of asking for it. In reality, in windy situation is the ATCO who provides this information to the pilot.

Was there any aspect of the instrumentation that confused you or did not help at the time?

The flight director wasn't working. This was disturbing the pilot who decided to switch it off. He gave it another try when it was too windy but nothing changed so it was tuned off again. This does not happen in reality.

Was there something you might like to have seen on the cockpit instruments, or heard from ATC or your fellow pilot?

No.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Run 2 - Run 4 - Run 7 (03/05/2016)

CURVE

Where were you during the whole run? When did you move from one point to another?

Run 2: Relaxed for the whole run.

Run 4: performance went from Focused to Under Pressure. The run was challenging in the difficulty to steer the plane, the whole situation was a demanding flight. “Was told to fly it manually – DLH procedure is to fly autopilot. From here felt under pressure”.

Run 7: performance went from Under pressure to Struggling. The run was kind of easy about the aircraft handling but it was a demanding situation about the surroundings, the loss of localiser, the ATC, the remaining fuel, and for making the right decision under time pressure. “As soon as noticed the fuel level, was immediately under pressure. Struggling – not a big moment but a linear progression. Struggling when the localiser was moving away”.

GENERAL

Was the simulation realistic for you?

Run 2: Daily routine business, nothing special. A warm-up. Yes, it was realistic.

Run 4: Yes, it was realistic. We have this situation in Germany 2/3 times a year.

Run 7: It’s realistic, we have these situations every year.

Do you feel you reacted as you would have done if it had been real?

Run 2: Yes. Was a normal real situation. [In previous runs, he had some problem to familiarise with sidestick.] The sidestick is still a little uncomfortable, not like the real airplane but besides the handling quality of the simulator and the situation (to land in Hannover, the turbulences, crosswinds) was realistic.

Run 4: In reality we have a little time more for making plans or evaluating the situation to prepare the approach, to talk to the other pilots who didn’t approach previously in Hannover…so you don’t have to do that much in that smaller period of time. It was realistic but not 100% realistic.

Run 7: I think it was the safer course of actions to do in the scenario [decision to continue the approach without localiser guidance].

What is your general impression about the scenario?

Run 2: Nothing special, normal landing.

Runs 4 and 7: demanding in different ways – run 4 is about controlling the aircraft, run 7 more about making the right decision under time pressure.

Where you very much surprised by the events?

Future Sky Safety has received funding from the EU’s Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Run 4: When onto a Non-Precision Approach [no ILS guidance]. Wind changed a little. NDB have horizontal but not vertical guidance for descend. Was unexpected (had hoped to use other RWY which had ILS but crosswind). Was told to fly it manually – DLH procedure is to fly autopilot. From here felt under pressure. In real life we would never do a NPA with manual flight.

Run 7: Not really. When we lost the localiser this was of course a surprise but it was a failure and I had the plan B, my mental model was two steps ahead so I could compete with this surprise. At that point of the approach I don't have to plan anything else. When you have technical failures or medical issues, usually you start to think "I hope this is everything that's going to happen to me, the situation is now complete and there's nothing more added up to the struggling."

How did you feel you were more under pressure, like you focused more on something?

Run 4: You don't have any other chances. Have to be focused on the approach. It is mentally demanding. Even in the real airplane, you are under pressure in these kinds of approaches.

Run 7: As soon as noticed the fuel level, was immediately under pressure. Struggling – not a big moment but a linear progression. Struggling when the localiser was moving away. Normally it is a mandatory go-around. But in the handbook it states the pilot can do whatever he likes as long as he had a better idea. So, is it safer to do a go-around with minimal fuel, or to break a rule and continue approach without localiser guidance. Decided to continue the approach (go around uses around 800 kg of fuel). I don't think about the regulations because it is complicated, but because I broke one. At that point it was the safest situation.

Did you feel in control of the situation after the events occurred?

Run 2: Yes.

Run 7: Have these situations at least once a year. But always felt in control of the situation, and mental model was always in the loop.

Was there a particular moment where you felt less in control?

Run 2: No.

PERFORMANCE

In hindsight, what might you have done differently?

Run 4: No.

Run 7: No.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

Did you feel in a hurry to get any particular information from your colleague or from the cockpit? If that was the case, which one?

Run 2: No.

Run 4: Not so time critical as we had plenty of fuel.

Run 7: Was in a hurry to get info from ATC as we started in a low fuel state – normally would have known about this earlier – had to get the Met info and 'RWY in use' info quickly from ATC.

How was your interaction with:

- **The captain?**

Run 4: I would have had more precise information from the Captain (he said he didn't know, but in reality he would have known).

Was there any aspect of the instrumentation that confused you or did not help at the time?

Run 2: No.

Run 4: No real confusion but, for example in a storm, you are already prepared from the beginning. In a real approach, you are more aware of the limits/capacity of the aircraft, already at the briefing. In that case, I expect precise information from my colleague, if I ask. I expect he says the limit is 34 nodes.

Run 4: [Sometimes leave the sidestick off (let it go) for half a second, just to see how the aircraft is reacting. Something he does in real life, maybe learned it in training. Not stress-related (did it in run 2 as well)].

BEHAVIOURAL

How did you realise that your performance was declining?

I probably avoid any more discussions, even with ATC. When the situation is critical, I say what I want to do and there is no exception to that anymore. Once I decide on the course of action, I stick to that choice and expect everyone to stick to it too. There is no more time for any discussion.

Scenario 1: Run 5 - Run 8 (03/05/2016)

CURVE

Where were you during the whole run? When did you move from one point to another?

Run 5: Focused. "Had to make a Plan B, and there were no more options available. Kind of stressful situation when running out of options. Noise pretty loud. Like a small shock. You feel the heart rate:

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

something is wrong. Colleague said it was feedback from ATC, but in reality could ask TAC for a change of frequency. Adds up to the stress level, have to shout a bit. Stress from noise worst at the beginning. Once understood it was just a background noise, gets better, but stays really annoying. Takes a lot of concentration, just because of the noise".

Run 8: Between Focused and Under Pressure. "The run was demanding and the wind shift was quite a surprise. It started the same, low on fuel, no more options available close to an airport, weather conditions that were ok but demanding".

GENERAL

Was the simulation realistic for you?

Run 5: As realistic as the previous scenarios. There is a little different in reality – if you are that low on fuel on APP to FRA, ATC would normally have been told. Rest of scenario was realistic. At the beginning, noticed was low on fuel (used to it by now). But would also notice it in the real situation as would be in amber or red. In plane would get a prediction based on inserted landing and how much fuel will have after landing. In sim was like guessing how much it will be.

Run 8: started the same, low on fuel, no more options available close to an airport, weather conditions that were ok but demanding. Asked the controller if the weather is getting worse or improving or stay the same. If it had been getting worse, would have diverted to Nuremberg, but as was the same, stuck to the plan.

Where you very much surprised by the events?

Run 5: Noise pretty loud. Like a small shock. You feel the heart rate (see below).

Run 8: Has been flying to 8 years, has had 20-25 times when landing just within the crosswind landing limits. A new colleague might have changed controls with the Captain. Was demanding. In this scenario, wind shift was quite a surprise. In Germany, winds at this speed tend to stay the same, e.g. from West at 30kts, so saw it during my scan. Would say this is not realistic.

What was your feeling during the events of the scenarios (Were you stressed, with high workload etc.)?

Run 5: Had to make a Plan B, and there were no more options available. Kind of stressful situation when running out of options. Noise pretty loud. Like a small shock. You feel the heart rate – did we lose an engine, or an abnormal, fire on board. Something is wrong. Colleague said it was feedback from ATC, but in reality could ask ATC for a change of frequency. Adds up to the stress level, have to shout a bit. Stress from noise worst at the beginning. Once understood it was just a background noise, gets better, but stays really annoying. Takes a lot of concentration, just because of the noise. Is pretty realistic. When fly one of the old A320s, noisy from the very beginning, raises your flight level right from the gate, ears ringing after the flight. Was looking at the Captain during the noise, trying to read the words from the lips.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Run 8: Felt I had to push the Captain more and ATC to give us a better position in the landing sequence, to get information etc. In real life a good controller would calm you down, tell you that you were number 5, etc. Normally they give you all the information you need, then leave you quiet. Stressful to always be asking for the information.

Did you feel in control of the situation after the events occurred?

Run 5: Yes. The stress was higher because of the noise that surprised me in the first seconds, was a really bad experience, but I didn't lose the control of the situation.

Was there a particular moment where you felt less in control?

Run 5: No.

Were you able to maintain the best course of actions?

Run 8: Low vis and localiser interference, became a problem. Had to think what is the safest course of action. Is it safer to continue to a known airport (no obstacles)? So, broke the rule, because low on fuel.

PERFORMANCE

In hindsight, what might you have done differently?

Run 5: No.

Run 8: No.

Did you feel your performance was affected by the previous run?

Run 8 following Run 5 did not really impact. In reality, you do this. If arrive in a thunderstorm, you then leave in one.

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

Did you feel in a hurry to get any particular information from your colleague or from the cockpit? If that was the case, which one?

No.

How was your interaction with:

- **The captain?**

Run 8: No.

- **The ATC?**

Run 8: No.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Was there any aspect of the instrumentation that confused you or did not help at the time?

Run 5: No.

Run 8: Localiser failure – have not experienced this – would be good if could skip to another type of approach with a single button (NDB, GPS approach etc.). In case you need to switch the landing RWY in a short period of time – would be a cool feature if could to this with 2 or 3 clicks on the FMS.

Was there something you might like to have seen on the cockpit instruments, or heard from ATC or your fellow pilot?

Run 8: See above.

BEHAVIOURAL

Did you realise that you changed your tone of voice?

Run 8: Not really, I didn't noticed. If I had a change in my voice would have been due to stress, you become more alerted. [At a certain point was talking to Captain in a very curt way.] Yes, in a minimum fuel situation, no longer the time to be friendly, just minimal dialogue, clear information. In a critical situation come back to the basic language, just commands.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Appendix A.2 Pilot 2

Run 5 - Run 6 - Run 8 (02/05/2016)

CURVE

Where were you during the whole run? When did you move from one point to another?

Run 5 and 6: between Focused and Under Pressure.

Run 8: "I was Struggling. Not all the time but most of scenario".

GENERAL

Was the simulation realistic for you?

Yes. There were few things that were different e.g. track angle, something in flight director, but mostly yes. The visual was pretty good, but I've never experienced turbulence so strong. It was extreme.

In general, the simulator was pretty closed to reality.

Do you feel you reacted as you would have done if it had been real?

Mostly yes. I reacted as I normally do.

What is your general impression about the scenario?

Low fuel situations: excessive low fuel. I think will not happen in real life that you have such little fuel and you wouldn't realise before. Usually, you can see the situation that is coming and you can plan what to do. Normally you have a warning. Then the approach is still the same.

Where you very much surprised by the events?

Run 8: Low fuel -> when I realise I was in time-critical situation, no more looking around, just get straight to the approach asap. As soon as we were inbound in final approach, I knew that there was no more we can do. All my capacities were focused on the approach. I was stressed (not at maximum, but 80%) and totally concentrated on flight.

High wind shifts -> the work just kept going, we had to work really hard to stay established and I was closed to losing control at one point. I've never had these feeling of being established. The localiser and glideslope took almost of all of my capacities and didn't think to fuel situation anymore because I know I was doing what I could.

Localiser interference is very unusual. I noticed twice. It was very stressful and effortful because it was not accurate. It lasted about 20 seconds. I spoke to colleague to decide if it was better to do a go around.

Did you feel in control of the situation after the events occurred?

Run 8: See above.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

PERFORMANCE

In hindsight, what might you have done differently?

I would have called the Mayday earlier, in both occasions. I would have used better the Flight director. I used it just when Captain suggested (Run 6) and on the final approach in third attempt (Run 8). I don't really trust the FD because it had a different behaviour from real life. I would have used the autopilot most of the time because it would have reduced a lot the workload, but it was part of the simulation. More briefing too, but there was lot of time pressure.

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

Did you feel in a hurry to get any particular information from your colleague or from the cockpit? If that was the case, which one?

Because of loud noise, I had difficulty in getting some information from colleague, but general no.

How was your interaction with:

- **The captain?**

Nothing really, a bit unrealistic because in real life you have the debriefing to know a little the captain, but here not. Some different procedures that have changed during the past two years since he's been retired. He supported me but I had the feeling the he wanted to trick me into something (part of the scenario).

- **The ATC?**

ATC communication was quite realistic but there are much more chatter.

Was there any aspect of the instrumentation that confused you or did not help at the time?

Flight director and MCDU (Multifunction Control Display Unit) the character was small and was hard to read. Usually we have big letter in real life. EFB (Electronic Flight Bag) that switched on and off was annoying. Eye-tracker on the nose was annoying. Sidestick was different a little bit.

Track indications would be very helpful especially with lot of wind.

Flight path vector (didn't work properly?)

Pitch angles weren't as I would have expected them.

Was there something you might like to have seen on the cockpit instruments, or heard from ATC or your fellow pilot?

Everyone would have more direct control and feedback from the Airbus, especially hard dynamic feedback through the yoke. It doesn't exist in the Airbus. You have to ask the computer to do stuff, because it flies.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Run 2 - Run 4 (02/05/2016)

CURVE

Where were you during the whole run? When did you move from one point to another?

Run 2: Between Focused and Under Pressure. The situation was common and in general ok, but I struggled with the aircraft control. For the most part was Focused and then sometime went to Under Pressure.

Run 4: Between Struggling and Failing.

GENERAL

Was the simulation realistic for you?

Run 4: Very definitively demanding, it was really hard to stay focused on the situation because I know that this will never happen in real life. It's hard to act as professionally as you normally would but you know that would never come this far.

Do you feel you reacted as you would have done if it had been real?

Run 4: In a real situation, I would have used the autopilot and prepare the approach even in the turbulence. The last part with the manual flying, I was afraid to fly with strong crosswind.

What is your general impression about the scenario?

Run 4: Unpleasant, many unexpected things, I was running from a prompt to the next one. Struggling with the airplane control, with turbulences and the late response to my inputs, etc. Decided to do the NDB (Non-Directional Beacon) approach, mistake to get information from the database. No guidance for the approach which he's never seen in real life. The entire time struggle with Flight Director really offset. The heading really not useful at all. Only distracting and get extra workload because you have to ask PM to set certain values (usually set altitude and that's it). I couldn't use the autopilot as I would, this is not realistic at the end. Lot of workload just to compensate turbulences instead of focusing on the approach. I had no time to do it. In real life a flight like this would be a disaster. Totally unsatisfactory because I was so busy controlling the aircraft and started to configure the approach a bit late. I felt the aircraft had nothing to help me, just me to control, no guidance. I didn't know what happens, we had the wrong ILS, I don't know why (Captain made a mistake, to check if is correct). Normally, when you insert the approach you have to set and check ILS but didn't work. Things didn't match: the information I saw on the screen didn't match my mental picture and what I saw outside. That was really disturbing. At some point, I relied on what I were seeing outside and forgot about all of indications except for altitude and speed.

Most of time was stressing, I'm not afraid of the wind normally but it was just a feeling that I was performing way behind where I want to be. You don't know how bad it is going to get. That is the stress. Started to move his head very frequently towards the EFB on the right and towards the instrumentation

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

around him (very quickly). Struggling with instrumentation and in trouble with automation. “I was looking for cues, for answers”. High workload and stress.

Run 2: Pretty much as expected. The situation is common.

What was your feeling during the events of the scenarios (Were you stressed, with high workload etc.)?

Run 4: I felt stressed and losing the control, the SA because what do I do if the most important things don’t work?

PERFORMANCE

In hindsight, what might you have done differently?

Run 4: Slowing down the approach even more.

Did you feel your performance was affected by the previous run?

Run 4: A bit of fatigue impact the performance.

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

Did you feel in a hurry to get any particular information from your colleague or from the cockpit? If that was the case, which one?

Run 4: See above

Was there any aspect of the instrumentation that confused you or did not help at the time?

Run 4: See above

BEHAVIOURAL

Did you realise that you moved and repositioned on the seat?

Run 4: Body repositioning -> He didn’t realise but supposed to be more focused on the situation and felt stressed.

Future Sky Safety has received funding from the EU’s Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Run 1 - Run 3 (03/05/2016)

CURVE

Where were you during the whole run? When did you move from one point to another?

Run 1: between Relaxed and Focused

Run 3: Under Pressure

GENERAL

Was the simulation realistic for you?

Run 1 and 3: Yes, were realistic.

What is your general impression about the scenario?

Run 1: Happy, nothing to report. Simulation was ok.

Run 3: More pleasant, better than yesterday [runs 5-6-8 and 2-4]. The airport was familiar.

What was your feeling during the events of the scenarios (Were you stressed, with high workload etc.)?

Run 3: Struggled with stability on vertical flight path, lots of gusts, lack of feeling, up and down-drafts are pretty strong, have to concentrate. It was hard. Demanding. A little stressed. Due to yesterday [he had difficulty with Run 4] heart rate was up, apprehensive. With Lufthansa recurrent training, if fail in sim twice in a row, get fired, so always a little apprehensive [post-note: Pilot 3 on Day 3 said that with recurrent training you usually know more or less what is coming, so there is some stress, but the HPE ones are more demanding]. Also you know you are being watched. But once you are in the sim and flying, it goes away.

PERFORMANCE

In hindsight, what might you have done differently?

Run 3: Ideally he'd have had better control of the vertical flight path.

Did you feel your performance was affected by the previous run?

Run 3: Due to yesterday [he had difficulty with Run 4] heart rate was up, apprehensive. Fatigue influenced a little the performance.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

Was there any aspect of the instrumentation that confused you or did not help at the time?

Run 3: Flight Director (FD) worse than yesterday. The EFB [Electronic Flight Bag] is such a drag to use. Would like to have more of a say – three pages, have to look at each one each tie, a dozen taps before I can use it. When it's time critical it's a problem, end up head down for a long time. When I have to babysit the FD, might as well do all the calcs manually. Info is there (EFB), but hard to get. When I meet new captains, we quickly end up ranting about the EFB. For example, when flying to FRA (Frankfurt), we get the info way too late, so usually we get all three approaches ready just in case. EFB is not flexible.

Was there something you might like to have seen on the cockpit instruments, or heard from ATC or your fellow pilot?

Tactile feedback on the EFB would be good, e.g. for some common functions, so can do it by feel. We used to have the key info on paper on a clipboard – was better – now the key info is buried.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Appendix A.3 Pilot 3

Run 3 - Run 1 (03/05/2016)

CURVE

Where were you during the whole run? When did you move from one point to another?

Run 3: Under pressure. However, in a real situation (without simulator problems – he mainly refers to the flight directors) he would be Focused. “I was not really confident with the behaviour of the simulator so I had a little bit of problem at the beginning of the simulation”.

Run 1: Relaxed. No particular comments about Run 1. All positive.

GENERAL

Was the simulation realistic for you?

Yes.

Do you feel you reacted as you would have done if it had been real?

No, because the behaviour of the simulator was a little bit confusing.

Where you very much surprised by the events?

Yes: “The flight director was off and then comes several things together: altitude and headings, speed control and there was one moment when I wasn’t really satisfied about my performance”.

What was your feeling during the events of the scenarios (Were you stressed, with high workload etc.)?

He realised that he was losing control because he lost the focus on the numbers.

Was there a particular moment where you felt less in control?

“I had to look at three things at one time and this was the point where I lost two of them”…“The colleague gave me hints to check heading, altitude…”.

PERFORMANCE

In hindsight, what might you have done differently?

“I increased the level [of workload] by myself when I switched off the flight director. This is maybe not the best decision”. When he lost focus, this might have been very helpful.

Do you think the way you performed could have had an impact on the safety of the flight?

“Switching off the flight director is maybe a downgrade of safety”.

Did you feel your performance was affected by the previous run?

Future Sky Safety has received funding from the EU’s Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

He started with run 3 hence we don't have any carry over effect. However, he highlighted that he needed to get used to the simulator so his performance was also due to this aspect: "I was not really confident with the behaviour of the simulator so I had a little bit of problem at the beginning of the simulation".

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

Did you feel in a hurry to get any particular information from your colleague or from the cockpit? If that was the case, which one?

He missed the first landing clearance because of high workload: "I was really concentrated on manual flying and I didn't get the landing clearance".

Was there any aspect of the instrumentation that confused you or did not help at the time?

The flight director. He turned it off and re-engaged it on the ILS profile.

Was there something you might like to have seen on the cockpit instruments, or heard from ATC or your fellow pilot?

Nope.

BEHAVIOURAL

How did you realise that your performance was declining?

"Because I lost the focus on the numbers"

Scenario 1: Run 4 (03/05/2016)

CURVE

Where were you during the whole run? When did you move from one point to another?

Run 4: From Relaxed at the beginning of the run to Under Pressure in the final part. "Not really struggling but under pressure. Demanding. The most difficult part was the alignment with RWY track and then starting the descent, not easy to follow the NDV needle with heavy crosswind".

GENERAL

Was the simulation realistic for you?

Run 4: Yes, was realistic; maybe not the combinations of the approach with the crosswinds but the situation overall was realistic.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

What is your general impression about the scenario?

Run 4: Not really struggling but under pressure. Demanding. The most difficult part was the alignment with RWY track and then starting the descent, not easy to follow the NDB needle with heavy crosswind.

PERFORMANCE

In hindsight, what might you have done differently?

Run 4: I was cheating a little bit because I gave the control to the captain and prepared my approach charts, FMS, etc.. to reduce my workload management and gave more capability for the approach planning (to focus on the approach); then I took back the control. That's what I do in normal business.

Did you feel your performance was affected by the previous run?

[He did 3 runs and ½] More used to the simulator behaviour so this makes it easier the manual flying part. It was difficult at the first run, but now is not a big deal anymore.

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

How was your interaction with:

- **The captain?**

It was good. We were close to the maximum crosswind limit, (both focused on it) he told me that we were fine with the wind and the descent profile.

- **The ATC?**

Fine. Just once he didn't respond but all the communication was interrupted and forgot to push the button already start talking, but it was fine [maybe a problem with the communication between ATC radio and simulator]

Was there any aspect of the instrumentation that confused you or did not help at the time?

Run 4: Issues with instrumentation? Navigation display, EFB, etc..? It was useable. It's a Non-Precision thing, it's difficult to check the flickering and the NDV needle so I was missing the track line but the captain told me that's a simulator thing.

Was there something you might like to have seen on the cockpit instruments, or heard from ATC or your fellow pilot?

[He flew a Boeing years ago] I feel more comfortable with the navigation display on the Boeing than the Airbus. For me there are too many information displayed too small (small signs for the localiser, green track line..), too small information in a too small place. In my opinion, the Boeing has a better design for the navigation display.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Run 8 (04/05/2016)

CURVE

Where were you during the whole run? When did you move from one point to another?

Run 8: From Relaxed at the beginning to Focused on final approach. Everything was almost as I expected so when we declared the emergency we get a direct radar for the final… I would say that was not more than Focused. (See answer to question "Was the simulation realistic for you?")

GENERAL

Was the simulation realistic for you?

Run 8: At the beginning not much fuel on board, but had more than half an hour. As he cleared us to the Charlie holding, there comes the idea that we won't make this time, so looked at the situation, entered the time on fuel remaining for less than half an hour, and made the decision to declare an emergency. When I realised, it was ok because once we declare an emergency we get a direct radar and from Charlie hold to FRA is 15 mins, so situation was tense but not really stressful. Was really focusing when localiser interference happened, and put my own gear down as Captain was busy. Never happened to me, but can be a real situation.

Do you feel you reacted as you would have done if it had been real?

Run 8: Reacted as in real life. Still of opinion that to move the gear retraction on my own was the right decision.

Where you very much surprised by the events?

Run 8: Not really surprised by the low fuel. Divert to Nuremburg would have been worst case if FRA runway blocked by an aircraft. But FRA was closest available and can be on ground in 15 minutes, 4km of runway and 3 landing runways, so just to have a plan B in mind when something unexpected happened, can divert to an alternate field.

What was your feeling during the events of the scenarios (Were you stressed, with high workload etc.)?

Run 8: See above.

Did you feel in control of the situation after the events occurred?

Run 8: Felt in control of the situation.

PERFORMANCE

In hindsight, what might you have done differently?

Run 8: Nothing.

Did you feel your performance was affected by the previous run?

Compared to runs yesterday, get more used to FD, controlling etc., easier than the first run yesterday.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

[Post-note – was not aware of the low fuel – was prompted by ATC. Also unaware of wind shift. Did not last long because experimenter wanted to shorten the scenario and avoid go-around.]

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

Did you feel in a hurry to get any particular information from your colleague or from the cockpit? If that was the case, which one?

Run 8: We were both focused on the fuel, so both had the same picture in mind.

How was your interaction with:

- **The ATC?**

ATC ok: had traffic into FRA, minimum separation and traffic inbound into FRA, and already some others declaring fuel emergency.

Was there any aspect of the instrumentation that confused you or did not help at the time?

Run 8: Behaviour of FD a little irritating, but get used to this behaviour.

Was there something you might like to have seen on the cockpit instruments, or heard from ATC or your fellow pilot?

I would appreciate if you have 45 minutes fuel remaining can flicker amber on the warning display – on the display everything is fine, but it is not. Need a visual reminder. Does it exist in real life? Never seen it.

Run 5 (04/05/2016)

CURVE

Where were you during the whole run? When did you move from one point to another?

Run 5: Relaxed for the whole run.

GENERAL

Was the simulation realistic for you?

Run 5: It was realistic. He was expecting something more. He was ready to shift from a picture to another one.

Do you feel you reacted as you would have done if it had been real?

Run 5: He would have reacted the same in reality. Everyday business.

What is your general impression about the scenario?

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Run 5: Loud noise was nothing special. I didn't feel any disturb in my approach. Nothing really stressful situation, maybe just focused on the fuel things because I missed today in the other run.

What was your feeling during the events of the scenarios (Were you stressed, with high workload etc.)?

Run 5: Relaxed for the whole run.

Did you feel in control of the situation after the events occurred?

Run 5: Control: "I was always ahead of the aircraft".

PERFORMANCE

Did you feel your performance was affected by the previous run?

Run 5: This run was the easiest. I expected something more from it, because of the previous one (run 8).

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

Did you feel in a hurry to get any particular information from your colleague or from the cockpit? If that was the case, which one?

Run 5: He did not feel in a hurry to get any information.

How was your interaction with:

- **The captain?**
- **The ATC?**

Run 5: All the interactions were good.

Was there any aspect of the instrumentation that confused you or did not help at the time?

Run 5: Aircraft did not have problem. Weather was good.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Appendix A.4 Pilot 4

Run 7 - Run 4 (03/05/16)

CURVE

Where were you during the whole run? When did you move from one point to another?

Run 7: Between Relaxed/Focused and Struggling/Failing. "The low point of the run (most struggling) was probably during the first approach when I saw I couldn't concentrate on being in the loop, aircraft didn't handled as easy as I thought, I fight to get it back. I couldn't plan because was too busy to execute the task".

Run 4: Between Relaxed/Focused and Under Pressure/Struggling. "The lowest point was when I tried to prepare Non-Precision and I had to think what to do for NP flight, flying manually… a lot of stuff at the same time".

GENERAL

Was the simulation realistic for you?

Run 7 and 4: Visual is ok, except for the landing which is offset, and the power settings and the control stick. Especially the power settings.

Do you feel you reacted as you would have done if it had been real?

Run 4 and 7: Reacted more or less the same.

What is your general impression about the scenario?

Run 4: I felt lot better than in the first one (run 7) because little bit more used to the simulator, to the captain. I felt the right mood to dealing with difficult scenario. I was more activated and was easier to stay ahead, in the loop. The approach was very difficult with flying manually, flight director didn't work, etc. but I did the best I can do.

Where you very much surprised by the events?

Run 7: I was getting a bit surprised and drowsy after the break, a bit more tired than I expected and got thrown into a difficult situation.

Captain informed me there was a possible fuel leak. Surprised me because I hadn't seen it. It was a bit time-critical. Then strong wind reported. Then the go-around because it became very unstable, then tried again with a short turnaround. Fuel leak – surprised, a bit lost for a moment, trying to figure it out. Had it been real would have been a shock. Looked at the fuel page, seemed to be stable, about 30 seconds.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Run 7: The low point of the run (most struggling) was probably during the first approach when I saw I couldn't concentrate on being in the loop, aircraft didn't handle as easy as I thought, I fight to get it back. I couldn't plan because was too busy to execute the task.

Run 4: The lowest point was when I tried to prepare Non-Precision and I had to think what to do for NP flight, flying manually… a lot of stuff at the same time.

Was there a particular moment where you felt less in control?

Run 7: Flight guidance didn't act as expected, so a lot of attention on how it is going to work out, and how I can fly the go around. Had to figure out how the simulator is working. I didn't get too stressed because it wasn't my fault but was the simulator.

PERFORMANCE

In hindsight, what might you have done differently?

Run 7: Nothing particular (would do differently). Did a go-around because was very unstable. Would have liked to fly it better, but was just fighting with the simulator.

Run 4: Yes, I take it a bit slower, was not ready. ATC said to speed up and the captain said to take out the power. Get the time to mentally prepare for really difficult approach. Do the preparation using the autopilot or to control the charts, time to look it over again.

Did you feel your performance was affected by the previous run?

Own performance – if had been mentally more ready for doing really hard work (had to hit the ground running). Was too relaxed for the situation.

The first (Run 7) made me awake and mentally prepared to the second one (Run 4). In the first run I wasn't prepared, I didn't expect so hard work.

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

Did you feel in a hurry to get any particular information from your colleague or from the cockpit? If that was the case, which one?

Run 7: Not in a hurry to get information, info was coming from them without me having to ask.

How was your interaction with:

- **The captain?**

Run 7: Captain was very straightforward, sometimes a bit pushy, trying to get the scenario going in a particular direction, in real life would be a bit different.

Run 4: Captain was more relaxed because we didn't have problem with fuel

- **The ATC?**

Run 7: ATC was normal. Once or twice I didn't get an answer.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Was there any aspect of the instrumentation that confused you or did not help at the time?

Run 7: Flight Director is horrible, worst I've seen, engine instruments are very difficult, the rest on the PFD is fine. Biggest problem is the power and trying to find a sensible power setting. EFB turned off quite a few times. Just had to touch it to come back on (On power save? Or stand by?).

During NPA you need a practical profile of the approach with all the information (tracks of the approach, tracks of the RWY, etc.). It's quite difficult to have it on the side, out of view. You have to make an effort to turn and takes a lot of time.

Was there something you might like to have seen on the cockpit instruments, or heard from ATC or your fellow pilot?

Can't think of anything additional.

BEHAVIOURAL

How did you realise that your performance was declining?

Run 7: Tighten the lips during demanding situations. He didn't notice.

Run 1 - Run 6 - Run 8 (03/05/16)

CURVE

Where were you during the whole run? When did you move from one point to another?

Run 1: From Relaxed to Focused. "Just waiting for something but nothing happened".

Run 6: From Relaxed to Focused/Under Pressure. "Warning about the localiser being bad, increased WL a bit, but all in an area where it was easy to fix".

Run 8: From Focused to Failing. "The low point was very short final, the last 10-15 seconds, was a speed drop, then got too high. Too little time to do a proper instrument check. Lasted until touchdown".

GENERAL

What is your general impression about the scenario?

Run 1: Nothing, no problem.

Run 8: It was helpful that got used to the simulator first, the gusting and handling the aircraft. Workload mainly ok (2-3 ISA).

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Where you very much surprised by the events?

Run 8: Discovered the fuel, had discussion about how we could go, then the speed, then the noise part, this surprised me, thought it was a technical problem, then noticed in 10s that everything seemed to be going fine, so was just an annoyance. I was alarmed though. Then could ignore it.

What was your feeling during the events of the scenarios (Were you stressed, with high workload etc.)?

Run 8: Captain mentioned sink rate. I was a bit more stressed, but all over too quick to get too stressed. I was more concentrated.

Did you feel in control of the situation after the events occurred?

Run 6: Felt in control.

PERFORMANCE

In hindsight, what might you have done differently?

Run 8: I would probably try to do one more cross-check of the vertical speed and instruments. Might have noticed earlier was going high.

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

How was your interaction with:

- **The captain?**

Captain pretty good, good inputs, he's from FRA I'm from MUN, so he had better knowledge.

- **The ATC?**

ATC – quite a few of the ATC calls were impossible to hear.

Was there any aspect of the instrumentation that confused you or did not help at the time?

Run 6: Instrument – usually use the flight director and the track triangle, but both weren't working. Don't know diversion from the course because just eye-balling it.

Was there something you might like to have seen on the cockpit instruments, or heard from ATC or your fellow pilot?

Run 8: What would help to have the wind in the scan, so that when you get a dramatic change in wind, otherwise it is in the corner. You don't have the time to look at the instruments during the flare, but that is for Pilot Monitoring to do, which he did. If the system would monitor that, it would help (a warning), e.g. diverting from a normal trajectory.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Run 5 - Run 3 (04/05/16)

CURVE

Where were you during the whole run? When did you move from one point to another?

Run 5: Between Relaxed and Focused. "Was a very relaxed scenario and I managed to get myself activated and focused. I pushed myself a little bit to concentrate and nothing has changed definitely at some point between being relaxed and being focused". To scan instruments could help to stay focused in a too relaxed situation.

Run 3: Between Relaxed/Focused and Under Pressure towards the end. Always a lot to do with turbulence, instrument scan, manual control. Was concentrating that he stayed relaxed and not too stressed. Towards the end was more under pressure, almost towards struggling.

GENERAL

Was the simulation realistic for you? Do you feel you reacted as you would have done if it had been real?

Run 5 and 3: Two runs were more realistic, getting used to the simulator. Landing was a bit weird (was the simulator). Otherwise, I reacted as would have done in real life. Really had to push down which would not happen in real life, was the simulator. Everything else pretty much as in real life. This much fight with the aircraft never in real life.

Where you very much surprised by the events?

Run 5: Was surprised by the noise again, annoyed me a bit more this time.

What was your feeling during the events of the scenarios (Were you stressed, with high workload etc.)?

Run 5: I noticed the fuel was low again. In real life would be more stressed, would notice much earlier. Stayed in control. Had spare time to have a regular look at it as was a normal approach.

Did you feel in control of the situation after the events occurred?

Run 5: Did not declare emergency. Yes, something could have gone wrong, e.g. problems with the flaps etc. Was aware of this. Was in the back of my mind. If there's another problem, would have declared an emergency straight away. I wasn't stressful.

Was there a particular moment where you felt less in control?

Run 5: No.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

PERFORMANCE

In hindsight, what might you have done differently?

Run 5: Happy with performance, would not have done something differently.

Did you feel your performance was affected by the previous run?

In the break between the two runs the first one was fairly easy, so had to stay alert for the second one.

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

Did you feel in a hurry to get any particular information from your colleague or from the cockpit? If that was the case, which one?

Run 5: No. Nothing non-standard I needed.

Run 3: Captain asked for the wind, missed it, knew he had missed it so asked again.

How was your interaction with:

- **The captain?**
- **The ATC?**

Run 5 and 3: Nor problems with Captain or ATC (both runs).

Was there any aspect of the instrumentation that confused you or did not help at the time?

Flight Director (FD) instrument in the simulator is useless. Also miss the track diamond (not available in this simulator). Heading instrumentation worked for first run but does not work well in turbulence (run 3).

Was there something you might like to have seen on the cockpit instruments, or heard from ATC or your fellow pilot?

Maybe, I would have liked wind information in the PFD (Primary Flight Display).

BEHAVIOURAL

How did you realise that your performance was declining?

Run 5: With noise, changed from English to German to answer ISA (level of workload): he said “drei” instead of three. He realised that and said that it was easier to pronounce with the noise, not because were more familiar.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Appendix A.5 Pilot 5

Run 6 - Run 5 (09/05/16)

CURVE

Where were you during the whole run? When did you move from one point to another?

Run 6: from struggling to focused at the end. It was the easier one, but not very comfortable because of the simulator. So I started struggling and I ended focused because I get more comfortable with the simulator. The different performance feeling was not due to the scenario, it was more a personal feeling related to the simulation (not well flown). Then during landing a little bit higher – at the end.

Run 5: I was all the time focused.

GENERAL

Was the simulation realistic for you?

Overall the simulator was realistic, I had two main problems:

- the thrust lever in IDLE (0 position) it is very hard to get outside, you have to push harder than the normal airplane, and then you get more thrust and then you have to bring it back;
- And the same is for the pedals

But to fly with it was good.

Do you feel you reacted as you would have done if it had been real?

Run 6: I would have flown a go around because I was not established on the criteria we have in Lufthansa; in fact, at 1000 feet we have to be ready with all, with speed, power, all stuff but it wasn't too much so, but we decided to continue as the Capitan said "we continue", so I went on, but I wouldn't in a real situation. I did a mistake here, I descent too fast and too high for the descent, not aligned with Lufthansa standards.

But for the main part of the scenario yes, as it had been real.

What is your general impression about the scenario?

Run 5 was an unrealistic situation due to the low fuel condition since the beginning of the scenario (starting with 1500 tons); "I found myself come to that point", it was obvious we were running out of fuel and we discussed that, we had a touch down with 1h fuel remaining, we were below the standards. We were put in a situation, and in that situation I think it was a realistic scenario, but the starting point was unusual.

Run 6 was a realistic approach to Frankfurt.

Where you very much surprised by the events?

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Run 5: I was surprised, even if in Frankfurt we have from time to time problems like that (not so loud however). I thought it was a failure of the simulator, because the captain said something about the noise. However, I was not afraid, it was just a loud noise, I just continued the approach.

No event surprised me in Run 6. For the localiser interference, I expected a NOTAM, as I usually consult that for such kind of event.

What was your feeling during the events of the scenarios (Were you stressed, with high workload etc.)?

Run 5: I was surprised at the beginning and a bit distracted by the loud noise. It was good that the flight was stable, that there was no influence to the airplane, it was just into my ears. The problem was on the communication with the captain, difficult to understand the landing clearances. But I was not afraid or something like that.

Run 6: no feelings about that, normal situation. The weather forecast on visibility were worse than the actual visibility. The reported low ceiling was not there, this was a bit surprising.

Was there a particular moment where you felt less in control?

Run 5: less in control no.

Run 6: just at the beginning. But I was always in control.

PERFORMANCE

In hindsight, what might you have done differently?

Run 6: I would have established earlier, to have reduced speed earlier, and then to have all the other charts ready (not only ILS but also the VOR and RNav) so I could switch just to have a look, because the captain said something about the VOR approach at the minima because of the low visibility and something with the localiser

Run 5: I would have informed the ATC earlier about the low fuel situation; 1250 or so is officially the minimum to declare emergency, so I did but it was too late, while we should have said something to the ATC around 2000 tons.

Do you think the way you performed could have had an impact on the safety of the flight?

No.

Did you feel your performance was affected by the previous run?

I improved, I knew that I had to improve and I was successful in that.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

Did you feel in a hurry to get any particular information from your colleague or from the cockpit? If that was the case, which one?

The charts. The communication was good with the captain, he was good PM. The communication with ATC was sometimes was broken, so we had to repeat the ATC instructions 3 or 4 times, but nothing that was dangerous. We had enough time to repeat the instructions. Cross-interference and instructions repetition may happen in real life too, and it might be dangerous especially in Frankfurt (high density area).

Was there any aspect of the instrumentation that confused you or did not help at the time?

No. It was not a real A320 cockpit, so you couldn't use the speed management function but it was not a problem for me.

Was there something you might like to have seen on the cockpit instruments, or heard from ATC or your fellow pilot?

No.

BEHAVIOURAL

How did you realise that your performance was declining?

Run 6: Yes, I realised I was improving because I felt more conformable, the cross-check with the instruments was better (get accustomed with the glasses), I was more focused and able to look more and better around.

Speak about specific BM [Mentioned breathing, movements during the loud noise.]

Not noted the breathing, but it is an effect of the arousal – that I realised that I was more comfortable during the second run. I realised that I had the feeling that it was easier and I was more relaxed during the second run than the first run (more stressful). But I couldn't say that I listened to my breathing. I just realised that it was different between the two runs. I sweat more during the first run, it was warmer, I had a feeling on the stomach, I moved more. Then I knew that I didn't want to have it, I didn't feel perfect in the situation, I knew that I wanted to perform better and I didn't want the stomach feeling. Also, the pulse was upper in Run 6 than in Run 5. I tried to relax myself between the two runs, I focused to get the stress out.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Run 7 - Run 4 (10/05/16)

CURVE

Where were you during the whole run? When did you move from one point to another?

Run 7: from between Under pressure and struggling to Under pressure after the go around (the situation improved). This was better at the end, after the go around the weather changed, the wind was better (no strong tail wind) so it was easier.

Run 4: started relaxed (in the approach), then when started the descent moved to focused, a bit more than focused during landing.

GENERAL

Do you feel you reacted as you would have done if it had been real?

Run 7: yes, get into the corner with the low fuel. I asked several times for the fuel situation, for the remaining time. I decided to do the Go Around because I was not aligned to the centre line (indicated by the magenta arrow in the localiser), we were offset with a large incorrect angle, and I saw the tail wind. Also the localiser was not that much, so I was surprised that we were so far away from the centre line. Because the centre line displacement on the instruments was not that much, but when I looked outside with the visual together with the incorrect angle there was a large offset. Then I tried to regain the centre line and with the addition of the tail wind I decided for the Go Around. I didn't feel very comfortable, I felt tingling my fingers as I started the approach with the low fuel, then it went worse as it become more difficult with the Go Around and few fuel remaining. Then I felt I was breathing hard. I felt I was losing a bit the control, I tried to get back into control focusing on the breathing for short time, and then focus on the basic flight skills leaving the other activities to the PM ("keep flying and let the other person supervising the rest. Keep the aircraft stable and fly it safely through the air"). And then I felt more relaxed, it happened after the go around – which went well. And after the feeling was better, I had everything more under control, more prepared.

Was there a particular moment where you felt less in control?

Run 7: Just very short time after the localiser interference, I just cross-checked the descent point to start the descent and the localiser went away and normally we are not allowed to decent when the localiser is not intercepted. If it goes outside the 2 dots we are not allowed to descent, because it could be a wrong one. So I said "What is that?" and I asked the captain to ask the controllers for previous reports. But then after the captain said the NDB, GPS and primary, as we have additional means to cross-check that we are on localiser and safety descent. So there was this small event and I started the descent but then I was not allowed to, so after that I felt less in control so I said "ok, start again". Then it came back very soon.

Then one time I went out of heading, but I think it was a simulator issue, I flew through the heading. And it was very struggling, and then I missed this heading as well so I knew "ok, let's focus more!".

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Where you very much surprised by the events?

No, during Run 4 no because we had fuel, the turbulences were there so it was a bit hard to control the speed but it was no difficult. There was nearly no change in the wind, so we knew what it was going on. The landing was not good, it was off of the centre line due to wind; there was something with the radar, I was controlling the radar, I started to brake that we were on the centre line and then I looked at the radar for the alignment and then the drift started to the left. So we ended with an incorrect angle and on the ground we did the flare to pull up the nose, hit the radar for the centre line and we were not correcting for the wind to the right. If you break too early you shift to the left, and it happened actually, even if we stayed to the runway, but on the left of the centre line. On the other side, if you brake too late it will be a hard landing. It is always difficult with the wind. Especially for me, as I don't have much experience with cross-wind landings.

PERFORMANCE

In hindsight, what might you have done differently?

I'd have used more the autopilot in both runs, especially in Run 7 to have time to calm down and focus on my stuff, because I was not that focused. I'd have used definitely on downwind to calm down. We should have used the ground speed (of wind) from the tower for the landing instead of looking at the wind indication on the NAV display (which calculates the wind that the aircraft thinks is happening around the aircraft). We usually compare the wind we have on board with the ground wind to anticipate a potential wind shear during the landing phase, but I didn't have the tower wind into my mind. The wind has an impact on the power you have to put on the aircraft to land, and you have to anticipate that. During the run we talked about the wind, but I didn't really get this information.

Run 4 I would have done a go around, we have fuel and the wind was strong and it would have been safer to do a second approach. However, it would have been a late go around. The landing was not dangerous however.

Did you feel your performance was affected by the previous run?

In Run 4 my performance was better, because I focused. I knew that I had to perform better, so there was a positive influence of the first run to the second run.

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

Did you feel in a hurry to get any particular information from your colleague or from the cockpit? If that was the case, which one?

I would have heard a wind check from the ATC. The PM could have asked for it.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Was there something you might like to have seen on the cockpit instruments, or heard from ATC or your fellow pilot?

It would be great to have on board the indication of the wind from the ground, to compare it directly and don't have to ask again.

BEHAVIOURAL

Speak about specific BM [Mentioned absence of talk, no movements, no looking around in Run 4]

I did it on purpose to be focused on the events and on the main flying skills. I didn't reply to PM jokes to focus on turbulences, I knew it was a challenging approach.

In Run 7 when I decided for the go around I had this stomach feeling.

Run 8 - Run 3 - Run 1 (10/05/16)

CURVE

Where were you during the whole run? When did you move from one point to another?

Run 8: I moved from Under pressure to Focus, as the weather was better.

Run 3: Between relaxed and focused; I stayed a bit more than relaxed, in a condition of high arousal.

Run 1: I was relaxed all the time.

GENERAL

Do you feel you reacted as you would have done if it had been real?

Yes

What is your general impression about the scenario?

Realistic. Run 3 and Run 1 are just daily operations, and Run 8 could happen, is realistic.

Where you very much surprised by the events?

In Run 8 the noise surprised me, then I decided. I didn't get the wind shift, I didn't have the opportunity to look at the wind direction.

What was your feeling during the events of the scenarios (Were you stressed, with high workload etc.)?

More routine, I felt focused and doing the job. Normal ops nothing very special.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Run 8: I had this low fuel in mind, but we talked about it and we knew that we could do a second round, even a second go around would be possible so at the end I was confident. The captain was faster than me to recognise the wind shear, I recognised the broken speed and large power setting but I didn't look at the wind. But the outcome was high power and low speed which is a result of the wind, so I realised in another way. I didn't have a bad feeling however. It was good that I solved the speed drop that was important.

Run 3: the turbulence in the simulator were strong and gave me uncomfortable feeling all the time. This is why I was in this status of high arousal.

Did you feel in control of the situation after the events occurred?

Yes. In Run 3 I thought I missed something, I was looking from something else a part from the turbulences.

In Run 8 I realised the low fuel situation instantly more or less, 70% of my daily work is in Frankfurt so I feel comfortable with the airport, I didn't need charts or other things.

PERFORMANCE

In hindsight, what might you have done differently?

Run 3 was pretty fast, everything happened fast, I tried to fly a bit more conservative in that run. I'd have used the gear earlier and be established earlier.

Run 8: no, I just speak about the missed approach, it was fine because I briefed it, so I would do it again. The go around was needed because there was a wind shift, the speed drops below our approach speed and the power was very high, so the configuration was not optimal.

My performance did not affect the safety of the flight. Steering could have been more accurate, it was all within limits, not unsafe.

Did you feel your performance was affected by the previous run?

I get accustomed to the events.

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

Did you feel in a hurry to get any particular information from your colleague or from the cockpit? If that was the case, which one?

I asked twice for the wind but I had it. Sometimes is useful if the tower informs about the previous reports from other aircraft so you can anticipate the issues and eventually decide for a go around earlier. If I was the PM I would have reported the wind shift to the PF couple of times to be sure he get it, but it is not a SOP it is just the strategy that I adopt. However, I'd prefer to have the tower wind on the screen, just in the approach phase.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Was there something you might like to have seen on the cockpit instruments, or heard from ATC or your fellow pilot?

I'd like to have runway state information on board, if there is snow, if it is warm enough of if it has accumulated during time, and its impact on the breaking action. I'd like to have a continuous monitoring on the screen, and then if something changes it could use a different colour or something like that. So, runway condition and wind have the same instruments for all the aircraft (Boeing and Airbus).

I'd like to have support for information prioritization, in particular in case of attentional tunnelling.

The cabin check ready for landing can be sent to the cockpit only at certain times, if they are ready earlier this information doesn't come through, they have to make a phone call to inform the cockpit.

It would be nice if the thrust lever is reported somewhere in the FDM, beside the speed rate to facilitate the correlation between speed and power settings, because the speed goes always with power (and pitch also) to make the aircraft stable. The cross-check will be faster.

BEHAVIOURAL

How did you realise that your performance was declining?

Performance was better at the end, during the easier run. No, the tingling fingers were not there, no high temperature, nothing special.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Appendix A.6 Pilot 6

Scenario 1: Run 8 – Run 6 (09/05/16)

CURVE

Where were you during the whole run? When did you move from one point to another?

Run 8: Focused at the beginning (because of turbulences and differences with respect to real aircraft), then going back and forth between under pressure and struggling. He started to feel uncomfortable (moving to under pressure and then to struggling) when he noticed about fuel shortage and ATC told they were 15 in the sequence, with delay vectors. It improved after they declared emergency, as they were then 1st in sequence and some stress went down ("a relief"). The loud noise did not impair his performance and not worsened his status.

Run 6: he remained always focused. Maybe he had a short moment in which situation worsened, when he realised there was tail-wind. But after he understood it was within limits, then he got back to focused – so it was just a moment. . It was not as uncomfortable as Run 8, as the situation often happens in Frankfurt. He had a better overview of the situation, not that many negative contributing factors.

GENERAL

Was the simulation realistic for you?

Run 8 was a bit unrealistic because of captain's decision making (was not supportive and very neutral – not really collaborating to decision) and fuel situation. Also the captain did not feel uncomfortable – so he was wondering whether he was the one being wrong in feeling stressed.

Run 6 was realistic. Also in this run the captain remained neutral and not very supportive. He thinks he flew the scenario in the same way he would do in reality.

Do you feel you reacted as you would have done if it had been real?

For both runs: yes, most likely. Looking back, the course of action looked he had seems very reasonable. However, because he missed captain support (like confirmation on decided course of actions), he cannot be 100% sure.

Where you very much surprised by the events?

Run 8, apart fuel problem, was very surprising for me) the loud noise – he thought of engine problem or sudden decompression and ii) for the problems in stabilising the approach when out of the clouds, as he was a bit too high and offset. The problems in approach all contributed to the not-perfect stabilisation on 1000 feet.

What was your feeling during the events of the scenarios (Were you stressed, with high workload etc.)?

Run 8: definitely stressed, mostly after fuel realisation. Keeping the airplane stable in gusts was challenging, but not stressing as fuel shortage. Loud noise was nothing compared to these other problems.

Did you feel in control of the situation after the events occurred?

Run 8: He felt a reduction of his capacity of controlling the situation – e.g. was not able to calculate times and crosswinds. He was lagging behind and not able to cross-check. Captain was fast and he was slow and not accurate instead. So he had to totally rely on captain's calculations. He had no complete overview of all the relevant information.

Was there a particular moment where you felt less in control?

Run 8: he missed an ATC communication about winds, so he had to ask again as he could not realise numbers in his head.

PERFORMANCE

In hindsight, what might you have done differently?

Run 8: he thinks that he could have managed better the flight in terms of rate descent and speed control – even though he thinks that the sub-optimal performance was probably caused by the differences in the simulator he had to quickly adapt to. He also had to use speed brakes. He would have liked to give control to the captain and check better the situation as he was unsure and had no real understanding of what was going on (no situational awareness). Landing also, was not a nice one. He is not satisfied with his performance.

Run 6 was managed better even though he had to use speed brakes, as the aircraft was decelerating slower than desired – something he tries to avoid generally. All in all, the flight was better stabilised compared to Run 8. He was satisfied with his performance.

Do you think the way you performed could have had an impact on the safety of the flight?

Run 8: he was feeling safety was at danger – i.e. if something more adds to the situation (mostly wind gusts barely at the limit and fuel shortage) then things go out of control. He felt close to the border of loss of safety. His feeling about safety improved just few seconds before landing – he was worried about safety because he was too high and offset

Did you feel your performance was affected by the previous run?

Yes. Because of all the troubles in Run 8 he was expecting catastrophic things in Run 6, something that it was also taking off some of his capacity.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

How was your interaction with:

- **The captain?**

Both runs: he was expecting some more participation to the situation and decision from his side. Captain was too neutral. There was no hint from him about what was the better course of action.

Scenario 1: Run 1 - Run 7 – Run 4 (09/05/16)

CURVE

Where were you during the whole run? When did you move from one point to another?

Run 1: always relaxed. Flight was just a routine one.

Run 7: between focused and under pressure. There was a go-around so for that there was an increase of workload that was anyway OK to handle.

Run 4: lot closer to struggling from the final approach onwards. He was having problem in staying aligned with the runway and keeping the proper speed. Also he needed a lot of attention because it was non-precision approach and there was a lot to do.

GENERAL

Was the simulation realistic for you?

Run 1: yes, very similar to reality.

Run 7: he just had two go-arounds in his career, instructed by ATC because of loss of separation. So, he did not experience a go-around because of unstabilised approach. Here he was 20 knots above limits – one time he had wind shift in Geneve, but still it was within limits. So, not realistic considering his real life experience.

Run 4: never experienced something like that – a change of runway because of sudden and strong wind change.

Where you very much surprised by the events?

Run 1: no surprise as nothing occurred.

Run 7: he was fully aware of the consequences of the wind shift, but did not realise the wind shift was the reason. He just though he had too much power. He was surprised by how much difficult it is to keep the parameters OK in manual flight. But wind itself was not a surprise.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Run 4: not surprised but experiencing a lot of workload, mostly due to strong winds and non-precision approach. Wind itself was not a surprise, theoretically one can expect a change of RWY. He was not actually expecting it – in fact he wanted to go for the approach and see how it would go. But then ATC informed of RWY change.

What was your feeling during the events of the scenarios (Were you stressed, with high workload etc.)?

Run 1: no particular feeling

Run 7: never felt like lagging behind or pure reactive.

Did you feel in control of the situation after the events occurred?

Run 1: always.

Run 7: He was feeling more in control when compare to first slot of runs. Got constantly a better overview of situation. Even though there was less fuel, he was not as upset as in the first run, as they were not number 15 in the sequence, he knew there was another RWY available etc. So he was not as worried as before.

Run 4: was in control, but it was difficult. He needed to delay the final landing checklist in order to have time to get full control of the aircraft. He was reminded two times by captain but he needed to delay it.

Was there a particular moment where you felt less in control?

Run 1: no

Run 7: not really.

Run 4: At some point the help of the captain was fundamental to stay "on track" and not go beyond the limit. The help was by the captain in form of altitude and distance indications – something really necessary to keep control.

PERFORMANCE

In hindsight, what might you have done differently?

Run 1: all was managed in an optimal way.

Run 7: he would have decided earlier for the go-around but the captain delayed that. Also he would have been more conservative, by reducing speed earlier and enter final configuration earlier he would have probably avoided the go-around. Not sure though, as there was a substantial wind shift. The start of simulation out of the blue did not allow a real approach briefing to be done. So he would have done this if he had time (NB not a problem regarding the way he managed, but a constrain coming from the way the simulation was set). This had an impact on his performance – e.g. he miscalculated the point of descent.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Run 4: as above, the start of simulation out of the blue does not allow a real approach briefing to be done. So he would have done this if he had time (NB not a problem regarding the way he managed, but a constrain coming from the way the simulation was set).

He needed some time to calculate the wind component, as he is not used to do it for head and tail – as crosswinds component are more important.

Did you feel your performance was affected by the previous run?

He was expecting bad things to happen when starting this second block. So, on first run at least he was quite alerted.

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

Did you feel in a hurry to get any particular information from your colleague or from the cockpit? If that was the case, which one?

Run 1: no

Run 7: no

Run 4: no

How was your interaction with:

- **The captain?**

Run 7 and Run 4: Communication was OK. Not much but he said the right things at the right time.

- **The ATC?**

Nothing particular to say

Was there something you might like to have seen on the cockpit instruments, or heard from ATC or your fellow pilot?

Run 1: no

Run 7: no

Run 4: no

BEHAVIOURAL

How did you realise that your performance was declining?

Run 1: none

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Run 7: none

Run 4: maybe something in here, but he is not sure whether it was because of the nose running and coughing. At some point he got warm and was sweating.

Run 5 (10/05/16)

CURVE

Where were you during the whole run? When did you move from one point to another?

Run 5: He was focused all the time, never under pressure.

GENERAL

Was the simulation realistic for you?

It was realistic.

Do you feel you reacted as you would have done if it had been real?

Yes – all the limits of the simulation apply, i.e. no possibility to make a full debriefing before approach.

What is your general impression about the scenario?

He felt stressed when he realised about the fuel.

Where you very much surprised by the events?

He realised quite fast about the fuel. It was the second thing he checked when the scenario started (after winds). Therefore, he was not very surprised.

He got a bit surprised by the loud noise, but he was not affected by that.

What was your feeling during the events of the scenarios (Were you stressed, with high workload etc.)?

He was asked about the Plan B only after the Captain asked for it – this happened after declaring the emergency. He realised he did not have any Plan B and he started figuring it out. Also he was probably too confident because the approach was to his home-base (Frankfurt).

Did you feel in control of the situation after the events occurred?

He was always in control

Was there a particular moment where you felt less in control?

Never, he was smoother in the approach when compared to the previous runs with strong winds.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

PERFORMANCE

In hindsight, what might you have done differently?

None. On the opposite, he started to reduce speed way earlier and made more use of the flight director – this left him with more capacity to face other potential issues.

Do you think the way you performed could have had an impact on the safety of the flight?

None

Did you feel your performance was affected by the previous run?

He was expecting bad things to happen, especially fuel. However, he also built some more "ability" in manoeuvring the aircraft, thanks to experience in the previous days.

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

Did you feel in a hurry to get any particular information from your colleague or from the cockpit? If that was the case, which one?

No. He never found himself waiting for critical information or asking himself what the colleague is doing. He always had the information he needed on time.

How was your interaction with:

- **The captain?**

OK. Captain asked ATC information at the right time.

- **The ATC?**

See above

Was there any aspect of the instrumentation that confused you or did not help at the time?

None.

Was there something you might like to have seen on the cockpit instruments, or heard from ATC or your fellow pilot?

None.

BEHAVIOURAL

Speak about specific BM

None – at least none he could remember.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Appendix A.7 Pilot 7

<u>Run 5 - Run 3 (10/05/16)</u>

<u>CURVE</u>

Where were you during the whole run? When did you move from one point to another?

<u>Run 5</u>: I felt a bit under pressure when I realised the fuel situation (what to do next? Getting to know the colleague…), then I felt struggling for a while, estimating the situation (I didn't know what I had in terms of support from my colleague, and about other things that could have had an impact on the decision for a go around). Then without the Go Around I get back to under pressure. The situation all in all was quite tense.

<u>Run 3</u>: it was basically just the approach, so I stayed more or less focused all the time.

<u>GENERAL</u>

Was the simulation realistic for you?

From 0 to 10, it was realistic 7, because the simulator is very different from a real aircraft in all the small details, especially the fuel prediction. Fuel prediction numbers are very important and you look at them for your calculations, and in this simulator it was just a rough estimate.

Then the entrance to the scenario was a bit strange. In real life you probably stay long in a holding and be prepared to the situation. In this simulation you go from 0 to 100 altogether, in real situation never happens like this, you never forget about fuel and realise that you are running out of fuel all in a sudden. The low fuel condition was not much realistic. You might come to a situation with little fuel but you know why, you know what you did, you already are in the mental mood.

The realistic component – the realistic feeling of a critical situation was not 100% there. The turbulences were realistic, but in real life I have a lot of adrenaline, more than in this simulation.

Do you feel you reacted as you would have done if it had been real?

<u>Run 5</u>: I did it as I would have done in real life. I went immediately with a Mayday, immediately set for the approach and nothing else mattered. I wouldn't have done anything else a part flying directly towards the airport no matter what.

The same for <u>Run 3</u>, it was a standard approach to me.

Where you very much surprised by the events?

<u>Run 5</u>: I was really surprised by the low fuel. The loud noise was ok, and there was anything else that surprised me. But the fuel situation was annoying.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

What was your feeling during the events of the scenarios (Were you stressed, with high workload etc.)?

Run 5: I was a bit disappointed by the fuel, I felt like I missed something. I would have realised it by myself, while actually the captain pointed out the fuel situation. He pointed out and then I realised the situation, and that was the entrance of the scenario for me. I didn't realise by myself because I thought we were in a normal approach and in a normal approach you never are in this situation. I didn't think about the possibility of being low in fuel. If (the scenario) started a bit more in advance and with a proper briefing it could have been different. But with the short briefing so close to the ground and with manual flying I just missed out the fuel; in a normal briefing you check out the fuel condition. The possibility of a Go Around made me feel quite tense.

Did you feel in control of the situation after the events occurred?

Run 5: I felt even more than in control because I had the feeling that the captain was not as keen on getting to land as I was. I think I even took control, command and everything (declare emergency and all the rest) because I had the feeling that he didn't want to proceed more. I took control to immediately get to the airport, so I felt very much in control, more than usual. The loud noise was annoying, it made it difficult to communicate, but it didn't affect my control of the situation; the fuel was my main focus.

Run 3: I was in control.

PERFORMANCE

In hindsight, what might you have done differently?

No. In real situation I would have had probably a better feeling, here I was not much prepared for this airport and approach. It was an unfamiliar approach to me and the preparation briefing was too short to me. I couldn't decide about what kind of missed approach I wanted, because I was not familiar with the terrain situation. I missed probably the clearance to an altitude and therefore I didn't switch the altimeter, but I didn't realise when it happened, I realised it in the final.

Do you think the way you performed could have had an impact on the safety of the flight?

A part from the fuel situation, which was risky but didn't depend on me, I think that the decision of going immediately back to the airport and the fact that I turned towards the airport without the ATC clearance could have had an impact on the safety of the flight, especially in a scenario with a lot of surrounding aircraft. I'm not sure I'd have done it in real life too.

Did you feel your performance was affected by the previous run?

Yes, because during the second run I was tense waiting for something to happen (and then nothing else happened a part from the turbulence).

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

Did you feel in a hurry to get any particular information from your colleague or from the cockpit? If that was the case, which one?

I would have had more support from the PM. I was surprised that he took the situation so calmly and didn't immediately call for Mayday.

How was your interaction with:

- **The captain? The ATC?**

The interaction with the captain was fair in the end, he worked fine, he gave suggestions, he was just very passive. I expected different reaction from his side due to the gradient of hierarchy in the cockpit (the captain is the captain). In this case he would have called the direction. But a part from that he was a good co-pilot, he did everything ok, he gave me all the information.

Was there any aspect of the instrumentation that confused you or did not help at the time?

The real FMS from the real Airbus would have helped a lot, in particular the Airbus Fuel Prediction page. There was a fuel prediction in the simulator as well, but it was just a button but it didn't tell you anything. The regular aircraft system in this situation would have helped a lot. The Electronic Flight Bag would have helped as well.

Some instruments in the simulator were different, I didn't have the exact distance from the aircraft to the runway thresholds. The ILS DME was missing, and it is quite helpful in the approach phase.

BEHAVIOURAL

How did you realise that your performance was declining?

When you have a bad approach once you are at the airport you realise the adrenaline that was in. You don't realise it while you are in the air as you are focused.

Just a mental feeling that I missed something (Run 5).

Speak about specific BM

In my case, I can tell that most of the time I lose the oral and audio channel first, and it was probably what happened with the QNH. The loud noise was annoying (that's why the deep breaths) but the real problem was the fuel.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Run 1 - Run 6 - Run 8 (11/05/16)

CURVE

Where were you during the whole run? When did you move from one point to another?

Run 1: between focused and relaxed.

Run 8: started with radar vector for direct approach, so we were probably under pressure; then we moved back to focused because we had a plan (we went for a direct approach) but in the approach itself the localiser went off and we went between under pressure and struggling, and we made the missed approach. However, as everything went quite fast we realised that we had enough fuel for a second approach and even for a third one, so after the completion of the go around and the new approach we started to get back in between under pressure and focused. It was pretty close to the fuel emergency but still manageable.

Run 6: we started relaxed, during the localiser shift we became a bit under pressure and therefore I missed to reduce the aircraft a little bit, because the attention was on the localiser and on keeping the right altitude. We were just right upon in limits for the missed approach (10.000 feet point), and after that there was no pressure anymore, and I ended between relaxed and focused.

It's hard for me to tell about the feeling situation, because I'm really focused on the actual task and on the instruments. I don't think about how I feel, I don't notice how I feel.

GENERAL

Was the simulation realistic for you?

Yes, they could happen.

Do you feel you reacted as you would have done if it had been real?

Yes.

Where you very much surprised by the events?

I was surprised, I didn't know they were coming even if I expected that something could happen. For the localiser interference I wandered if it was a simulator issue or supposed to be a real problem.

Did you feel in control of the situation after the events occurred?

In Run 8 the decision for the go around was the closest point to losing control of the situation, because there were several things together (the noise, the localiser off, the difficult communication). The go around initially was not super calm at the beginning, it was a high workload situation, but overall after that the situation improved. The go around helped to get out of a confused situation, and we could make another controlled approach. Just before the go around I was not 100% sure to be able to make a correct

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

approach because of the disrupting factors. And the low fuel condition had quite an impact on the sequence of events.

In Run 6 I felt much more in control of the situation (than in Run 8) because we only lost the localiser, we could still communicate and we could still track the altitude and the distance without the use of the localiser so we still navigated along the navigation green line. Since we had more time to talk about the situation, the situation was not so confused, I did not feel in loss of control. I didn't realise about the wind shift, we gained a lot of speed (I put a lot of power in) but again I didn't know if it was a simulator issue. It impacted the stability of the flight in the end, but we didn't mention the wind as a factor. I just mentioned that we had too much speed, but I thought it was mainly because of my reaction with the thrust lever. I didn't realise it was due to outside conditions. I basically reacted to the thing but I didn't mention it verbally.

PERFORMANCE

Do you think the way you performed could have had an impact on the safety of the flight?

I would have got established earlier in the Run 6, because the distraction of the localiser made me forget about the configuration of the aircraft. So we prolonged the configuration of the aircraft and therefore we went close to our limit of 10.000 feet. I would have wanted to do earlier, in time.

For Run 8 I think that the go around was necessary in that situation, as it was quite severe. In that particular case, with no localiser, no ground contact, almost no communication I suppose I would have done a Go Around even in real life.

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

Did you feel in a hurry to get any particular information from your colleague or from the cockpit? If that was the case, which one?

From the ATC it would have been good to have information about the localiser instability, because if you know what to expect is easier to react, to be mentally prepared to that situation (have an alternative plan, have another type of approach etc.). Sometimes this information is given in the NOTAM (for example if the localiser signal is weak) or reported from previous aircraft.

Was there any aspect of the instrumentation that confused you or did not help at the time?

Not really.

Was there something you might like to have seen on the cockpit instruments, or heard from ATC or your fellow pilot?

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Visual information on terrain situation, like in google map, on a reliable system. Terrain is our main risk, this is why we have to stay on the localiser, the localiser provides safety areas and ensures that we are safe during the descent. That's the only reason why we are worry if we are not on the localiser. If we could have another back-up system that shows you the obstacle situation or if you are still on the safe space terrain-wise it would be good in these situations.

In Run 8, I needed exactly what was provided by the PM.

BEHAVIOURAL

How did you realise that your performance was declining?

In Run 8, with the low fuel, when I realised that we were in emergency I had the feeling that I needed to be more assertive, more in command, to push more for a quick return. It would have been interesting to see the reaction of the crew if both pilots were in the same situation (both subjects, and not subject + confederate). In this case it would be possible to see the evolution of the situation with the real roles, he would have been the captain (in charge for the aircraft) and I would have been the first officer; that might have changed my behaviour a little bit, as in emergency situation the teamwork goes down a bit, the captain can ask for my opinion but then decide very fast about what to do.

In this special scenario I had the feeling that I needed to be the active member and decide what to do, which is quite unusual for the actual hierarchy. A regular line captain there would have pressed very much for the approach, and I would have supported as good as I could.

Speak about specific BM [Mentioned tone of voice, content of communication, seat movements]

I don't realise that I change my behaviour, or that I move. I don't have thoughts about feelings, about passengers, or whatever, I don't worry about crashing, but I'm really tense and focused on the key parameters (speed, glide slope, localiser) – that is all I have in my mind. But then, as soon as we touch down, I realise that the breathing was tense, that maybe I hold breaths. But I realise that only once I land, not during the approach.

Run 4 - Run 7 (11/05/16)

CURVE

Where were you during the whole run? When did you move from one point to another?

Run 4: I moved from relaxed to under pressure, and I stayed under pressure more or less during the whole approach because the approach was really demanding because of the wind and the NDB (weak instrument in this case). I moved because of the increase of the task demands.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Run 7: the approach was all the time in the area between focused and under pressure, it was not as demanding as some of the others because we started the approach pretty soon and we were above the final reserve. The localiser deviation was not that bad, it came back few second later. After the localiser deviation stopped I went back to focused.

GENERAL

Where you very much surprised by the events?

Yes, I was surprised by the NDB (Run 4) as I expected an ILS. Maybe dissatisfied more than surprised.

Did you feel in control of the situation after the events occurred?

Yes, I felt in control of the situation in both runs. It was just harder to maintain the control with the NDB approach, it was more workload.

PERFORMANCE

In hindsight, what might you have done differently?

Run4: We asked once for a prolonged downwind to get a little bit more time to prepare the aircraft (asked 3 more miles). Maybe it would have been the case to slow down the approach and take more time to be established, some more miles would have been better.

Do you think the way you performed could have had an impact on the safety of the flight?

Not really. The NDB approach was not super stable, that could have been performed a little bit better but it was not risky. And the visual was good, so the landing was not a problem; with different weather conditions (low ceiling) an alternate airport would have been better, but with that visual it was not a problem.

I was quite used to the low fuel situation, so I was more calm than the other times, we could handle it a little bit better. Also, we reacted really early to that condition, we asked for a direct approach and therefore we were in a good track.

Did you feel your performance was affected by the previous run?

For sure my performance in these runs was affected by the previous runs. First of all I get used to the simulator, then I experienced two runs with low fuel therefore the handling was better – probably more relaxed. Also the experience with the localiser interference had an impact, I just mentioned that there was the interference and then waited for having it back. I cross checked with other instruments, we were on track and safe from the altitude point of view. It was not that challenging. Also probably the two runs were easier for me because I'm familiar with Hannover airport.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

Did you feel in a hurry to get any particular information from your colleague or from the cockpit? If that was the case, which one?

No, I think the runs were not that complicated and I got all the information I needed.

How was your interaction with:

- **The captain? The ATC?**

In the second run (Run 7) the ATC was a bit strange, as if he was a trainee.

Was there any aspect of the instrumentation that confused you or did not help at the time?

Just simulator specific stuff, unusual and not so helpful like in a real aircraft.

Was there something you might like to have seen on the cockpit instruments, or heard from ATC or your fellow pilot?

I cannot think to anything. If I have to think about more information for Run 4 I'd think to an ILS.

BEHAVIOURAL

How did you realise that your performance was declining?

No, I didn't realise anything, mostly focused on the main task. I don't have this self-awareness.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Appendix A.8 Pilot 8

Run 7 (10/05/16)

CURVE

Where were you during the whole run? When did you move from one point to another?

At the beginning between under pressure and struggling, just to get used to the simulator (it behaved differently from a real aircraft). Then in the end I was quite close to losing it, because I had the fuel situation that was very tense, you have to land because you only have one option, also the weather situation was likely to maintain the same. Finally, the localiser interference was really unexpected from my side. Finally we were stabilised on the approach, but I lost the tail wind component on the approach. I was a bit behind.

GENERAL

Was the simulation realistic for you?

He had to get used to the simulator, as it is a bit different from reality. Thrust levers were quite difficult to move. The indications did not seem to be reliable as in reality. Coming to the scenario, he never found himself in such a situation for what concerns weather at least. The simulation was lacking some realism as there is time enough in reality to realise about the bad evolution of certain situations like fuel.

Do you feel you reacted as you would have done if it had been real?

Yes, even though he had to get used to the simulator and this required some time.

What is your general impression about the scenario?

Loss of LOC was quite distracting, but fuel was the most impacting factor. He did not always feel with full situational awareness during these two events.

Where you very much surprised by the events?

He felt quite lost and surprised by the loss of localiser – was close to lose the whole situation. The loss of LOC affected his trust in the glideslope, so he doubted about the whole ILS reliability. Another breaking point was the realisation about the fuel situation.

What was your feeling during the events of the scenarios (Were you stressed, with high workload etc.)?

He realised about the wind component going from head to tail, but he was quite target fixated on the approach itself – he felt like being a "bit behind", he was losing the overview of the whole situation. He was not able to carry on secondary tasks, like calculating better wind impact in his mind and he just had in mind "rough numbers".

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Did you feel in control of the situation after the events occurred?

Not always. During the loss of LOC and during the approach he was above the glideslope and then a bit too much down – in both cases he felt close to lose the control and he had wrong power settings, though speed was OK (he was bumping through the glideslope).

Was there a particular moment where you felt less in control?

See above.

PERFORMANCE

In hindsight, what might you have done differently?

He did not have too much time to think about the descent profile – so it was not optimal, but again a continuous descent would have been a "nice to have" but not necessary considering the situation.

Do you think the way you performed could have had an impact on the safety of the flight?

He thought that many things had to go wrong to find himself in such a situation (bad weather, low fuel etc.). Therefore, he was expecting bad things, especially in the moment in which he realised about the fuel.

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

How was your interaction with:

- **The captain?**

He would have wished the Captain to ask for info about the LOC interference, to understand whether there was any big trouble with the ILS overall. His trust in the glideslope was affected, so he felt compelled to check the minimum safe altitude.

He could have also monitored better the fuel situation.

- **The ATC?**

Information from ATC was basic to say the least, not very helping.

Was there something you might like to have seen on the cockpit instruments, or heard from ATC or your fellow pilot?

Thinking about the tailwind component, each manufacturer and each airline has limits for landing. It would be great to have better visual indication when the limit is being approached. A flashing indication would be great to catch attention of the pilot.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Another thing which he has seen on the Embraer, is the visualisation of the optimal descent profile on the vertical plan, with a symbol for the airplane moving and showing if it adheres to the best profile.

Run 4 - Run 1 - Run 8 – Run 3 (10/05/16)

CURVE

Where were you during the whole run? When did you move from one point to another?

Run 3: Struggling shortly before 1000 feet gate as sink rate and pitch were not good and he was going "sinusoid" through the descent profile – he was not well stabilised. He was feeling like failing at the moment of go-around as he was "target-fixated" and wanted to land the aircraft, afterward focused. But then again almost under pressure, because he was not on a standard pattern when he re-tried the approach so he also needed to use speed-brakes.

Run 4: Starting somehow more than focused and then moving to the under pressure area. He had to fly a manual NDB approach that is quite unusual for him. He had to fight with NDB indication as it is not a usual way for him to reason. There was a go-around but he was feeling more in control as the decision fully came from him (while in Run 3 there was "suggestion" from captain).

Run 8: From focused to failing during the scenario. He did not realise the wind shift – many things happened! Fuel was the biggest concern again. He was under pressure as soon as he realised about the fuel – which was not immediate realisation.

Run 1: always relaxed, no wind, nothing bad happened.

GENERAL

Was the simulation realistic for you?

Runs were not realistic for what concerns knowing about fuel situation. Also, with these winds they would not fly manual. NDB would be also flown with the help of automation, not fully manual. Run 1 was very realistic, maybe even too much fuel.

Do you feel you reacted as you would have done if it had been real?

In general, yes.

What is your general impression about the scenario?

Run 3: from the intercept of glideslope he adopted a defensive strategy, by setting earlier the final speed, to control better the aircraft despite the strong winds and not bust any limit. He was fully aware of the situation.

Where you very much surprised by the events?

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Run 8: the loud noise was like a strong wake up call. He was fixated with the LOC problem, looking for a good way to maintain correct pitch and vertical speed. Fact is that his trust in glideslope was affected by the sudden disappearance of LOC – so he started to focus on what could have been affected by an unreliable glideslope.

Was there a particular moment where you felt less in control?

Run 3: it was the worst as he lost situational awareness due to his fixation on landing the aircraft. The situation was saved by the captain who instructed the go-around. In this situation he was more in reactive mode and he was not able to plan ahead, relying on the captain for the most.

Run 4: on first NDB approach, between 1000 feet gate and go-around he was acting reactively.

Run 8: in this situation he was fiddling a lot with thrust levers as he was struggling a bit with them and the sidestick as well, probably due to the differences with reality of the simulator itself. Also he felt like he has to increase the visual scan after the bang sound, but was not able to make meaning of all the information he perceived on the instruments.

PERFORMANCE

In hindsight, what might you have done differently?

Energy management was quite problematic in all three cases, the simulator characteristics being different. It was a bit problematic to stabilise the aircraft at the 1000 feet gate, especially with the turbulence and wind shift.

Do you think the way you performed could have had an impact on the safety of the flight?

Biggest concern for safety was the lack of fuel. For the rest he thinks the flights were not jeopardized by his actions.

Did you feel your performance was affected by the previous run?

Run 8: he did not immediately realise about fuel because in the previous run (Run 4) fuel was fine. Moreover, in the last run he was quite exhausted. Also, relaxing with Run 1 at the end was bad, because he suffered from the increase of difficulty in the last run.

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

Did you feel in a hurry to get any particular information from your colleague or from the cockpit? If that was the case, which one?

Run 8: he would have liked to understand why delay vectors were given by ATC, as they were out of the blue on the top of a low-fuel situation.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Was there something you might like to have seen on the cockpit instruments, or heard from ATC or your fellow pilot?

Run 8: He struggled to find a good correction angle considering the wind as there was no track information with the wind was available on PFD (it something present even in case of electrical failure!). He could not remember how to calculate the correction angle and he was flying in manual mode. If a similar information could be accessible in manual mode, anyway it should be accessible in 2/3 "clicks" otherwise it would be too distracting. As an alternative, calculation to correct angle could be done by PM.

Run 4: For NDB approach, it would be great to have a visualisation of the deviation from vertical profile (Embraer-like).

In all situation with low visibility: it would be great to have a picture of the runway – e.g. through infra-red camera – while being in the clouds, so that you already have expectation in your mind on what you will find out of the cloud base.

BEHAVIOURAL

How did you realise that your performance was declining?

He did not feel like having any symptoms of performance decline.

Run 6 - Run 5 (11/05/16)

CURVE

Where were you during the whole run? When did you move from one point to another?

Run 6: Always focused

Run 5: Started under pressure, than moved between under pressure and struggling during the loud noise. Finally when the noise ended went back in between focused and under pressure. The overall performance improved after that event.

GENERAL

Was the simulation realistic for you?

Run 6: Yes, it is not unusual to start approach with tail wind, especially in Frankfurt, even though he did not experience very late changes. Therefore, it was "80% realistic".

Run 5: Not so realistic, as mental picture of fuel situation builds up during the whole flight.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Do you feel you reacted as you would have done if it had been real?

Yes / Yes

What is your general impression about the scenario?

Run 6: He was quite aware of the whole situation, except about the tail wind, even though he was kind of primed by the previous runs. Situation not too complex even though there was a localiser failure. His trust in glideslope was not affected. In the yesterday run, his "faith" in glideslope was affected. Overall, approach was stable all the time.

Run 5: Overall it was quite OK. There were not environmental factors influencing the flights, the biggest problem being the lack of fuel – something that always puts stress on him.

Where you very much surprised by the events?

Run 6: Wind shift was somehow surprising but not too much, but not the localiser failure as he was somewhat primed. He never thought of aborting the approach, as profile was normal all the time.

Run 5: Low fuel is always a problem – despite when realising about it, stress comes in. "Bang" noise was not too much of a trouble, even though it was somehow more stressful than the one experienced the day before despite being already encountered in the run (NB: not clear why despite a number of questions).

What was your feeling during the events of the scenarios (Were you stressed, with high workload etc.)?

Run 6: Not stressed and not impaired for what concerns situation awareness. Some increase in workload though.

Run 5: He felt stressed because of the low fuel situation.

Did you feel in control of the situation after the events occurred?

In both runs he thinks he was always in control, with the exception of being late in Run 5 with the configuration of the aircraft prior to the 1000 ft gate, as he was disturbed by the loud noise.

PERFORMANCE

In hindsight, what might you have done differently?

None

Do you think the way you performed could have had an impact on the safety of the flight?

Run 5: In Run 5 he had the feeling that safety of the flight was possibly jeopardized not because of his action, but because of the low fuel situation that was definitely the most stressing factor. His idea of safety in approach is "landing with at least 1000 kg of fuel".

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Did you feel your performance was affected by the previous run?

Run 6: He was primed by the wind shifts already experienced.

Run 5: He was primed by the bang sound already experienced. Fuel problem was already known as well, but still he affected his stress situation.

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

Did you feel in a hurry to get any particular information from your colleague or from the cockpit? If that was the case, which one?

Run 6: In addition, he would have liked the Captain to provide information about the tail wind component, as the Captain realised that but did not provide any information in real time about that. This in fact could have impaired the configuration of the aircraft, i.e. he could have arrived late in properly doing all the actions needed to configure

Run 5: He would have liked to get a "better" treatment from ATC, to be put straight into the sequence to land earlier. In addition, he would have liked track miles information earlier and in general to be informed more completely about ATC plans for his flight.

How was your interaction with:

- **The captain?**
- **The ATC?**

See above

Was there something you might like to have seen on the cockpit instruments, or heard from ATC or your fellow pilot?

Run 6: He would have liked to get a repetition of the RVR information by Tower ATCO – which did not occur.

Run 5: At some point he had the impression both ATC and Captain were at a different level of awareness – like they were not worried

BEHAVIOURAL

How did you realise that your performance was declining?

Run 5: He realised that during the bang noise he forgot to command the plane configuration properly – he did not go for flaps 3 and flaps 4. So, it was not stressing, but disrupted his course of actions – and he could not afford to miss anything as he could not go for a go-around. He resumed the course of action during the sound was still on.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Appendix A.9 Pilot 9

Run 8 (11/05/16)

CURVE

Where were you during the whole run? When did you move from one point to another?

Run 8: In the beginning he was quite relaxed and stood there for a while. Situation started to change when he realised about how fuel – moving towards focused/under pressure. Loss of LOC definitely moved him to "Under pressure".

GENERAL

Was the simulation realistic for you?

He notices some differences with real aircraft – throttles and sidestick, the feeling of them was not the same as real. There was no interpolation on the Flight Director display, so curves were represented as 90 degrees corners. Also, out-of-blue fuel situation was not realistic. For the rest it was realistic even though he never experienced the events of the scenario all together in reality.

Do you feel you reacted as you would have done if it had been real?

Yes.

Where you very much surprised by the events?

- Fuel situation was the most surprising event. He wondered whether there was a leak – this was at top of his mind.
- Loud noise was surprising, but the effect lasted for just some seconds. He did not feel impaired by the sound.
- Localiser goes down and in reality he would have gone for a go-around immediately – but again the low fuel was quite worrying. After some discussion he decided to go-around, after the captain convinced him of the quantity for another short approach. He also realised that he could have gone also for a second go-around.

What was your feeling during the events of the scenarios (Were you stressed, with high workload etc.)?

He was looking for causes of the low fuel and ruling out easily the leak. He though the fuel was not enough, but the captain convinced it was OK to give a try.

When the LOC went down, his trust in the glideslope was not affected anyway.

He was feeling a bit reactive when he realised the loss of localiser, but in general he thinks he kept a good look-ahead.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Did you feel in control of the situation after the events occurred?

He was a bit unsure in the whole scenario, because of the small discrepancies between simulator and his usual reality (A320 cockpit) – like lack of interpolation, speed indication etc.

Was there a particular moment where you felt less in control?

He felt in danger when he discovered about the low fuel and also trying a go-around with that amount – i.e. he was worried about the quantity that was going to remain.

PERFORMANCE

Do you think the way you performed could have had an impact on the safety of the flight?

It would have been nice to see earlier the low amount of fuel – especially considering they flew for a long time on delay vectors.

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

How was your interaction with:

- **The captain?**

He did not expect any information about the fuel from the captain. He was just surprised and worried. It was unusual to speak in English. He switched to German for some seconds. He was surprised the captain was not concerned and was not aroused. He did not find him to be very supportive – he was too relaxed considering the situation. He did not think he was the strange one, but he felt he was one who needed to lead the situation. All in all, the behaviour of the captain added some more stress.

Was there something you might like to have seen on the cockpit instruments, or heard from ATC or your fellow pilot?

He would have liked to understand why there was so little fuel.

Thinking about new information, he would have liked to get from the instruments information about the consumption of fuel depending on different choices of trajectory, e.g. in case of go-around, in case of being delayed by ATC – it is hard to estimate for pilot because you have to keep on mind a lot of data.

Also, when on FAF / outer marker it would be great to know from the cockpit in advance what is the amount of fuel needed to perform a go-around.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

BEHAVIOURAL

How did you realise that your performance was declining?

He was feeling heart rate going up and realised he switched to different languages for some seconds – French and Spanish – which it is something he learned a bit during his last vacation. He realised this was a symptom of stress.

Run 4 - Run 6 (12/05/16)

CURVE

Where were you during the whole run? When did you move from one point to another?

Run 4: quite close to the focused phase a bit in between with "under pressure".

Run 6: he was focused and remained there for the whole scenario.

GENERAL

Was the simulation realistic for you?

Run 4: he has flown 10 NDB approaches in his life, so in this sense it was realistic. But the problem is that no briefing was done at the beginning of the scenario.

Run 6: it was realistic.

What is your general impression about the scenario?

In general, he was satisfied of his performance.

Where you very much surprised by the events?

Run 4: situation was rushed so he had to shorten the briefing and focus more on the situation itself. Overall it was a bit stressful.

Run 6: he did not realise about the wind shift; he noticed some change in the speed but did not associate that to the wind shift.

What was your feeling during the events of the scenarios (Were you stressed, with high workload etc.)?

Run 6: he was OK with the go-around, not much negative feeling. He felt a lot of workload but something bearable.

Did you feel in control of the situation after the events occurred?

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Run 6: he always felt he had enough time in front of him. He was neither lagging behind nor feeling reactive.

Was there a particular moment where you felt less in control?

Run 6: they were stabilised only at 970 feet and descent rate was a bit high and other small things were not good, so they decided to go-around.

PERFORMANCE

In hindsight, what might you have done differently?

Run 4: he would have been great to receive more updates from ATC on the wind.

Run 6: in real life he would have done the briefing earlier but here he had to do it very late. There some minor things not perfect but overall it was OK. The loss of localiser made him loss some seconds needed to configure the aircraft on time. But no big deal.

Did you feel your performance was affected by the previous run?

Not really felt affected by previous runs. Just checking soon the fuel in Run 4 which was the first of the day after Run 8.

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

How was your interaction with:

- **The captain?**
- **The ATC?**

Run 4: he would have been great to receive more updates from ATC on the wind.

Run 6: he would have liked some more wind updates from ATC. He likes a lot when ATC is supportive.

Was there something you might like to have seen on the cockpit instruments, or heard from ATC or your fellow pilot?

Run 4: it would be cool to have indications of the winds that are updated in real time. No audio, just visualisation. Also, track indication is always a big support especially in case of strong cross-winds. Functioning PAPI indication also would have been appreciated (Note: these are no novel tool).

Run 6: he would have liked a HUD in a scenario like that to avoid too much eye movements. In particular, considering the reduced visibility, there is a need to focus more on the instruments. A HUD would allow to group more closely together all the needed information. Info he would like: speed, attitude, altitude, glideslope, flap settings, winds. He would like this information to appear before final approach.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

BEHAVIOURAL

How did you realise that your performance was declining?

Run 4: None.

Run 6: None

Run 1 - Run 3 (12/05/16)

CURVE

Where were you during the whole run? When did you move from one point to another?

Run 1: Always relaxed as there was no problem at all.

Run 3: In between relaxed and focused due to wind gusts.

GENERAL

Was the simulation realistic for you?

Yes – Run 3 was quite peculiar due to the winds.

Do you feel you reacted as you would have done if it had been real?

Yes

What is your general impression about the scenario?

Both of them easy, compared to the previous ones.

Where you very much surprised by the events?

Run 1: There was no event at all.

Run 3: In between relaxed and focused due to wind gusts.

What was your feeling during the events of the scenarios (Were you stressed, with high workload etc.)?

Run 3: He had to be focused for the whole duration of the flight due to winds. He realised he was scanning the instruments a bit more compared to normal.

Did you feel in control of the situation after the events occurred?

Always, in both scenarios. In Run 3 he was just more focused.

Was there a particular moment where you felt less in control?

Never. He was always prepared in time for the 1000 feet gate. Never feeling of lagging behind.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

PERFORMANCE

In hindsight, what might you have done differently?

None. Approach was just slightly unstable but always under control. Never thought of doing a go-around.

Did you feel your performance was affected by the previous run?

Not really, in both cases. He had no particular expectations. He was just generically expecting some events to occur in Run 3 after the "break" experienced in Run 1.

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

Did you feel in a hurry to get any particular information from your colleague or from the cockpit? If that was the case, which one?

No particular need as there were no particular events.

Was there something you might like to have seen on the cockpit instruments, or heard from ATC or your fellow pilot?

Run 3: He was missing the track indication (NB this is a simulator feature) which is very useful with strong winds – not used just to fly with heading in these situations.

Moments before touchdown, when looking outside because you are visual on runway, you need to give back and forth short looks to the speed – so it would be cool to have speed on HUD (no need for head movements).

BEHAVIOURAL

How did you realise that your performance was declining?

Run 3: because of focus needed to control the airplane with the gusts:

- Some heartbeat up;
- Sweaty palms (a bit);
- Some movement of the lips.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Appendix A.10 Pilot 10

Run 6 - Run 1 (11/05/16)

CURVE

Where were you during the whole run? When did you move from one point to another?

Run 6: at the very beginning, just relaxed and focused – like usual flying – and starting the approach it went to slightly struggling - before struggling, not really struggling; we were under pressure because we had problems and we had to find a solution (no localiser, nothing). Then slightly high as we had to come back to the glide, I felt a bit under pressure because my performance was focused on that and I couldn't focus on much things a part from that, I could still execute any emergency thing (go around and so on) but I could not focus on small details. As soon as we were on the glide and on the localiser, stabilised, I went back to focused. I realised I was moving towards struggling because information that was not relevant to the task I didn't treat anymore. For example, Air traffic controller said something about the wind, but I didn't pay much attention on that information. The wind was roughly in limits, so I didn't analyse – process that information more than that ("I put a tick in the box"); and when I do that I know that I'm under pressure with something. Thus, I focus only on flying the aircraft, and I rely on my Pilot monitoring to take care and to give inputs if he has the feeling that something is going wrong. I focus on the main task and pay attention only to task-related information.

Run 1: between Relaxed and Focused all the time.

GENERAL

Was the simulation realistic for you?

The simulator was not realistic to me, as I had to put more energy in flying the aircraft, and some of the things I'm used to have as power settings, pitch of the aircraft, were not correct. So I had to spend bit more energy on that, which means less energy for something else. For the first run, I had the feeling that the localiser was not that much a problem, but the problem was that some things that you are used to do were still automated but not as in the usual aircraft so it was not perfectly realistic.

Localiser not coming never happened to me but it is something that can happen in real life, and it is still an event that doesn't create much pressure because you know what you should do if it doesn't come. But the situation which comes from the localiser is not that realistic, because we don't descent if we don't find the localiser. And the next situation – which is being too high – was very realistic; for many reasons it happens often that we are too high.

Do you feel you reacted as you would have done if it had been real?

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

With both runs, we both reacted as a crew, as we would have done in real life. We reacted as a crew, we tried to find a solution to the localiser interference, in particular because descending without being sure of being on the localiser is not safe. We tried to give time to the ILS, to the localiser to come.

What was your feeling during the events of the scenarios (Were you stressed, with high workload etc.)?

I felt comfortable, there was no challenge, I was doing my job, which includes finding a solution for the problems you have. No particular feelings, just working.

Did you feel in control of the situation after the events occurred?

Always.

PERFORMANCE

Do you think the way you performed could have had an impact on the safety of the flight?

Never had that feeling. It was Frankfurt, I'm familiar with the airport and I know the environment around the airport, I know that everywhere around is flat, no mountains, so you can go down without problems, if you alone (without traffic in the vicinity of the TCAS) you can go down without any terrain. So I never felt like we were affecting the safety.

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

Did you feel in a hurry to get any particular information from your colleague or from the cockpit? If that was the case, which one?

For Run 6, I found challenging for a pilot to do the primary task but you still had to focus on the rest, any other pieces of information on what could go in the wrong direction (is the wind becoming too strong? Is the rate of descent too strong? Is the terrain becoming too close?). I'd like to have such kind of information, but not at warning level, I'd like to have a help that tells you that you are still safe but you are going in a "dangerous" direction (e.g. "reduce the rate of descent right now"). That's what I expect from my colleague or from the cockpit.

Let's take the wind as example. If the wind becomes too strong from the side you can still do the approach but you are not going to land because it's over the limits of the aircraft. Or tail wind for instance, you don't notice it until the landing. I'd appreciate that the aircraft – or the colleague put these data into a system that gives me a small warning telling me "pay attention, you have tail wind", as it affects my approach, with tail wind is going to be more difficult than you expect. At the moment, the system doesn't tell you that, the system only shows you the wind but it doesn't show you the link between tail wind and the approach phase. I'd love the system to recognise if we are too high and too fast and we still have tail wind condition, and tell me before to pay attention as in this situation it is going to be difficult.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

The wind is up to now is the most difficult problem, you cannot get down the tail wind; it is a factor that affects severely the approach. A system that is able to recognise this situation in advance is something that typically a pilot wants. It also something that the colleague could tell you, as he/she knows the track in the entire situation.

[In other phases of flight] Turbulences during climb with temperature. During climbing you could expect turbulences in specific situations with temperature (i.e. when the temperature rises or drops quickly) and at the same time you have a certain wind condition. As a human, is difficult to predict turbulences because you have to pay attention to a lot of information. However, the conditions and the patterns for turbulences are known, so it would be nice to have a system that informs you that turbulences can be encountered. I don't want something that predicts the future, but a system that gives you advices on things that could happen. It allows you to focus only on flying the aircraft, as even if you have energy and resources to concentrate on all the things aspects such as temperature, wind, may be skipped.

On modern aircraft, I have the feeling that the thresholds that are set for the alarms are the last line of defence. Which means that when the thresholds are overcome is too late to remain in a normal situation, it's time to act as an emergency situation. It would be good to set thresholds a little bit earlier – "soft thresholds" – to have a kind warning before the situation becomes too dangerous ("don't be alarmed, we are still safe, but this pattern is known and through this pattern something can happen, the situation may become critical" – "pay attention to this"); then, if you continue in that direction you have the real warning (for example "terrain terrain"). It should be a two-stages system (similar to the TCAS with traffic advisory and resolution advisory): soft warning – hard warning that forces you to act. It's useful because I can see things coming, so I can set my mind on that.

Was there any aspect of the instrumentation that confused you or did not help at the time?

No, but I'm very used to the interface. There are still things that I don't like on Airbus, as for example you don't have the speed brake information right in front of you, it's on the side. Speed brake can lead to stall so it's a very important information that you have to be aware of what you are doing with the speed brake; if this information remains out, at some point you will stall. It's a primary information and should not be on the side.

With respect to the interface, I love the colours – green, magenta. They are important, and there are not too many colours, just the relevant ones. And it's very "clean", you don't have too many information, you only have what's necessary.

How was your interaction with:

- **The captain? The ATC?**

The ATC I didn't have to interact, do anything. No major information came from the ATC, it was a standard ATC, just "clear to land" and weather information. And with the captain we didn't have much to say

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

actually. I had the feeling that he knew what we had to do, he knew what he had to do during the approach and I knew what I was doing so there was no much to interact, or decisions to meet.

Run 4 - Run 7 - Run 3 (11/05/16)

CURVE

Where were you during the whole run? When did you move from one point to another?

Run 4: I started focused, then during the NDB approach I was a bit under pressure; it was due to the NDB, as the NDB approach is quite demanding – without any help from the aircraft – but I saw the runway.

Run 7: started focused and moved to under pressure, "I was under pressure but I don't have the feeling I degraded my performance, the performance was still there despite the load of work". I was a bit under pressure because I knew that if something went wrong it could go really wrong and I knew that a go around could be possible. I knew it could go very quickly to struggling, so that's the reason I position myself on "under pressure" even if my performance was stable (didn't get bad).

Run 3: From focused to struggling. Turbulences were really strong. At the very beginning of the approach I was always focused as it was a basic flight through turbulences. But at some point it increased very much and when I started the descent, when I intercepted the glide, I started to move from focused to struggling. The more we get close to the ground the more I struggled because the margins started to reduce.

GENERAL

Was the simulation realistic for you?

Yes, all the three.

Do you feel you reacted as you would have done if it had been real?

Yes.

Where you very much surprised by the events?

Not really surprised, but it was a bit challenging.

In Run 4 change your mind from ILS to NDB is challenging because it is another type of approach, even if it was not much of a problem.

Run 7: I found the low fuel condition really demanding because you don't know what to expect, you don't know what the ATC will do, you don't know if it is going to be long, you don't have a plan. You just know "if I fly the go around it will cost me so much time to come back that it's a real emergency situation". It's

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

demanding because you are very very close to falling down, to call mayday and to land in an emergency situation because you don't have fuel, you don't see much. I found this run quite demanding.

Run 3: it was demanding because it was flying, really flying.

Did you feel in control of the situation after the events occurred?

I felt in control of the situation except at the end of Run 3. Maybe because the simulator was not realistic, but in that situation you do your best but you never know with the wind, you just do your best. If you go beyond certain limits you have to go around, but you never know. It was not a major problem to land in Run 7 with low fuel and low visibility (which was not a problem), but the wind is unpredictable and this is what makes you feel not in control of the situation.

PERFORMANCE

In hindsight, what might you have done differently?

Run 4: I wouldn't have changed a thing.

Run 7: maybe we should have decided together to tell to the ATCO that we didn't have much fuel. It was not a mayday situation but we should have communicated a bit more with the ATCO. We should have informed him that if anything had happened we were in a critical situation. I've noticed that most of the time, when you are low on fuel in your cockpit you don't come to tell that. I don't know why, but it's difficult to tell on the frequency "we are not yet low on fuel, but we don't have much fuel, if anything happens…". That's what I would have done differently.

Run 3: no, I just flew against the wind.

Do you think the way you performed could have had an impact on the safety of the flight?

No.

Did you feel your performance was affected by the previous run?

It was affected by the simulator actually. The more you fly that simulator, the more you learn how the simulator is flying, the more you find new things to counteract the programme of the simulator. My experience with the simulator affected the way I flew in the three runs.

These are small scenarios, so you get activated and you have to react soon. While I prefer to warm up a bit before activating my flying skills (it's like a warm pattern). I feel better if the situation is windy and requires flying skill, if I had it warmed. So if it's very windy I take the control, I shout down the autopilot very early before the glide intercept in real life. In this way I can feel the aircraft, I get warmer and I'm ready for the next steps. But I have colleagues that do the opposite, that don't do anything until the end (like going from 0 to maximum) and I've noticed that their landings are very poor. But in Lufthansa, in windy conditions, we tend to take the control early to feel the aircraft, because every wind is different. In

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

the case of the simulation, I think I've used the first two runs to warm up and be ready for the windy run. Thus, the first two runs affected my performance in the third one.

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

Did you feel in a hurry to get any particular information from your colleague or from the cockpit? If that was the case, which one?

No, I felt like I had to tell two or three times "fuel is our priority" (in Run 7) because I had the feeling that for the PM the situation was fine. I don't know if he was playing the role, but to me the situation was still fine but he had to be aware that my main concern at the moment was not the visibility but it was the fact that Hannover was the only option, if both runways closed we had to land on the taxiway. It was our only option. It was my main concern, I wanted to be sure he understood the situation and he was aware of the problem.

How was your interaction with:

- **The captain? The ATC?**

I think I brought him to the point I wanted him to be, I think I made it clear that our plan B in case of go around was still Hannover. We agreed that there was no other option and we agreed that our main concern was the fuel. I don't know if the interaction was good or bad, but I achieved what I wanted to achieve. From CRM point of view I would say that the crew knew where it wants to go.

I didn't notice anything from the ATC.

Was there any aspect of the instrumentation that confused you or did not help at the time?

In the simulator you cannot feel the energy level of the aircraft. In the air you can feel, you can predict what in two seconds will happen with the aircraft (considering the load of the aircraft, the pitch and the power). You can increase/decrease power just feeling the aircraft going up or going down. That's a thing that you acquire with the experience, and that's the kind of thing that is very difficult to feel when you don't see anything outside, a for example when you are flying an ILS and you are in the clouds (you don't feel that). I'd appreciate to have a bar that tells you "now with the wind drifting in the next two seconds with this power you are going to gain energy", so you can adapt the power. However I think this kind of communication is difficult to present on an interface. And probably is too much information for a pilot.

BEHAVIOURAL

How did you realise that your performance was declining?

I didn't notice anything, and I did act as I was supposed to do.

Speak about specific BM [Mentioned movements on the seat during Run 4]

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Maybe I moved a bit, and it was probably the most challenging situation for a pilot, flying an NDB, because there are a lot of things you have to think about. This time the weather was good and it helped, but it was quite challenging. Maybe I moved a bit because it was challenging. In fact, at the very end I realised I didn't put my seat correctly. At the beginning everything was fine but I got closer to the runway and I realised I was too lean on my seat. It had nothing to do with my seat position, it had to do with the approach, it was maybe connected to the stress level.

I paid attention to the stick grasp, I do it as well. I grasp the stick tight when the situation is challenging or demanding. For example, during the wind situation or during the NDB situation. Not during the low fuel situation because it was not a flying skill thing. I realised it because I paid attention to that, and I released the pressure as soon as I realised that, and I flew better. So I grasp a bit more the side stick when the situation depends on my flying skills. I can use it to realise when the situation is getting tense, and I can step back a bit.

Run 8 - Run 2 (12/05/16)

CURVE

Where were you during the whole run? When did you move from one point to another?

Run 8: I started under pressure because of the fuel, then there was the wind and I was still under pressure, then I found the aircraft difficult to fly because of the wind and it was a demanding ILS and the performance was not very good, so I was struggling. I struggled because of the wind, the ILS, and the fact that it was not a stable condition; then there was the noise and the localiser, so it was very demanding so I moved roughly here (he points the curve between struggling and failing), actually not failing but before failing because the performance became really low. I was just trying to find a solution, to know at which altitude I should have to stop, which was 2250 because below that altitude without the localiser we don't know any more if we are safe. I never came back to the normal performance until the landing due to the wind and the fact that we were never really aligned to the runway, and if I did a correction it never really worked properly.

Run 2: I was always between relaxed and focused.

GENERAL

Was the simulation realistic for you?

Yes it was. The simulator was not realistic to fly but the simulation was realistic.

Do you feel you reacted as you would have done if it had been real?

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Yes, exactly in that way. In the first one (Run 8) we did our best to switch off the noise and then the localiser disappeared, and in that moment I did the best that I could to continue because doing a go around was not really what I wanted in that situation. So I tried to stay on the ILS.

What is your general impression about the scenario?

It was a demanding scenario (Run 8), and the environment was very realistic. It was a sum of several small things that made the situation demanding for me. Loud noise and localiser disappearance never happened to me in real life, actually such a situation never happened to me, with the wind, with the weather, with the fuel… Too many factors at one time.

What was your feeling during the events of the scenarios (Were you stressed, with high workload etc.)?

The feeling was that I was not at my best level, but I tried to keep the aircraft within the limit of flying, just stay on the localiser as long as possible with the wind, stay on the glide slope as long as possible, and I felt really uncomfortable because nothing was stable in that situation, everything was moving all the time. You could concentrate and focus on one thing which is the primary task, but at one point you your primary task is not that one anymore because you miss the communication with your colleague… every time the primary task was a new one and you have to be sure that you achieve the new primary task. For example the second primary task was to have the gear down and I could not have the gear down because I could not talk with the PM. So the only thing I could do was to use my hand to tell him I wanted the gear down. There were many primary tasks, the hierarchy changed, I had to reconsider the situation every time and I still had to fly the aircraft.

Did you feel in control of the situation after the events occurred?

Run 8: I felt not really in control of the simulator because I found it difficult to fly, but I felt in control of the situation because I knew what to do, it was not perfectly done but I knew what I had to do. I never felt we didn't know what to do, as a crew we knew what we had to do and we did that.

Run 2: yes, all the time, it was a standard ILS.

PERFORMANCE

In hindsight, what might you have done differently?

Nothing, I performed exactly as I would have done in real life, including asking the PM if he felt well in that situation (Run 8).

Do you think the way you performed could have had an impact on the safety of the flight?

Flying an ILS like that… not an impact on safety… The speed was moving a lot, we were making a continuous descent, then the glide dropped down. We were in the limits, but the limits were very close. It was not the best ILS I flew because of the conditions. It was safe but we were really close to a go around.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Did you feel your performance was affected by the previous run?

I was active during the second run (Run 2), I was waiting for something to happen. I recognise that I'm active because my posture is different when I wait for something to happen, I move a bit my head, I feel like "looking for cues" with my eyes, my eyes move a little bit faster because I look if something is changing or not.

PERFORMANCE / POSSIBLE INSIGHTS FOR DESIGN

Did you feel in a hurry to get any particular information from your colleague or from the cockpit? If that was the case, which one?

No.

How was your interaction with:

- **The captain? The ATC?**

Yes, with the captain was really good (Run 8) because we couldn't communicate but he understood what I wanted, to get the aircraft configured for the landing. I asked if he could hear me, I wanted to be sure he could hear me, I wanted to know what his state was because I was confident we would make it with the sound but I was not sure he was fine, he felt safe and I didn't want him to feel uncomfortable and not be able to say anything. If I have this situation in real life too, when something happens and I feel in control of the situation but I don't know if the colleague feels well I ask him, it's the way I act. During landing I asked the PM "do you feel safe?".

Was there any aspect of the instrumentation that confused you or did not help at the time?

After all the runs I started to find the wind indication not in the right place. The wind indication is on the up left side of the display, but it is really away from my standard eye pattern. Normally it is not such a big problem because the wind is something that doesn't move much during the approach. But in every run it was difficult to fly the ILS because they were switching wind, it was moving all the time not like in real life. I know it was done on purpose to increase the workload and I had the feeling that this indication, that was important, was too far from my eye pattern.

Was there something you might like to have seen on the cockpit instruments, or heard from ATC or your fellow pilot?

Maybe a display showing that we were cleared to land could be useful. A written message instead of the voice message, something that you can read again if you don't remember it. When you receive the clearance you set your mind to land and after 20 seconds it may happen that you find yourself wondering "did I get the clearance?". In real life, in the second run, if I didn't remember the clearance I would have asked my colleague "I didn't hear the clearance, would you mind to ask again?". But in the first run, you don't have that much capacity. It is not an information that you process actually. You know more or less

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

you've got the clearance and you just land. Anyway in Run 8 I wouldn't have waited for the clearance to land, it was not my problem, I just wanted to land.

The written clearance is something that I'd like to have in the cockpit. Actually not for every situation (altitude etc.) otherwise you receive too many messages, but for the landing clearance, in that particular situation (Run 8 – with that quantity of events) it would have been great to have it. When the workload is very high I'd love to have something written, a visual cue.

BEHAVIOURAL

How did you realise that your performance was declining?

I didn't notice anything, I didn't have enough mental capacity to pay attention to my markers or my status. I believe the workload was so high I had nothing in my head left to check my behaviour.

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Appendix B. SCENARIO 2 SPECIFIC PERFORMANCE INDICATORS

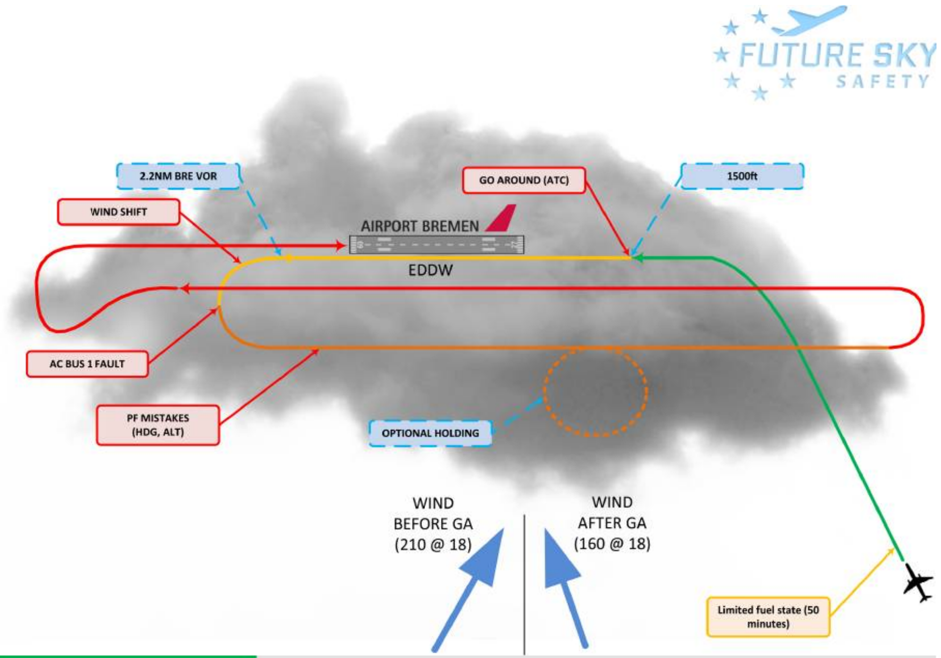
			General	Descent & Approach 27	Go around	AC BUS FAULT	Second LAPA calc (RWY)	Approach
Situational Awareness	Identifies and assesses accurately the state of the aircraft(systems), aircraft position, environment, and of people involved and anticipates on what could happen in the future.	1 Exceeds	Develops 'what if' scenarios and plans for contingencies, including further anticipation of the far future.	Interprets and what-if scenarios, options, updates throughout the descent	Notices fuel status, wind shift and PF errors. Interprets and what-if scenarios, suggests courses of action	Combines fuel status with AC fault,verbalises landing distance, realises it is a OM-B inop problem and its consequences	Actively monitors and includes weather (wind)observations in flight progress, combines with technical status of the aircraft. QRH awareness (extra checklist) Mentions CAT1 landing	Identifies potential scenarios that can evolve as a result of condition (Mentions extra space in front of the runway, and potential nose wheel steering issues), e.g. runway overrun and subsequent evacuation
		2 Meets	Has an awareness of the aircraft state in its environment (including people); projects to near future and anticipates changes.	Mentions fuel/endurance in minutes	Notices fuel status, wind shift and PF errors.	Verbalises consequences for Landing dist.	Accepts additional inputs from ATC without actively seeking new information. QRH awareness, Mentions CAT1 landing	Acknowledges runway state, fuel state, weather. Mentions potential nose wheel steering problem
		3 Below	Spends time searching for irrelevant information, incomplete assessment of the situation.	Mentions fuel/endurance in kilos/tons	Does not notice one of the three events	Does not mention fuel, and/or landing dist	Slow in accepting new information. Misses QRH and/or CAT1 issue.	Requires extra time to complete the picture of the situation
		4 Unacceptable	Does not or incorrectly identify the state (changes), does not seek updates.	No mention of fuel status	Does not notice two of three events, or all	Misses two of the three issues or all	Ignores new information	Misses urgency of the situation

			General	Descent & Approach 27	Go around	AC BUS FAULT	Second LAPA calc (RWY)	Approach
Decision making	Identifies the problem and diagnoses, identifies risks, generates options, decides, resolves problems, monitors, and evaluates.	1 Exceeds	Anticipates future states, effects and risks, is pro-active.			Demonstrates knowledge (Realises it is a OM-B inop issue) of AC BUS fault consequences	Anticipated possible weather deterioration in relation with landing runway	Prepares for potential contingencies. Requests full help of emergency vehicles, prepares cabin, dedicates for landing etc.
		2 Meets	Evaluates (potential) problems, identifies risk, considers alternatives and selects the best course of action. Continuously reviews progress and adjust plans.			Understands the consequences once read from ECAM/OM. Realises it is a OM-B inop issue	Combines technical failure with changes weather situation, decided on landing 09	Accepts information and identifies the criticality of the situation: low fuel, relatively short runway, bad weather
		3 Below	Evaluates the problem poorly. Makes decisions based on incomplete information.			Does not combine the consequence with operational status (landing dist, 27 potentially too short with tailwind, 09 longer flying time considering low fuel)	Passive in decision making, leaves the decision to ATC or PF	Misses elements, e.g. briefs go-around i.s.o. dedicating to landing
		4 Unacceptable	Does not identify there is a problem. Does not indicate what must be done. Does not adjust plan where necessary.			Proposes course of action not taking into account both anomalies	Does not decide on runway 09 as best option for landing	Suggests wrong actions, e.g. not using maximum stopping performance, possibly making a go-around

			General	Descent & Approach 27	Go around	AC BUS FAULT	Second LAPA calc (RWY)	Approach
Application of procedures	Identifies and applies accurately procedures in accordance with published operating instructions and applicable regulations, using the appropriate knowledge. Only if a higher degree of safety is achieved, deviation from standard procedures might be necessary.	1 Exceeds	Identifies and applies at the correct moment the procedures according to published operating instructions.	Mentions threats and errors, defines mitigating measures		Demonstrates knowledge of AC BUS fault procedure, anticipates time available vs procedure length	Correctly calculates LAPA at an early stage, identifies consequences early. Uses Mayday call	Mentions threats and errors, defines mitigating measures. All preparations and checklists completed well in time
		2 Meets	Execute the prescribed procedures and operate the systems during normal and abnormal circumstances.	Briefing, approach preparation incl landing dist calc, approach checklist	Performs actions and call-outs (gear, flaps, ATC comms)	call for ECAM actions	Correctly calculates landing distance. After discussion identifies consequences. Uses Mayday call	Briefing, approach preparation, approach checklist. Time management in relation with time pressure
		3 Below	Applies in general the procedures according to published operating instructions. Unnecessarily skips procedure steps.	Misses steps, caught by approach checklist	Slow in actions, gear, flaps, ATC comms	Slow in call for ECAM actions	Calculates LAPA, but initially with some errors that are corrected	Misses steps, caught by checklist or ECAM Status. Takes extra time to complete the procedure
		4 Unacceptable	Follows the wrong procedure. Skips important procedure steps. Or follows the procedure in such a way that the result is influenced negatively.	Misses essential threats (RWY length, weather, fuel)		Forgets to fly the aircraft, misses procedure steps (status page, OM ref)	Incorrect calculation, wrong result	Misses essential threats (runway length, runway state, fuel status, weather) in the briefing

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

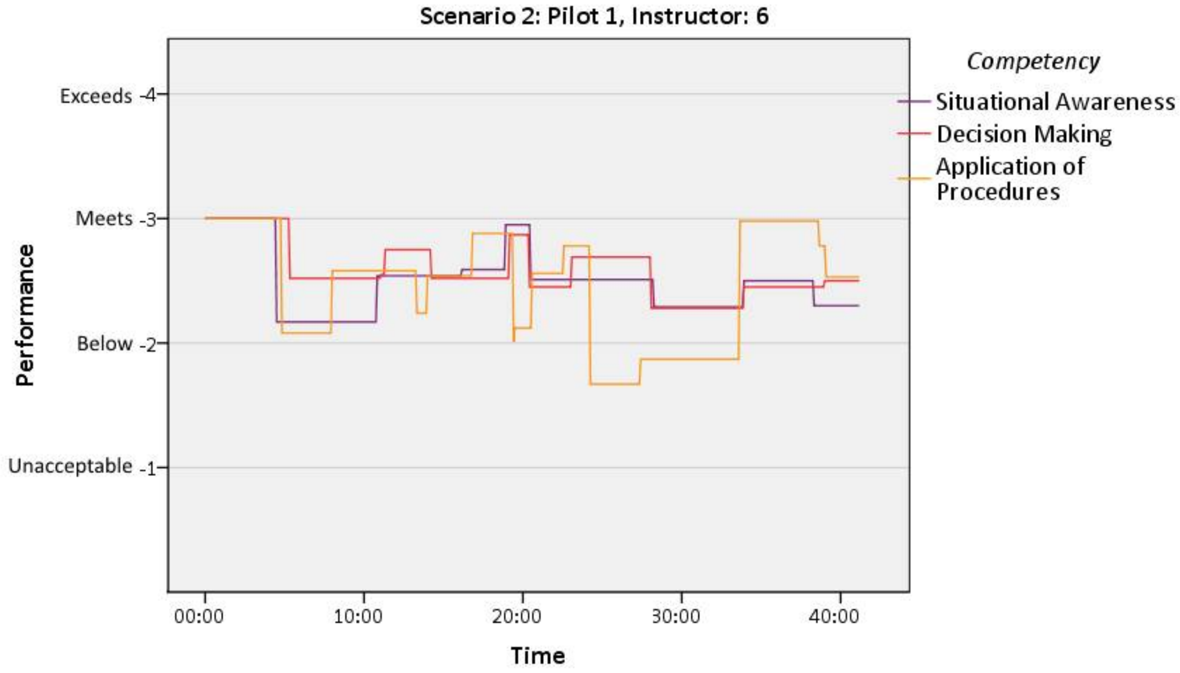
Appendix C. SCENARIO 2 FLIGHT PHASES AND EVENTS

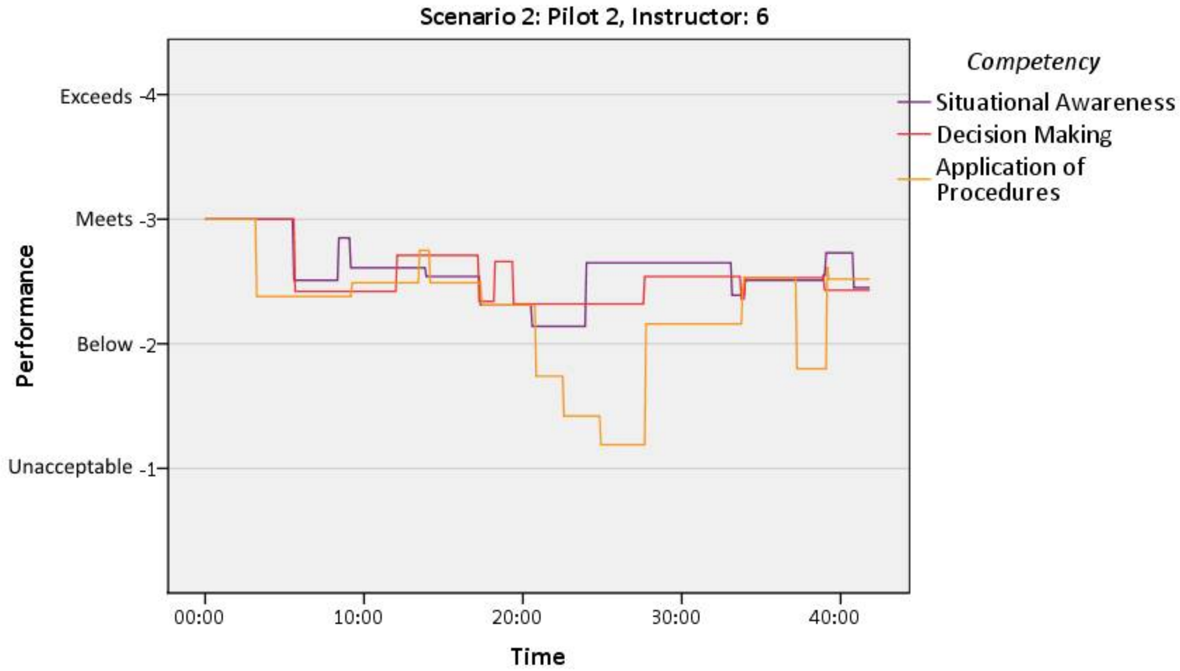


#1 Descent (RWY27)	This phase is a relatively quiet and normal approach into BRE **RWY 27**. The PF does ask the PM (candidate) to perform several tasks, including a **LAPA calculation**. This is a relatively calm phase of the flight. They are **a little tight on fuel**.
#2 Go around	ATC request a **Go-Around**, this phase is a little busier than the previous, but is also not exceptionally complex (a GA is still a relatively common event). However, at the end of the GA the occurrence of the **AC BUS 2 FAULT** makes this a non-standard situation.
#3 AC BUS 1 FAULT	When this failure occurs the **PM's workload** increases significantly with ECAM & procedure actions. At the same time the **PF is making several small flying errors** to observe how the PM maintains situational awareness.
#4 Second LAPA calc (RWY 27)	Before they can land the PM must redo the **LAPA calculations** as the **wx has changed** and they have a failure with possible side effects. It is important that the PF assume a passive role so that we can observe the PM's diligence and decision making in this process. The LAPA should conclude that **RWY27 is unsafe**. RWY09 is best option, but requires **another LAPA**. The PF will ensure that at the end the crew will go for **RWY 09**.
#5 Third LAPA calc (RWY 09)	The **Third LAPA** calculation for RWY09 will only resolve when "**emergency landing**" is selected (this uses rwy overrun lengths). In addition, the **PM must refer to the OM-B** to discover that **automatic rollout is not possible, CAT2 limit, PF w/o window heating**. Again the PF should assume a passive role. The PF will ensure that the aircraft remains within a survivable range of the airport.
#6 Second approach (RWY 09)	The last phase is the **approach to RWY09**. The PF is initially passive, observing whether the PM initiates briefings and covers all important approach considerations. If this does not work, the PF will ensure that **the crew does a control handover** (due to window icing), and are **committed to land** with emergency vehicles standing by.

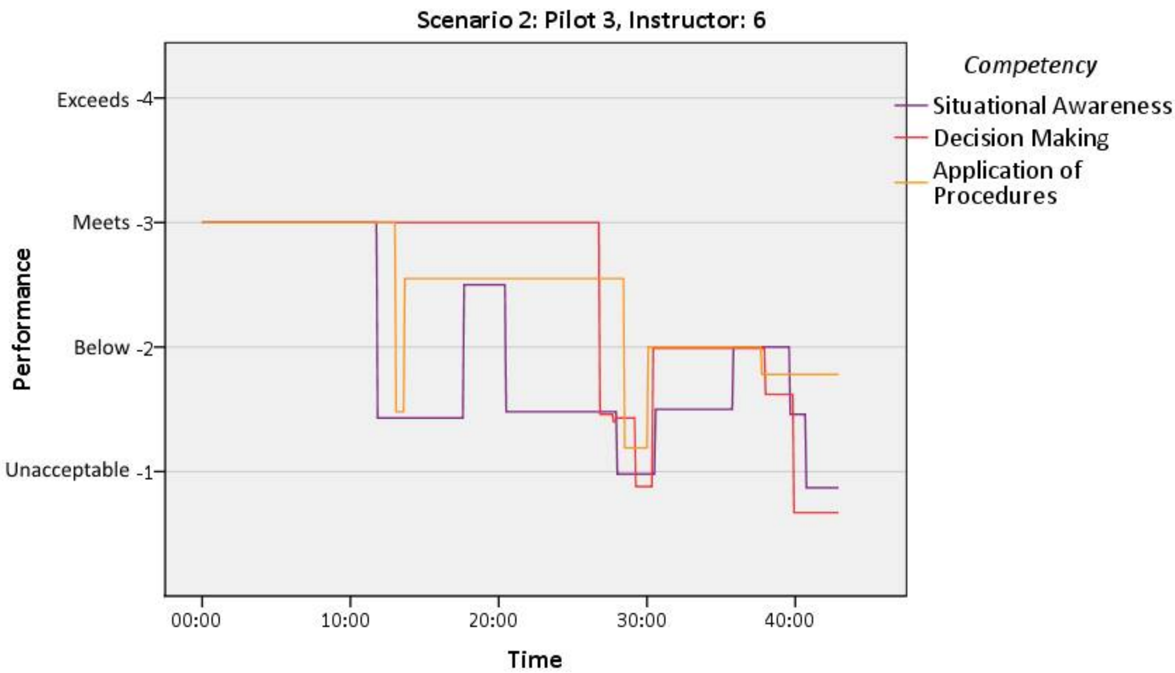
Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

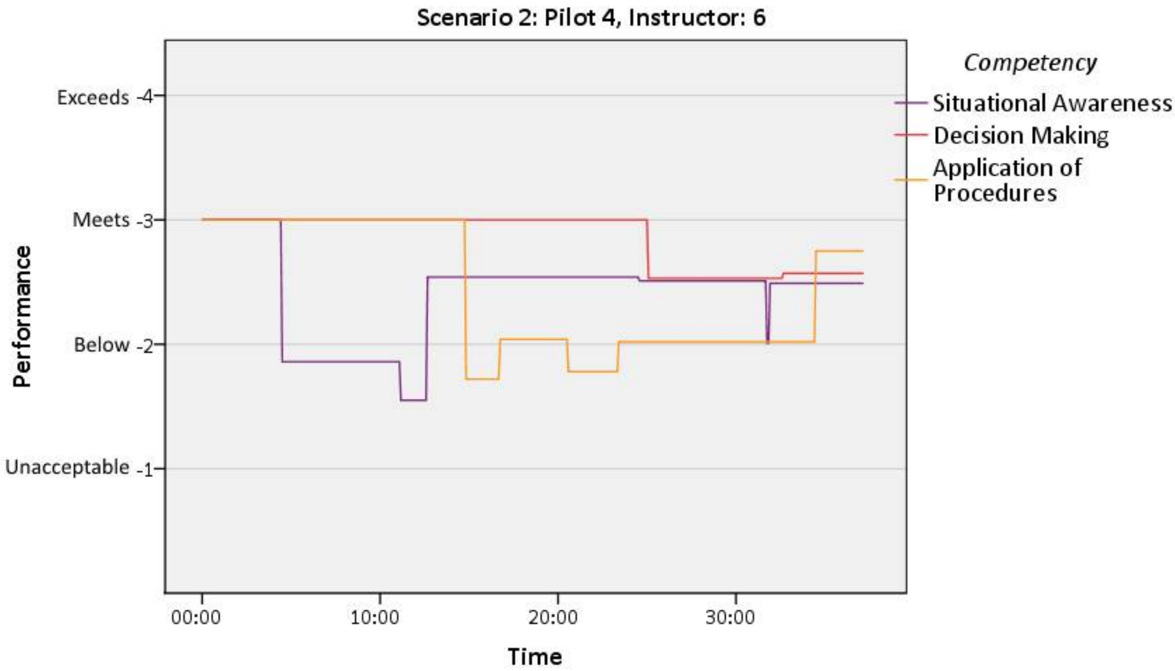
Appendix D. PILOT MONITORING PERFORMANCE



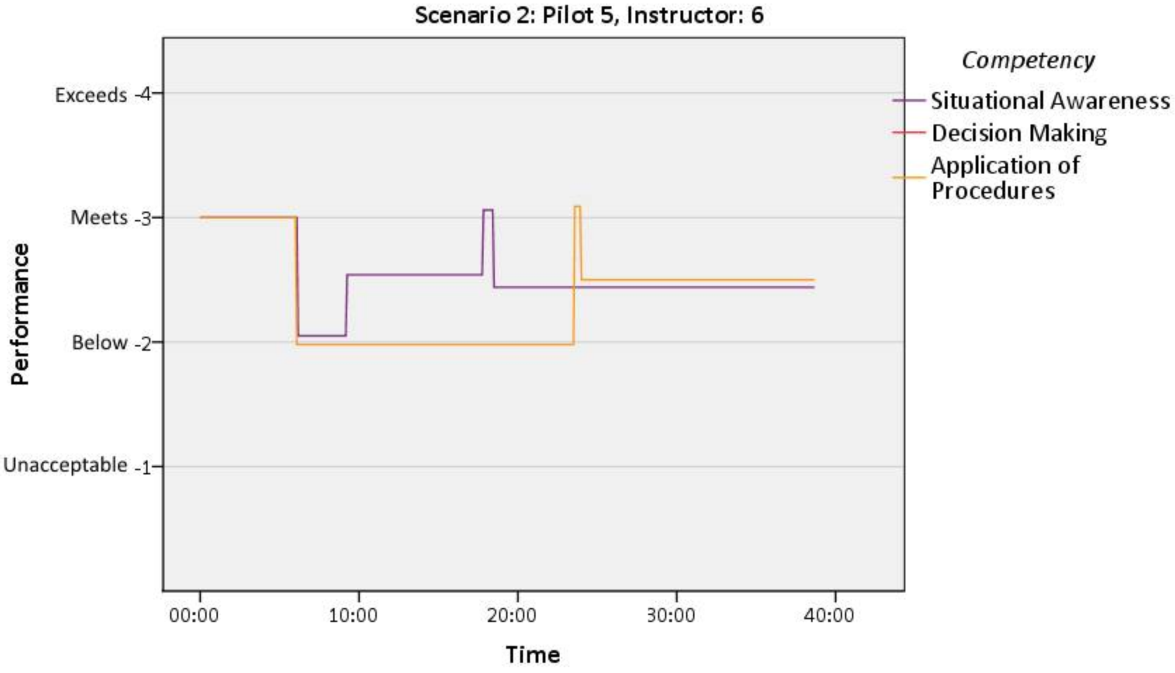


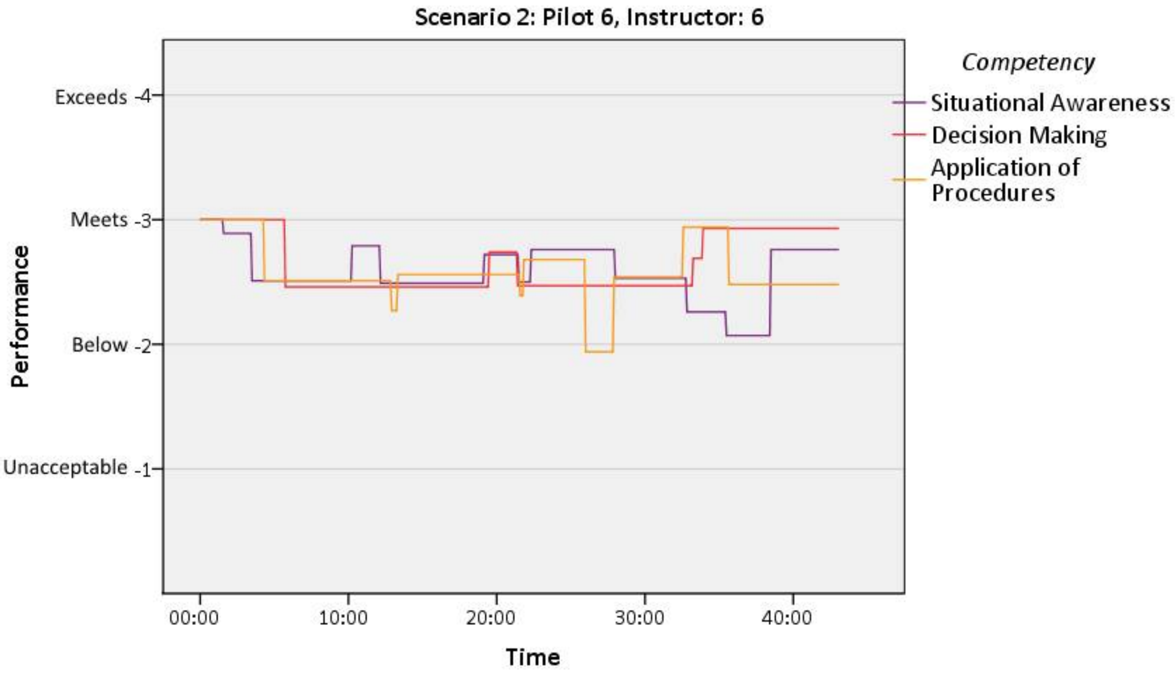
Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

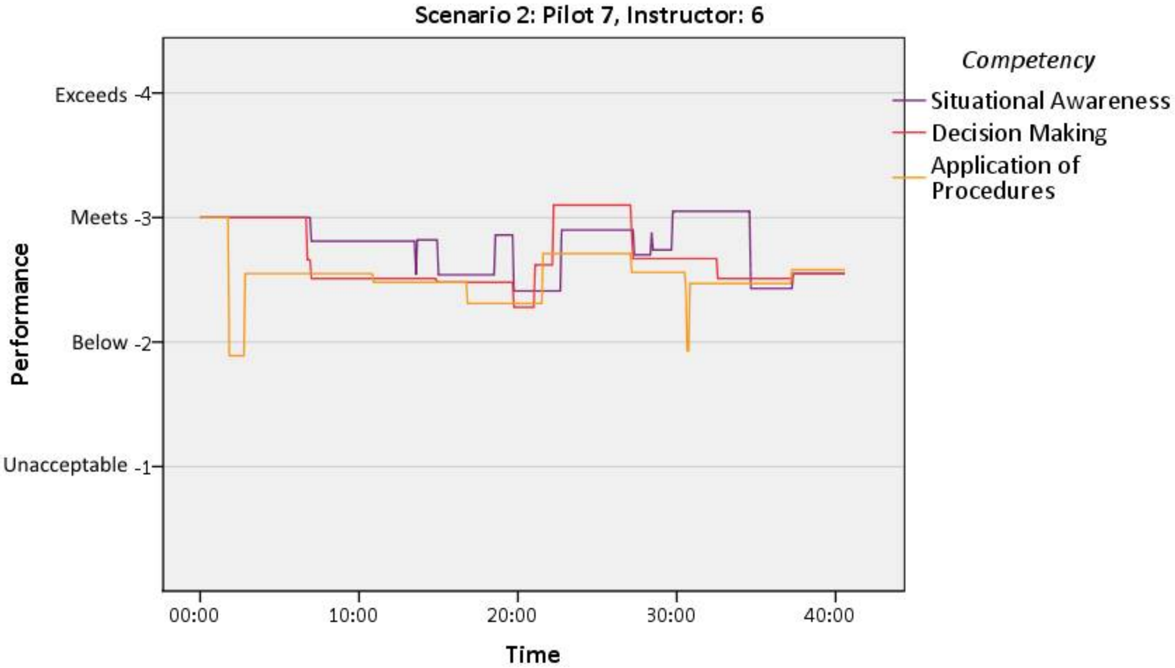


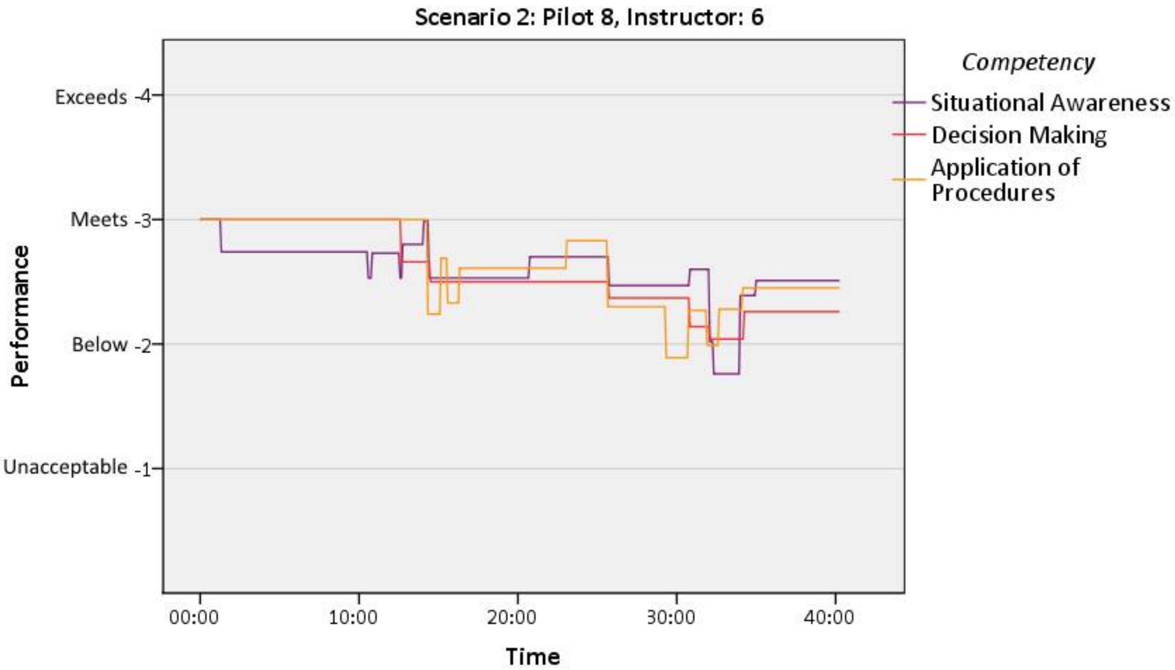


Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

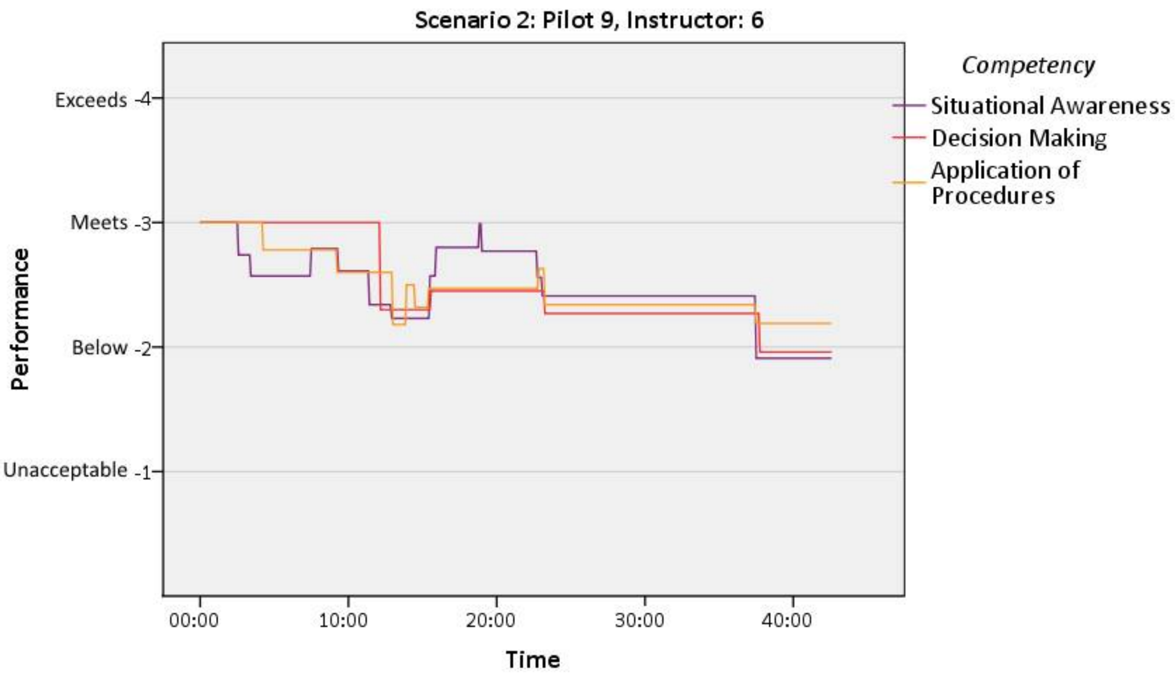

Scenario 2: Pilot 5, Instructor: 6
Performance
Exceeds -4
Meets -3
Below -2
Unacceptable -1
00:00
10:00
20:00
30:00
40:00
Time
Competency
Situational Awareness
Decision Making
Application of Procedures

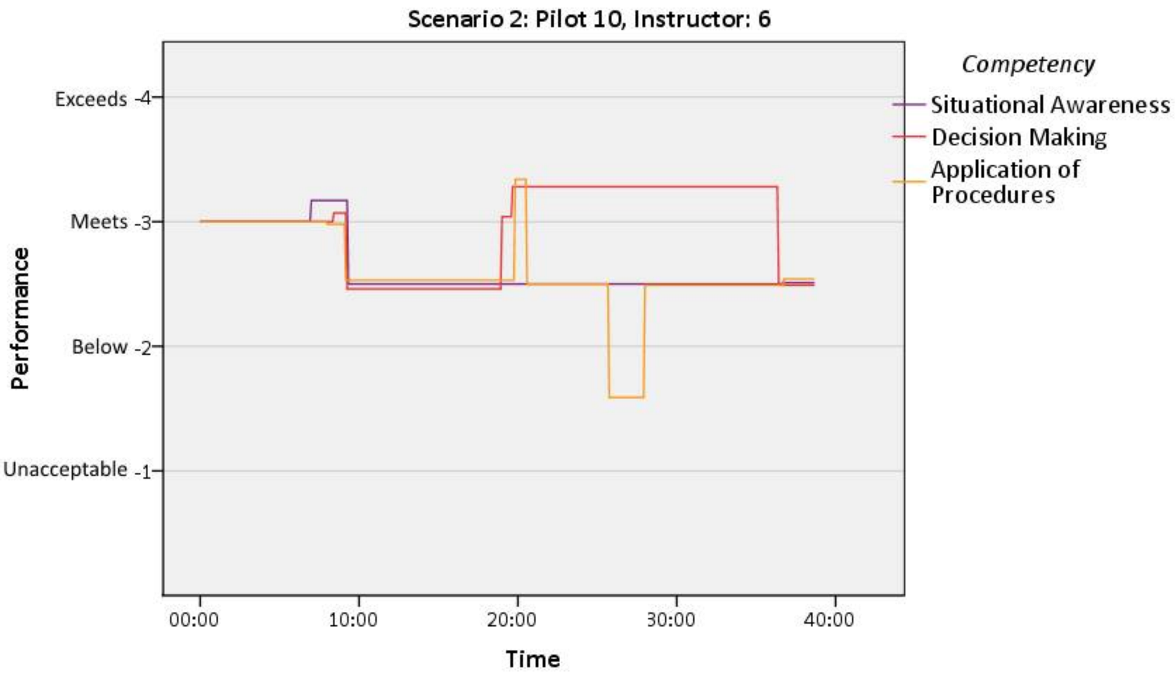

Scenario 2: Pilot 6, Instructor: 6
Performance
Exceeds -4
Meets -3
Below -2
Unacceptable -1
00:00
10:00
20:00
30:00
40:00
Time
Competency
Situational Awareness
Decision Making
Application of Procedures





Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.





Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Appendix E. GUIDELINES AND QUESTIONNAIRES OF THE COGNITIVE WALKTHROUGH

Pilot X	**PHASE 1 DESCENT - First LAPA Calculation**	
Description	This phase is a relatively quiet and normal approach into BRE RWY 27. The PF does ask the PM (candidate) to perform several tasks1. First LAPA calculation. It is important that we use this phase and phase two to set a baseline performance. AC Registration: D-AIUL, GW 64,0 tWeather BRE: Info J: RWY 27, Wind 210/ 18, 2000m , B/400 O/2000 Temp. 1/0 QNH 1013, RWY wet (2mm), tempo vis 600m, light snow	
Trigger event	"Descent"	
BASELINE	OBSERVED	QUESTIONS FOR COGNITIVE WALKTHROUGH
Cues ?? Announced		- Do you have any preoccupations about this landing at the beginning of the sim? - Is it starting rough already ? If yes, why are you looking for first ? - Are you mostly focusing on time and fuel management ?
What has to be understood Calculations have to be made = LAPA Take into account specific: • Runway • Weather • Aircraft status • ...		
What are the options Decide for landing on that specific runway And that the RW length is sufficient for this situation		
What has to be done LAPA calculations		

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Pilot X	**PHASE 2 - Go around by ATC**	
Description	This phase is a little busier than the previous, but is also not exceptionally complex, a GA is still a relatively common event. However, at the end of the GA the occurrence of the AC BUS 1 FAULT (phase three) makes this a non-standard situation. **Go-Around due to Friction Measuring on RWY in progress**	
Trigger event	"Go around"	
BASELINE	OBSERVED	QUESTIONS FOR COGNITIVE WALKTHROUGH
Cues ?? Announced		- What did the go around mean for you? What di dit change in your perception of the situation ? - You hear « Go Around » and then, what are you doing ? - Does it still a standard procedure at this point ? - Did you need to check special informations at this point ?
What has to be understood - Cannot land as is - Prior preparation are not good anymore		
What are the options Standard missed approach procedure has to be performed		
What has to be done MAP to perform		

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Pilot X	PHASE 3 - BUS 1 FAULT	
Description	When this failure occurs the PM's workload increases significantly. At the same time the PF is making several small flying errors to observe how the PM maintains situational awareness. ECAM: PM does ECAM Actions and reads Status OM-B Chapter 3: PM reads AC Bus 1 Fault: Other INOP System (no relevant other items mentioned)	
Trigger event	BUS 1 FAULT	
BASELINE	OBSERVED	QUESTIONS FOR COGNITIVE WALKTHROUGH
Cues ?? ALARM: Explicit by system **What has to be understood** - This becomes a "Non standard situation" - A lot of preexisting data are not good anymore. - Specifics of failure are not always available on ECAM and need to be looked into the EFB (OM-B) **What are the options** - Must perform ECAM procedure to figure out the abnormal and enter them in the EFB. (Abnormal, failure and weather) - Prioritization have to be done **What has to be done** - ECAM procedure: PM does ECAM Actions and reads Status - Has to look into OM-B Chapter 3: PM reads AC Bus 1 Fault: Other INOP System		- What is your fuel time when the electrical failure hit ? - What is your first goal after the electrical failure ? - How do you check that you are in control of the plane ? - Did you look into the ECAM in order to have an idea of how bad was the failure ? - What was the clues to determine the severity of the failure ? - How the failure will influence the landing ? - Do you think that you know how long it will take to take care of that failure - Are you thinking that you need to deal with that failure right away ? - After you managed to partially resolve the failure, did you have an estimate of what was left about dealing with the failure ? - Are you thinking of how it will change your landing and the procedures you will have to follow ? - If yes, how are you using this information, and what for ? - Did you think about the extension of the runway 09 at that moment ? - Did you think of other factors that could make the situation worst ?

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Pilot X	PHASE 4- Second LAPA Calculation (RW27)	
Description	Before they can land the PM must redo the LAPA calculations as the WX has changed and they have a failure with possible side effects. It is important that the PF assume a passive role so that we can observe the PM's diligence and decision making in this process. 2. LAPA calculation: PM should conclude that RWY27 is unsafe. RWY09 is best option, but requires another LAPA. The PF will ensure that at the end the crew will go for RWU 09.AC Registration: D-AIUL, GW 64,0 tWeather BRE: Info K: RWY 09, Wind 160/ 18, RVR 500, 600, 550m, B/200 O/400, Temp 1/0, QNH 1013, RWY covered with 4mm slush, light snow	
Trigger event	Weather Change (inconspicuous) - While doing new LAPA calculation - Cannot land on RW 27	
BASELINE	OBSERVED	QUESTIONS FOR COGNITIVE WALKTHROUGH
Cues ?? - Could get it on printer but because of failure it is not working -> Must ask ATC - on EFB, the LAPA HMI provides info (highlighted when there is a problem)		- How did you get the informations needed for the LAPA ? (ATC, kind of failure, full landing, informations displayed on the lower ICAM) - For which runway are you doing the LAPA ? - Why was the runway 27 not finally possible? Why can't you land there ? (tailwind) - Does the tailwind information is enough to make you realize landing on runway 27 is not possible ? - What are you thinking at that time ? Did you make a calculation or was it a decision made thanks to your experience ?
What has to be understood Results of LAPA calculations depend on weather change or not		
What are the options - Weather change is witnessed before and LAPA calculations are correct. (-> skip this phase) - LAPA done without asking for weather before (wrong calculations) - LAPA has to be done again		
What has to be done - LAPA calculation with new weather DATA - Change RW from 27 to 9		

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Pilot X	PHASE 5- Third LAPA Calculation (RW09)	
Description	Weather like above3. LAPA calculation analyzes RWY09 but will only be possible with given Abnormal and Weather and indicating "emergency only" in EFB. In addition, the PM must refer to the OM-B to discover that only a manual rollout is possible (no automatic roll out). Again the PF should assume a passive role. The PF will ensure that the aircraft remains within acceptable range of the airport.	
Trigger event	"Could not land on RW 27..." well - "Cannot land on runway 9 as is"	
BASELINE	OBSERVED	QUESTIONS FOR COGNITIVE WALKTHROUGH
Cues ?? - on EFB, the LAPA HMI provides info (highlighted when there is a problem) - have to do an "automatic landing", but... - ...in the OM-B, "warning no auto roll out possible"		- Why did you do a third LAPA calculations ? (too much wind crosswind, no automatic rollout possible) - How does that change your vision of what you have to do ? - Is this a situation very peculiar ?
What has to be understood - LAPA calculation analyzes RWY09 but will only be possible with : • Given Abnormal • Weather change • "Emergency only" in EFB - in the OM-B, PM must realize that only a manual rollout is possible (no automatic roll out, so the pilot will have to manually leave the runway after landing).		
What are the options - Declare "emergency" - Land elsewhere (not played in this scenario)		- Do you have a lot of practice in this kind of procedure ?
What has to be done - Declare "Emergency" in EFB Redo the LAPA with: • Given Abnormal • Weather change • "Emergency only" in EFB - The PM must refer to the OM-B to discover that only a manual rollout is possible and prepare for that.		- Does it still manageable ?

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Pilot X	**PHASE 6 - PM checks QRH**	
Description	QRH: PM must check Quick Reference Handbook (OM-B -> Airborne Equip. req. for CATII/III). On page 2, Windshield Heat (L or R Windshield) is mentioned. No landing by PF allowed.(Remark: PF is maybe not capable of performing the landing due to frozen windshield.) PM must prepare to do the landing.	
Trigger event	If the pilot is highly trained he will look for extra information in the QRH: - "QRH update on limitation due to window deicing feature"	
BASELINE	OBSERVED	QUESTIONS FOR COGNITIVE WALKTHROUGH
Cues ?? - In QRH, difficult to find and not easily sorted out in EFB documentation		- Why did you think about reading the QRH ? - What were you looking for in it ? - Did you think about the time it would take you ? - What did you find in the QRH ? Was it a vitally important information ?
What has to be understood - PF is not capable of performing the landing due to frozen windshield and the lack of deicing option. - Thus PM will do the landing.		
What are the options		
What has to be done - PM must prepare to do the landing. - The crew is committed to land!		

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Pilot X	PHASE 7 - Second Approach (First Approach onto RW09)	
Description	The last phase is the approach to RWY09. The PF is initially passive, observing whether the PM initiates briefings and covers all important specialties. PF performs the CATII/ III briefing.If no QRH check by PM: PF indicates that his window is frozen and his sight is limited. Control change over in 400 ft due to low visibility of PF. Actual visibility is 500m, clouds B/200ft.	
Trigger event	Second approach - "Window is frozen on PF side"	
BASELINE	OBSERVED	QUESTIONS FOR COGNITIVE WALKTHROUGH
Cues ?? - "Ice on window" - Predictable through the exploration of OM-B: "limitation of deicing features"		- Did you hear about the icing on the window ? - Why does the PF can't do the landing ? (visual clues are necessary to do manual landing) - Did you look in the EFB ? Did you read « window hit number 1 » ? - Did you finished all the required procedures ?
What has to be understood PF is not capable of performing the landing due to frozen windshield and lack of deicing option.		
What are the options PM has to do the landing		
What has to be done 1- PM has to initiate briefing for PF to do the landing and then change when PF cannot do it due to the frozen window 2- Control change 3- PM has to do the landing		

GENERAL QUESTIONS
- What would have been the best other options to manage the landing of this flight ?
- Is there something you would have like to have to be best prepared to deal with the landing ?
- Do you think that the LAPA is time consuming in this kind of situations ? (preset of the LAPA)
- Was the mission demanding globaly?
- Did the mission make sense for you ? Was it an actual possible situation ?

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Appendix F. EVENTS FOR PILOTS 2 TO 10.

Pilot 2	Video time	Scenario time (VT-1'35")	FOB (kg)	Time remaining (regarding FOB)	stickers for representation
Start of the recording	00:00	NA	NA	NA	NA
Actual start	01:35	00:00	2200	00:55:00	(55'00''/2200kg) Actual start
Start LAPA 27	03:32	01:57	2122	53:03	(53'03''/2122kg) Start LAPA 27
Finish LAPA 27	05:00	03:25	2063	51:35	(51'35''/2063kg) Finish LAPA 27
Clear ILS for RWY27	06:33	04:58	2001	50:02	(50'02''/2001kg) Clear ILS for RWY27
PM mentions the fuel	09:27	07:52	1885	47:08	(47'08''/1885kg) PM mentions the fuel
Go Around (ATC mentioned the friction car 2 times)	11:08	09:33	1818	45:27	(45'27''/1818kg) Go Around (ATC mentioned the friction car 2 times)
PM notices the no climb right away	13:36	12:01	1719	42:59	(42'59''/1719kg) PM notices the no climb right away
AC Bus Failure	13:58	12:23	1704	42:37	(42'37''/1704kg) AC Bus Failure
PF begins ECAM procedures	14:29	12:54	1684	42:06	(42'06''/1684kg) PF begins ECAM procedures
PM suggests to reduce speed (good)	15:39	14:04	1637	40:56	(40'56''/1637kg) PM suggests to reduce speed (good)
PM continues ECAM procedures	15:52	14:17	1628	40:43	(40'43''/1628kg) PM continues ECAM procedures
Status is clear	18:31	16:56	1522	38:04	(38'04''/1522kg) Status is clear
Begins the overall LAPA	18:55	17:20	1506	37:40	(37'40''/1506kg) Begins the overall LAPA
Consider new airport	19:11	17:36	1496	37:24	(37'24''/1496kg) Consider new airport
PM ask captain to ask ATC about Hamburg & Hannover weather	19:29	17:54	1484	37:06	(37'06''/1484kg) PM ask captain to ask ATC about Hamburg & Hannover weather
Airports denied, ATC offers to give vectors back to Bremen	20:10	18:35	1456	36:25	(36'25''/1456kg) Airports denied, ATC offers to give vectors back to Bremen
ATC offers to give new weather. PM writes it down	20:53	19:18	1428	35:42	(35'42''/1428kg) ATC offers to give new weather. PM writes it down
Back to Bremen	21:33	19:58	1401	35:02	(35'02''/1401kg) Back to Bremen
Continues LAPA without new wind (PM has it but he doesn't put it in)	21:34	19:59	1400	35:01	(35'01''/1400kg) Continues LAPA without new wind (PM has it but he doesn't put it in)
Doesn't remember the correct failure	21:51	20:16	1389	34:44	(34'44''/1389kg) Doesn't remember the correct failure
Emergency declared	25:02	23:27	1262	31:33	(31'33''/1262kg) Emergency declared
PM asks for the new weather again, and put it in LAPA this time. He notices RWY09 is better right away	26:16	24:41	1212	30:19	(30'19''/1212kg) PM asks for the new weather again, and put it in LAPA this time. He notices RWY09 is better right away
They decide to go on RWY09, PM requests it	26:56	25:21	1186	29:39	(29'39''/1186kg) They decide to go on RWY09, PM requests it
PM understands they can make an automatic approach but not an automatic rollout	28:17	26:42	1132	28:18	(28'18''/1132kg) PM understands they can make an automatic approach but not an automatic rollout
Enters RWY09 in FMS	29:21	27:46	1089	27:14	(27'14''/1089kg) Enters RWY09 in FMS
End of LAPA, PM knows he can land with emergency	29:41	28:06	1076	26:54	(26'54''/1076kg) End of LAPA, PM knows he can land with emergency
PF asks to go check INOP systems	31:51	30:16	989	24:44	(24'44''/989kg) PF asks to go check INOP systems
PM mentions the fuel (less than 25 minutes left)	32:53	31:18	948	23:42	(23'42''/948kg) PM mentions the fuel (less than 25 minutes left)
PF asks to enter runway in FMS	34:47	33:12	872	21:48	(21'48''/872kg) PF asks to enter runway in FMS
Clear to land, ILS 09	35:48	34:13	831	20:47	(20'47''/831kg) Clear to land, ILS 09
PF mentions the ice on window	36:27	34:52	805	20:08	(20'08''/805kg) PF mentions the ice on window
They discuss about the ice, PM seems to understand he will have to do the landing	37:18	35:43	771	19:17	(19'17''/771kg) They discuss about the ice, PM seems to understand he will have to do the landing
Briefing by the PF	37:54	36:19	747	18:41	(18'41''/747kg) Briefing by the PF
Manual landing doesn't seem clear for PM	38:10	36:35	736	18:25	(18'25''/736kg) Manual landing doesn't seem clear for PM
PM checks in OMB and find the 80" switch	39:13	37:38	694	17:22	(17'22''/694kg) PM checks in OMB and find the 80" switch
PM knows they can gain some runway by diving under the glide slope	41:35	40:00	600	15:00	(15'00''/600kg) PM knows they can gain some runway by diving under the glide slope
PF switch	41:45	40:10	593	14:50	(14'50''/593kg) PF switch
AP/AT disengaged	42:12	40:37	575	14:23	(14'23''/575kg) AP/AT disengaged
Touchdown	42:25	40:50	566	14:10	(14'10''/566kg) Touchdown

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Pilot 3	Video time	Scenario time (VT-4'14")	FOB (kg)	Time remaining (regarding FOB)	stickers for representation
Start of the recording	00:00	NA	NA	NA	NA
Actual start	04:14	00:00	2050	51:15	(51'15''/2050kg) Actual start
Start LAPA 27	06:50	02:36	1946	48:39	(48'39''/1946kg) Start LAPA 27
Finish LAPA 27	07:45	03:31	1909	47:44	(47'44''/1909kg) Finish LAPA 27
Briefing for RWY27	08:23	04:09	1884	47:06	(47'06''/1884kg) Briefing for RWY27
Go Around	13:08	08:54	1694	42:21	(42'21''/1694kg) Go Around
Should declare emergency	13:08	08:54	1694	42:21	(42'21''/1694kg) Should declare emergency
Weather requested by PM but not transmitted	16:31	12:17	1558	38:58	(38'58''/1558kg) Weather requested by PM but not transmitted
AC Bus Failure	16:33	12:19	1557	38:56	(38'56''/1557kg) AC Bus Failure
PF begins ECAM procedures	18:03	13:49	1497	37:26	(37'26''/1497kg) PF begins ECAM procedures
PM continues ECAM procedures	18:50	14:36	1466	36:39	(36'39''/1466kg) PM continues ECAM procedures
ATC request climb	19:08	14:54	1454	36:21	(36'21''/1454kg) ATC request climb
PM notices the no climb	19:34	15:20	1436	35:55	(35'55''/1436kg) PM notices the no climb
Ends ECAM procedures	21:40	17:26	1352	33:49	(33'49''/1352kg) Ends ECAM procedures
PM suggests to start APU	21:56	17:42	1342	33:33	(33'33''/1342kg) PM suggests to start APU
Starting APU	22:19	18:05	1326	33:10	(33'10''/1326kg) Starting APU
Weather asked, recalled by PF (not transmitted before)	22:40	18:26	1312	32:49	(32'49''/1312kg) Weather asked, recalled by PF (not transmitted before)
MINIMUM FUEL	25:29	21:15	1200	30:00	(30'00''/1200kg) MINIMUM FUEL
Weather transmitted	25:30	21:16	1199	29:59	(29'59''/1199kg) Weather transmitted
RWY27 not possible	25:58	21:44	1180	29:31	(29'31''/1180kg) RWY27 not possible
PM requests vectors back to Bremen	26:37	22:23	1154	28:52	(28'52''/1154kg) PM requests vectors back to Bremen
Beginning of LAPA 09	26:54	22:40	1143	28:35	(28'35''/1143kg) Beginning of LAPA 09
Consider new airport	30:38	26:24	994	24:51	(24'51''/994kg) Consider new airport
Ask for weather in Hannover and Hamburg	31:39	27:25	953	23:50	(23'50''/953kg) Ask for weather in Hannover and Hamburg
Consider going to Frankfurt	32:51	28:37	905	22:38	(22'38''/905kg) Consider going to Frankfurt
Aware about low fuel	33:05	28:51	896	22:24	(22'24''/896kg) Aware about low fuel
Declare emergency	33:48	29:34	867	21:41	(21'41''/867kg) Declare emergency
Message to cabin	35:25	31:11	802	20:04	(20'04''/802kg) Message to cabin
Finish final LAPA for RWY09 (final parameters)	37:19	33:05	726	18:10	(18'10''/726kg) Finish final LAPA for RWY09 (final parameters)
PM talks about the 300m extra meters and the possibility to gain some distance	37:27	33:13	721	18:02	(18'02''/721kg) PM talks about the 300m extra meters and the possibility to gain some distance
PM goes to OM-B	40:03	35:49	617	15:26	(15'26''/617kg) PM goes to OM-B
Clear ILS09	40:46	36:32	588	14:43	(14'43''/588kg) Clear ILS09
Nothing found in OM-B	44:55	40:41	422	10:34	(10'34''/422kg) Nothing found in OM-B
Complete briefing by PF	44:56	40:42	422	10:33	(10'33''/422kg) Complete briefing by PF
Ice on window	45:12	40:58	411	10:17	(10'17''/411kg) Ice on window
PF switch	45:37	41:23	394	09:52	(09'52''/394kg) PF switch
Touchdown	46:17	42:03	368	09:12	(09'12''/368kg) Touchdown

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Pilot 4	Video time	Scenario time (VT-0'49'')	FOB (kg)	Time remaining (regarding FOB)	stickers for representation
Start of the recording	00:00	NA	NA	NA	NA
Actual start	00:49	00:00	2000	00:50:00	(50'00''/2000kg) Actual start
Start LAPA 27	03:29	02:40	1893	47:20	(47'20''/1893kg) Start LAPA 27
Finish LAPA 27	06:03	05:14	1790	44:46	(44'46''/1790kg) Finish LAPA 27
Go Around (ATC mentioned the friction car & asked to reduce speed)	10:29	09:40	1613	40:20	(40'20''/1613kg) Go Around (ATC mentioned the friction car & asked to reduce speed)
PM mentions the fuel (1.5T)	13:01	12:12	1512	37:48	(37'48''/1512kg) PM mentions the fuel (1.5T)
AC Bus Failure	13:44	12:55	1483	37:05	(37'05''/1483kg) AC Bus Failure
Ask for vectors to stay in Bremen area (good)	14:52	14:03	1438	35:57	(35'57''/1438kg) Ask for vectors to stay in Bremen area (good)
PF begins ECAM procedures	15:27	14:38	1414	35:22	(35'22''/1414kg) PF begins ECAM procedures
PM continues ECAM procedures	17:09	16:20	1346	33:40	(33'40''/1346kg) PM continues ECAM procedures
PM doesn't notice the no climb	17:22	16:33	1338	33:27	(33'27''/1338kg) PM doesn't notice the no climb
Discussions about ECAM status finished	19:30	18:41	1252	31:19	(31'19''/1252kg) Discussions about ECAM status finished
PM considers new airport	19:36	18:47	1248	31:13	(31'13''/1248kg) PM considers new airport
PF suggests to stay in Bremen and to declare emergency	20:37	19:48	1208	30:12	(30'12''/1208kg) PF suggests to stay in Bremen and to declare emergency
Back to Bremen	21:07	20:18	1188	29:42	(29'42''/1188kg) Back to Bremen
PM starts fast calculations for RWY27	21:21	20:32	1178	29:28	(29'28''/1178kg) PM starts fast calculations for RWY27
Fast LAPA finished, PM knows he can stay in Bremen	22:15	21:26	1142	28:34	(28'34''/1142kg) Fast LAPA finished, PM knows he can stay in Bremen
ATC offers to give new weather	22:36	21:47	1128	28:13	(28'13''/1128kg) ATC offers to give new weather
New LAPA with new wind	23:11	22:22	1105	27:38	(27'38''/1105kg) New LAPA with new wind
PF mentions the RWY09 (unusual)	23:26	22:37	1095	27:23	(27'23''/1095kg) PF mentions the RWY09 (unusual)
PF suggests to ask for Hannover's & Hamburg's wind (unusual)	23:40	22:51	1086	27:09	(27'09''/1086kg) PF suggests to ask for Hannover's & Hamburg's wind (unusual)
Correct LAPA for RWY09	25:25	24:36	1016	25:24	(25'24''/1016kg) Correct LAPA for RWY09
PM goes into OMB to check the crosswind limits	27:18	26:29	940	23:31	(23'31''/940kg) PM goes into OMB to check the crosswind limits
Correct crosswind limit found, auto landing with manual roll-out	28:03	27:14	910	22:46	(22'46''/910kg) Correct crosswind limit found, auto landing with manual roll-out
First clue for the ice on the window	29:25	28:36	856	21:24	(21'24''/856kg) First clue for the ice on the window
PM realizes on his own that he might have to make the landing	29:32	28:43	851	21:17	(21'17''/851kg) PM realizes on his own that he might have to make the landing
Briefing by the PF	29:56	29:07	835	20:53	(20'53''/835kg) Briefing by the PF
Talking about the chances of overrun and diving below the glideslope	30:49	30:00	800	20:00	(20'00''/800kg) Talking about the chances of overrun and diving below the glideslope
Clear to land, ILS 09	32:12	31:23	744	18:37	(18'37''/744kg) Clear to land, ILS 09
PM knows they are committed to land	32:44	31:55	723	18:05	(18'05''/723kg) PM knows they are committed to land
PM is really aware he will have to do the landing due to ice	32:47	31:58	721	18:02	(18'02''/721kg) PM is really aware he will have to do the landing due to ice
PF mentions fuel (600kg)	33:57	33:08	674	16:52	(16'52''/674kg) PF mentions fuel (600kg)
PF switch	36:37	35:48	568	14:12	(14'12''/568kg) PF switch
AP/AT disengaged	37:01	36:12	552	13:48	(13'48''/552kg) AP/AT disengaged
Touchdown	37:11	36:22	545	13:38	(13'38''/545kg) Touchdown

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Pilot 5 (37'10")	Video time	Scenario time (VT-2'25")	FOB (kg)	Time remaining (regarding FOB)	stickers for representation
Start of the recording	00:00	NA	NA	NA	NA
Actual start	02:25	00:00	2000	00:50:00	(50'00"/2000kg) Actual start
Start LAPA 27	05:19	02:54	1884	47:06	(47'06"/1884kg) Start LAPA 27
Finish LAPA 27	06:08	03:43	1851	46:17	(46'17"/1851kg) Finish LAPA 27
Go Around	11:13	08:48	1648	41:12	(41'12"/1648kg) Go Around
PM mentions the fuel (1.6T)	14:25	12:00	1520	38:00	(38'00"/1520kg) PM mentions the fuel (1.6T)
AC Bus Failure	14:40	12:15	1510	37:45	(37'45"/1510kg) AC Bus Failure
PF begins ECAM procedures	15:33	13:08	1474	36:52	(36'52"/1474kg) PF begins ECAM procedures
PM continues ECAM procedures	17:08	14:43	1411	35:17	(35'17"/1411kg) PM continues ECAM procedures
PM notices the no climb right away	17:33	15:08	1394	34:52	(34'52"/1394kg) PM notices the no climb right away
Status cleared	19:37	17:12	1312	32:48	(32'48"/1312kg) Status cleared
PM wants to come back to Bre, ask for heading to 27	19:59	17:34	1297	32:26	(32'26"/1297kg) PM wants to come back to Bre, ask for heading to 27
Second LAPA 27	20:39	18:14	1270	31:46	(31'46"/1270kg) Second LAPA 27
New weather gave by ATC	21:58	19:33	1218	30:27	(30'27"/1218kg) New weather gave by ATC
ATC asks if they want 27 or 09	22:15	19:50	1206	30:10	(30'10"/1206kg) ATC asks if they want 27 or 09
ATC repeats the wind a second time	23:16	20:51	1166	29:09	(29'09"/1166kg) ATC repeats the wind a second time
09 is better than 27 (PM)	23:26	21:01	1159	28:59	(28'59"/1159kg) 09 is better than 27 (PM)
Request runway 09 (PM)	23:50	21:25	1143	28:35	(28'35"/1143kg) Request runway 09 (PM)
Low fuel (PF)	24:31	22:06	1116	27:54	(27'54"/1116kg) Low fuel (PF)
Emergency (PF)	25:42	23:17	1068	26:43	(26'43"/1068kg) Emergency (PF)
Diversion not possible (PF)	26:45	24:20	1026	25:40	(25'40"/1026kg) Diversion not possible (PF)
Discussing about the non standard landing	30:30	28:05	876	21:55	(21'55"/876kg) Discussing about the non standard landing
Final LAPA for RWY09	31:52	29:27	822	20:33	(20'33"/822kg) Final LAPA for RWY09
PF mentions fuel (900kg)	32:00	29:35	816	20:25	(20'25"/816kg) PF mentions fuel (900kg)
Discussions about final LAPA	33:31	31:06	756	18:54	(18'54"/756kg) Discussions about final LAPA
PF talks about the landing and the limit for AP switch off	35:48	33:23	664	16:37	(16'37"/664kg) PF talks about the landing and the limit for AP switch off
PM doesn't know about the limitation, and he doesn't go check in OM-B	36:56	34:31	619	15:29	(15'29"/619kg) PM doesn't know about the limitation, and he doesn't go check in OM-B
Committed to land (PM)	37:22	34:57	602	15:03	(15'03"/602kg) Committed to land (PM)
First cue about the ice on window	37:39	35:14	590	14:46	(14'46"/590kg) First cue about the ice on window
PM knows he has to do the landing if his captain doesn't see	37:53	35:28	581	14:32	(14'32"/581kg) PM knows he has to do the landing if his captain doesn't see
PF switch	38:54	36:29	540	13:31	(13'31"/540kg) PF switch
AP/AT disengaged	39:22	36:57	522	13:03	(13'03"/522kg) AP/AT disengaged
Touchdown	39:35	37:10	513	12:50	(12'50"/513kg) Touchdown

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Pilot 6 (42'21'')	Video time	Scenario time (VT-0'40'')	FOB (kg)	Time remaining (regarding FOB)	stickers for representation
Start of the recording	00:00	NA	NA	NA	NA
Actual start	00:40	00:00	2200	00:55:00	(55'00''/2200kg) Actual start
Start LAPA 27	03:28	02:48	2088	52:12	(52'12''/2088kg) Start LAPA 27
Finish LAPA 27	04:38	03:58	2041	51:02	(51'02''/2041kg) Finish LAPA 27
Go Around	10:03	09:23	1824	45:37	(45'37''/1824kg) Go Around
PM asks for new wind (good) Not transmitted	13:06	12:26	1702	42:34	(42'34''/1702kg) PM asks for new wind (good) Not transmitted
AC Bus Failure	13:12	12:32	1698	42:28	(42'28''/1698kg) AC Bus Failure
PF begins ECAM procedures	13:53	13:13	1671	41:47	(41'47''/1671kg) PF begins ECAM procedures
Climb 4000 (PM is still the pilot flying!)	14:42	14:02	1638	40:58	(40'58''/1638kg) Climb 4000 (PM is still the pilot flying!)
PM continues ECAM procedures	16:11	15:31	1579	39:29	(39'29''/1579kg) PM continues ECAM procedures
Captain says to PM that he didn't pull. The roles are reversed!	17:22	16:42	1532	38:18	(38'18''/1532kg) Captain says to PM that he didn't pull. The roles are reversed!
ECAM finished	17:47	17:07	1515	37:53	(37'53''/1515kg) ECAM finished
New weather (PF)	18:37	17:57	1482	37:03	(37'03''/1482kg) New weather (PF)
PM suggest to try a diversion	19:32	18:52	1445	36:08	(36'08''/1445kg) PM suggest to try a diversion
PM mentions the fuel (1400kg)	19:49	19:09	1434	35:51	(35'51''/1434kg) PM mentions the fuel (1400kg)
Ask weather in Hannover & Hamburg (PM)	20:17	19:37	1415	35:23	(35'23''/1415kg) Ask weather in Hannover & Hamburg (PM)
Vectors back to Bremen (PM doesn't seem confident they can land here, but PF insists)	21:53	21:13	1351	33:47	(33'47''/1351kg) Vectors back to Bremen (PM doesn't seem confident they can land here, but PF insists)
Declare emergency (PM)	22:09	21:29	1340	33:31	(33'31''/1340kg) Declare emergency (PM)
PM realizes just with the new wind that 09 is better	22:37	21:57	1322	33:03	(33'03''/1322kg) PM realizes just with the new wind that 09 is better
PM request RWY09	22:51	22:11	1312	32:49	(32'49''/1312kg) PM request RWY09
PF suggests to go check extended procedures (OM-B-chap3)	24:57	24:17	1228	30:43	(30'43''/1228kg) PF suggests to go check extended procedures (OM-B-chap3)
Correct LAPA for RWY09 (PM asks for new weather again, just in case)	29:41	29:01	1039	25:59	(25'59''/1039kg) Correct LAPA for RWY09 (PM asks for new weather again, just in case)
PF gives the first clue for the ice on window. PM switches on the anti ice	30:18	29:38	1014	25:22	(25'22''/1014kg) PF gives the first clue for the ice on window. PM switches on the anti ice
PM knows about the 300 extra meters, they know they can land on 09	32:16	31:36	936	23:24	(23'24''/936kg) PM knows about the 300 extra meters, they know they can land on 09
PM suggests to do a manual landing to dive below the glide slope	32:56	32:16	909	22:44	(22'44''/909kg) PM suggests to do a manual landing to dive below the glide slope
PM goes into OM-B to check CWC limitations	34:48	34:08	834	20:52	(20'52''/834kg) PM goes into OM-B to check CWC limitations
Briefing by the PF	37:03	36:23	744	18:37	(18'37''/744kg) Briefing by the PF
Clear to land, ILS 09	37:32	36:52	725	18:08	(18'08''/725kg) Clear to land, ILS 09
Ice on the screens. PM remembers the windshield heating	38:43	38:03	678	16:57	(16'57''/678kg) Ice on the screens. PM remembers the windshield heating
PM is aware he will have to do the landing due to ice	38:52	38:12	672	16:48	(16'48''/672kg) PM is aware he will have to do the landing due to ice
PF mentions fuel (600kg). Committed to land	39:21	38:41	652	16:19	(16'19''/652kg) PF mentions fuel (600kg). Committed to land
PF switch	42:33	41:53	524	13:07	(13'07''/524kg) PF switch
AP/AT disengaged	42:41	42:01	519	12:59	(12'59''/519kg) AP/AT disengaged
Touchdown	43:01	42:21	506	12:39	(12'39''/506kg) Touchdown

Pilot 7	Video time	Scenario time (VT-1'01'')	FOB (kg)	Time remaining (regarding FOB)	stickers for representation
Start of the recording	00:00	NA	NA	NA	NA
Actual start	01:01	00:00	2200	00:55:00	(55'00''/2200kg) Actual start
Start LAPA 27	04:50	03:49	2047	51:11	(51'11''/2047kg) Start LAPA 27
Finish LAPA 27	07:09	06:08	1954	48:52	(48'52''/1954kg) Finish LAPA 27
Go Around (ATC mentioned the friction car)	10:21	09:20	1826	45:40	(45'40''/1826kg) Go Around (ATC mentioned the friction car)
PF suggest to ask for radar vectors for new approach	13:01	12:00	1720	43:00	(43'00''/1720kg) PF suggest to ask for radar vectors for new approach
AC Bus Failure	13:47	12:46	1689	42:14	(42'14''/1689kg) AC Bus Failure
PM mentions the fuel (1600kg)	14:44	13:43	1651	41:17	(41'17''/1651kg) PM mentions the fuel (1600kg)
PF begins ECAM procedures	15:15	14:14	1630	40:46	(40'46''/1630kg) PF begins ECAM procedures
PM continues ECAM procedures	18:50	17:49	1487	37:11	(37'11''/1487kg) PM continues ECAM procedures
PM notices the no climb (50 sec)	19:43	18:42	1452	36:18	(36'18''/1452kg) PM notices the no climb (50 sec)
End of ECAM status	20:18	19:17	1428	35:43	(35'43''/1428kg) End of ECAM status
PM considers new airport	20:20	19:19	1427	35:41	(35'41''/1427kg) PM considers new airport
Weather for Hannover	21:43	20:42	1372	34:18	(34'18''/1372kg) Weather for Hannover
PM notices crosswind in Hannover, concludes that they are committed to Bremen. Declares emergency	22:24	21:23	1344	33:37	(33'37''/1344kg) PM notices crosswind in Hannover, concludes that they are committed to Bremen. Declares emergency
Back to Bremen	22:42	21:41	1332	33:19	(33'19''/1332kg) Back to Bremen
PM goes into OMB to check extended procedures	24:16	23:15	1270	31:45	(31'45''/1270kg) PM goes into OMB to check extended procedures
New wind, asked by PM after ATC ask if they want 27 or 09	27:05	26:04	1157	28:56	(28'56''/1157kg) New wind, asked by PM after ATC ask if they want 27 or 09
PF suggests to use 09 because of the wind (unusual)	27:48	26:47	1128	28:13	(28'13''/1128kg) PF suggests to use 09 because of the wind (unusual)
First clue for the ice on the window	29:18	28:17	1068	26:43	(26'43''/1068kg) First clue for the ice on the window
Starting LAPA 09	29:37	28:36	1056	26:24	(26'24''/1056kg) Starting LAPA 09
Mentions fuel (500kg, but maybe before a certain limit)	30:57	29:56	1002	25:04	(25'04''/1002kg) Mentions fuel (500kg, but maybe before a certain limit)
End of first LAPA 09	33:04	32:03	918	22:57	(22'57''/918kg) End of first LAPA 09
Correction of LAPA by PF	31:38	30:37	975	24:23	(24'23''/975kg) Correction of LAPA by PF
PF says that without N/W steering, they can't do an auto roll out	33:51	32:50	886	22:10	(22'10''/886kg) PF says that without N/W steering, they can't do an auto roll out
Cross wind warning on LAPA, end of LAPA 09	35:18	34:17	828	20:43	(20'43''/828kg) Cross wind warning on LAPA, end of LAPA 09
PM reminds ice on window and ask how they are going to do if PF can't see	36:29	35:28	781	19:32	(19'32''/781kg) PM reminds ice on window and ask how they are going to do if PF can't see
Clear ILS 09	37:43	36:42	732	18:18	(18'18''/732kg) Clear ILS 09
PF talks about doing a manual landing with CATII. PM knows you have to disconnect auto pilot at 80''	36:55	35:54	764	19:06	(19'06''/764kg) PF talks about doing a manual landing with CATII. PM knows you have to disconnect auto pilot at 80''
Difficult manipulation of the EFB by PM	40:10	39:09	634	15:51	(15'51''/634kg) Difficult manipulation of the EFB by PM
PF says he can't see anything	40:45	39:44	610	15:16	(15'16''/610kg) PF says he can't see anything
Pilot switch	40:52	39:51	606	15:09	(15'09''/606kg) Pilot switch
AP/AT disengaged	41:07	40:06	596	14:54	(14'54''/596kg) AP/AT disengaged
Touchdown	41:22	40:21	586	14:39	(14'39''/586kg) Touchdown

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Pilot 8	Video time	Scenario time (VT-2'15'')	FOB (kg)	Time remaining (regarding FOB)	stickers for representation
Start of the recording	00:00	NA	NA	NA	NA
Actual start	02:15	00:00	2000	00:50:00	(50'00''/2000kg) Actual start
PM mentions the fuel (approx.2T)	05:16	03:01	1879	46:59	(46'59''/1879kg) PM mentions the fuel (approx.2T)
Start LAPA 27	05:54	03:39	1854	46:21	(46'21''/1854kg) Start LAPA 27
Finish LAPA 27	07:39	05:24	1784	44:36	(44'36''/1784kg) Finish LAPA 27
Go Around (ATC mentioned the friction car & asked to reduce speed)	12:49	10:34	1577	39:26	(39'26''/1577kg) Go Around (ATC mentioned the friction car & asked to reduce speed)
AC Bus Failure	16:09	13:54	1444	36:06	(36'06''/1444kg) AC Bus Failure
Control the plane	16:34	14:19	1427	35:41	(35'41''/1427kg) Control the plane
PF begins ECAM procedures	17:24	15:09	1394	34:51	(34'51''/1394kg) PF begins ECAM procedures
PM continues ECAM procedures	18:55	16:40	1333	33:20	(33'20''/1333kg) PM continues ECAM procedures
PM notices the no climb (35 sec) because he wanted to stay at 3000.	19:51	17:36	1296	32:24	(32'24''/1296kg) PM notices the no climb (35 sec) because he wanted to stay at 3000.
PM wants to start APU. PF agrees	20:20	18:05	1276	31:55	(31'55''/1276kg) PM wants to start APU. PF agrees
PF suggests to ask for vectors for Bremen	20:51	18:36	1256	31:24	(31'24''/1256kg) PF suggests to ask for vectors for Bremen
Back to Bremen	21:20	19:05	1236	30:55	(30'55''/1236kg) Back to Bremen
End of ECAM procedures	22:32	20:17	1188	29:43	(29'43''/1188kg) End of ECAM procedures
PF asks to go check extended procedures - Chap 3	22:48	20:33	1178	29:27	(29'27''/1178kg) PF asks to go check extended procedures - Chap 3
Begins LAPA RWY27, doesn't remember the failure	23:13	20:58	1161	29:02	(29'02''/1161kg) Begins LAPA RWY27, doesn't remember the failure
End of first LAPA, begins OM-B	25:37	23:22	1065	26:38	(26'38''/1065kg) End of first LAPA, begins OM-B
End of OM-B	27:58	25:43	971	24:17	(24'17''/971kg) End of OM-B
PF suggests to get new wind	28:10	25:55	963	24:05	(24'05''/963kg) PF suggests to get new wind
PM notices 09 is better	29:26	27:11	912	22:49	(22'49''/912kg) PM notices 09 is better
Begins a LAPA for 09	29:34	27:19	907	22:41	(22'41''/907kg) Begins a LAPA for 09
Declares emergency (thanks to PF)	29:59	27:44	890	22:16	(22'16''/890kg) Declares emergency (thanks to PF)
First clue for the ice on the window	31:09	28:54	844	21:06	(21'06''/844kg) First clue for the ice on the window
Understand the warning message from LAPA	34:41	32:26	702	17:34	(17'34''/702kg) Understand the warning message from LAPA
PM knows he can use the extension of the RWY	35:08	32:53	684	17:07	(17'07''/684kg) PM knows he can use the extension of the RWY
Correct LAPA	36:22	34:07	635	15:53	(15'53''/635kg) Correct LAPA
Second cue for ice on window by PF	36:59	34:44	610	15:16	(15'16''/610kg) Second cue for ice on window by PF
PM suggests he will have to do the landing	37:11	34:56	602	15:04	(15'04''/602kg) PM suggests he will have to do the landing
PF asks when they have to disengage AP. PM knows about the 80''	38:14	35:59	560	14:01	(14'01''/560kg) PF asks when they have to disengage AP. PM knows about the 80''
Clearance ILS 09	38:23	36:08	554	13:52	(13'52''/554kg) Clearance ILS 09
PM remembers the call to the cabin	41:05	38:50	446	11:10	(11'10''/446kg) PM remembers the call to the cabin
Final talk about the ice, PM understands he has to take over	41:14	38:59	440	11:01	(11'01''/440kg) Final talk about the ice, PM understands he has to take over
PF switch	41:29	39:14	430	10:46	(10'46''/430kg) PF switch
AP/AT disengaged	41:49	39:34	417	10:26	(10'26''/417kg) AP/AT disengaged
Touchdown	41:58	39:43	411	10:17	(10'17''/411kg) Touchdown

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Pilot 9	Video time	Scenario time (VT-2'08'')	FOB (kg)	Time remaining (regarding FOB)	stickers for representation
Start of the recording	00:00	NA	NA	NA	NA
Actual start	02:08	00:00	2200	00:55:00	(55'00''/2200kg) Actual start
Start LAPA 27	06:01	03:53	2044	51:07	(51'07''/2044kg) Start LAPA 27
Finish LAPA 27	08:03	05:55	1963	49:05	(49'05''/1963kg) Finish LAPA 27
Go Around (ATC mentioned the friction car & asked to reduce speed)	11:17	09:09	1834	45:51	(45'51''/1834kg) Go Around (ATC mentioned the friction car & asked to reduce speed)
AC Bus Failure	14:42	12:34	1697	42:26	(42'26''/1697kg) AC Bus Failure
PF begins ECAM procedure	15:56	13:48	1648	41:12	(41'12''/1648kg) PF begins ECAM procedure
PM notices the no climb (11 sec)	17:37	15:29	1580	39:31	(39'31''/1580kg) PM notices the no climb (11 sec)
PM mentions the fuel (1500kg)	17:54	15:46	1569	39:14	(39'14''/1569kg) PM mentions the fuel (1500kg)
PM continues ECAM procedures	18:17	16:09	1554	38:51	(38'51''/1554kg) PM continues ECAM procedures
End of ECAM procedures	19:43	17:35	1496	37:25	(37'25''/1496kg) End of ECAM procedures
PF asks to go check OM-B extended procedures	19:50	17:42	1492	37:18	(37'18''/1492kg) PF asks to go check OM-B extended procedures
PM suggests to stay near Bremen before doing OM-B	20:48	18:40	1453	36:20	(36'20''/1453kg) PM suggests to stay near Bremen before doing OM-B
Open OM-B but doesn't check it	21:06	18:58	1441	36:02	(36'02''/1441kg) Open OM-B but doesn't check it
PM suggests to consider another airport (ask weather for Hannover)	21:16	19:08	1434	35:52	(35'52''/1434kg) PM suggests to consider another airport (ask weather for Hannover)
Weather at Hannover is not good	22:20	20:12	1392	34:48	(34'48''/1392kg) Weather at Hannover is not good
Back to Bremen	22:31	20:23	1384	34:37	(34'37''/1384kg) Back to Bremen
New information kilo at Bremen (ATC)	23:27	21:19	1347	33:41	(33'41''/1347kg) New information kilo at Bremen (ATC)
First clue for the ice on the window	24:02	21:54	1324	33:06	(33'06''/1324kg) First clue for the ice on the window
PM finally has some time to go into OMB extended procedures	24:17	22:09	1314	32:51	(32'51''/1314kg) PM finally has some time to go into OMB extended procedures
End of OM-B, he didn't check again the INOP systems but knows he has no TCAS and what may have caused the failure	26:28	24:20	1226	30:40	(30'40''/1226kg) End of OM-B, he didn't check again the INOP systems but knows he has no TCAS and what may have caused the failure
Emergency declared (PF points out the low fuel situation)	26:34	24:26	1222	30:34	(30'34''/1222kg) Emergency declared (PF points out the low fuel situation)
Start new LAPA for 27	27:19	25:11	1192	29:49	(29'49''/1192kg) Start new LAPA for 27
LAPA 27 finished with wrong wind	30:08	28:00	1080	27:00	(27'00''/1080kg) LAPA 27 finished with wrong wind
PF asks if the wind has changed. PM asks a second time wind to ATC, even though he wrote it down	30:24	28:16	1069	26:44	(26'44''/1069kg) PF asks if the wind has changed. PM asks a second time wind to ATC, even though he wrote it down
LAPA 27 with good wind, PF points out they have tail wind and need 09	30:54	28:46	1049	26:14	(26'14''/1049kg) LAPA 27 with good wind, PF points out they have tail wind and need 09
Request 09	31:12	29:04	1037	25:56	(25'56''/1037kg) Request 09
LAPA 09	31:54	29:46	1009	25:14	(25'14''/1009kg) LAPA 09
LAPA 09 finished	33:20	31:12	952	23:48	(23'48''/952kg) LAPA 09 finished
Committed to land (PF)	36:52	34:44	810	20:16	(20'16''/810kg) Committed to land (PF)
Clearance ILS 09	38:53	36:45	730	18:15	(18'15''/730kg) Clearance ILS 09
Second clue for ice on window	40:14	38:06	676	16:54	(16'54''/676kg) Second clue for ice on window
PF switch	43:26	41:18	548	13:42	(13'42''/548kg) PF switch
AP/AT disengaged	43:49	41:41	532	13:19	(13'19''/532kg) AP/AT disengaged
Touchdown	44:07	41:59	520	13:01	(13'01''/520kg) Touchdown

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.

Pilot 10	Video time	Scenario time (VT-7'00'')	FOB (kg)	Time remaining (regarding FOB)	stickers for representation
Start of the recording	00:00	NA	NA	NA	NA
Actual start	07:00	00:00	2200	00:55:00	(55'00''/2200kg) Actual start
Start LAPA 27	08:00	01:00	2160	54:00	(54'00''/2160kg) Start LAPA 27
Finish LAPA 27	09:11	02:11	2112	52:49	(52'49''/2112kg) Finish LAPA 27
PM mentions the fuel (1.9T)	13:35	06:35	1936	48:25	(48'25''/1936kg) PM mentions the fuel (1.9T)
PM considers Hannover	13:46	06:46	1929	48:14	(48'14''/1929kg) PM considers Hannover
PF gives a paper on which amount of fuel is needed to go to another airport (unusual)	13:57	06:57	1922	48:03	(48'03''/1922kg) PF gives a paper on which amount of fuel is needed to go to another airport (unusual)
Go Around (ATC mentioned the friction car)	15:53	08:53	1844	46:07	(46'07''/1844kg) Go Around (ATC mentioned the friction car)
PM wants to decide now if they divert or not	17:31	10:31	1779	44:29	(44'29''/1779kg) PM wants to decide now if they divert or not
AC Bus failure	19:30	12:30	1700	42:30	(42'30''/1700kg) AC Bus failure
PF begins ECAM procedure	20:25	13:25	1663	41:35	(41'35''/1663kg) PF begins ECAM procedure
PM continues ECAM procedure	21:55	14:55	1603	40:05	(40'05''/1603kg) PM continues ECAM procedure
PM notices the no climb (10 sec)	22:07	15:07	1595	39:53	(39'53''/1595kg) PM notices the no climb (10 sec)
End of ECAM	23:57	16:57	1522	38:03	(38'03''/1522kg) End of ECAM
Check weather in Hannover	24:08	17:08	1514	37:52	(37'52''/1514kg) Check weather in Hannover
PF asks to go check OM-B. PM asks for a heading back to Bremen first	26:07	19:07	1435	35:53	(35'53''/1435kg) PF asks to go check OM-B. PM asks for a heading back to Bremen first
Back to Bremen	27:02	20:02	1398	34:58	(34'58''/1398kg) Back to Bremen
ATC offers to give new weather (because they are back to Bremen)	27:54	20:54	1364	34:06	(34'06''/1364kg) ATC offers to give new weather (because they are back to Bremen)
PM understands they need 09 because they need CATII	28:54	21:54	1324	33:06	(33'06''/1324kg) PM understands they need 09 because they need CATII
PM goes through OM-B and doesn't find nothing more	29:21	22:21	1306	32:39	(32'39''/1306kg) PM goes through OM-B and doesn't find nothing more
LAPA for 09	30:34	23:34	1257	31:26	(31'26''/1257kg) LAPA for 09
PM declares emergency (with agreement of PF)	30:54	23:54	1244	31:06	(31'06''/1244kg) PM declares emergency (with agreement of PF)
First clue for the ice on the window	31:22	24:22	1225	30:38	(30'38''/1225kg) First clue for the ice on the window
PM request RWY09 (should have done it already)	33:33	26:33	1138	28:27	(28'27''/1138kg) PM request RWY09 (should have done it already)
Clear ILS CATII 09	38:38	31:38	934	23:22	(23'22''/934kg) Clear ILS CATII 09
Ice on the screen, PM is ready to take control	42:35	35:35	776	19:25	(19'25''/776kg) Ice on the screen, PM is ready to take control
Control switch	44:27	37:27	702	17:33	(17'33''/702kg) Control switch
Safety position call	44:29	37:29	700	17:31	(17'31''/700kg) Safety position call
AP/AT disengaged	44:42	37:42	692	17:18	(17'18''/692kg) AP/AT disengaged
Touchdown	44:57	37:57	682	17:03	(17'03''/682kg) Touchdown

Future Sky Safety has received funding from the EU's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 640597.