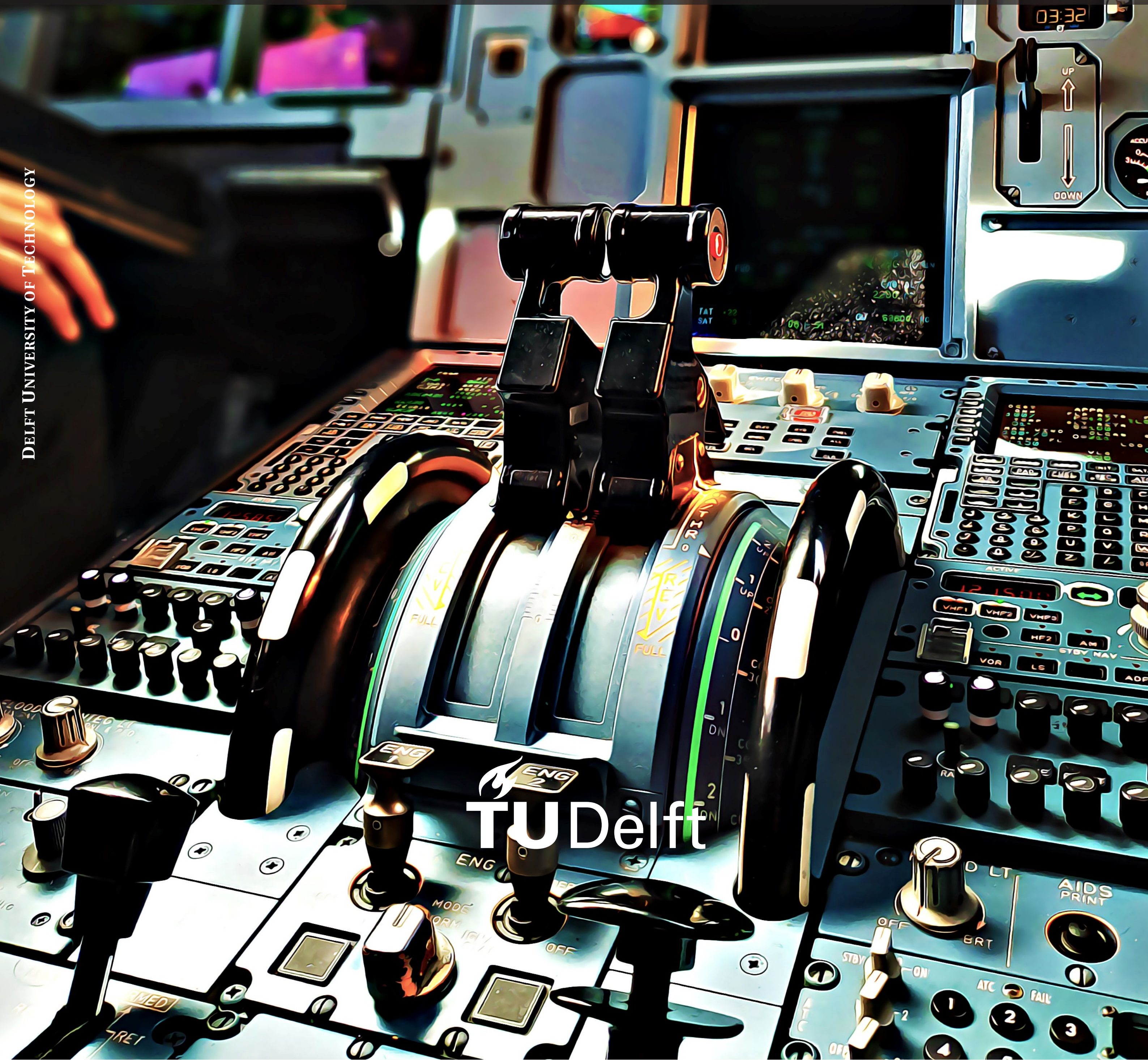
Training Startle and Surprise Management on The Flight Deck

Matteo Piras
4550552



Delft University of Technology

Training Startle and Surprise Management on The Flight Deck

by

Matteo Piras

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Friday October 28, 2022 at 01:00 PM.

Student number:	4550552	
Project duration:	March 12, 2021 - October 28, 2022	
Thesis committee:	Prof. dr. ir. M. Mulder,	TU Delft
	Ir. O. Stroosma,	TU Delft
	Dr. ir. J.C.F. de Winter,	TU Delft
	Dr. H.M. Landman,	TU Delft, TNO
	Dr. ir. M.M. van Paassen,	TU Delft

TU Delft
Delft University of Technology

Preface

This Thesis summarises the work which has been accomplished in the past year (and half!), and, more impressively, it marks the conclusion of a chapter of my life which started in Delft back in 2016. Reaching this goal would have not been possible without the support of some great people who I had the luck to meet throughout my journey and for whom spending few words of gratitude is now a must.

First, I would like to thank my supervisors for their bright guidance during my Thesis project: it is thanks to your precious inputs that I could successfully bring this project to a conclusion. Thank you Olaf for always providing me with responsive feedback and helping me out with all the questions, problems and doubts that I had about DUECA! Renè, it was great fun to test my simulations on SIMONA once they actually got to work: thank you for your patience during the numerous and, at times frustrating, test sessions that we had. And it is especially thanks to Annemarie that I could find all the participants to run my experiment! Your continuous support in setting up the experiment and in the follow up analysis was greatly appreciated. Max, thank you for always finding the time to answer my questions and for giving me a direction when the project was getting to a halt. On a similar note I would like to thank Hans Mulder, Alexander in't Veld, Herman van Veen and Jeroen van Roij for their contribution in finding airline pilots willing to join my research, for helping out with the testing and for the inspiring talks about flying!

Besides those who directly helped me with the project, I would like to address here all the friends from Delft and from Italy who I had the pleasure to spend these years with. It would definitely take more than a paragraph in the preface of my Thesis to properly thank you all: so I will keep it short and I will just say that growing up together and sharing with you all the highs and lows throughout this time made my life simply beautiful! I am quite confident that our friendship will be a long lasting one.

Finally, I want to especially thank my sisters and my parents for the love that they have always shown. Chiara, thank you for always supporting me with your kind words. Francesca, as promised, thank you for all your help with accessing the documentation that I needed for my research. Marta and Sofia, you are funny! Thank you for cheering me up with all your jokes. The biggest note of appreciation goes to my parents, who never stopped believing in me: thank you for supporting me when things were not so easy, and for always being by my side in my life choices. Dad, I am looking forward to eventually go flying together!

Matteo Piras
Delft, 14 October, 2022

Contents

List of Figures

List of Tables

Nomenclature

Abbreviations

ANC	Aviate Navigate Communicate
ATC	Air Traffic Control
$ATQP$	Alternative Training Qualification Programme
CWS	Control Wheel Steering
EBT	Evidence Based Training
$EGPWS$	Enhanced Ground Proximity Warning System
$ICAO$	International Civil Aviation Organisation
ISA	Instantaneous Self Assessment
$LOC-I$	Loss of Control In-flight
NLR	Nederlands Lucht- en Ruimtevaartcentrum
NM	Nautical Miles
$NOTAM$	Notice To Airmen
$NTSB$	National Transportation Safety Board
ORM	Operational Risk Management
PF	Pilot Flying
POH	Pilot Operating Handbook
RPM	Revolutions Per Minute
$RSME$	Rating Scale of Mental Effort
SEP	Single Engine Piston
$TCAS$	Traffic Collision Avoidance System
TRA	Temporary Reserved Area
VFR	Visual Flight Rules
VOR	Very High Frequency Omnidirectional Range

Part I

Paper

Training Startle and Surprise Management on The Flight Deck

Author: M. Piras; Supervisors: O. Stroosma, H.M. Landman, M.M. van Paassen, M. Mulder
Control & Simulation Section, Faculty of Aerospace Engineering
Delft University of Technology, Delft, The Netherlands

***Abstract*—Startle and surprise training interventions developed and tested in previous research have shown how falling back to mnemonic aid procedures when experiencing an unexpected event during flight can ameliorate the effects of startle and surprise responses. The current research follows up on the recommendations provided by Van Middelaar et al. [1] on the implementation of the *COOL* procedure which, in some cases, was also found too demanding to execute and somewhat distracting from ensuring a safe flight path. A new procedure featuring the steps *Aviate Breathe Check (ABC)* was developed to explicitly allow for task prioritisation (i.e. aviating over troubleshooting) and to be shorter in its execution while still addressing stress management related steps. We tested the *ABC* procedure with an experiment involving 25 airline pilots divided into an experimental group (n=13) and a control group (n=12) both of which experienced the same simulation scenarios designed to evoke startle and surprise. No significant effects of the type of training intervention on flying performance, on stress levels and on mental effort were found. Following events involving a sudden upset, the *ABC* procedure was overall shown to have a higher implementation rate compared to the *COOL* procedure evaluated in previous research in relation to the same test scenarios. This suggests that the *ABC* procedure could be more easily implemented on the flight deck. In an operational environment the ABC procedure could hence be integrated in recurrent pilot training programmes and translated into operational practice, without requiring extensive training. It is recommended to further test the procedure in a follow up experiment involving multi-crew operations and/or pilots with a more diverse training background.**

I. INTRODUCTION

Despite the routine procedures that airline flying involves, every flight is inevitably different and always features a degree of uncertainty which comes with possible unexpected situations. Not rarely pilots on the flight deck are faced with startling and/or surprising events, which, most of the times, are however consequence free and arise from routine factors such as ATC interactions, environmental conditions and incursions or in flight conflicts with other aircraft rather than extraordinary events [2].

Nonetheless startle and surprise events have the potential to compromise flight safety as many incidents and accidents even in recent history have shown [3], [4], [5]. Although the adverse outcome of a flight is never the result of a single cause but rather follows from the concatenation of many events, human factors including startle effect, surprise and distraction are often reported as at least contributing to loss of control accidents in flight [6], [7], this being the accident category repeatedly recording the highest number of fatalities.

The importance that startle and surprise have in affecting flight safety has to be found in the implications that these could have on the flight deck. Although startle and surprise are conceptually different and independent responses, as the former involves an unconscious startle reflex to a sudden stimulus, while the latter arises from a mismatch between expectations and reality [8], these can be characterised by similar effects encompassing both physiological and cognitive effects which could originate from the same event [9].

Temporary motor impairments following the startle reflex have been shown to last until ten seconds for complex motor tasks [8], inevitably causing the interruption of the ongoing activity and hence possibly posing a threat to the safety of a flight if manual input from the pilot is required to control the flight path. At the same time, cognitive skills such as situation awareness and decision making abilities might deteriorate from the disorientation and brief confusion following the stimulus, consequently creating the potential for the onset of further startle and surprise occurrences [8].

Similarly to startle, the mismatch arising from surprise results in the interruption of the task in progress, with the length of the interruption being a function of the schema-discrepant event [8]. It follows that the bigger the mismatch, the longer the task interruption and consequently the bigger the implications on the flight deck activities. Throughout the process of trying to solve such mismatch (i.e. reframing), the temporary absence of a suitable frame within which information is processed, or the implementation of an incorrect frame, might compromise the pilot ability of predicting future aircraft states, therefore enabling anticipatory actions. As a consequence, the pilot behaviour might become sequential and reactive rather than anticipatory and proactive [9].

One of the most discussed cases from recent flight history showing the extent up to which startle and surprise effects can impair a flight crew relates to the case of Air France flight 447. The blockage of the pitot tube due to the aircraft flying in a ice crystal environment caused an inconsistency between the measured speeds, therefore making the airspeed readings unreliable. Although both pilots correctly identified the loss of valid speed indications ("...*We haven't got a good*

display...", " We have lost the speeds ..."), neither of them referred to the right procedure for the current situation.
Despite the continuous triggering of the stall warning, providing explicit startling stimuli through aural and haptic alerts, the pilots reframed the problem as overspeeding, even though such frame was inconsistent with the high nose attitude and the high rate of descent [3]. Throughout the last part of the flight the pilot flying gave therefore mainly erratic and extreme pitch up and roll inputs (up to hitting the stop limits) leading the aircraft to reach angle of attack values of about 40 degrees.

Although this and other cases can be analysed in detail and provide insights in the implications that startle and surpise might have on the flight deck, researchers seem to agree on the fact that startle and surprise events are not associated to a defined set of causes, namely startle and surprise can be triggered by potentially any unexpected event [10].

The challenge arises of developing a startle and surprise training intervention method which is not tailored towards the resolution of specific failures. Following the rising need of improving startle and surprise management training within the often (too) predictable airline training programmes [11], researchers studied training interventions which focused on using unpredictability and variety in pilot training to improve performance in surprise situations as well as at developing and testing dedicated mnemonic aid procedures [12], [1], [10], [13]. The current research focuses on the latter type of training intervention.

A number of mnemonic aids such as COOL (Calm Down, Observe, Outline, Lead) [1], URP (Unload, Roll, Power), TAP (Time, Attitude, Power) [13] and BAD (Breath, Analyse, Decide) [14] have been developed and their effectiveness in recovering from startle and surprise events has been investigated. Overall, previous research showed that the implementation of these aids can be beneficial and also that pilots are in general willing to implement them in their operational routine [1], [10], [13].

In particular, the application of stress management related steps such as deep breathing and muscle relaxation was concluded to positively correlate with the task of collecting information, enhancing situation awareness [10]. Focusing on stress management (Calm Down in COOL, Unload in URP and Breath in BAD) and on collecting information through the senses prior to explicitly diagnosing the problem are considered the core and strength of these procedures. However, applying the technique in situations where the flight path is not yet under control could defeat the purpose of a safe recovery.

Van Middelaar et. al. [1] concluded that their COOL procedure might have a distraction effect, causing pilots to de-focus from the primary task of flying the aircraft. In addition, participants in the study indicated that the procedure itself could be too extensive and time demanding to be applied when little time is available [1]. Landman et al. [15] suggested that such procedure could be improved by firstly focusing on immediate issues and by simplyifing the COOL procedure itself [15].

Based on these considerations, a further startle and surprise management procedure was defined. It aims to allow for task prioritisation, with the first priority being ensuring a safe flight path, and to support pilots in recognising and controlling the physiological and cognitive effects resulting from startle and surprise. An experiment was designed to assess the effects of such procedure on pilots' flying performance, stress and workload management in comparison with a control group who received the same training without learning the procedure. In addition, the experiment also allowed for a comparison with an experimental group implementing the existing COOL procedure, for which raw data was retrieved from a previous experiment conducted by Van Middelaar et al. [1].

The paper is structured as follows: first, the existing startle and surprise management procedures are briefly reviewed and the *ABC* procedure is introduced in section II. Section III provides an overview of the experiment design and of the experimental set up, introducing the groups of participants joining the research, describing the apparatus used for the experimental activities and explaining the experiment procedure in details. The hypotheses and dependent measures are also included in this section, which is concluded with subsection III-K summarising the data analysis process related to these measures. The following section presents the results from the data analysis. In particular, subsection IV-A and subsection IV-C give an overview of the data that have been discarded in the analysis and report the results in relation to the main hypotheses respectively. The results from further scenario based analyses are reported in subsection IV-D, subsection IV-E and subsection IV-F respectively. The presented results are discussed in section V, while recommendations for future projects are provided in subsection V-E. Finally, the conclusions to this research work are presented in section VI.

II. Procedure Development

A. Brief review of existing procedures

Previous research has investigated different procedures to support pilots in recovering from startle and surprise events on the flight deck. A preliminary background study focused on revising few existing startle and surprise management procedures from literature with the goal of highlighting commonalities and differences among them. These include: *TAP (Time Attitude Power)* [13], *URP (Unload Roll Power)* [10], *COOL (Calm Down Observe Outline Lead)* [1] and *BAD (Breathe Analyse and Decide)* [14]. Their steps and related functions are categorised in Table I.

TABLE I: Comparison of startle and surprise management procedures.

	TAP	URP	COOL	BAD
Aviate	Attitude, Power	-	-	-
Stress Management	-	Unload	Calm Down	Breathe
Information Collection	-	Roll	Observe	Analyse
Situation Analysis	-	Roll	Outline	Analyse
Decision Making and Execution	-	Power	Lead	Decide

The general focus of the *TAP* procedure lies in providing pilots with specific steps involving basic known pitch and power settings for pilots to regain basic control of the aircraft. Instead, the other procedures considered aim at coping with possible motor and cognitive impairments by focusing on stress management, regaining situational awareness and promoting decision making.

In particular, the COOL and URP techniques address stress management by implementing a specific breathing and muscle relaxation technique finalised at counteracting the physiological effect of startle and surprise, while, by contrast, the BAD technique includes only deep breathing as a stress management step. At the same time, the TAP procedure does not focus on stress management at all. The implementation of these steps has been shown to have a positive influence on the task of collecting information [10], enhancing situation awareness, and are considered to be the core of such procedures.

When considering the complexity of each procedure it is observed that BAD and TAP involve fewer steps and are less elaborate than the URP and COOL techniques. The execution of the first ones is hence probably more time efficient than the latter ones.

The COOL procedure in particular appears to be the most extensive technique among the ones discussed: in contrast with the *"Roll"* step from URP , COOL makes a distinction between the *"Observe"* step and the *"Outline"* step, therefore allocating the task of collecting information and the task of giving meaning to the situation to two explicitly different moments in time. If on one side, such distinction allows for a more refined discretization of tasks, therefore providing more explicit guidance in the application of the procedure, on the other hand it also lengthens the procedure itself, making it demanding to execute and possibly distracting in situations requiring flight path control following an upset [1].

From these considerations it follows the need for a startle and surprise management procedure that aims at task prioritisation (i.e. aviating over troubleshooting) and stress management while allowing for an easy implementation of the procedure.

B. ABC procedure

The *ABC* procedure was developed to support three different goals:

- to ensure a safe flight path over problem identification and troubleshooting related tasks
- to recognize and control the physiological and psychological reactions resulting from startle and surprise events
- to support the reframing process.

The items encompassing these goals are defined as follows [15]:

- ***Aviate***: recalling the *Aviate*, *Navigate*, *Communicate* order of priorities which is well known to pilots, this item is meant to tailor the first course of action of the pilot towards ensuring control of the aircraft at all times. Although the meaning of "*Aviating*" could be arguably broad, two elements in particular need to be addressed, namely flight path control and energy management. Once these aspects are being taken care of, the other priorities can also be tackled. The actions taken when performing this step are situation specific: it is therefore not intended to provide guidelines on how to aviate, but rather to set pilots in the mindset of taking the required actions to stabilise the aircraft to the best of their abilities. In particular the aim is to attain a steady horizontal flight path with wings level as much as the situation allows.
- ***Breathe***: this second step deals with stress management to cope with the physiological response originating from the startle and surprise responses. At this stage the pilot has to detach from the current problem being faced and focus on his/her own breathing, muscular tension and application of control forces. This step greatly borrows from the successful stress management technique implemented in previous similar studies [1], [10] and requires sitting up-right on the seat, relaxing arms legs and shoulders, deeply breathing for a couple of seconds and focusing on the application of the control forces when exhaling as a conclusion of the breathing cycle.
- ***Check***: to conclude the procedure, this step involves the acknowledgement of the primary instruments (airspeed, attitude, altitude, direction indication and throttle settings) and secondary instruments (engine indications, flaps and gear settings) by calling their status (e.g. "*Primary Instruments are good*" or "*speed is low*"). The difference with the initial "Aviate" step lies in the more conscious and systematic scanning that is to be performed at this stage compared to the first unconscious response performed to ensure the control of the flight path.

III. Method

A. Experiment Design Summary

The experiment featured a between-participant design involving 25 airline pilots. They were divided into an experimental group, implementing the ABC procedure, and a baseline group undertaking the same training and facing the same scenarios, without being introduced to such procedure. Furthermore the data related to the experimental group from the *COOL* study were also included in the current study [1]. Performance measurements were based on the calculation of the total time spent outside a safe envelope defined by attitude and speed boundaries, while mental effort and stress levels were measured respectively using a Rated Scale of Mental Effort [16] and different 1-10 Likert scales [17], which participants used to provide the related ratings after the familiarisation, pre-test and test scenarios respectively.

To show compliance with the overall experiment design, the experiment itself was pre-registered on the Center for Open Science platform where a summary of the research plan can be accessed [1].

B. Hypotheses

It was hypothesised that the experimental group would show a significantly better flying performance compared to the control group and *COOL* group over the test scenarios (H1) and that the control group would show a significantly higher increase in mental effort and stress from the initial uneventful familiarisation scenarios to the post test scenarios in comparison with the experimental group (H2). In addition to these main hypotheses, the research group looked into significant differences in stress and mental effort between the *ABC* and the *COOL* groups, with the former group being expected to show significantly higher ratings (H3).

The main rationale behind the mentioned hypotheses lies on the fact that explicitly letting pilots recall the order of priorities which they should comply with (i.e. aviating over troubleshooting) through the *ABC* procedure would result in these pilots being always firstly concerned with flying the aircraft. Hence they would be expected to perform better on this aspect against the pilots from the control or *COOL* groups for which such priority is not explicitly addressed. Furthermore, the inclusion of a stress management technique in the *ABC* procedure in combination with the reduced length of the procedure itself and of its execution is believed to ease the implementation of such procedure with consequent beneficial impact on workload management and stress levels in relation to what experienced by the *COOL* and control groups.

C. Participants

The 25 participants were divided into a control and experimental group. The statistics describing both sample groups and the participants from the *COOL* study are reported in Table II.

TABLE II: Descriptive statistics on participants' experience.

	ABC (μ, σ)	Control (μ, σ)	*COOL* (μ, σ)
Flight Experience [Hours]	(9750 , 6899.2)	(10160 , 5732.1)	(7437 , 5617)
Age [Years]	(41 , 9.9)	(44.5 , 8.3)	(37.1 , 12.72)
Airline Experience [Years]	(17.6 , 11.2)	(14 , 9.5)	(13.54 , 10.75)
Trait Anxiety Scores	(31.4 , 5.3)	(26.7 , 3.3)	(21.1 , 12.3)
Captains	6	7	4
First Officers	6	4	6
Second Officers	1	1	2
TRI / TRE	2	1	Unknown
Male Pilots	11	12	12
Female Pilots	2	0	0

[1]the research plan can be accessed at: https://osf.io/juz8m

Although some participants undertook an Upset Prevention and Recovery Training (UPRT) throughout their career, none of them had an extensive aerobatic experience or a military background. Excluding subjects with such profile was an experimental design choice as recurrently practicing unusal attitudes during active flight duty could bias the results. Although all pilots had experience on multi-engine piston aircraft from the initial training stage of their career, only one participant had recent experience on this class as a former flight instructor.

No significant differences were found between the three groups when performing an ANOVA test on the total flight experience (*F(2,34) = 0.727, p = 0.491*), age (*F(2,31) = 0.953, p = 0.396*), or total years spent working for an airline (*F(2,34) = 0.996, p = 0.380*).

Trait anxiety of each participant was assessed with the State-Trait Anxiety Inventory (STAI) test [18] before the experiment. A significant difference in STAI scores was found between the groups (*H(2) = 9.152 , p = 0.010*), with the experimental *ABC* group showing significantly higher scores than the *COOL* group and the control group.

D. Apparatus

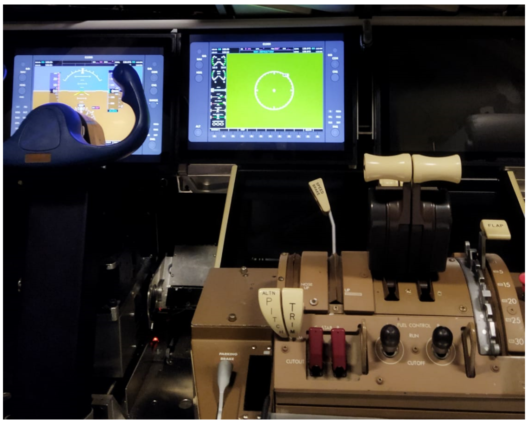
The aircraft model is a Piper Seneca PA-34-III developed by De Muynck and Hesse [19]. The PA-34 is a light twin engine propeller aircraft developed by Piper Aircraft which firstly entered production in 1971.

Given the two engine configuration, it is necessary for all participants to hold (or have held) a Multi Engine Piston (MEP) rating. No type rating is required to fly the PA-34 aircraft, however, as it is an experimental pre-requisite to make the developed startle and surprise intervention training non type-specific.

Similarly to previous experiments performed on the topic of startle and surprise management, the model was implemented in the TU Delft SIMONA Research Simulator (SRS).

SIMONA (shown in Figure 1) is a full motion flight simulator featuring six hydraulic actuators and allowing the pilot a field of view of 180 degrees over a collimation mirror placed in front of the cockpit. However, during the experiment the right visuals were not working due to the malfunction of one of the projectors, limiting the field of view to 120 degrees only. Its systems are comparable to a level D simulator, even though SIMONA does not qualify as such due to the different configurations that it can feature to support research in the aerospace but also in the automotive field. The cabin layout resembles the one of a standard airliner and it significantly differs from the cockpit of the PA-34. For the experiment a column with electric pitch trim and rudder pedals were installed on the left seat while throttle and flaps controls were available to the pilots on a pedestal in-between the Captain and First Officer seats.

(a) External view of SIMONA



(b) Internal set-up used for the experiment

Fig. 1: SIMONA Research Simulator (SRS)

The avionics consisted in a simulation of the G1000 Primary Flight Display (PFD) and Multi Function Display (MFD) developed by Van Leeuwen [20]. No interaction with these avionics was however required from the participants and these were only used to gather primary indications on speed, altitude and attitude, and secondary indication on engine parameters and gear settings.

A similar experimental set up was used for the previous *COOL* study, except that the latter involved the use of digital instruments based on a Cessna Citation 550, including flap indications which the Garmin 1000 used for the current study did not feature.

E. Procedure

The experiment procedure is summarised in Figure 2. During a first ground briefing each participant was explained the sequence of experimental activities and introduced to the SIMONA simulator and its features. The pilot was hence lead to the simulator for a familiarisation flight consisting of two scenarios where he/she was tasked to fly a standard left-hand circuit at 1,000 ft in Amsterdam Schipol Airport, on runway 18C, respectively with and without crosswind. Throughout the scenarios the pilot was guided in the demonstration of the stall warning, gear and flaps deployment, and go around procedure and was shown the key reference points of the circuit. This experimental session was then concluded with a pre-test consisting of a left engine failure during final approach to the same runway used for the familiarisation flights: in this occasion the pilot performance was assessed to check the balancing of the groups.

Upon the completion of the pre-test scenario, each participant was asked to leave the simulator for a ground briefing on startle and surprise, followed by another simulator session consisting of four training scenarios where different startling and surprising events were introduced. Such session was meant to expose both groups to a wide range of failures, and in particular for the experimental group to practice with the startle and surprise management training procedure previously introduced in the ground briefing.

All experimental activities performed up to this stage aimed at preparing each participant for a last test session featuring three scenarios throughout which pilot performance was measured. The experiment was finally concluded with a de-briefing where the details of each scenario covered were unveiled to the pilot, leaving also the opportunity for the researcher to discuss more in depth the scope of the current research and for the participant to provide feedback on the training received and on the experiment in general.

The procedure here described was developed after the experiment design from the *COOL* study, with the exception that the familiarisation round in this case involved also the execution of the go-around procedure in preparation for one of the test scenarios which was not part of the *COOL* study itself.

F. Ground Training

The ground training focused on presenting to all participants the conceptual differences between the startle and surprise responses, highlighting the physiological and cognitive effects and their implications on the flight deck activities by means of few examples. While the first part of the ground training was common to both groups, only the experimental group was introduced to the "*ABC*" (*Aviate*, *Breathe*, *Check*) procedure, the goals and steps of which were explained in detail.

The participants from the *ABC* group were asked to call out-loud each step of the procedure and to report the application of its steps on a dedicated questionnaire at the end of each test scenario.

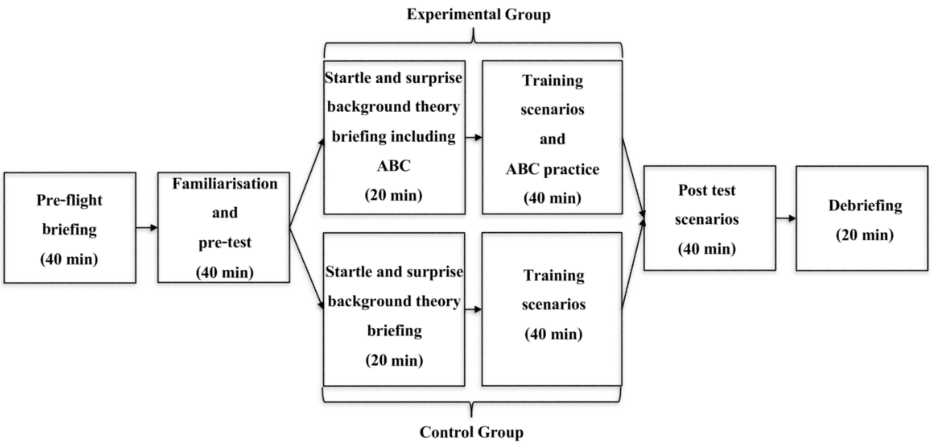


Fig. 2: Experiment procedure.

G. Training Scenarios

Four training scenarios were implemented in the training phase, each of which featuring a different failure. The only exception was represented by one uneventful scenario being identical to the first familiarisation round. This scenario was needed to firstly let the experimental group practice the *ABC* procedure: participants were hence required to fly the same circuit as the one previously practiced with and to apply the procedure several times throughout the downwind and final legs. The control group also flew this scenario to equalise training time.

The second training scenario involved a false stall warning triggered 45 seconds from the start, with the aircraft being already configured for landing on final approach to runway $18C$ at about $700ft$ and $5NM$ from the threshold. The pilot was tasked with landing the aircraft on the runway in strong cross wind conditions ($19knots$ of cross wind from $090deg$). This scenario differs from the one of the previous *COOL* study which instead featured the loss of rudder authority. The reason for replacing this failure with a spurious stall warning is that the latter provides an explicit warning compared to the former failure which could be unnoticed. For the same reason, the *COOL* study included an additional training scenario involving a RPM indication failure which, however, was discarded for the purpose of this training.

The same set up characterizes also the third scenario, which featured a sudden rudder bias when descending below 600 ft, imparing directional control and causing a marked yaw motion.

The training was finally concluded with an engine failure occurring slightly after rotation on runway 18C, at 90 knots, with the pilot being tasked with flying the usual circuit pattern at 1,000ft.

H. Test Scenarios

The formal assessment of pilot performance was conducted on the basis of three scenarios, two of which were implemented from the previous *COOL* study. These scenarios did feature, respectively, a flap asymmetry when deploying the flaps on baseleg, and a center of gravity shift upon rotation causing a sudden upset in pitch [1]. These scenarios were selected from the previous study for comparison with the current study as they feature an upset that destabilises the flight path, hence calling for aviating actions to recover from the upset itself. Furthermore, in the mass shift scenario the application of the *COOL* procedure was reported to be particularly distracting as for more than half of the participants from the experimental group the *COOL* procedure itself was applied too soon, namely before ensuring a safe flight path [1].

The third test scenario was developed to feature few novel elements. The pilot was tasked to fly a visual approach to runway 24 in Rotterdam The Hague airport during which a landing gear extension failure occurred, hence forcing a go-around. Contrary to the previous cases, throughout this scenario ATC communication recordings were played simulating other traffic in the proximity of the airport. Such communications were scripted and recorded to be played during the scenario itself. Some of these communications concerned the participants who were required to respond and comply with ATC as they would normally do when flying. After the triggering of the failure and during the execution of the go-around procedure, a series of non flight essential requests were made to the pilot including reporting the amount of fuel remaining, the number of People On Board (POB) and readability checks, the response to which was expected to be delayed to a later moment, therefore hypothetically leading the pilot to prioritize flying the aircraft over communicating with ATC.

I. Dependent measures

Pilot performance as well as subjective workload, stress, and level of startle and surprise have been assessed.

1) Flying performance: The assessment of the former follows from the definition of "aircraft upset" as "an event that unintentionally exceeds the parameters normally experienced in flight or training". The parameters here considered are:

- Bank angle $\phi > 40\,\mathrm{deg}$
- Pitch angle $\theta > 20\,\mathrm{deg}$ or $\theta < -10\,\mathrm{deg}$
- Speed $V_{TAS} < 80 knots$

with $80 knots$ being the minimum control speed of the Piper Seneca PA-34 used for the experiment. Based on this definition, flight performance over all the test scenarios is measured as the total time spent by the pilot outside of the safe envelope defined by the above mentioned parameters. Such variable serves as an indicator of the overall performance throughout the test session, this being hence independent of the scenario specific failure.

2) Subjective ratings: The level of stress, startle and surprise were measured using three separate 1-10 Likert scales [17] to be filled in by the participants at the end of each test scenario and after the completion of the pre-test. Similarly, the assessment of the perceived workload was performed based on the Rated Scale of Mental Effort (RSME) [16] which pilots used to provide the related rating before commencing a new simulation.

Pilots from the experimental group were also asked to indicate which steps of the procedure they implemented in relation to a specific test scenario and how helpful the trained *ABC* procedure was to cope with the scenario specific event using a 1-10 Likert scale [17].

3) Additional Measures: To gain insight in the performance of the participants, a number of observations were collected during the experiment as well as secondary measures such as the number of times the safe envelope was exceeded.

J. Data Exclusion Criteria

Although all participants were briefed on how to fly each scenario, the test scenarios themselves still allow for some freedom, potentially leading to unexpected actions that could compromise pilots' performance. In nominal situations (i.e. prior to the occurrence of the scenario specific failure), whenever a pilot was found not to follow the standard circuit or to deviate from the prescribed use of flaps and gear in the circuit, the related data was excluded from the analysis.

Furthermore, technical faults that could spoil the scenario specific failure did also represent a reason for data exclusion. Similarly, data were excluded for those participant who did notice the failure characterizing the scenario.

K. Data Analysis

Normality and equality of variances of the raw data used for the analysis were assessed using respectively the Shapiro-Wilk normality test and Levene's test. For multivariate analysis the equality of covariance matrices was checked with the Box's test. Parametric tests were performed when the assumptions of normality and equality of variance were complied with.

A Kruskal-Wallis test was performed over the flap asymmetry and mass shift scenarios, to assess possible performance differences between the *ABC*, *COOL* and Control groups.

Furthermore, the difference between each post test scenario and the familiarisation scenario was taken to investigate possible differences between the *ABC* and control groups in terms of increase in RSME and stress ratings respectively. A Mann-Withney test was hence performed on the obtained differences in relation to each post test scenario.

To compare the experimental groups from both studies in terms of the same performance aspects, a MANOVA analysis was performed including the *ABC* and *COOL* groups as between participant independent variables and the flap asymmetry and mass shift scenarios as within participants independent variables. While the Likert scales used for measuring stress levels provide ordinal data, researchers have shown that parametric tests are robust also when this type of data are used [21]. As the *COOL* experiment used a 0-10 Visual Analogue scale for measuring stress instead of a Likert scale, the comparison between the *ABC* and *COOL* groups for this measure was based on the zscores obtained by standardising all the ratings (from the control and experimental groups) collected for the *ABC* and *COOL* experiments respectively.

In addition, further scenario based analyses were performed using ANOVA testing for parametric data or Kruskal-Wallis or Mann-Whitney tests for non parametric data.

IV. RESULTS

A. Exclusion of Participants

The data gathered from some of the participants had to be discarded for statistical analysis purposes as the related performance did not respect the design objective that a specific scenario was developed upon. In particular, one participant from the experimental *ABC* group got lost and did not notice the flap asymmetry failure possibly due to the reduced visibility, consequently affecting the subjective ratings and flying performance. On the same scenario another participant from the same group deployed the flaps on short final, hence noticing the failure right before landing and consequently did not have time to apply the procedure. Similarly, four and six participants respectively from the *ABC* and control groups did not notice the landing gear failure in the last test scenario. In addition, two participants from the experimental group experienced technical issues with the simulator in the mass shift scenario as a consequence of which the simulation had to be stopped after the occurrence of the failure, hence compromising their performance over the whole scenario.

Table III summarises the data that have been discarded in each scenario. The results presented in the next sections have been obtained without including the unreliable data here presented.

TABLE III: Data disregarded per test scenario

	ABC	Control	reason for disregarding the data
Flap Asymmetry	2	-	*got lost due to reduced visibility (1)* *deployed flaps on short final (1)*
Mass Shift	2	-	*technical issue with the simulator*
Landing gear failure	4	6	*did not notice the failure*

B. Group Balancing Assessment: pre-test scenario

To check for group balancing between the *ABC*, control and *COOL* groups a preliminary analysis is performed on the flying performance data and subjective ratings gathered for the pre-test scenario before the training. A Kruskal-Wallis test performed on the time spent outside the envelope does not show any significant difference between the groups (*H(2)= 0.175,p = 0.916*). Similarly, no between group differences are found when performing a MANOVA on stress and RSME scores (*F(4,68) = 2.183, p = 0.080*). However, univariate testing shows that the groups significantly differ in terms of RSME ratings (*F(2,34) = 4.107, p = 0.027*). Further pairwise comparison based on Bonferroni adjustments shows a significant difference between the ABC and COOL groups on RSME scores (*p = 0.024, C.I.= 2.194,39.357*), with the former reporting higher scores ($\mu = 80.69$) compared to the latter ($\mu = 59.91$).

C. Preliminary Results

1) Time outside of the envelope: Figure 3 shows the boxplots of the three groups for the total time outside of the envelope: few outliers are detected, at around $300s$, $800s$ and $150s$ respectively. From the scatter plot in figure Figure 4 it can be also shown how overall pilots spent more time outside the envelope in the mass shift scenario than in the flap asymmetry scenario.

The results of a Kruskal-Wallis test performed on the total time spent in an upset condition over the first two test scenarios (flap asymmetry and mass shift) does not show a significant difference among the *ABC*, control and *COOL* groups (*H(2)= 1.029,p = 0.598*).

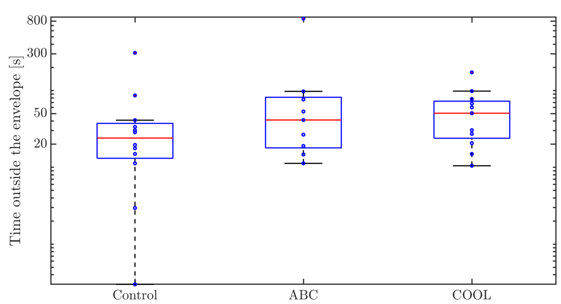


Fig. 3: Total time spent outside of the safe envelope over the flap asymmetry and mass shift scenario.

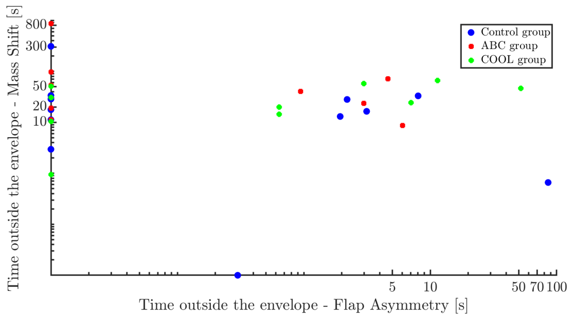


Fig. 4: Time outside the envelope for the mass shift and flap asymmetry scenarios.

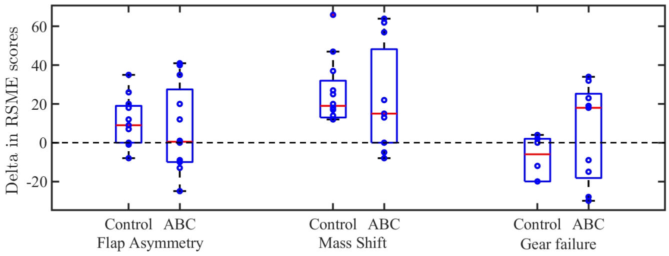
2) Stress and Workload: Boxplots showing the RSME and stress score distribution in relation to the hypotheses H2 and H3 are presented in Figure 4 and Figure 5 respectively.

From Figure 4 it is observed that, over the test scenarios, the increase in RSME and stress from the familiarisation scenario do not seem to differ much between the *ABC* and control groups. At the same time, however, the former group shows a higher increase in stress levels in relation to the flap asymmetry (respectively μ = 1.27 and $\mu = 1.08$), mass shift (respectively μ = 3.54 and $\mu = 2.58$), and landing gear failure scenarios (respectively μ = 1.77 and μ = -1.37). From the last scenario it can also be observed that the mean increase in stress is negative for the control group (hence, on the average, experiencing a decrease in stress levels) while the mean increase in stress for the *ABC* group is instead positive. The same occurs when considering the increase in RSME scores for this scenario, in relation to which the *ABC* and control group show an average increase of 4.88 and -7.66 respectively.

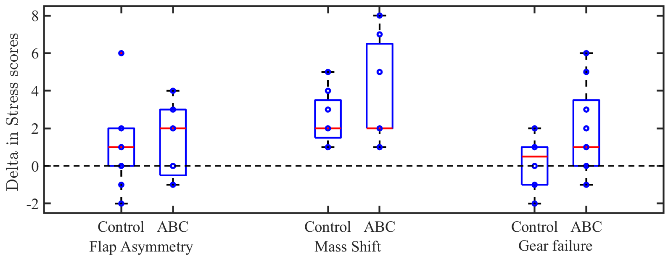
Furthermore, the mass shift scenario shows a higher increase in both measures compared to the other test scenarios. Few participants for both groups show also a decrease in RSME and stress levels over the scenarios, hence reporting the familiarisation scenario to be more stressful and effortful than the specific test scenarios.

The analysis performed to assess possible group differences (*ABC* versus control) in increasing RSME ratings showed such differences to be not significant in the flap asymmetry scenario *(U = 59.5 , p = 0.688)*, mass shift scenario *(U = 54.0, p = 0.460)*, and landing gear scenario respectively *(U = 19.0 , p = 0.345)*. Similarly, no significant increase in stress levels is found between groups in relation to the flap asymmetry scenario , *(U = 58.5, p = 0.640)*, mass shift scenario *(U = 57.0, p = 0.565)* and landing gear scenario *(U = 17.5, p = 0.256)* respectively.

The further MANOVA analysis performed to assess the difference in stress and subjective workload experienced between the two experimental groups over the flap asymmetry and mass shift scenarios reveals that overall there is no significant main effect of the type of training intervention on the two dependent variables *(F(2,18) = 0.909, p = 0.421)*. However, following up univariate testing shows that a significant difference in RSME scores exists between the flap asymmetry and the mass shift scenario *(F(1,19) = 7.104, p = 0.015,* η_p^2 *= 0.272)*, these reporting a mean score of 75.167 and 68.375, respectively.

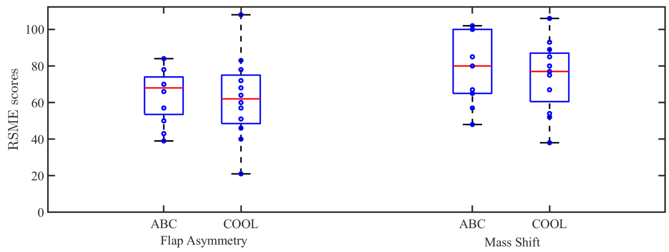


(a) Increase in RSME scores from the familiarization to the post test scenarios.

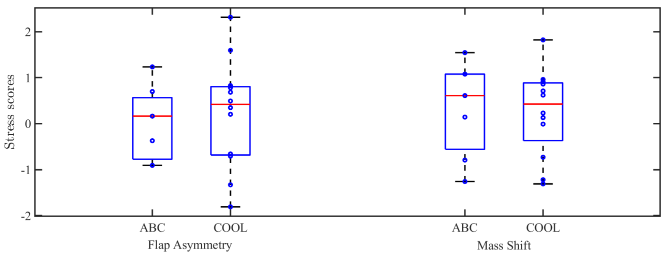


(b) Increase in stress scores from the familiarization to the post test scenarios.

Fig. 4: Increase in RSME and Stress scores in relation to the *ABC* and control groups.



(a) RSME scores over the flap asymmetry and mass shift scenarios



(b) Stress scores (zscores) over the flap asymmetry and mass shift scenarios.

Fig. 5: RSME and stress scores in relation to the ABC and COOL groups.

A summary of the test results in relation to the main hypotheses is presented in Table IV and Table V. Although the non significant results here presented lead to reject the main hypotheses, the significant effects found from the MANOVA analysis in addition to the number of outliers detected when assessing the total time spent in an upset condition as shown in Figure 3 provides some evidence for a more in depth analysis to be performed. This analysis aims at gaining more insight in possible group differences on a scenario based case and at highlighting within participant differences among test scenarios. To do so a number of observations collected during the experiment is presented next, followed by a further testing on the subjective ratings collected.

TABLE IV: Summary of test results in relation to H1 and H3.

Hypothesis	Test Statistics	Significance
H1	$H = 1.029$	$p = 0.598$
H3	$F = 0.909$	$p = 0.421$

TABLE V: Summary of test results in relation to H2.

	Stress		RSME	
Scenario	Test Statistics	Significance	Test Statistics	Significance
Flap Asymmetry	$U = 58.5$	$p = 0.640$	$U = 59.5$	$p = 0.688$
Mass Shift	$U = 57.0$	$p = 0.565$	$U = 54.0$	$p = 0.460$
Gear Failure	$U = 17.5$	$p = 0.256$	$U = 19.0$	$p = 0.345$

D. Scenario related observations

1) Pre-test: This scenario features four crashes per group with the addition of one pilot from the experimental group who probably survived a precautionary landing in a field. One pilot belonging to the control group landed on RWY 18R instead of the designated RWY 18C. All pilots but one pilot from the control group could successfully identify what the failure was about.

2) Flap Asymmetry: Four out of eleven participants from the experimental group showed difficulties in figuring what the issue was about. Of these participants two barely noticed that something was off nominal but were not particularly upset from the failure while two others reasonably guessed that the problem was aileron related. Not being able to frame the problem correctly seems to correlate with the consequent decision making as three out of four pilots here considered decided to land with full flaps and one with flaps 25. The remaining pilots from the experimental group who correctly identified the problem performed a flapless landing or landed with flaps 25. A similar pattern is seen in the control group although to a less severe extent as only three participants seemed to disregard the problem (*" a lot of power needed ... continuing"*) or framed it as an engine failure. Of these, two landed with full flaps and one with flaps 25. These observations are summarised in Table VI.

When coming to the application of the *ABC* procedure for this scenario, it is assessed that the *Breathe* and *Check* steps were, respectively, not applied in two and one instances, hence only nine out of eleven participants from this group fully complied with the *ABC* steps. One of the pilots who did not apply the procedure in its full length also belonged to the group of pilots who did not frame the problem correctly.

TABLE VI: Observations for the flap asymmetry scenario.

	ABC	Control	COOL
Crash	-	-	unknown
Different landing loc.	-	-	unknown
Stalls	-	-	unknown
Landing conf: flaps up	5/11	3/12	2/12
Landing conf: flaps 25	3/11	6/12	6/12
Deployed flaps full	8/11	5/12	4/12
Landing conf: flaps full	3/11	3/12	4/12
Aviate	11/11	N/A	N/A
Breathe	9/11	N/A	N/A
Check	10/11	N/A	N/A
Calm Down	N/A	N/A	11/12
Observe	N/A	N/A	12/12
Outline	N/A	N/A	12/12
Lead	N/A	N/A	12/12

3) Mass Shift: The mass shift scenario features a sudden failure which pilots struggled more to cope with than the previous flap asymmetry scenario. In only one occasion a pilot from the experimental group crashed while attempting to land on RWY 18R, while a pilot from the control group landed on RWY 27. Most participants seemed to understand that the failure relates to a pitch problem but keeping the aircraft under control was still demanding: four instances of stall occurred in the control group against one instance from the experimental group. Table VIIsummarises the descriptives and results here presented.

TABLE VII: Observations for the mass shift scenario.

	ABC	Control	COOL
Crash	1	-	unknown
Different landing loc.	1	1	unknown
Stalls	1/11	4/12	unknown
Landing conf: flaps up	9/11	9/12	9/12
Landing conf: flaps 25	2/11	0/12	3/12
Deployed flaps full	0/11	4/12	0/12
Landing conf: flaps full	0/11	3/12	0/12
Aviate	11/11	N/A	N/A
Breathe	11/11	N/A	N/A
Check	11/11	N/A	N/A
Calm Down	N/A	N/A	10/12
Observe	N/A	N/A	12/12
Outline	N/A	N/A	10/12
Lead	N/A	N/A	11/12

Furthermore, while all participants from this group applied the full *ABC* procedure, one of them particularly focused on the aviate step: the pilot significantly deviated from the circuit altitude by climbing to 5,000ft and struggled to take effective actions to control the aircraft (*"Trying to find a way to keep the aircraft under control but how ?? "*). A similar behaviour is observed for one participant in the control group who also left the circuit pattern and reached approximately 2,400ft before levelling off.

4) Landing Gear Failure: This scenario featured a straightforward failure which, however, went unnoticed for four and six participants in the experimental and control groups respectively. The reason behind not seeing the failure was reported to be either related to the warning sound not being loud enough, or misunderstood for an autopilot disconnect transmitted over from another aircraft, or accidentally caused by ATC distracting the pilots from scanning the instruments. Of those participants who noticed the failure one participant from each group cycled the gear on final, while by contrast one pilot per group did not cycle the gear at all and landed with the gear up.

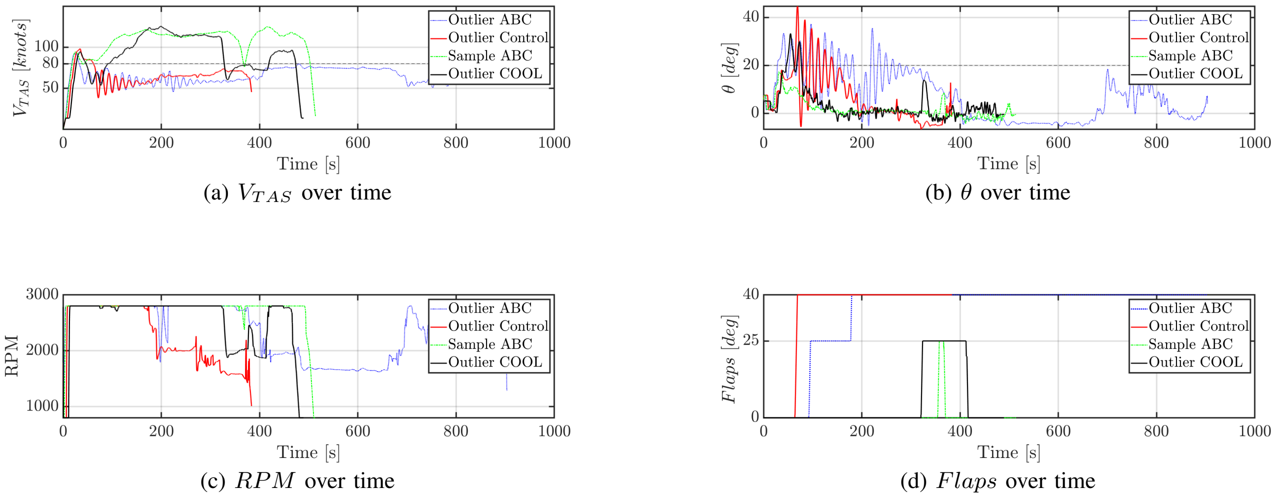
When coming to prioritising the execution of the go around over responding to ATC, six participants in total (four and two participants from the *ABC* and control groups respectively) delayed the response to the controller in at least one occasion, mainly reporting to stand by for the amount of fuel remaining or delaying/ disregarding the later landing clearance. The initial request asking for the number of POB or asking for a radio check were replied to on the spot as they did not require a thorough scanning of the instruments. Overall pilots found the ATC environment easy to deal with compared to the busy airspaces they usually fly into. Only one participant reported to be particularly stressed because could not spot the traffic at the beginning of the scenario, situation which however solved itself as soon as the conflicting traffic left the control zone. The full procedure was applied in this scenario by eight out of nine participants, while one participant applied only the aviate step.

TABLE VIII: Observation for the landing gear scenario.

	ABC	Control
Delayed response to ATC	4/9	2/6
Different circuit pattern	2/9	3/6
Does not cycle gear	1/9	1/6
Aviate	9/9	N/A
Breathe	8/9	N/A
Check	8/9	N/A

E. Further Analysis on Flying Performance

A scenario-based analysis was performed to assess possible differences in time spent in upset among the groups using a Kruskal-Wallis test. Neither the flap asymmetry scenario ($H(2)= 1.814,\ p = 0.404$) nor the mass shift featured a significant main effect of the training intervention on the time spent outside the envelope ($H(2)= 3.229, p = 0.199$).



(a) V_{TAS} over time

(b) θ over time

(c) RPM over time

(d) $Flaps$ over time

Fig. 6: Flight data from the outliers of the *ABC*, control and *COOL* groups recorded in relation to the mass shift scenario.

When focusing on the mass shift scenario, it is noticed that the outliers from the preliminary analysis are relatable to this specific test scenario. To gain further insight in possible behavioural differences among these participants, their flight performance is described in Figure 6 in terms of V_{TAS}, θ, propeller RPM values and flap selection choices. The mass shift scenario occurred soon after rotation, hence causing a sudden increase in pitch. It can be firstly observed how less than three minutes into the scenario both pilots from the *ABC* and control group decided to fully extend the flaps, maintaining such configuration until landing. The consequent ballooning effect caused the pitch attitude to oscillate around $20\,\text{deg}$ and the speed to drop and oscillate around a value 60 $knots$. The plateau which can be observed in pitch values between $400s$ and $700s$ when focusing on the outlier from the experimental *ABC* group, shows that by reducing power (hence decreasing RPM values) pitch attitude reduces with a consequent increase in speed to safer values. The same behaviour is shown at an earlier stage by the outlier from the control group.

A different approach was taken by another pilot from the experimental group (Sample ABC in Figure 6), whose data are plotted in Figure 6 against the one of the other participants and was hence taken as a sample reference of those pilots from the experimental group showing a safer performance. In this last case, the choice of the pilot to keep the flaps up after take off allows for the speed to slightly fluctuate above $80 knots$, despite the engines being spooled almost constantly to full power. Attempting to lower the flaps to $25\,\text{deg}$ caused a sharp drop in speed and local increase in pitch values, which the pilot promptly counteracted by retracting the flaps again. The outlier from the *COOL* group shows a similar decision making process by deploying flaps around a similar time frame but to a longer extent. The effect in terms of speed drop and pitch angle increase are however attenuated by a power reduction. From this performance comparison it can be observed how the outliers from the different groups show a similar decision making process in terms of flaps choice and power settings over the scenario.

At a group level the number of occurrences that the flight envelope was exceeded was also analysed as an overall indication of decision making taken by the pilots to recover from the upset. A Kruskal-Wallis test performed in relation to the mass shift scenario shows no significant differences among the groups (*H(2)= 4.931,p = 0.082*).

F. Subjective Ratings

A boxplot of the RSME and stress scores is shown per scenario in Figure 7, hence giving an overview of both measures at a scenario based level and over the test scenarios: from this figure the mass shift scenario seems to feature higher ratings for both measures compared with the other two test scenarios, which is in line with what observed in subsection IV-C. Furthermore at a scenario based level the experimental *ABC* group shows higher stress ratings compared to the control group in all scenarios.

A one way repeated ANOVA performed on the three test scenarios including the *ABC* and control groups shows a significant main effect of the test scenario on RSME scores *(F(2,22) = 8.363, p = 0.02,* η_p^2 *= 0.432)*. Following pairwise comparisons based on Bonferroni adjustmenst shows the mass shift scenario to significantly differ from the flap asymmetry *(p = 0.018, 95% C.I. = 3.063,33.604)* and the landing gear scenarios, respectively *(p = 0.034, 95% C.I. = 1.730,46.270)*.

Similarly, a significant main effect of the test scenario is found on stress levels using a Friedman test *(*$\chi^2(2)$ *= 14.476, p = 0.001)*. A following post hoc analysis conducted using

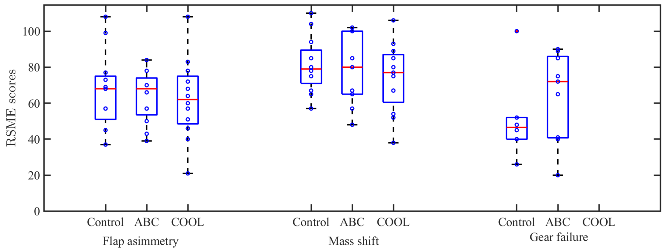
Wilcoxon signed-rank tests found the mass shift scenario to be significantly different from the flap asymmetry scenario *(Z = -3.580, $p < 0.01$)* and the landing gear scenario *(Z = -3.097, p = 0.02)*.

Furthermore, a one way independent ANOVA test assessing the effect of the training on mental effort was additionally performed on the flap asymmetry and mass shift scenarios separately. However, neither the former test, *(F(2,32) = 0.138, p = 0.871)* nor the latter *(F(2,32) = 0.441, p = 0.647)* lead to a significant difference among the *ABC*, control and *COOL* groups. An independent t test was performed to compare the *ABC* and control groups in relation to the landing gear scenario without highlighting any significant main effect *(t(13) = -0.930, p = 0.369)*. A non parametric analysis conducted using a Kruskal-Wallis test on the stress scores did not lead to any significant result in relation to the flap asymmetry *(H(2) = 1.151, p = 0.563)* and mass shift scenarios *(H(2) = 2.585 , p = 0.275)*. Similarly, no between group differences were found when conducting a Mann-Withney test comparing the *ABC* and control groups on the same dependent measure *(U = 21.5 , p = 0.506)* in relation to the landing gear scenario.

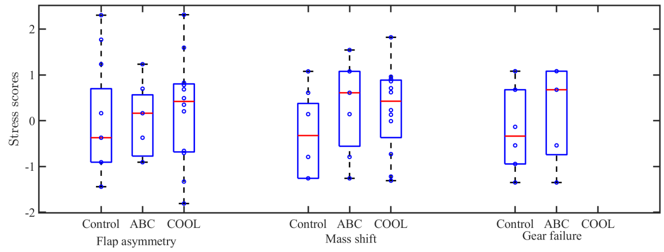
When considering startle levels over the test scenarios a significant main effect of the test scenario is found on this dependent measure *(F(2,22) = 36.739, $p < 0.01$, $\eta_p^2 = 0.770$)* when performing a one way repeated measure ANOVA, consistently reporting the mass shift scenario to significantly differ from the other two test scenarios. A significant main effect of the test scenario flown is also found when conducting Friedman test *($\chi^2(2)$ = 11.850, p = 0.003)* on the surprise scores. Following post hoc analysis based on Wilcoxon signed-rank test showed surprise levels in the mass shift scenario to significantly differ from the those related to the flap asymmetry scenario *(Z = -2.537, p = 0.01)* and gear failure scenario *(Z = -2.753, p = 0.006)*.

To obtain an insight in how useful the two procedures have been perceived, a Mann-Whitney test was performed for each test scenario. Results do not show any significant difference in ratings between the *ABC* and *COOL* groups in relation to the flap asymmetry *(Mann-Withney U = 54.0, p = 0.449)* and mass shift scenarios *(Mann-Withney U = 0.540 , p = 0.100)* respectively. When comparing the rated usefulness of the *ABC* procedure over the test scenario it is observed that the average ratings for this measure in relation to the mass shift scenario ($\mu = 7.17$) is higher than for the flap asymmetry ($\mu = 6.17$) and gear failure scenarios ($\mu = 5.66$). A Friedman test performed in relation to this measure does not show any significant difference among scenarios (χ^2 = *1.0 , p = 0.607*).

Compared to the *COOL* procedure which features an application rate of 91% and 83% for the flap asymmetry and mass shift scenario respectively, the *ABC* procedure has been applied to its full extent by all participants in the mass shift scenario and by 81% of the participants in the flap asymmetry scenario.



(a) RSME scores per group and scenario.



(b) Stress scores per group and scenario

Fig. 7: RSME and Stress scores per group and scenario.

V. Discussion

A. *Training Comparison*

Although most of the analysis results highlight non significant differences between the groups, some valuable insights about the trained *ABC* procedure still arise when comparing data at a scenario based level and from a within participants analysis.

Firstly, significant differences in mental effort, stress, startle and surprise provide evidence that the mass shift scenario is more challenging than the other test scenarios. What makes such scenario overall demanding is the combination of a sudden upset in combination with the cause of the upset itself not being identifiable, consequently calling for a troubleshooting process. The problem faced here is initially time critical as the failure happens soon after rotation, however as soon as the pilot manages to gain altitude, time becomes less of a critical factor.

Despite the non significant differences between the *ABC* and the control groups, remarks and feedback provided by the participants from the former group can be related to the higher stress ratings found for the *ABC* group in the preliminary analysis and in particular in relation to the mass shift scenario. In this occasion, the application of the *Breathe* step was in seen as a forced step and was additionally reported not to be useful by a pilot because of the high pressure situation as a consequence of not being able to fully control the aircraft. The same pilot later reported that if he had managed to better control the aircraft he would have also felt consequently more relaxed. It could hence be argued that successfully stabilising the flight path after the occurrence of an upset provides already stress relief and could also be seen as a preliminary step to the following formal stress management related steps. Furthermore, for some participants the procedure was seen as “another task to do”, hence increasing the workload instead of decreasing it.

When comparing the *ABC* and *COOL* groups, the non significant differences between groups reported from the preliminary analysis and from the following up scenario-based analysis provide support to the argument that the effectiveness of both of the exprimental trainings on alleviating mental workload and stress was comparable. It can furthermore be argued that the core of both methods laid in the application of the same stress management technique consequently explaining the non significant differences found in terms of the dependent variables here considered.

Based on the fact that all participants applied the full procedure in this scenario it cuold be argued that, compared to *COOL*, the *ABC* procedure could be successfully more easily implemented by pilots and it could be hence less time demanding to execute. Furthermore, one of the pilots from the experimental group reported to appreciate *ABC* more than the *ROC* (*Reset Observe Confirm*) learnt during his training as he felt the *ABC* procedure to be better defined and useful to breakdown his thinking even if he was successfully managing the problem. On the other hand, as remarked by another participant, the *ABC* procedure should fit within the execution flow of the other flight deck procedures. Hence the easiness with which such procedure can be implemented in practice is also dependent on the airline specific training framework.

B. Applicability of the ABC procedure

When encompassing the other two test scenarios in the discussion, the lower application rate of the procedure in response to the specific event leads to the consideration that the *ABC* procedure is somewhat less suitable. This is also in line with the procedure being, on the average, rated less useful for the flap asymmetry scenario and landing gear scenario compared to the mass shift scenario. Although startle and surprise events are context independent and hence startle and surprise management training should not be tailored towards a specific set of failures, it is worth here to consider the differences in scenarios to understand why the *ABC* procedure might not always be suitable or helpful. This further analysis is important also in light of the fact a participant reported that it was not always clear when to apply the procedure when an upset is not involved.

Compared to the mass shift, the flap asymmetry is a noticeable yet subtle upset (as there is no flap indication provided) which however did not let the pilots struggle too much to keep the aircraft under control, to the extent that some participants disregarded or were never fully aware of the issue. Therefore, some participants were not triggered to go through the procedure in all its steps or thought it was not necessary to do so. On the other hand, it was interesting to notice one participant to apply the procedure multiple times in response to the reduced visibility and after losing track of her position. In comparison with the flap asymmetry, the landing gear failure is clearly shown on the multifunction display although no upset follows from the failure itself. For this specific event, calling out explicitly the Aviate step felt unnecessary as the problem can easily be framed, the following up course of action is clear and the flight path is never destabilised by the failure itself. The go-around procedure which is instructed by ATC is also a routine procedure that pilots are used to and although ATC proactively tried to higher the workload during this phase, none of the pilots found the ATC environment challenging. The lower usefulness of the procedure as rated by the *ABC* group for this scenario could hence be relatable to the higher increase in stress and RSME from the familiarisation scenario compared to the control group.

Extrapolating these observations to a scenario independent level, one might conclude that those situations which are non nominal but which pilots are very familiar with or which are particularly time pressing and do not involve a clear upset not only are less likely to trigger the application of the *ABC* procedure but it also less likely for the procedure itself to be helpful when applied.

Overall *ABC* was commented by the participants to be a useful procedure in those situations when time is available. Furthermore, as suggested by one participant, *ABC* could be in general seen as a “toolbox” to refer to in case of necessity rather than a strict sequence of items to comply with. Due to the infinite scenario possibilities that pilots can face in real flight, it might not always be needed to apply all the steps of the procedure: as an example, calling out explicitly the aviate step when the flight path is clearly under control might be seen as a forced step. In other words the execution of the procedure could be more flexible, without yet changing the procedure itself.

C. Implications of the Aviate step within the ABC procedure

The presence of outliers in both the experimental and control groups in combination with higher time values from the *ABC* group could suggest that the implementation of the *ABC* procedure and in particular of the *Aviate* step could have an influence, although not significant, on the flying performance. However, when further analysing the flight data related to the marked outliers, it can be argued based on the results presented in subsection IV-E that the actions taken by these participants is comparable and relatable to an inappropriate choice of flaps an thrust settings causing the flying performance to deteriorate. Complementary to these results, the different decision making process followed by the sample pilot from the *ABC* group who refrained from selecting flaps suggests the difference in performance among these samples being based more on personal intuition and experience rather than on a tangible effect from the implementation of the procedure. Analysing the occurrences spent outside the envelope at a group and assessing non significant differences among the groups also gives an indication that the set of actions taken to cope was comparable among all participants, hence further supporting this reasoning. Within this perspective one could hence conclude that there is no evidence for the *Aviate* step to have supported or hindered pilots in coping with the failure.

D. Factors possibly influencing the results

To critically look at the results obtained from the analysis it is also important to consider the source of the data on which the data analysis has been performed, namely the participants. A first point of discussion relates to the companies that the participants work for. The majority of the pilots fly for the same company: during their recurrent training these pilots are regularly exposed to startle and surprise management training and were hence already familiar with stress management techniques before the experiment. Only two pilots flying for a different airline were new to the topic of startle and surprise management training. This uniformity in training background might have affected the significance of the results as participants from the control group might also have implicitly recurred to known stress management techniques.

Furthermore, although the training and familiarisation time was considered to be sufficient by the majority of the participants, three pilots from the experimental group reported that too little information was available about the aircraft and that more time should have been spent on familiarising with the aircraft itself, and they consequently felt there was too much of a gap between the familiarisation scenarios and those scenarios featuring failures. This unbalance could be related with the significant difference found in STAI scores between the *ABC* and control groups in the familiarisation scenario and consequently to the higher stress ratings found for the former group in the test scenarios.

Another factor influencing the performance of the participants and consequently the results of the experiment is the type of aircraft used: all but one pilot (former multi-engine piston instructor) were not used to fly multi-engine propeller aircraft anymore and found the model of the PA-34 really responsive compared to jet flying. Most of them hence argued that if they had flown an airliner (even though not necessarily their type) the performance would have been different. This point has not to be disregarded especially because all scenarios involved manual flying and were not concerned with autopilot related issues. Finally when comparing the results from the two experiments one has also to consider that the experiments themselves were independently conducted by two different researchers: as a consequence the way the briefing was delivered or other non-accountable differences might have played a role in affecting the results.

E. Recommendations for future work

To prevent uniformity in training background to possibly affect the results, it is recommended for future work to include participants from different airlines, possibly undertaking different training programmes, or to invite private pilots, as these are more likely not to have been exposed to startle and surprise management training.

Furthermore, to increase the likelihood of finding a statistical main effect of the type of training intervention, it is also advised to revise the test scenarios used for the experiment. The landing gear failure scenario was shown not to be challenging enough for the procedure to make a difference between the groups, hence suffering from a number of design flaws, which, had they been accounted beforehand, would have possibly changed the outcome of this scenario. First, the ATC environment experienced by the pilots was reported by all participants not to be particularly demanding. The inclusion of ATC for this type of experiment is a design option which can have a lot of potential for creating surprise and enhancing workload: for future purposes it is however recommended to increase the volume of traffic in the scenario by including more aircraft, some of which with a similar call sign, therefore possibly creating confusion. ATC could also be used as an artefact to interrupt the implementation of the training procedure as it could happen in real life or to deliberately deliver erroneous information. In addition, the other major issue with this scenario is related to the broad action space left to the pilots, some of which deviated from the expected flight path after informing ATC. Such risk could be prevented by better restricting pilots in the actions by, for example, limiting the type of request made to ATC.

It is also recommended to consider the implementation of a different aircraft model for future experiments if airline pilots are considered as participants: although it is important to develop a training intervention which is not type specific, flying a class of aircraft which pilots are not familiar with might increase the chance of introducing bias in the results. If pilots were to fly a multi-engine jet aircraft rather then a multi-engine piston for manual flying operations, the resulting performance would be expected to be different. As a final remark, the possibility to test the *ABC* procedure in multi-crew operations could also be considered in a follow up experiment.

VI. Conclusions

This research investigated the effectiveness of the *Aviate Breathe Check (ABC)* procedure on improving flying performance and reducing stress and mental effort in startle and surprise situations on the flight deck. The results from the experiment lack to show a significant main effect of the training intervention on flying performance overall, which is also shown when addressing a scenario based analysis. No evidence was found that the implementation of the *Aviate* step to explicitly call out and address the main priority of controlling the flight path can support or hinder pilots in the process of recovering from an upset.

When coming to assessing the increase in RSME and stress scores from the familiarisation to the post test scenarios, no significant differences were found between the experimental and control groups from the current experiment. However, such lack of significance could also be relatable to the fact that most participants fly for the same airline and are recurrently exposed to startle and surprise management training.

Furthermore, despite the non significant differences in terms of mental workload and stress between the ABC and COOL groups in the mass shift scenario, the higher application rate of the former trained procedure in the mass shift scenario suggests that ABC could be more easily implementable than COOL in events involving a sudden and marked upset. It is hence argued that such procedure could be integrated into recurrent pilot training programmes and implemented in operational practice without requiring extensive training. On the other hand, the ABC procedure itself was deemed not to be clearly implementable in off nominal situations not involving an upset and which are well familiar to pilots.

REFERENCES

[1] S. H. Van Middelaar, *Startle & Surprise Intervention Training Evaluating Training Effects on Pilot Performance in Off-Nominal Operations (MSc Thesis)*, 2018.

[2] J. A. Kochan, E. G. Breiter, and F. Jentsch, "SURPRISE AND UNEXPECTEDNESS IN FLYING: DATABASE REVIEWS AND ANALYSES."

[3] Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile, "Interim Report n°3 - On the accident on 1st June 2009to the Airbus A330-203registered F-GZCPoperated by Air Franceflight AF 447 Rio de Janeiro - Paris," Tech. Rep., 2009.

[4] Gobierno Bolivariano de Venezuela, "Expediente JIAAC-9-058-2005 i," Tech. Rep., 2005.

[5] National Transportation Safety Board, "Aircraft Accident Report - United Airlines, INC. McDonnel-Douglas, DC-8-61, N8082U," 1978.

[6] International Air Transport Association, "Loss of Control In-Flight Accident Analysis Report Edition 2019 Guidance Material and Best Practices," Tech. Rep., 2019.

[7] A. Talone and F. Jentsch, "Evaluating startle, surprise, and distraction: An analysis of aircraft incident and accident reports," 2015.

[8] J. Rivera, A. B. Talone, C. T. Boesser, F. Jentsch, and M. Yeh, "Startle and surprise on the flight deck: Similarities, differences, and prevalence," in *Proceedings of the Human Factors and Ergonomics Society*, vol. 2014-January. Human Factors an Ergonomics Society Inc., 2014, pp. 1047–1051.

[9] A. Landman, E. Groen, M. M. van Paassen, W. Bronkhorst, and M. Mulder, "Dealing With Unexpected Events on the Flight Deck: A Conceptual Model of Startle and Surprise," *Human Fcators*, pp. 1161–1172, 2017.

[10] J. Field, E. Boland, J. van Roij, J. Mohrmann, and J. Smeltink, "Startle Effect Management," Tech. Rep., 2018.

[11] S. M. Casner, R. W. Geven, and K. T. Williams, "The effectiveness of airline pilot training for abnormal events," *Human Factors*, vol. 55, no. 3, pp. 477–485, 6 2013.

[12] P. Van Oorschot, *Using Unpredictability and Variety in Pilot Training to Improve Performance in Surprise Situations (MSc Thesis)*, 2017.

[13] M. Gillen, *A Study Evaluating if Targeted Training for Startle Effect can Improve Pilot Reactions in Handling Unexpected Situations in a Flight Simulator (Phd Thesis)*, 2016.

[14] W. Martin, "Developing Startle and Surprise Training Interventions for Airline Training Programs."

[15] A. Landman, *Managing startle and surprise in the cockpit (Phd Thesis)*, 2019.

[16] A. Ghanbary Sartang, M. Ashnagar, E. Habibi, and S. Sadeghi, "Evaluation of Rating Scale Mental Effort (RSME) effectiveness for mental workload assessment in nurses," *Journal of Occupational Health and Epidemiology*, vol. 5, no. 4, pp. 211–217, 10 2016.

[17] R. Likert, *A Technique for the Measurements of Attitudes*, 1932.

[18] C. D. Spielberger, R. L. Gorsuch, R. Lushene, P. R. Vagg, and G. A. Jacobs, *Manual for the State-Trait Anxiety Inventory*. Consulting Psychologists Press, 1983.

[19] R. De Muynck and M. Hesse, "The a priori simulator software package of the piper pa34 seneca iii," 1990.

[20] J. van Leewen, *The Effectiveness of Simulating Active Problem Solving in Pilot Training to Deal with Automation Surprises (MSc Thesis)*, 2020.

[21] G. Norman, "Likert scales, levels of measurement and the "laws" of statistics," *Advances in Health Sciences Education*, vol. 15, no. 5, pp. 625–632, 12 2010.

Part II

Preliminary Report

Note: this report has been graded under AE4020

1
INTRODUCTION

The rapid technological advancements in the aviation industry throughout the past century not only led to performance achievements that one hundred years ago were not even considered plausible but also allowed a significant increase in safety standards.

Despite the continuous effort in improving aviation safety, unfortunately incidents and accidents still occur, and not rarely human factors are involved as contributors [13]. While it would be too simplistic to put all the blame on the pilots as aviation accidents are often the result of a concatenation of causes as it is well exemplified by the notorious Swiss cheese model [14], the systematic analysis performed by Belcastro et al. found inappropriate crew action/inaction to be a direct precursor of vehicles upset, which would eventually lead to Loss Of Control accidents, these contributors typically being worsened by weather disturbances and poor visibility [13].

Multiple Statistical analyses conducted within the span of decades have repeatedly confirmed Loss of Control In-flight (LOC-I) to be the leading category in terms of fatalities [15],[16],[17],[1]: Figure 1.1 shows that within a frame of 10 years, from 2008 to 2018, the LOC-I category, although not being the most frequent accident category, it is the one which recorded the highest number of fatalities, reaching a peak of 2462 fatalities.

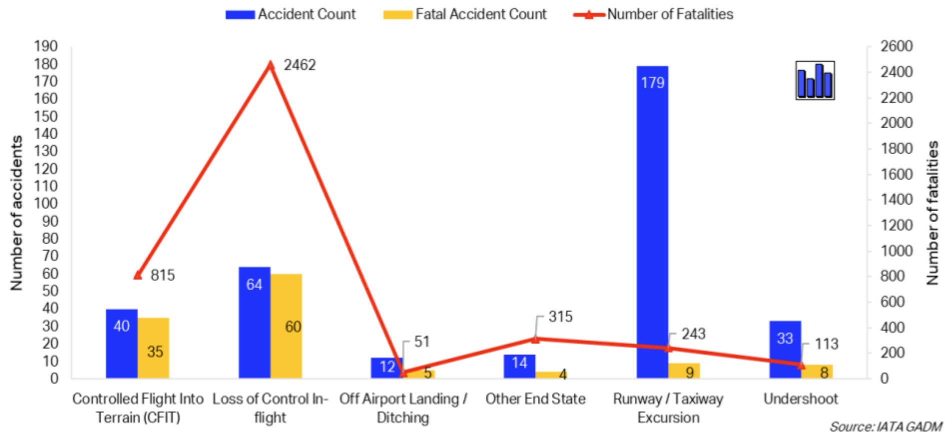


Figure 1.1: Number of accidents and fatalities as a function of accident category [1]

When coming to the contribution of human factors to this accident category, startle effect, surprise, distraction and lack of situation awareness are often cited along [1],[18]. It follows therefore the importance of understanding the role and features that startle and surprise have on the flight deck.

Following the rising need of improving startle and surprise management training within the often too predictable airline training programmes [19], research has been studying training interventions focused on using unpredictability and variety in pilot training to improve performance in surprise situations as well as at developing and testing dedicated mnemonic aid procedures [20], [9], [21], [22]. The current research focuses on the latter type of training intervention.

PROBLEM STATEMENT

Few mnemonic aids such as COOL (Calm Down, Observe, Outline, Lead), URP (Unload, Roll, Power), TAP (Time, Attitude, Power) and BAD (Breath, Analyse, Decide) have already been developed and their effectiveness in recovering from startle and surprise events has been investigated upon. Overall, previous research on the topic of startle and surprise management in aviation seem to agree on the fact that the implementation of these aids can be beneficial and has also found that pilots are in general willing to implement them in their operational routine [9], [21], [22].

In particular, the application of stress management related steps such as deep breathing and muscle relaxation was concluded to positively correlate with the task of collecting information, therefore enhancing situation awareness [21]. Focusing on stress management (Calm Down in COOL, Unload in URP and Breath in BAD) and on collecting information through the senses prior to explicitly diagnosing the problem are therefore considered the core and strength of these procedures. However, applying the procedure in situations where the flight path is not yet under control could defeat the purpose of a safe recovery and even worsen the situation.

As a matter of fact van Middelaar et. al. concluded that the specific COOL procedure might have a distraction effect causing the pilots to de-focus from the primary task of flying the aircraft. Furthermore, the procedure itself could be too extensive and time demanding to be applied when little time is available [9].

RESEARCH OBJECTIVE AND RESEARCH QUESTIONS

From the discussed considerations it follows the need of developing and testing a startle and surprise intervention training aiming at supporting pilots in addressing the right priorities (i.e. aviating first over troubleshooting and decision making) when dealing with startle and surprise as well as at improving the time efficiency in the execution of the procedure.

The following research objective is therefore defined for the current research:

To improve startle and surprise management on the flight deck by further developing and testing the COOL mnemonic aid procedure .

In relation to this objective the following main research question arises:

How can the COOL procedure be adapted and tested to improve task prioritization and workload management ?

For the purpose of this research task prioritisation relates to to the main goal of correctly applying the procedure in its steps while complying with the main priority of aviating the aircraft first.

The following research sub-questions are therefore to be addressed during the preliminary phase of this assignment:

- *How can the pilot be supported in adhering to the main priority of flying the aircraft before addressing troubleshooting related tasks when dealing with startle and surprise on the flight deck ?*
- *Which steps of the existing COOL procedure can be optimised to allow for a natural and efficient execution flow of the procedure itself?*

Once a new startle and surprise management procedure is developed, the current research aims at assessing its efficacy as a mean to recover from startle and surprise events. To this purpose the following additional research sub-questions will be addressed:

- *Under manual flying conditions, does the application of the new procedure support pilots in timely preventing aircraft upset conditions when experiencing a sudden startling disturbance ?*
- *Does the application of the new procedure allow to minimise pilots subjective workload when experiencing surprise in a high workload situation ?*
- *Is the application of the new procedure effective in supporting problem solving (on a secondary complex task) while not compromising the primary task of aviating the aircraft ?*

Report structure

The report is structured in three main chapters.
Chapter 2 aims at introducing a more general definition of startle and surprise, therefore discussing their main traits and influencing factors. Within the context of the conceptual model of startle and surprise developed by Landman et al. [6] the concept of frame and re framing is explained, and the influence of stress on human performance is discussed. Furthermore this chapter introduces the role and impact of startle and surprise in aviation, elaborates on the implications that these responses have on the flight deck activities and discusses a couple of major aircraft accidents which startle and surprise have been determined to be a contributor for. The chapter is finally concluded by digressing on startle and surprise management training in both civil and military contexts.

The existing startle and surprise management procedures are reviewed in Chapter 3, therefore allowing to summarise the main rationale behind such procedures. section 3.1 explains the role of mnemonic aid procedures in aviation, while the following sections address the rationale, training and assessment and the main findings from previous studies related to the TAP, URP, COOL and BAD procedures respectively. A qualitative comparison of the procedures follows. The chapter is then concluded by highlighting some key procedural goals and introducing the ABC procedure.

Finally Chapter 4, elaborates upon the experiment design, therefore defining the experiment goals and discussing the simulation scenarios that will be implemented during the experiment itself. Hypotheses and dependent variables for each scenario as well as a contingency plan to account for the risks related to the experiment execution are presented in section 4.4 and section 4.5 respectively. section 4.6, section 4.7, section 4.9 respectively define the number and required qualifications of the participants, elaborate on the apparatus which will be used during the experiment and discuss the experimental constraints and limitations.

2
Literature Review on Startle and Surprise

The following chapter aims at providing a more general and exhaustive review of startle and surprise. Section 2.1 discusses the characteristic traits of these two responses highlighting the differences between them and is followed by a brief summary of the factors influencing the severity of their responses. section 2.3 reviews the conceptual model relating startle and surprise developed by Landman et al. and digresses on the concept of frame and the importance of reframing. The influence of stress on human performance and reframing is finally discussed in section 2.4.

Lastly, the following chapter aims at at outlining startle and surprise characteristic responses specifically in relation to this field (section 2.5). Furthermore section 2.6 provides and overview of how startle and surprise management is being trained for in the civil as well as in the military context.

2.1. Distinctive traits of startle and surprise

Despite sharing some commonalities, startle and surprise are characterised by physiological and conceptual differences [3], and hence these terms shall not be used interchangeably. It is therefore intended in the following subsections to illustrate how these differ from each other.

2.1.1. Startle

The term startle refers to a complex combination of physiological, emotional and cognitive reactions to a sudden stimulus[23],[24]. The very first startle response has the function of directing the attention towards the source of the stimulus itself and it consists of a reflex in the form of instinctive muscle contractions (e.g. eye-blinking) such to rapidly prepare the body for possible adverse consequences [23]. It is triggered via the sensory thalamus by a rapid appraisal in the amygdala, with a latency period of less than 100 ms and its duration ranges from 0.3 to 1.5 s depending on the stimulus severity [23]. Meanwhile, the increase of physiological parameters such as heart rate, breathing frequency and blood pressure prepares the body for the fight or flight response [24]: this stress response can be regarded as an evolutionary mechanism which prepares us for either facing a potential threat or for avoiding it by fleeing.

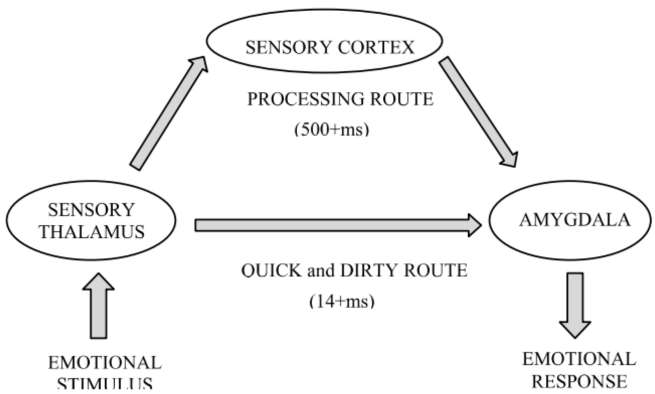


Figure 2.1: startle processing routes [2]

While the startle reflex is meant for a responsive reaction away from a potential threat, it doesn't allow for assessing whether the stimulus is actually threatening or not. Concurrently to this response, a slower process through the neocortex takes place (see Figure 2.1), therefore processing the perceived information to a deeper level and resulting into an assessment on whether immediate action is required [21].

Temporary motor and cognitive impairment could result following the initial startle reflex . As a matter of fact, motor performance can be disrupted for up to 10 s when performing a complex motor tasks, while also impacting information processing therefore influencing situation awareness and decision making abilities[23].

Cognitive and motor impairment from the startle response have been shown to exacerbate whenever the perceived stimulus is associated with a threat, therefore causing the initial startle response to further worsen in a fully developed stress reaction known as fear potentiated startle, in which the described startle effects are bigger in magnitude and duration [24].

2.1.2. Surprise

Surprise is a cognitive-emotional response to unexpected events resulting from a mismatch between personal expectations and the way the environment is perceived [21], [23] and, in contrast with startle, it can be triggered by both, the presence, and the absence of a stimulus.

It follows therefore that startle and surprise are independent responses since the occurrence of one of them doesn't imply the other. Still they could both be triggered by the same event. Consider for example the noise from a pistol shot: whenever the shot is unanticipated one can be startled and surprised at the same time. On the other hand, research has shown that foreknowing the exact time when the shot will take place can completely eliminate the element of surprise [23].

On the contrary, surprise can be very subtle and arise from a slow appraisal that what expected does not match reality: in this case, the absence of salient and sudden stimuli would not cause any startle response, while the element of surprise could still result in comparable effects.

As a matter of fact, similarly to the effects of startle, surprise can also result in physiological responses such as increase in blood pressure and heart rate as well as cognitive responses involving attention narrowing, with consequent impact on situational awareness and decision making [23].

On the other hand, the impairments resulting from surprise last generally longer than for startle, with the related duration being dependent on the magnitude of the mismatch [23]. The process of solving the surprise mismatch is most often referred to as reframing, it is potentially effort-full and, as such, it could result in further confusion and loss of situation awareness. Reframing will be further discussed in section 2.3.

2.2. FACTORS INFLUENCING SEVERITY OF STARTLE AND SURPRISE

The response severity to startle and surprise is influenced by a number of factors, including level of arousal, emotional influence and specific individual differences.

It is indeed assessed that the degree of influence of the emotional component characterising the specific startle and surprise event has also a direct influence on the response: as anticipated, if the perceived stimulus is considered threatening, a full stress response known as " fear potentiated startle" can develop. The subjective emotional memories related to a specific stimulus influence its assessment and hence the likelihood of the stimulus itself to be classified as a threat [21].

In addition, subjective differences do exist which make some individuals more prone to adversely react to startle stimuli: these group of people belong to the category of "hyperstartlers", compared to the " low reactors " category which instead refers to those individuals who show a less intense response to stimuli [21],[24]. Furthermore, both high and low level of arousal/stress can exacerbate the level of severity of the startle response specifically [21], [23], [3].

The degree of unexpectedness does also contribute to the severity of the surprise: as a matter of fact consequential effects of surprise are more marked with an increasing discrepancy between reality and expectations and hence the bigger the discrepancy, the bigger the surprise and its effects [23].

2.3. MODELLING STARTLE AND SURPRISE: FRAMING AND REFRAMING

While startle and surprise related causes and effects have been deeply investigated upon in the past, the current research managed to identify only few models attempting to define a conceptual relation between them. The core of the conceptual model proposed by Landman et al. [3], recalls to the perceptual cycle introduced by Meyer et al. [25] and consists in a continuous repetition of stimulus perception, appraisal and consequent action selection and execution. As introduced in section 2.1 and pictured with the left cycle of Figure 2.2, startle could result from the perception of a sudden intense stimulus, which could further cause a fully developed stress response whenever the stimulus itself is identified as a threat. As it will be discussed in section 2.4 stress and anxiety can affect the elements of the perceptual cycle and impact the reframing process following surprise.

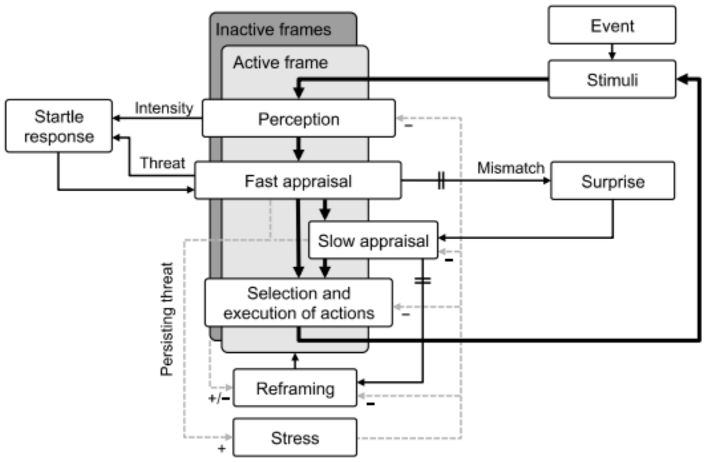


Figure 2.2: Conceptual model of startle and surprise [3]

In Figure 2.2 the two parallel vertical lines following perception and appraisal mark a discrepancy between the active frame the reality, therefore giving rise to surprise, which can eventually trigger the process of reframing. The so called "*frame*" is a key concept in the understanding of the cause-effect relationship that characterises startle and surprise and it is therefore important to spend few lines to introduce its definition and to address the related *reframing* process.

A frame can be considered as an explanatory mental structure, such as a story, a map or a plan which synthesizes the perceived information and draws a mental picture of the on-going situation. Frames are therefore an internal representation of the external reality , and, as such, they are subjective, experience dependent and specific to what is being perceived [26]. Furthermore the implementation of a specific frame is also goal oriented and it is influenced by the person's stance (e.g. workload, fatigue level and commitment to an activity).

According to Klein et al. [27], a specific frame is initially inferred from few salient data features referred to as "*anchorages*" or "*anchor points*", which do support its construction, while the frame itself filters, connects and hence accounts for the incoming data.

The process of matching data and frame most often occurs in a automatic and pre-conscious manner, based on pattern matching and recognition [27]. On the other hand, the active frame can be further elaborated upon and expanded based on what we learn form the environment, doing which could instead require a slower and more knowledge based appraisal [3],[27].

As long as no in-congruence exists between what was expected according to the established frame and what was perceived, the processes of perceiving and appraising data, and consequent decision making follow each other in a continous perceptual cycle [25],[3].
However, as soon as a surprising event is experienced, the current active frame is questioned upon. As Klein et al. explain [27], questioning a frame is an aspect of sense making which does imply the acknowledgement of a mismatch between data and frame, but not necessarily the understanding of the reasons behind this mismatch.

Solving the mismatch might simply require discarding inconsistent or unreliable data, therefore allowing to preserve the current frame: the same outcome could result from perceiving data incorrectly [3], a situation which can be addressed with minor efforts by paying a closer attention to the perceived stimuli.
On the other hand, whenever correct data highlights major inconsistencies leading consequently to a frame breaker, discarding the active frame might be inevitable: at this point it is therefore necessary to radically re-define the way the problem is understood by possibly reconsider the previously discarded data [27]. This process takes the name of "*reframing*" or "*frame switching*".

The temporary absence of a suitable frame within which information is processed or the implementation of an incorrect frame might compromise the ability of pilot of predicting future states, therefore enabling anticipatory action. As a consequence, his/her behaviour might become sequential and reactive rather than anticipatory and proactive [3]. It follows that reframing could become an effortfull process potentially involving knowledge-based behaviour [28].

In order to regain an understanding of the situation, a possible strategy is to compare different available frames against the mismatch, which might be particularly useful when a high degree of uncertainty is involved. It makes sense to argue that experience plays a significant role in solving the mismatch by frame comparison: the more experiences one has built in the past, the bigger the range of frames available for comparison. However, the over abundance of hypothesis that these knowledge based frames can offer could also cause indecision and therefore slow down the reframing process [29]. This is especially dangerous when a rapid frame switching is required, as a consequence of which it is instead valuable to be able to discard all the unnecessary frames such to be able to focus only on the main task.

Not being able to find a suitable replacement, most often people do have to seek for new anchors and construct a new frame[**?**] in order to re-enable sense making. Klein et al. suggest that the development of these new frames relies mainly on local cause-effect relationships rather than on comprenshive models (just in time mental models) [27].

When facing frame breaking, another option to consider is to give up on attempting to understand the reasons of the mismatch and prioritise new goals by adopting a well known frame [29].

2.4. THE INFLUENCE OF STRESS ON PERFORMANCE AND REFRAMING

As discussed, startle and surprise can result in acute stress, which, especially in the presence of threat-related stimuli, can have a detrimental effect on cognitive and motor performance [3]. It is therefore intended in this section to summarise all the major traits that characterise the effect of stress on human performance. While subsection 2.4.1 describes the relationship between stress and performance, subsection 2.4.2, elaborates upon the different classifications of stress. Section 2.4.3, subsection 2.4.4 and subsection 2.4.5 discuss the influence of stress on attention, motor performance and decision making respectively.The different strategies and coping mechanisms that are needed to tackle stress are finally addressed in subsection 2.4.6.

As a remark for the reader, in this section the terms "stress" and "anxiety" are used interchangeably. Strictly speaking conceptual differences exist between these two emotional responses, with stress being triggered by an external stressor, while anxiety resulting from persisting feelings of worry even without the presence of such stressor [30]. However, stress and anxiety show very similar symptoms, these being typified by fatigue, muscle pain, difficulties in decision making, anger etc. [31],[32]. Furthermore, they can also be addressed with similar coping mechanisms, which will be discussed about in subsection 2.4.6. For the purpose of this research, no specific distinction is therefore made between the two.

2.4.1. RELATION BETWEEN STRESS AND PERFORMANCE

What stress is and how it can be defined has been subject of numerous debates over the years. Traditionally two main models found consensus among the scientific community, namely the stimulus-based model and the response-based model [4].

The stimulus-based model relates to the engineering definition of mechanical stress and is based on the assumption that specific conditions influenced by exogenous variables are universally stressful (e.g. heat, noise, workload etc.) [4], without therefore taking into account individual differences. On the contrary, the response based approach argues that stress involves responses which can be triggered also by endogenous variables.

Besides these two, a third model known as transactional model was later developed, which sees stress as the interaction between the enviroment and the individual. From this approach the most widely accepted definition of stress is derived, the one proposed by Lazarus [33], based on which stress generates following the subjective assessment that the demands to cope with a specific situation are greater than the perceived capacity of coping with the situation itself [34].

If on one side such perceived discrepancy could motivate the research of additional resources [3], on the other hand the same rising anxiety could also negatively impact performance. As a matter of fact, while a degree of stress could be beneficial as it leads to an increased state of vigilance, therefore enhancing information processing abilities, excessive stress could instead lead to a state of hyper-vigilance , which, on the contary, could negatively impact such ability [35].

In literature a very well-known relationship between arousal and performance was theorised in 1908 by Yerkes and Dodson as an inverted U shaped curve [36]:

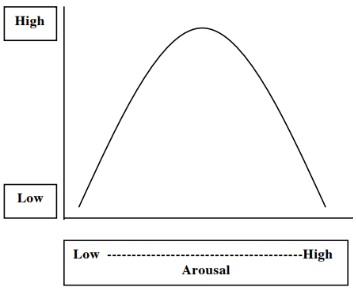


Figure 2.3: Yerkes - Dodson law relating arousal and performance [4]

As shown in Figure 2.3, the level of performance increases with arousal up to an optimum point, passing which a further increase in arousal level would result in a performance degradation. Where the optimum lies in this curve varies from a person to person basis: as a matter of fact stressors are cumultive and additive [8]. Each one of us has therefore a subjective limit to the capacity of handling stress which is influenced by personal physiological and psychological characteristics [8].

2.4.2. STRESS CATEGORIES

Different categorisations of stress can be addressed: in particular, in their extensive literature review Shahsavarani et al. [7] suggest that such categorisation could differ depending on the nature of the stressor, the influence of stress on the individual, and the time of exposure to the stressor itself. Table 2.1 summarises the different categories of stress as reported by Shahsavarani et al. A brief description is provided here below for each of the classes included in Table 2.1

Stress Category	**Classes within Stress Category**
Based on the nature of the stressor	Physiological Stress
	Psychological Stress
Based on influence of stress on the individual	Eustress
	Distress
Based on time of exposure to the stressor	Acute Stress (short term)
	Chronic Stress (long term)

Table 2.1: Stress Categories based on the literature review performed by Shahsavarani et al. [7]

PHYSIOLOGICAL AND PHYSICAL STRESS

A physiological stress response is naturally elicited by the confrontation with a situation which could cause physical harm, pain, discomfort, or simply a feeling of danger [37]. On the other hand psychological stressors involve facing situations where self esteem, personal reputation and success are at stake.

EUSTRESS AND DISTRESS

As introduced in subsection 2.4.1, stress does not always have a negative impact on individuals: on the contrary, the rise of stress could facilitate the employment of the extra resources needed for an effective response to challenging circumstances [3]. This beneficial type of stress can also be addressed to as eustress. Opposite to it, distress is instead generated when the perceived capacity to cope with a certain situation is lower than the perceived demands. To be noted is that the balance between eustress and distress is much dependent on one's self-efficacy, namely the self judgement of how a person can carry out a specific task [38]. Therefore, people having a low self-efficacy will perceive demands to be less eustresful than distressful.

ACUTE STRESS AND CHRONIC STRESS

Acute stress and chronic stress are often referred to as short term and long term stress respectively [7]. As a matter of fact, the former type arises following the impossibility to accomplish a time-bound task, which would consequently generate a physiological stress response as described in section 2.1. Acute stress is associated with the alarm and resistance phase of the General Adaptation Syndrome (GAS) formulated by Hans Selye [39], the first one being related to an immediate fight or flight reaction as a result of the stressor exposure, while the latter resulting from a continous exposure to such stressor over time, therefore failing to return to homeostasis [40].

On the contrary, chronic stress is associated with the exhaustion phase of GAS, namely when the individual has not managed to adapt to the prolonged stress exposure and has depleted the resources to keep coping with it [40]. Once reaching this stage, a person might experience tiredness, anxiety and depression and could run the risk of developing stress related health conditions [40]. Furthermore, from the literature review performed by Senanayake et al., it follows that chronic stress doesn't solely result from failing to return to homeostasis, but also from a repeated activation of a stressor or from a low and slow adaptation to the stressor itself [39].

2.4.3. INFLUENCE OF STRESS ON ATTENTION

In general, stress causes attention to focus on the main task, while fading peripheral cues . This phenomenon is often referred to as attention tunneling: clearly, depending on whether such cues are relevant for the accomplishment of the main task, performance is positively or negatively affected [4].

According to the attention theory developed by Eysenck et al.[41], anxiety is known to impair attentional control by reducing the influence of the goal directed attentional system and causing an imbalance towards the stimulus driven attentional system. The first type of attentional system is goal oriented, it is influenced by knowledge and expectations and it allows for a top-down control of attention, while the second one mentioned drives the attention to follow a bottom up process [41]. Therefore, when stressed, we tend to be more easily distracted by task-irrelevant stimuli and this tendency becomes especially marked when threat related stimuli are involved.

Such tendency has been proven experimentally by a number of experiments such as the one conducted by Ohman et al. [42], involving the detection of a discrepant picture (either a flower or a snake) within a matrix of pictures related to each other (mushrooms): it was concluded that subjects were faster in detecting the unrelated picture when the discrepancy was threat related (snake) rather than when the discrepancy did not provide any fear relevant stimuli (flower).

The mentioned imbalance results specifically from the impairment of two main functions of our working memory and particularly of the control executive, these being inhibition and shifting, which are respectively responsible for suppressing the interference from task irrelevant stimuli and repeatedly switching from one task to another. Depending on the level of task demand on the central executive and the impairment of these functions, the processing efficiency, namely the "*relationship between the effectiveness of performance and the effort of resources spent in task performance*"[41], is consequently affected.

Besides the influence on attentional control as described in the work of Eysenck et al. [41], Nieuwenhuys proposes an integrated model of anxiety and perceptual motor performance which also takes into account interpersonal and behavioural effects of anxiety, highlighting the fact that attentional control impairments are consequent to effects of anxiety on perception and action selection.

It is indeed argued that anxiety has an influence on how we do perceive the environment, causing the interpretation of a stimulus as a threat to be more likely. In turn, the perception of action possibilities results to be altered as compared to a non anxious state, which can further influence the selection of action possibilities itself [34]. It can therefore be argued that anxiety alters the current active frame.

The effects which are being reported here are well exemplified by an experiment conducted by Nieuwenhuys [43], and aimed at assessing the shooting behaviour of a group of police officers under pressure. The participants were tasked with the decision of shooting or not shooting to appearing suspects who could be armed and hence potentially represent a threat to the officers' life, or not armed and willing to surrender. Anxiety was induced by introducing a shoot back of small plastic bullets: it was assessed that the officers under the influence of this threat were more prone to quickly respond with shooting based on the expectation of being shot back rather than on the visual clue confirming whether or not the suspect was armed. The experiment shows therefore that under anxiety we might not be able to perceive all task relevant cues and, as a consequence our course of action is also impacted.

2.4.4. INFLUENCE OF STRESS ON PERCEPTION AND MOTOR PERFORMANCE

Perceptual motor performance has overall been shown to deteriorate under stress: numerous experiments lead to the conclusion that manual dexterity and in particular fine motor skills can especially deteriorate as a consequence of the exposure to stressors of different nature such as noise, temperature changes, fear, time pressure and workload [4].

Making approximations to cope with the demands resulting from a time bound task and hence accepting inaccuracies in the task execution (think about inaccuracies in manually flying an ILS approach due to overload) is an example of skill deterioration caused by stress: on the same line, omission of a particular action (e.g. lowering the landing gear during final approach) or incorrectly responding to a given stimulus are also performance detrimental effects of stress which could have similar or even worse consequences on safety [8].

By contrast, perceptual motor skills are more resilient to stress than high order cognitive processes [4].

2.4.5. INFLUENCE OF STRESS ON DECISION MAKING

From subsection 2.4.3 and subsection 2.4.4, it is concluded that stress alters perception and consequently the process of (re)framing and action selection [3]. It also follows that stress has an impact on decision making at individual and at team level as it can disrupt crew communication and coordination [44].

Overall, under stress decision making is more prone towards the implementation of non-compensatory and simpler strategies: in contrast to rational decision making which is characterised by weighing the pros and cons of each option at hand, quicker forms of decision making aim at finding the most feasible solution for a given situation and time constraints rather than defining an optimal one [8].

As a consequence, under stress fewer options and decision factors are considered and the decision process itself becomes more rigid [4]. While this adaptation in decision making allows to possibly better cope with the rising of acute stress, the biases and short cuts that our brain recurs to this purpose could also lead to decision errors.

In particular, few of the major information assessment biases that the Civil Aviation Authority lists in the flight-crew human factors handbook include the tendency of being influenced by the most recent information (recency), overlooking information which could explain evidence more reliably, relying on the most easily accessible information (availability) and inductive reasoning (accepting small samples) [8].

Similarly, well known mechanisms for shortening decision choices involve confirmation bias and anchoring to known experience retrieved frames when no reference frame can be easily established.

INFLUENCE OF TIME PRESSURE ON DECISION MAKING

Major implications on decision making follow from rising time stress: especially under time pressure, solutions need to be found which are workable, timely and cost-effective [45]. In particular, Klein stresses in his model of rapid decision making how recognitional decision making strategies are more appropriate under time pressure than analytical ones. Based on this model, decision-makers focus on a serial evaluation of options to find a suitable solution for the specific situation rather than comparing strengths and weaknesses to define an optimal one.

The rising time pressure resulting from imposed external or internal constraints inevitably leads one to become aware of the passage of time and of the consequent need for time management [46]. More formally, time perception is said in this case to occur under prospective conditions, as a result of which our attention is divided between the task of time estimation and other non temporal-related tasks. It follows that time estimation competes for attentional resources with the other on-going tasks from our working memory.

THE ROLE OF EXPERIENCE IN DECISION MAKING UNDER TIME PRESSURE

It can also be argued that most decisions taken by pilots in their daily duties seem to easily follow from past experience even in apparently difficult or complex situations [8]. Based on the findings reported by Klein, under time stress and ambiguity expert decision makers apply recognitional decision strategies to identify a suitable course of action at the first attempt, namely they rarely have to recur to a secondary solution. In other words, expert decision makers use the available time to assess the feasibility of a specific solution, projecting how the solution itself would be implemented and possibly re adapting it if needed. When time is limited, a solution is implemented which is the most likely (but not necessarily the most optimal) to be successful, based on the experience of the decision makers. In general recognitional decision strategies are considered more appropriate than analytical ones under time pressure and ambiguity: however, it is also concluded that such strategies could fail to result in a desirable outcome when the decision maker lacks the necessary experience to identify an effective course of action, when failing to figure out the pitfalls or when not optimising the chosen course of action as appropriate [45].

2.4.6. STRESS INTERVENTIONS

In literature two different approaches are recurrently discussed about to tackle and manage acute stress, namely stress inoculation training and stress management training.

The difference between these two approaches lies in the timing and purpose of the specific stress intervention: as a matter of fact while the former aims at making an individual more resilient to stress by acting a priori on the potential stressful condition, therefore psychologically preparing beforehand the individual to face stressful events, the latter supports stress management throughout the stressful event itself [4].

The key for a successful stress inoculation training is a realistic pre-exposure through training simulation, allowing consequently to build self confidence and improving personal sense of predictability and control [4]. The startle and surprise intervention training that this research is developing can therefore be considered as a form of stress inoculation training: by discussing the physiological and cognitive/emotional response

resulting from startle and surprise during ground training (the so called sensory phase as it has been addressed by Inzana et al. [47]), typifying the events that are likely to occur as a trigger and consequence of the startle and/or surprise (procedural phase) and by introducing and practicing in the simulator different coping strategies (instrumental phase), an effective recovery from such events is looked for, with a consequent enhancement on the safety of the flight.
The limitation of this type of intervention lies in the number of training sessions required to inoculate an individual: as a matter of fact Driskell et al. [48],[49], concluded that the number of sessions as well as the size of the training group is proportional to the size of the outcome. As it has been recommended from previous research [9], it is expected that recurrent startle and surprise training will help in enhancing the efficacy of the training intervention. Because such training involves the implementation of procedural steps, hence requiring cognitive effort, research has shown that retention will be shorter than for other physical related tasks, especially as the complexity of the training increases [50]. While it is difficult to estimate how often a specific training intervention should be trained to be effective in operational practice, it is here argued that such intervention should fit within the nominal scheduled recurrent trainings undertaken by pilots.

As mentioned, stress management techniques are typically implemented following the exposure to the stressor to alleviate its effect and to prevent the onset of further startles or surprises [4]. The flight-crew human factors handbook, recommends to take the following actions in order to start dealing with acute stress [8]:

- *Recognise the factors that are combining to cause acute stress.*
- *If stress is being produced by overload, pause to organise a list of priorities.*
- *Manage one's time: apportioning time to each item helps to develop a cycle of activity.*
- *When appropriate delegate duties and learn to off load.*
- *Learn how to mentally and physically relax. It may help to consciously relax one's muscles whenever feeling stressed or tense.*
- *Be positive and tackle responsibilities and problems as they occur. Avoid the tendency to put things off in the hope they will go away.*

Furthermore Table 2.2 summarises few main coping strategies that can lower the perceived task demand.

Coping Strategy	Purpose	Implementation
Action Coping	adjust/ change the stressful situation	removing the stressor
Cognitive Coping	reducing the emotional and physiological impact of stress	rationalisation and/or emotional and intellectual detachment from the situation
Relaxation Techniques	reduce anxiety and control tension	progressive muscle relaxation and use of mental imagery (e.g. meditation, autogenics)
Counselling Techniques	support cognitive coping and action coping by modifying the way the situation is perceived	professional counselling , receive personal support from friends or colleagues
System Direct Coping	removing symptoms of stress	physical exercise, others

Table 2.2: Coping strategies as reported in the flight-crew human factors handbook [8]

2.5. CHARACTERISTICS OF STARTLE AND SURPRISE ON THE FLIGHT DECK

As discussed, startle effect and surprise have contributed to the negative outcome of a number of aviation accidents. On the other hand, previous studies revealed that although surprise is actually a recurrent emotional response on the flight deck, it rarely leads to serious implications for the safety of a flight. Similarly, startling events are often consequence free, but compared to surprise they occur more sporadically [21], [18].

Nonetheless both startle and surprise or a combination of them have the potential to undermine flight safety as they could lead to cognitive and motor impairments, which, although temporary, could affect the ability of the crew to perform its in-flight duties.

The startle response is firstly typified by a reflex triggered by a sudden stimulus of mainly auditory and visual nature: when flying, the source of such stimulus can originate from within the cockpit for example in the form of a flashing warning light or an acute and sudden aural warning (e.g. stall warning), or from outside the cockpit (e.g. lightning or sudden appearance of another aircraft or obstacle within the field of view). Besides causing an involuntary and rapid muscle contraction therefore impacting motor skills, startle can also result in a temporary disorientation and brief confusion, which, in turn, can impair judgement and decision making abilities. While performing a motor or cognitive task, the direct consequence of startle is therefore the interruption of the ongoing task: as an example, such interruption can lead to disrupting the execution flow of a specific checklist or multi-step procedure, requiring therefore more time to resume the procedure itself [23]. In line with this example, impairment of decision making or problem solving skills could prevent pilots to identify and implement the correct checklist and course of action for the specific situation they are in.

As discussed in section 2.1, surprise does not necessarily follow from a stimulus but rather from the appraisal that the mental model set in place to make sense of the on going situation does not match what it is actually happening. It is therefore an emotional response which is conceptually independent from startle, even though they both could be triggered by the same event.

In literature, surprise is often explained by a mismatch between the current active frame or schema and reality [23]. Similarly to startle, such discrepancy results in the interruption of the task in progress, with the length of the interruption being function of the schema-discrepant event [23]. It follows therefore that the bigger the surprise the longer the task interruption and consequently the bigger the implications on the flight deck activities. The result of a study conducted by Kochan et al. to assess which types of situations pilots consider more surprising shows that overall unexpected events do not need to be unusual or rare to be truly surprising: on the contrary, most of the surprise related reports analysed throughout the study involved routine procedures [51]. As a matter of fact, the same study reports that surprise eliciting factors result from the aircraft state, ATC interactions, environmental conditions and incursions or in flight conflicts with other aircraft.

On the other hand, researchers from the field seem to agree on the fact that startle and surprise events are not associated to a defined set of causes, namely these responses can be triggered by potentially any event [21].

Nonetheless, on the flight deck most surprises are found to be elicited by automation [51]: these especially result from the pilots not having a sufficient insights in the advanced automation systems on board. As a consequence pilots might not fully understand such systems and could consequently take a course of action which is inappropriate to solve the issue they are facing.

The following cases exemplify the contribution of startle and/or surprise on recent aircraft accidents and they are indeed recurrently discussed in literature.

2.5.1. AIR FRANCE 447

One of the most discussed cases from recent flight history relating to automation surprise is the accident of the Air France Flight 447 while on a routine flight from Rio de Janeiro to Paris in 2009. A couple of hours after departing, the blockage of the Pitot tube due to the aircraft flying through an ice crystal environment (freezing rain) caused an inconsistency between the measured speeds, therefore making the airspeed readings unreliable. This malfunction further resulted in the disconnection of the autopilot and caused the switch from normal law to alternate law, as a consequence of which the aircraft had no longer an automatic stall protection [52].

Although both of the co-pilots correctly identified the loss of valid speed indications ("*...We haven't got a good display...*", " *We have lost the speeds ...* "), neither of them referred to the right procedure for the current situation.

Despite the continuous triggering of the stall warning, the pilots in the cockpit reframed the problem as overspeeding, even though such frame was inconsistent with the high nose attitude and the high rate of descent [5]. Throughout the last part of the flight the PF (sitting on the right seat), gave therefore mainly erratic and extreme pitch up and roll inputs (up to hitting the stop limits) leading the aircraft to reach angle of attack values of about 40 degress prior to the impact, as it can be seen from Figure 2.4.

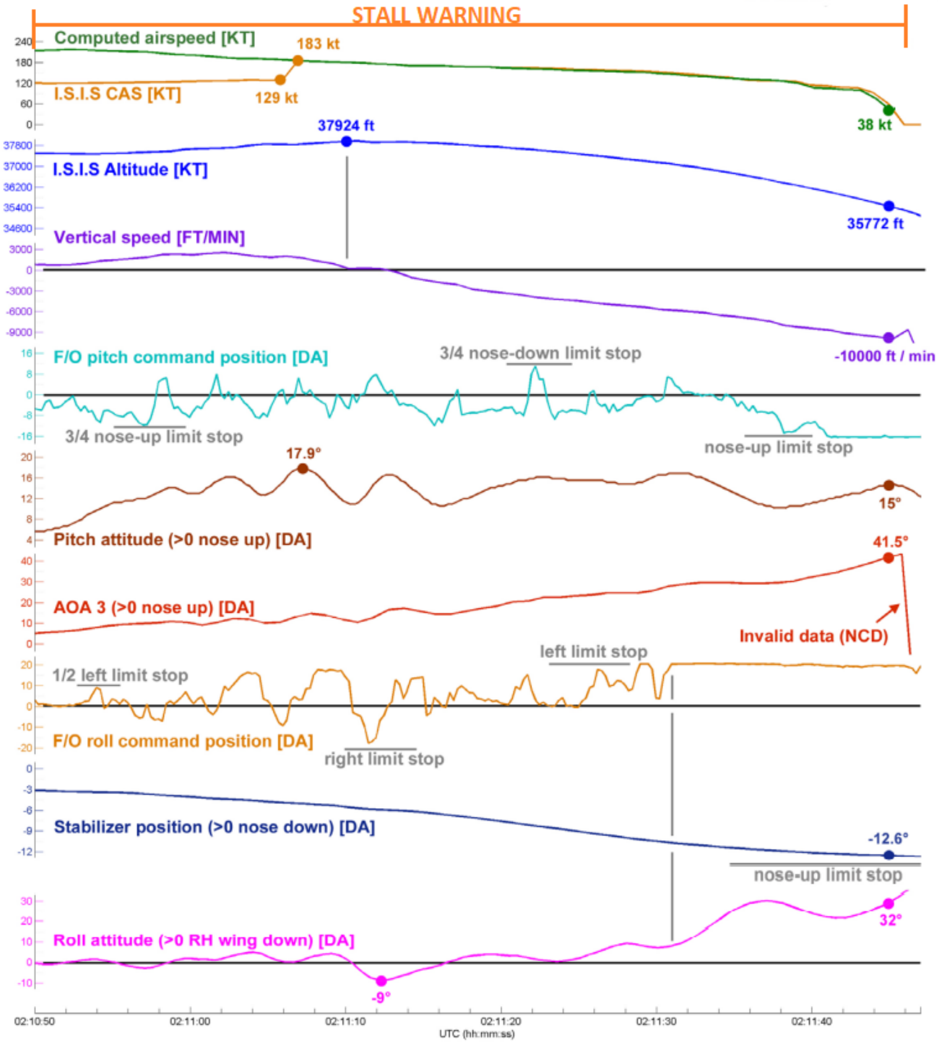


Figure 2.4: Flight data of the last minutes of the flight AF447 [5].

Despite being surprised by the ongoing event, the pilots were most likely also startled by the triggering of the stall warning (*"What is that ?"*), therefore impairing the initial decision making which eventually lead to the crash of the aircraft [22]. Furthermore, the absence of the Captain in the cockpit and hence the lack of a clear role division between pilot flying and pilot not flying contributed to enhance the degree of confusion in the cockpit ("*Is he coming or not ?*"), further exacerbating the situation [5].

2.5.2. WEST CARIBBEAN AIRLINES 708

The case of West Caribbean 708 represents also a clear example of frame mismatch followed by surprise .

On a routine flight from Panama City to Martinique International Airport the aircraft was unintentionally lead to a deep stall which the crew could not recover from. The captain erroneously mistook the engines drop in RPM caused by the activation of the anti-ice systems at too high altitude and the consequent loss of engine performance with an engine flame out, therefore reacting by increasing the wing angle of attack to increase the air flow to the rear mounted engines [53].

The action of the Captain exacerbated the situation as the speed slowly dropped to values within the so called "speed unstable" region of the power curve, namely where a decrease in speed results into an increase in required power, further causing a decrease in speed. Such speed drop went unnoticed and the aircraft entered a deep stall, recovering from which is particularly challenging for an aircraft featuring a T-tail configuration, such as the one of the MD-82 which the crew was flying.

The human factor analysis conducted by the Bolivian Authorities concluded that the attention of the Captain was channelled to the RPM indications, which caused loss of awareness of the other crucial cues that could have made the crew understand the real cause of the problem [53].

As it will be discussed in section 2.1, attention narrowing to the source of a stimulus is typical of a startle and surprise response: the Captain therefore based his problem solving and consequent decision making on a limited amount of information, which lead to implementing an incorrect active frame. The overall lack of Crew Resource Management and the inadequate decision making from the crew further contributed to exacerbate the situation to an unrecoverable extent [53].

2.6. TRAINING FOR STARTLE AND SURPRISE: CIVIL AND MILITARY APPROACH

To gain further insight on how startle and surprise can be trained for, civil and military approaches are in these section addressed. To this purpose Captain Thomas Jansen [1] responsible for the 777 and 787 type trainings at KLM and Colonel Miguel Saez Nievas [2], F-18 fighter pilot for the Spanish Air Force, have been briefly interviewed on the topic.

2.6.1. THE CIVIL APPROACH: TRAINING STARTLE AND SURPRISE AT KLM

KLM is currently evolving from legacy type of training to Evidence-Based Training (EBT), even though at the current stage ATQP (Alternative Training Qualification Program) is being implemented: this type of training focuses on the development of specific competences, the mastering of which would allow pilots to face potentially any in flight situation. This training philosophy opposes the standard scenario based training which instead is based on the recurrent training of specific scenarios. An often argued weakness of the latter is that simulator sessions might become predictable and lose training efficacy [19].

Especially when dealing with the unexpectedness featured by startle and surprise, EBT seems to offer a more robust approach than the one offered by a scenario based training: on the other hand, the successful implementation of EBT heavily relies on an effective analysis of training data collected throughout the simulator sessions and it might therefore take a few years before a fully developed EBT programme can be set into place.

KLM pilots undergo three simulator trainings and a formal check every year, each of which includes a dedicated section to startle and surprise: during the training sessions pilots are currently learnt to implement the URP (Unload Roll Power) recovery technique which has been designed and assessed during a project featuring the collaboration between KLM and KNLR [21]. A detailed explanation of what such procedure consists of and of this research in general will be provided in subsection 3.2.2. The task of designing the simulations to be implemented during the recurrent training and checks is yearly assigned to a committee of instructors who develop five type specific scenarios per aircraft type. The so developed yearly programme is then formally approved by the national aviation authorities.

At a general level, airline training programmes could benefit from taking into account pilot self reports of experienced startle and surprise. As Captain Jansen commented, often pilots are however too self complacent and would rarely admit they have been startled or surprised. Furthermore at the moment, no startle and surprise specific self reporting system is set into place at KLM.

[1] In person conversation on the 29th of October 2021
[2] Zoom meeting on the 1st of November 2021

2.6.2. THE MILITARY APPROACH TO TRAINING STARTLE AND SURPRISE

Contrary to civil airline training, flight training in the airforce features a more conservative approach, as simulator sessions are still designed mainly following the principles of scenario based training, even though elements of novelty are introduced throughout the sessions to prevent pilots from narrowing down their problem solving abilities to standard "pre-defined" solutions.

On the average a pilot flies 200 hours per year, of which up to 50 hours are allowed to be flown on a Full Flight Simulator. Of these hours 15 are fully dedicated to emergency training. Due to the wide range of operations that military flying involves, most of the training time is however spent on tactical training.

As a matter of fact, in military flying a distinction is well marked between training tactics, flight safety and operational safety: the former aims at enhancing the pilot ability to fly on the battlefield in conditions that might be different than the expected ones ("train how to play "), while the latter refers to all the relevant items related to the safety of the flight and is more task-specific (e.g. going through all the required checklists and procedures to successfully perform an ILS approach). Finally operational safety deals with anything which could compromise the safety of the mission (e.g. crossing the wrong airspace at the wrong time).

Before each mission, an operational risk assessment (formally known as Operational Risk Management - ORM) is performed to address all the possible risks that the specific mission could involve and a risk level is consequently defined: by doing so, the likelihood of encountering surprise is minimised. On the other hand, startle could still occurr despite the pre-flight preparation and cannot really be anticipated or prevented.

Similarly, a de-briefing is performed after each flight during which any issue experienced by the pilots is addressed. Interestingly, in the military there exist a strong feeling of team bonding: as often pilots lives are dependent on each other, learning from your own mistakes as well as from the mistakes of others is a pillar concept within the military training philosophy. Contrary to what happens in civil aviation, self reporting own experiences of startle and surprise and discussing them with the colleagues is therefore common and considered part of the responsibilities of a pilot.

2.7. SUMMARY AND CONCLUSIONS

This chapter introduced the concepts of startle and surprise, highlighting the differences between them: startle is always triggered by a sudden stimulus, while surprise is more subtle and arises from the appraisal of a mismatch between active frame and reality.

Despite being two conceptually different responses, these share however some physiological and cognitive characteristics such as increase in heart rate, breathing frequency, muscular tension and attention narrowing respectively, therefore impairing not only motor performance but also situation awareness and decision making. The severity of the response is influenced by a number of factors such as level of arousal, emotional influence, individual differences and level of unexpectedness of an event.

The relation between startle and surprise can be described through the conceptual model of startle and surprise as presented in Figure 2.2, within the context of which the definition of frame and the importance of reframing have been addressed.
Stress plays a major role in both startle and surprise responses: section 2.4 discusses how human performance and the reframing process itself are therefore affected. At an attentional system level, a shift from the goal directed attentional system to the stimulus driven attentional system takes place as a consequence of the inhibition and shifting function of our control executive being impaired.

Motor performance is also affected as manual dexterity can be significantly reduced. Furthermore action omission and inaccuracy acceptance becomes more common under stress. On the other hand, perceptual motor skills have found to be more resilient.

The influence of stress on decision making results into adopting non-compensatory and simpler strategies, aiming at finding a suitable solution for the specific situation rather than an optimal one following a rational action selection process.

Next, subsection 2.4.6 summarises the two major approaches to tackle acute stress, these being stress inoculation training and stress management training. While the first is more focused on stress prevention, therefore acting a priori on the potential stressful condition, stress management aims at providing a mean of recovery from the stressful event itself.

Finally, this chapter has introduced the concept of startle and surprise in relation to aviation. It was discussed how these responses are conceptually different from each other although they can both be triggered by the same event.

Research has shown that even though startle and surprise are frequently experienced by pilots on the flight deck without major consequences, these have the potential to seriously compromise the safety of the flight and have been found to contribute to numerous LOC-I accidents. Despite featuring different causes, the cases of AF447 and WCA708 both reveal how the cognitive ability to respond to well trained events can be impaired following startle and surprise.

Civil and military approaches to startle and surprise training have also been investigated upon: while the trend in airline training is to move towards Evidence Based Training, therefore focusing on training specific competences, the military approach seems to be more conservative and still related to scenario based training.

KLM in particular has being implementing the URP technique that was developed in 2018 during a project in collaboration with NLR [21]. Characteristic of the military environment is the sense of team belonging, especially due to the fact that often pilots lives are in the hands of each other. Consequently military personnel is more keen to share personal experiences of startle and surprise for others to benefit as well.

3
DEVELOPMENT OF A PROCEDURE TO MANAGE STARTLE AND SURPRISE

This chapter reviews the existing startle and surprise management techniques, with the purpose of defining the requirements and specific steps for further developing the *COOL* procedure.

Section 3.1 discusses the general rationale behind mnemonic aid procedures, after which the currently existing startle and surprise management techniques are presented in section 3.2 and the rationale behind the development of each is explained. The reviewed procedures are then compared against the theoretical principles discussed in Chapter 2, therefore allowing to outline a set of goals holding for the definition of a general a startle and surprise management technique. The *ABC* startle and surprise management technique is finally introduced in section 3.5 and each step of such technique is consequently discussed.

3.1. GENERAL ROLE OF MNEMONIC AID PROCEDURES

Those events in aviation that are most often trained for, are also those for which a well defined procedure exists [8]. On the other hand, pilots might find themselves to face unusual or novel complex situations for which no checklist or procedure seems to be readily implementable, and that might require them to undertake a rational decision making process, which often requires knowledge based behaviour.

Mnemonic aid procedures in the form of decision acronyms are of common use in aviation and provide therefore a mean to proceduralise such process by defining a logical sequence of steps to follow [8].
These aids are in particular deemed to be useful when enough time is available for the procedure to be fully applied. However, when time pressure plays a role or the solution to a problem is readily available through the existing procedures ,or, it is simply obvious, the implementation of such aids might be counterproductive. [8], [54].

In a study comparing different decision making aid procedures, Soll et al. [55] conclude that the core of most decision making acronyms involves a situation assessment step, risk assessment step and a step addressing selection among available options. Some of the addressed techniques (e.g. FOR-DEC) are furthermore defined in such a way that each step is related to a specific question: by doing so the concentration of the pilot is better focused on each and every phase of the decision making process.

3.2. REVIEW OF EXISTING STARTLE AND SURPRISE MANAGEMENT PROCEDURES

It is intended in this section to revise the existing startle and surprise mnemonic aid procedures with the purpose of highlighting and comparing the main requirements which such procedures are based upon. In particular, the specific steps involved in each of them are addressed, as well as the experimental context in which such procedures have been trained and tested.

3.2.1. TAP: TIME ATTITUDE POWER

In order to assess whether targeted training could improve decision making following startle and surprise, Gillen [22] conducted a study involving 40 crews operating in different U.S. airlines and flying on 7 different aircraft types.

RATIONALE BEHIND THE DEVELOPMENT OF THE PROCEDURE

The procedure proposed by Gillen aims at coping with the initial cognitive impairments resulting from startle effects. It focuses on time recognition and aircraft attitude control. The steps involved are:

- **Time** : the first element considered in the recovery process is time, leading the pilots to evaluate by means of altitude how much time they have at hand.

- **Attitude**: following the basic rule "Aviate", "Navigate", "Communicate", it is stressed how pilots should stabilise the aircraft before investigating upon any anomaly. Specifically it is instructed to keep a pitch between 3 and 5 degrees and to level the wings.

- **Power**: as required for the specific configuration.

The main rationale behind these steps is that, in case of a loss of flight instruments, the aircraft can be safely flown when knowing the correct pitch and power settings, or can at least be kept as stable as possible until it can be figured out what is going on.

TRAINING AND ASSESSEMENT

The study findings are based on both a questionnaire to pilots aimed at self assessing the pilot flying skills when experiencing startle as well as on performance data from a simulator assessement in a level D FAA approved full flight simulator. Such data consists in a collection of scores per participant in defined sub tasks for each simulation scenario. During training a briefing was provided to the participants on the correct power, pitch and bank angle that the pilots should attain in unusual attitudes by referring to the mnemonic previously developed.

Training and assessment addressed low altitude and high altitude scenarios. In particular the low altitude scenario assessment involved a malfunction of the landing gear during landing. The aircraft was furthermore simulated to be in low fuel conditions, therefore creating time pressure in the scenario. In resemblance of the AF447 flight accident the startle event in the high altitude scenario results instead from the loss of air speed data with the consequent disconnection of the autopilot. Furthermore, an engine aural warning was triggered shortly after this fault, causing additional distraction from the main task of flying the aircraft.
Prior to the training and assessment a questionnaire was handed to the pilots to assess their experience and the perception of their own flying skills following a startle event. This was done with the purpose of investigating possible correlation with the pilots performance in the simulator.

SUMMARY OF RESULTS

For both scenario types the crews undertaking the training achieved better performance than the FAA set standard in the performance criteria set for the specific scenarios (low and high altitude) did outstand the performance of those crews which were not trained for startle management. On the other hand, although no crew lost control of the aircraft, the untrained crews were found to be performing worse than such standards.

From the analysis of the factors characterising each event , problem identification was found to be a significant contributing factor to the success of the high altitude scenario. Similarly time management was also considered to be a significant factor in the low altitude scenario.

In general it was shown that startle training can significantly improve pilot's reaction to an unexpected event the consequence of which can in particular be mitigated by experience and proficiency in manually handling the aircraft.

3.2.2. URP: UNLOAD ROLL POWER - ROC: RESET OBSERVE CONFIRM

The Unload Roll Power technique (URP), also known as Reset Observe and Confirm (ROC) was developed within a study which involved the collaboration between NLR, EASA, and KLM [21]. In total 44 pilots from KLM took part in this study, of which 24 are type rated on the B747-400 and 20 are type rated on the B737-NextGen.

RATIONALE BEHIND THE DEVELOPMENT OF THE PROCEDURE

As for the TAP procedure, the focus of the URP technique is on recovery rather than on prevention. In particular, the goal of the training intervention is to address and inhibit the effects of the fight or flight response as well as supporting structured decision making process.

The URP procedure is the outcome of a strategy development based on three objectives. Firstly the strong emotions associated with startle and surprise events that could impair motor and cognitive performance should be addressed and controlled. Secondly, intuitive behaviour in response to the need of taking immediate action should be prevented, to avoid worsening the situation by the rising of new surprising events. Lastly, it is argued that pilots are more prone to accept a new training procedure if such procedure is comparable with the already existing practices that are current routine in their daily duties. The URP acronym reminds to the technique which pilots at KLM are trained with to recover from upset events: the same acronym which pilots are therefore familiar with from their training was hence adapted to respond to startle and surpise effects.

The specific steps involved in the URP procedure are as follows [21]:

- **Unload**: the "*Unload*" step focuses on addressing the physiological and phsychological effects of the startle and/or surprise response. To prevent attention narrowing on one cue it is firstly required to take physical distance from the unexpected event by sitting upright. Secondly, The physiological effects of startle such as increase in heart rate, breathing frequency, muscle tension are to be coped with by respectively deep breathing (inhaling via the nose and exhaling through the mouth) and consciously relaxing arms, legs and shoulders. This conscious way of relaxing reminds to the Jacobsen's relaxation technique, which is based on the assumption that relaxation of voluntary muscles can influence and reduce the activity in the central nervous system [56]. Finally, to ensure cognitive alignment with the other pilot, it is required to check upon him/her by calling his/her name, touching his/her shoulder and assessing his/her response.

- **Roll**: this step aims at (re)gaining situational awareness by focusing on what perceived through the senses, allowing therefore to mind the big picture rather than jumping into conclusions. Once a shared understanding of the situation is gained, the available time can be assessed as well as the possible options at hand can be defined.

- **Power**: with this step, pilots are encouraged to consider the effects of mitigating strategies as well as to critically reflect upon the possible errors committed. On the one side this step aims therefore at confirming that the situation has been correctly assessed and that no important information are still missing. On the other side, it is also intended to promote Threat and Error Management.

When coming to the application of the above mentioned steps it is mentioned that the procedure should be applied after guaranteeing a safe flight path and making sure that personal safety is not at imminent risk. Furthermore it is assumed that pilots are trained to face exceptional situations to a level that requires upset prevention and recovery.

TRAINING AND ASSESSMENT

The training intervention consisted in a 1.5 hours of classroom session followed by 1.5 hours of simulator training. In particular the classoroom session covered a brief back ground on startle and surprise and discusses few relevant accidents related to the topic. In this occasion the URP technique was introduced and practiced.

Simulator training was provided by training one element of the procedure at a time, therefore stopping the scenario after the completion of each step to allow for feedback from the instructor. Multiple short scenarios have been developed for training purposes based on the KLM's behavioural system such to cover different types of startling and surprising events.

Assessment was performed at two separate moments throughout the session, namely after the classroom training and after the simulator training: these were respectively dealing with an explosive decompression with consequent engine damage and a lightning strike during an automatic approach resulting in the malfunctioning of the mode control panel . Furthermore, a baseline assessment was performed for 20 pilots (uninformed group) at the start of the sessions itself: the startling element in this this case involved an eletronic failure during an approach causing the primary flight display and the navigation display to fail.

Physiological measurements (heart rate measurements, eyes tarcking) as well as on instructor observations were used to assess the performance of the pilots. The effects of the technique have been assessed by instructors based in the SHAPE behavioural markers [21].

Worth noticing is that the focus of the simulator training is on managing the response to startle and surprise events and hence on the application of the "*Unload*" and "*Roll*" step and not on the successful completion of the scenario following decision making (Power step).

SUMMARY OF RESULTS

The results from a questionnaire at the end of the session showed that overall the procedure was rated as useful by the participants who were able to include the different steps of the technique in their work flow. It was also concluded that pilots well managed to choose the timing when to apply the "*Unload*" step, even though remarkable difference was noticed between long haul and short haul pilots, with the last ones showing higher self-ratings. On the other hand some more experienced pilots showed more difficulty in choosing when to apply the technique and in the application of the technique itself.

However, performance improvement was not assessed. Instead one of the behavioural parameters measured, namely "collecting information", was observed to improve for those participant learning the procedure. About this result it was argued that cognitive performance as physical performance requires repetitive training before obtaining relevant improvements. Information collection was shown to positively correlate with the "*Unload*" step indicating that this step is beneficial to prevent cognitive tunneling and jumping to conclusions.

3.2.3. COOL: CALM DOWN OBSERVE OUTLINE LEAD

A study conducted by Landman et al. [3] aimed at assessing the effects of the COOL mnemonic procedure to manage startle and surprise on pilots abilities to aviate, to diagnose problems and to consequently take decisions. The study involved 24 line pilots, of which one half was assigned to the control group and the other half was assigned to experimental group.

RATIONALE BEHIND THE DEVELOPMENT OF THE PROCEDURE

Like the URP technique, also the COOL technique supports the pilots in dealing with the physiological and psychological aspects of startle and surprise by providing them with a structured way of managing stress and reframing.

The specific steps of the COOL procedure are as follows:

- **Calm Down**: comparably to the Unload step in the URP procedure, this step involves deep breathing, sitting upright and relaxing shoulders and hands, while becoming aware of the given control inputs. By doing so pilots are guided to take distance from the startling and/or surprising event therefore preventing to take immediate action.

- **Observe**: this step forces the pilot to read outloud the basic instrument readings, without yet formulating an hypothesis about the ongoing situation. By doing so the "big picture of the situation" is accounted for, therefore preventing to jump into conclusions.

- **Outline**: with this step, the pilot is explicitly requested to formulate a diagnosis of the situation such to make sense of what is going on.

- **Lead**: based on the assessment of the situation following the previous steps, the *"Lead"* step calls for a choice of a course of action, which has then to be executed.

TRAINING AND ASSESSMENT

The experiment was conducted within the SIMONA simulator using a model of the Piper Seneca PA-34 III. All the participants undertook a familiarisation flight where they had the opportunity to get acquainted with the characteristics of the aircraft. The participants were also subjected to a pretest and a post test such to be able to compare the performance of the control and experimental groups. The pretest scenario was characterised by a sudden engine failure before touch down while in the post test assessment pilots experienced four different scenarios.

The post test simulations included aircraft system malfunctions such as a flap malfunction and unreliable speed indications, as well as startling events resulting from the occurrence of external events, namely a bird strike during take off causing a spurious stall warning and a cargo shift in the cabin resulting in a sudden pitch up moment.

In all the simulations pilots were constrained to fly a circuit pattern, therefore allowing an easier performance comparison among the participants. Furthermore some of the scenarios featured elements of added complexity such as low visibility or flying in a different airport compared with training.

The training phase of the experiment consisted in a theoretical training on startle and surprise. For the experimental group only, a further briefing on the use of startle and surprise management checklists followed, during which the COOL procedure was introduced. In addition, this group received a further simulator training where the pilots could get acquainted with the implementation of the COOL procedure.
Pilot performance was assessed based on three different aspects, namely *Aviating*, *Problem Diagnosis* and *Decision Making*. Furthermore pilot ratings were needed to assess the usefulness of the COOL procedure.

SUMMARY OF RESULTS

From the experiment it was found that the experimental group performed significantly higher in decision making compared to the control group, leading to the conclusion that physical relaxation and the unbiased observation of the basic instruments and aircraft behaviour as prescribed by the *Calm Down* and *Observe* steps were beneficial in preventing cognitive tunneling and allowing an effective troubleshooting. As a matter of fact, these two steps were rated as the most useful by the participants, while the latter ones in the procedure were considered redundant, therefore unecessarily complicating the procedure.

Suggestions given by the participants on how to simplify the procedure involve calling flight instrument readings by their meaning instead of by their absolute value, to reduce the number of parameters to assess, and to leave out the *Outline* and *Lead* steps.

Another important result gathered from the experiment is that the experimental group performed significantly worse than the control group when considering the adherence to the criteria defined for the *Aviate* performance aspect: delayed pilot actions or inappropriate prioritisation of the procedure over ensuring a safe flight path suggests that the procedure itself might have a distraction effect.

3.2.4. BAD: BREATH ANALYSE DECIDE

The BAD procedure was designed by the pilot and human performance expert Martin Murray [1] [57] as a startle and surprise management technique, but its efficacy has not yet explicitly tested.

RATIONALE BEHIND PROCEDURE DEVELOPMENT

This technique was developed to provide a short mean of recovery which pilots should recur to following startle and surprise. The steps involved in the procedure are:

- **Breath**: the purpose of this step is to buy time following startle or surprise by structured breathing , such to allow the fight or flight response to dissipate. No specific breathing technique is however defined: since BAD has been designed as a short term technique, a more complete and effective management of the on-going stress would be best achievable by reframing and implementing other resources.

- **Analyse**: the purpose of this step is to lead our attention to the most critical pieces of information, such to regain situation awareness after startle and surprise. Therefore the main flight parameters such as speed, altitude and pitch are checked and called out loud.

- **Decide** : this step consists in defining the best course of action once situation awareness is regained.

[1] Conversation with the author via e-mail on the 21st of June 2021

Regarding the applicability of the procedure it is argued that, in order for the procedure itself to be effective as a mean of recovery from startle and surprise events, its implementation should be second to ensuring a safe flight path. In particular seven events call for recognised primed decisions and therefore for immediate actions [57]:

- EGPWS Warning
- Rejected Take Off
- Reactive Windshear
- Stall Warning
- Aircraft Upset
- Cabin Altitude
- TCAS RA

If the situation does not require an immediate response, then the BAD technique can be implemented.

TRAINING AND ASSESSMENT

Over several months of training in a B777 simulator, pilots were exposed to extensive testing on startle and surprise, were briefed on the BAD procedure and encouraged to use this technique when facing startle or surprise in the simulator. However, no specific experiment was run on the efficacy of this procedure, as such testing was limited to routine training sessions. Hence no specific conclusions can be drawn regarding the efficacy of the BAD technique.

3.3. PROCEDURES COMPARISON

For the procedures so far discussed Table 3.1 summarises the purpose of each of the related steps, therefore allowing to make an high level comparison among the procedures themselves.

Goal \ Procedure	TAP	URP	COOL	BAD
Aviate	Attitude, Power	-	-	-
Stress Management	-	Unload	Calm Down	Breath
Information Collection	-	Roll	Observe	Analyse
Situation Analysis	-	Roll	Outline	Analyse
Decision Making and Execution	-	Power	Lead	Decide

Table 3.1: Comparison of startle and surprise management procedures

When considering Table 3.1 it is firstly concluded that the general focus of the TAP procedure differs from the one of the other techniques: possible cognitive impairments are in this case faced with by providing the pilot with a mnemonic involving specific pitch and power basic instructions aimed at regaining attitude control of the aircraft. By contrast URP, COOL and BAD aim at coping with possible motor and cognitive impairments by focusing on stress management, regaining situational awareness and promoting decision making.

In particular it is observed that the COOL and URP techniques address stress management by implementing a specific breathing and muscle relaxation technique finalised at counteracting the physiological effect of startle and surprise, while, by contrast, the BAD technique includes only deep breathing as a stress management step. At the same time, the TAP procedure does not focus on stress management at all.

While including stress management in the procedure resulted in a positive impact on regaining situation awareness after experiencing startle and surprise, the question arises of what the benefit is of having a more extensive stress management technique as included in COOL and URP compared to the BAD procedure. It is argued however that, while for COOL and URP the stress management related steps have been considered a core element of the procedure by the participating pilots [58], with such element having a positive impact on the task of collecting information[21], deep breathing only as prescribed in the BAD technique might not result in an equally effective stress amelioration.

When coming to the complexity of the each procedure it is observed that BAD and TAP involve fewer steps and are therefore less elaborate than the URP and COOL techniques. It follows that the execution of the first ones is probably more time efficient than the latter ones.

The COOL procedure in particular appears to be the most extensive technique among the ones discussed: in contrast with the *"Roll"* step from URP , COOL makes a distinction between the *"Observe"* step and the *"Outline"* step, therefore allocating the task of collecting information and the task of giving the meaning to the situation to two explicitly different moments in time. If on one side, such distinction allows for a more refined discretization of tasks, therefore providing more explicit guidance in the application of the procedure, on the other side it also lengthens the procedure itself.

3.4. MAIN RATIONALE AND GOALS

Following up on the procedures review and comparison as discussed in the previous sections, it is now intended to discuss the main rationale behind startle and surprise management techniques and to define the main goals that a startle and surprise management procedure should address.

As introduced in section 2.5 startle and surprise events cannot be related to a limited set of causes, but they are rather context independent. From this consideration it follows the need of defining a high level procedure which pilots shall rely upon to deal with the vast majority of in flight startle and surprise events [21].

Following up on this point , emphasis during training should be spent on recovery rather than on prevention since the uncountable number of situations which could trigger a startle or a surprise would make difficult (if not impossible) to cover all possible scenarios during training. This is also one of the drivers at the basis of competency based training [59]. Behind this reason, it has also to be considered that the initial startle reflex and fight or flight response are instinctive evolutionary responses and, as such, are difficult to control or influence in a preventive manner. Similarly, it is also not feasible to assume that expectations will always be met: independently of how pilots are trained for preventing surprises on the flight deck, it is inevitable that their expectations won't be met at times[21].

Lastly, the procedure should promote desirable behaviour, this being generalised by staying calm, being communicative and thinking rationally and critically while keeping in mind the big picture of the situation [21]. To this purpose, the implementation of a stress management technique and a systematic assessment of the situation by checking and calling the main flight parameters could be beneficial.

In addition to the above mentioned considerations, which have also been mainly addressed by the procedures reviewed in section 3.2, another key point in the development of the procedure is related its applicability: one of the findings from previous research is that pilots sometimes do apply the technique while the primary concern should have been to aviate the aircraft [9]. It follows therefore the training need to support the pilots in choosing the right priorities for each event.

Based on this rationale, few key goals are identified :

- Supporting pilots in prioritising the correct sequence of tasks.

- Recognising and controlling the physiological and psychological reactions resulting from startle and surprise events.
- Preventing cognitive tunnelling and the instinctive implementation of an inappropriate frame which could result in further startle and surprise events (snowball effect).
- Supporting reframing following startle and/or surprise.
- Promoting and supporting decision making.
- Defining a mnemonic which is not counter intuitive and that follows a natural flow.

3.5. ABC: Aviate Breath Check

Based on the rationale and goals as defined in the previous section, a three step procedure is proposed based on the recommendations provided by Landman [60]:

- **Aviate**: as discussed, pilots shall always be concerned with flying the aircraft such to ensure a stable flight path. The actions taken when performing this step are situation specific, and could both imply manual control as well as the use of automation: it is therefore not intended to provide guidelines on how to aviate, but rather to set pilots in the mindset of taking the required actions to stabilise the aircraft to the best of their abilities. In particular the aim is to attain a steady horizontal flight path with wings level as much as the situation allows. By doing so, a reliable frame is established consequently allowing pilots to gain the necessary time to perform the rest of the procedure.

  Nonetheless, although the high level principle of Aviate, Navigate and Communicate clearly highlights the type of tasks that pilots should prioritise on the flight deck, one could still argue that the meaning of *"to aviate"* is too broad for practical puproses. It is therefore important to define more specifically what this high level task entails.

  It is commonly accepted as a general principle in aviation [2]that aviating encompasses two life critical activities, namely flying the aircraft and preventing the situation to deteriorate in case of urgency or distress. More in details the former primary task is further characterised by ensuring a safe flight path and correct energy management at all times during flight: if these two items are being taken care of, other critical duties can be addressed.

- **Breath**: this step is comparable with the *"Calm Down"* step from the COOL procedure and the *"Unload"* step from the URP procedure. Based on the positive feedback given by participants in previous research [9],[21], the implementation of a multi-step management technique is at this stage encouraged: sitting upright and consequently stretching and relaxing arms, legs and shoulders are the first actions to be taken. Deep inhaling, holding breath for 2 or 3 seconds and exhaling while focusing on the applied control forces are the next steps which naturally follow.

- **Check**: Similarly to the *"Observe"* and *"Outline"* steps in the COOL procedure, the *"Check"* step aims at firstly collecting all the necessarily information required to secondly outline the problem. This is done by focusing on all senses and by calling out loud what perceived. The primary (airspeed, attitude and altitude indications) and secondary (engines indications, flaps and gear settings) flight instruments can therefore be scanned through. Contrary to what done in the COOL technique the instruments are to be called by their meaning rather than by their numerical value (e.g. "*speed is good*", rather than "*speed is 105 knots*"): by doing so the workload of the pilots is expected to be reduced.

3.6. Summary and conclusions

This chapter aimed at reviewing the currently existing startle and surprise management procedures and at introducing the *ABC* technique, therefore discussing its steps and related rationale.

[2]personal coversation with Captain Thomas Jansen on the 29th of November 2021.

Overall it is concluded that mnemonic aid procedures supporting decision making are currently implemented in aviation in situations for which no clear procedure exists and time pressure is not an issue.

Furthermore the comparison of the existing startle and surprise management techniques lead to highlight the importance of stress management as an initial mean of recovery from startle and surprise. Together with an unbiased observation of the situation, the *"Calm Down"* and *"Observe"* steps and their related steps from the URP and BAD procedures have been deemed to be the most useful by the pilots participating in previous research.

Following the goals and princicples as drafted in section 3.4 a three step technique is defined based on the recommendations provided by Landman et al. [60]: the "*Aviate*" step is introduced not to provide the pilots with a specific set of actions to fly the aircraft, but rather to remind them that stabilizing the flight path and a proper energy management are always the key priorities to take care of before other issues are addressed. Secondly the "*Breath*" step corresponds to the "*Calm Down*" step from the COOL procedure while the "*Check*" is comparable with the "*Observe*", with the difference that the call out of the observations is performed by reading out loud their meaning rather than the related absolute value.

4
Experiment Design

To assess the benefit of implementing the ABC technique as a mean of recovery from startle and surprise, an experiment is designed aiming at evaluating pilot performance over a number of scenarios involving manual flying. Furthermore, with this experiment it is intended to highlight potential performance differences with respect to the results obtained from the previous research where the COOL technique has been tested.

This chapter is structured as follows: section 4.1 discusses the goals of the experiment. After listing the requirements driving a suitable scenario development process, the scenarios which will be flown during the experiment itself are introduced in section 4.3, and followed by a description of the related hypotheses and dependent measures in section 4.4. An overview of the experimental procedure, apparatus, and of the required qualifications of the participants involved is provided in section 4.8, section 4.7 and section 4.6 respectively. Lastly, constraints and limitations of the study are addressed in section 4.9.

4.1. Goals of the experiment

Given that the research questions defined in Chapter 1 address multiple research aspects on startle and surprise management on the flight deck, the experiment requires consequently to be designed for multiple goals.

A first goal of the experiment is to to assess whether the implementation of the ABC technique allows pilots to achieve a better performance in aviating the aircraft compared to the results obtained when testing the COOL procedure: as reported in subsection 3.2.3, pilots do sometimes tend to apply the procedure too quickly, therefore prioritising the application of the procedure itself over securing a safe flight path. By explicitly reminding the pilot to first take the necessary actions to aviate the aircraft after the occurence of a startling or surprising event, it is intended to evaluate possible improvements on specific performance criteria as it will be discussed in the next section.

Following up on the theme of performance comparison, another crucial aspect that is worth attention is time management in the implementation of the procedure: the ABC technique has been designed to be shorter and more efficient to execute compared to its former version and hence the question arises of whether workload management assessed in conditions of high workload is significantly improved when using the ABC technique versus when not implementing any startle and surprise management technique.

Finally, the experiment also aims at evaluating the potential benefits of using the ABC procedure to stimulate problem solving. With this respect the goal of the experiment is twofold: firstly it is intended to test whether the procedure supports the pilot in understanding the aircraft systems and the related malfucntions. Secondly, it is intended to assess whether the technique represents an obstacle to decision making when the pilot is subjected to time pressure conditions.

4.2. Main rationale and requirements behind the definition of the simulation scenarios

Making reference to the work of van Oorschot [20] the following requirements are established to guide the development and implementation of the experimental scenarios:

- **REQ1 -** Each scenario shall allow for performance comparison among participants:
  - **REQ1.1 -** Scenario shall be time bound.
  - **REQ1.2 -** Freedom in the scenario shall be limited by a discrete number of action possibilities.
- **REQ2 -** Positive outcome in the scenario shall be achievable.

- **REQ3 -** Scenario shall be realistic: extremely rare or complicated scenarios would possibly decrease the realism of the simulation and consequently make participants doubtful of the effectiveness of the procedure. The goal should be to test the pilots with challenging situations whose occurrence the pilots believe to be possible.

- **REQ4 -** Surprise in the scenario shall not be obvious to solve: in other words scenario shall allow for reframing to occur. Some situations are well trained, such as engine failures or control servo failure and hence they most likely won't result in an effortful reframing process for which the developed procedure could be useful.

- **REQ5 -** Training and test scenarios shall be unrelated to each other: as startle and surprise are context independent, training scenarios shall not influence the performance of the participants in the test scenarios.

- **REQ6 -** The set of training and test scenarios shall include scenarios with different priorities. This criterion is important especially in relation to the goal of assessing whether the training intervention is effective in supporting the pilot in prioritising aviating the aircraft over troubleshooting the problem.

- **REQ7 -** Scenarios shall start from a stable situation: in this way the pilots have the time to become acquainted with the situation before the startle/surprise is introduced.

- **REQ8 -** The specific scenario should be solvable by one pilot only.

In addition to the requirements mentioned so far, it is important to stress that, in order to successfully assess the benefits of the ABC technique as a mean of recovery from startle and surprise situations in the cockpit, significant startle and surprise have to be elicited.

This can be a challenging task as pilots are aware of flying in a simulator and hence no real threat exists. However, depending on whether startle or surprise or both have to be elicited, few considerations can be accounted for, such to make the scenarios as realistic as possible.

When the intention is to create surprise, the goal is to induce the pilot in a situation where his understanding of what is going on mismatches the real situation, therefore stimulating reframing [3]. The key aspect of such mismatch is that it should be preferably subtle, and it should not be easy to solve. Landman at al. [3] suggest that such an outcome can be achieved by making the pilot face novel situations different from the ones experienced during training or by explicitely introducing misinformation.

On the contrary, startle has to be salient, hence involving sudden perceptual stimuli, such as a loud bang or an unexpected aircraft motion [3].

Furthermore, in order to make the task of evoking physiological, behavioural and psychological responses associated with startle and surprise easier, it is important to reduce the so called "simulator mindset" and to provide a full immersive experience in the simulating scenario: this can be achieved not only by letting pilots face realistic tasks and distractions in the simulator, but also by stressing an overall attitude of professionalism, such for the pilots to behave as if they were flying in a real aircraft as much as possible. This could therefore include wearing harnesses, headset and even a uniform [61].

In addition to what previously mentioned, Cohen reports that one of the biggest challenges in reducing the simulator mindset is represented by providing realistic communications: in designing the test scenarios, this element shall therefore be taken into account.

4.3. SCENARIO DEVELOPMENT

To achieve the goals as defined in section 4.1, and based on the principles reported in section 4.2, four different test scenarios are considered for the experiment, each of which has been thought of to mainly address one of the previously mentioned goals.

4.3.1. TESTING FOR THE CORRECT APPLICATION OF THE PROCEDURE

In order to allow for a fair performance comparison between the application of the COOL procedure and the ABC technique, two former scenarios from the research conducted by Van Middelaar et al. [9] are considered for the current experiment, namely the mass shift scenario (MASS) and the flap asymmetry (FLAP) scenario. An overview of these is provided here below.

The MASS scenario involves a center of gravity shift upon rotation, this being caused by the sudden movement of a piece of cargo which has not properly been tied down in the cabin. The direct consequence of such shift is a sudden pitch up, which the pilot can recover from either by rolling away from level or by reducing thrust. Action has possibly to be taken before 20 degrees of pitch are exceeded, as exceeding this value would cause a significant reduction in controllability. Furthermore, the balloon effect caused by deployment of flaps could further worsen the situation and hence good airmanship would be demonstrated by landing with flaps up.

In the flap asymmetry scenario the pilot is tasked with flying a standard circuit in low visibility conditions and with 12 knots crosswind, see Figure 4.2. When selecting 25 degrees of flaps on downwind, the left flap remains up due to a malfunction, causing therefore a rolling and yawing motion, which is further diverging if 40 degrees of flaps are selected. Proper action would therefore consist in leaving the flaps at 25 degrees for landing or retracting them to the up position.

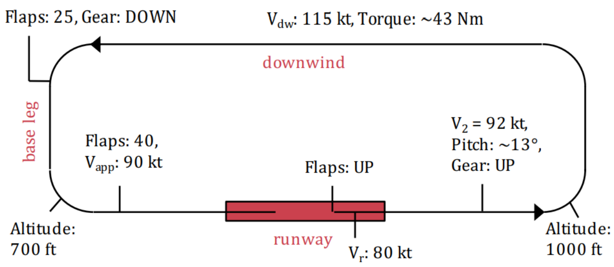


Figure 4.1: Circuit pattern flown in the flap asymmetry scenario [6]

4.3.2. TESTING FOR WORKLOAD MANAGEMENT

To test the impact of applying the procedure on workload management in a high workload situation, a scenario involving manual flying and a landing gear failure on final approach is proposed.
The mentioned scenario is based on the high altitude test scenario implemented by Gillen [22] as well as on two real aircraft accident cases, respectively United Airlines Flight 173 and Eastern Airline Flight 401, the story line of which is briefly summarised here below.

UNITED AIRLINES FLIGHT 173

Flight United Airlines 173, was a commercial flight scheduled on the 28th of December 1978, from New York to Portland, with an enroute stop in Denver [62].

The aircraft crashed as a result of fuel exhaustion, after the crew had spent too much time troubleshooting a landing gear problem and preparing for a possible emergency landing while flying a holding pattern for about an hour, therefore losing track of the current fuel state and fuel flow rate.

As a probable cause of the accident, the NTSB reports the lack of situational awareness from the Captain side to monitor the current fuel consumption, with the failure of the other two crew memebers to understand the criticality of the situation being an additional contributor.

EASTERN AIRLINES FLIGHT 401

A similar accident occurred few few years earlier, in 1972, when the crew of a Lockheed L-1011 on a routine flight from New York to Miami inadvertently disconnected the autopilot while trying to assess the correct deployment of the landing gear, which resulted into an uncontrolled and unnoticed descent to the ground [63].

After acknowledging a problem with the nose landing gear, which had failed to lock in the extended position, the Captain of the flight performed a missed approach: the ATC on duty instructed therefore to maintain the current altitude of 2000 ft and to initiate a left hand turn with the final goal of directing the aircraft on the downwind leg for landing. Approach and go around had been flown up to that point manually by the first officer, while the altitude hold mode function was engaged once on downwind in order for the crew to focus on the landing gear related problem. At this stage of the flight the Captain might have inadvertently hit the control wheel, therefore causing the disengagement of the autopilot and enabling the CWS to maintain a negative pitch which caused an 200ft/min descent to the ground that went unnoticed to the crew due to the dark night and the lack of ground lighting.

As in the previous case, the probable cause of the accident was the lack of situational awareness of the crew, which failed to monitor the flight instruments in the final minutes of the flight.

SCENARIO STORY LINE

Although the experiment involves single pilot simulations only, and hence team work and crew resource management is not applicable, the cases discussed above from literature feature several key elements that could be included in the design of this specific scenario.

- **The task**: manually flying the aircraft for landing is challenging, especially as landing is probably the most skill demanding phase of the flight, and hence high workload is involved.
- **The possibility for numerous distracting elements**: ATC communications, possible airspace congestions and close to minima weather conditions also involving startling events such as sudden lightnings are all stressors that can be included in the simulation to enhance the complexity of the scenario.
- **An interesting mechanical failure**: although a landing gear failure will probably result only in a mild surprise which can be identified and addressed by the crew in a short time frame, the occurrence of this failure in such a critical moment of the flight could result in a sudden workload increase, therefore causing additional strain on the pilot flying.
- **Fuel criticality**: the lack of awareness of the critically low fuel state caused the crash of the United Airlines Flight 173. By introducing low fuel levels in the simulation, time pressure is created, which could lead the participants into rushing for finding a quick solution to the situation they are in, hence increasing the chances of committing safety critical mistakes.

The above discussed elements can therefore be combined into one scenario as explained in the following story line (see figure Figure 2.1).

The scenario begins with the pilot flying an approach to the runway and he/she is tasked with landing the aircraft according to the SOP of the specific airport. The location chosen for the current scenario is Rotterdam The Hague Airport (EHRD).

Due to traffic congestion the pilot is instructed by ATC to fly an holding pattern at point MIKE for a couple of minutes. While performing the task a low fuel aural and visual warning is triggered, following which the pilot is cleared to descend and to land on the designated runway.

When lowering the landing gear on final approach, the right (or left) main landing gear fails to extend, therefore causing a yaw and roll motion due to the asymmetric pressure drag distribution, which has to be promptly corrected for by applying opposite rudder. As the gears are not properly extended for a safe landing, the pilot is forced to execute a go around and is instructed to fly a standard circuit pattern.

On downwind, the pilot is given the time to assess the situation, therefore retracting the landing gear and extending it again, this time with a positive outcome. Once the aircraft is properly configured for landing, the controller issues a new landing clearance. The scenario therefore ends when the pilot safely lands the aircraft on the runway.

Given this overview, the ABC technique is supposed to be applied at different stages of the scenario. Firstly when the landing gear fails to extend, the pilot is supposed to call for the "Aviate" step, therefore taking the necessary actions to compensate for the yaw motion, to execute a safe go around and to correctly join the downwind leg.

When flying downwind and with a safe flight path being ensured, the pilot can proceed with the execution of the "Breath" and "Check" step, after which he/she is supposed to cycle the landing gear and prepare the aircraft for a second landing attempt.

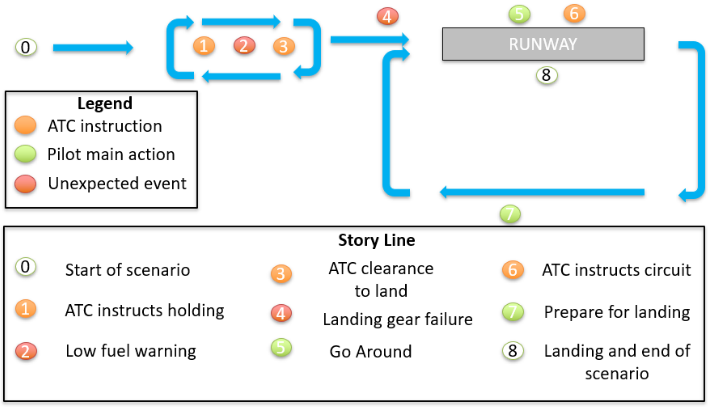


Figure 4.2: Sequence of events in the Landing Gear scenario

ATC COMMUNICATIONS IN THE SCENARIO

As mentioned in section 4.2, one of the key elements contributing to the realism of the simulation is the inclusion of ATC communications.

For the purpose of this specific scenario, such communications have also the goal of increasing the workload of the pilot, which can become especially heavy in the missed approach phase of the flight, when the pilot himself has to critically prioritise the task of flying the aircraft over responding to ATC.

Lastly, the requests from ATC to the pilot flying this scenario will purposely aim at inducing the pilot himself/herself into troubleshooting the technical issues encountered: by doing so it is meant to assess whether the participant adheres to the correct sequence of priorities (i.e. prioritising aviating over troubleshooting).

Given the relevance of the role of ATC for the current experiment, it is important to spend few lines describing the different types of communication occurences during this scenario.

A first remark is that the simulation will take place within the CTR of Rotterdam The Hague Airport (EHRD), and hence all radio communications are coordinated with the tower controller (TWR), who is responsible for managing the air traffic within the CTR itself. Furthermore, in Rotterdam the tower controller is also responsible for vehicles ground movements (aircraft and airport vehicles): in busier airports, such as Amsterdam Schipol, this task is instead delegated to a different controller (GROUND).

Figure 4.3 classifies radio communications in two different types, namely "routine" conversations between the controller and the air or ground traffic and "event" type of communications. The former refers to common and standard communications that characterise the daily routine of managing air traffic within the control zone. Communication of the latter type are instead related to non conventional situations, including non intentional communications, such as the case when a pilot erroneously presses the push to talk button, therefore transmitting non relevant information.

The second distinction which Figure 4.3 makes is between communications addressed to the pilot (which in this simulation is flying an aircraft with call sign PH-TUD) and communications from the controller to other traffic.

Routine ATC - PH TUD	Routine ATC - Other Traffic	Event ATC - PH TUD	Event ATC - Other Traffic
Holding instructions	Traffic crossing CTR	Report POB	Unintentional interruption from other pilot
Landing clearance	Taxi clearances	Request need of emergency services	Legend (Phase of Flight) HOLDING FINAL GO AROUND DOWNWIND BASE FINAL
Circuit instructions	Position report from a/c with similar call sign	Disturbances on frequency	
Expect late landing clearance	Report leaving frequency	Report remaining fuel	

Figure 4.3: ATC Occurences in the Landing Gear (LG) scenario

A detailed script of the ATC communications occuring in this scenario can be found in Appendix A.

4.3.3. TESTING FOR PROBLEM SOLVING AND COGNITION

The last experimental scenario aims at testing the efficacy of the procedure in relation to problem solving and decision making. Similarly to the scenario introduced in subsection 4.3.2 the following scenario is also meant to induce pilots into the troubleshooting loop, while possibly distracting them from the main task of flying the aircraft.

However, in contrast with the previous case, the current scenario features few major differences. Firstly and most importantly, the participants will experience a more subtle and complex surprise, forcing them to seek the related cause while being distracted from the primary task of ensuring a safe flight path and correct energy management.

To this purpose, a series of legs will be manually flown to an uncontrolled field with simulated low visibility and strong gusty wind conditions: the unfavourable weather has the potential to drift the aircraft off course if the flight path is not monitored carefully. During the flight, a spurious fuel warning is intermittently triggered at intervals of few minutes from each other. Furthermore one of the engines will be programmed to slowly decrease in RPM during the flight, therefore gradually increasing the thrust asymmetry and consequently causing a slow but steady bank which, if not corrected for, will cause a significant deviation from the flight path and in the worst case scenario it could end up in a spiral dive.

To be remarked is the fact that these technical issues are not related to each other: however, the idea is induce pilots in believing that there is a correlation between the two, therefore inducing some confusion and hopefully stimulating an effortful reframing process.

The specifics of the scenario can therefore be summarised as follows:

- **Task**: The primary task of the current scenario is to fly a few legs (inbound and outbound of the PAM (Pampus) VOR beacon) to Lelystadt (EHLE) airfield, which, for the purpose of this research is considered to be uncontrolled, where the pilot will fly a standard VFR circuit and land on runway 25. Figure 4.4 shows the planned route to the field. The simulation will be initialised few nautical miles north of point MIKE at 1000ft. A NOTAM will furthermore be provided to the pilot reporting that a TRA is activated at Almerestad from ground level up to 3000 ft due to an ongoing police operation.

- **Startle and Surprise factors**: Within a couple of minutes into the simulation the low fuel warning is triggered intermittently for few seconds at regular intervals of 1 minutes for the entire duration of the scenario. When such warning goes off for the first time, the RPM of the right engine will slowly start to decrease at a rate of 100 RPM drop per minute. These values will have eventually to be tuned when testing the scenario in the SIMONA simulator. In addition, few ATC communication can also be included with the purpose of increasing the pilot work by forcing him/her to look for other traffic in the vicinity.

- **Weather Conditions**: the weather will involve low visibility (about 4 Km) and easterly winds blowing at 20 - 25 knots with gusts of 28 knots. The ceiling will furthermore be limited at 1500 ft. Such conditions are therefore above VFR minima, but they are relatively challenging for VFR flying: especially the strong winds will make more likely for the pilot to be deviated off from the nominal course. Furthermore, as a consequence of flying VFR, the pilot is expected to focus more on outside visuals, therefore paying less attention to the ongoing situation within the cockpit.

Figure 4.4: VFR route to be flown in the Problem Solving Scenario

4.4. DEPENDENT VARIABLES AND HYPOTHESES

The performance of the participants can be assessed based on performance criteria which can specifically be defined for each of the scenarios described in section 4.3. For each scenario the related performance criteria are here addressed, and the related hypotheses are consequently introduced and explained .

4.4.1. MASS AND FLAP SCENARIOS: PERFORMANCE CRITERIA

The performance of the pilots in MASS and FLAP is assessed following the criteria already established in the previous research conducted by van Middelaar et al. [9]. In particular, performance assessment is based on three categories, namely aviating, problem diagnosis and decision making in relation to which the specific criteria per scenario are summarised in Table 4.1, Table 4.2 and Table 4.3 respectively [9].

Scenario	Action	Description
FLAP	Prevent excessive bank angle	After selecting flap 25, the pilot responds quickly enough to prevent the bank angle from exceeding 40 degrees.
FLAP	Maintained speed	After selecting flaps 25 on base leg, the pilot is vigilant enough not to let the speed drop below $V_{mca}(80knots)$.
MASS	Prevented excessive pitch angle	When the mass shift occurs, the pilot responds quickly enough not to let the pitch angle exceed 20 degrees.
MASS	Recovered quickly	When an excessive pitch angle occurs, the pilot responds quickly enough to bring the pitch angle back to below 20 degrees, within 10 seconds after the mass shift.

Table 4.1: Performance criteria defined for *"Aviate* as defined by van Middelaar et al. [9]

Scenario	Description
FLAP	Identified a flap asymmetry or malfunctions flaps
MASS	Identified a cargo or mass shift

Table 4.2: Performance criteria defined for *"Problem Diagnosis"* by van Middelaar et al. [9]

Scenario	Action	Description
FLAP	Refraining from selecting flaps 40	The pilot does not exacerbate the asymmetry and refrains from selecting flaps 40.
MASS	Configured early	Recognising that configuration changes may exacerbate the controllability issues, the pilot configures flaps and/or gear earlier and at higher altitude (before turning to base leg) or keeps flaps up.
MASS	Increased altitude	To increase the safety margin, the pilot flies at higher altitude. To limit advertent altitude increases, those who selected flaps in downwind are excluded as this is likely to cause an inadvertent altitude increase.
MASS	Selected flaps carefully	Recognising that the ballooning effect may again cause excessive pitch up, the pilot takes measures to prevent pitch from exceeding 20 degrees when selecting flaps. Those keeping flaps up are not included.
MASS	Increased final	To increase the safety margin, the pilot increases the time and distance at final, by turning to final at least 1500 m from the runway compared to the last familiarization pattern. This would require planning in downwind, and is therefore only applicable in MASS.

Table 4.3: Performance criteria defined for *"Decision Making"* as defined by van Middelaar et al. [9]

4.4.2. MASS SHIFT AND FLAP ASYMMETRY SCENARIO: PERFORMANCE CRITERIA

The following hypotheses are defined for the MASS and FLAP scenarios:

- Compared to the control group and the results obtained when testing the COOL procedure, the experimental group is expected to perform significantly better with respect to the performance criteria defined for *Aviate* (see Table 4.1): as a matter of fact, from the experiment conducted by van Middelaar et al. [9] it was found that pilots were delaying upset prevention and upset recovery related actions with a consequent impact on the flying performance. By explicitly prioritising these tasks through the sequence of steps "*Aviate*", "*Breath*", "*Check*" it is expected the experimental group to be more reactive to the experienced failures in the MASS and FLAP scenarios.

- The experiment is expected to lead to similar results for *"Problem Diagnosis*" and "*Decision Making*" as the ones obtained by van Middelaar et al. [9]: this hypothesis follows from the consideration that the "*Breath*" and *"Check"* steps are conceptually the same as the *"Calm Down"* and *"Observe"* steps from the *COOL* procedure, therefore leading the consideration that stress management and information collection related tasks would similarly influence the outcome in *Problem Diagnosis*. Similarly the *"Lead"* step from the *COOL* procedure was concluded to naturally follow from the previous steps of the same procedure [9] and hence no significant difference in performance is expected.

4.4.3. LANDING GEAR FAILURE SCENARIO: HYPOTHESES

This scenario has been designed to allow the measurement of different performance aspects. As discussed in section 4.1, workload management and aviating are the main points of interest of this scenario and several different parameters can therefore be considered to assess whether the implementation of the ABC technique has a significant impact on these performance aspects.

Two subjective workload measures are considered to sample the perceived workload by the participants.

- The Instantaneous Self-Assessment (ISA) requires to rate the perceived workload on a scale from 0 to 100. For its simplicity, this type of assessment can be used to measure workload at different stages through the scenario. In particular, three measurements are defined: a first assessment is taken when the pilot is flying the holding pattern and before the low fuel warning is triggered (point one on Figure 4.5)and serves as a baseline measure of the workload experienced by the participants before introducing any out of the ordinary event. The second workload measure (second purple marker in Figure 4.5) is instead asked for when climbing after the execution of the go around procedure , therefore aiming at assessing the subjective workload following the landing gear failure and the missed approach. The last assessment is finally performed at the end of downwind, when the aircraft has been configured for landing.

  The timing of these measurements is chosen such not to interfere with the story line of the scenario and, at the same time, to capture workload after the occurrence of significant events, when the different steps of the ABC technique are possibly meant to make the difference in terms of performance and subjective workload when compared to the control group.

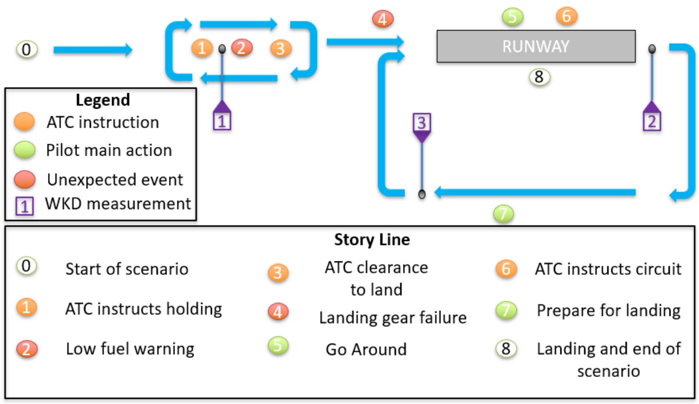


Figure 4.5: Timeline of the Workload measurements in the LG scenario

- A second subjective measurement is performed using the Rating Scale of Mental Effort at the end of the scenario, with the purpose of assessing the overall workload experienced by the participants throughout the scenario itself. Furthermore, the need of such additional scale arises from the attempt of balancing the disadvantages of Instantaneous Self-Assessment ratings. As a matter of fact, while ISA allows for quick ratings which can be efficiently performed "on the spot" during the simulation, therefore preventing the risk of our memory altering the initial workload perception in a post simulation assessment, this type of assessment doesn't take into account the fact that not only people have a different perception of workload, but they could also use different ranges of the same scale. By providing a verbal description of the ratings associated with the scale , the RMSE aims at tackling such disadvantage. Furthermore another advantage of such scale is that it is easier and less time consuming to implement when compared to the NASA-TLX scale but at the same time it has been proven to be equally effective in measuring mental workload [64].

While these measures are representative of the subjective workload of the participants, they do not allow to draw conclusions about their flight performance: it is therefore important to define few criteria which also make a performance comparison between experimental and control group possible. Such criteria are summarised in Table 4.4 and they have been adapted from the FAA-S-ACS-11 Airman Certification Standards [11] and make reference to the POH of the Piper Seneca [10].

Phase of the Flight	Performance Criterion / Required Skill
Holding	Report correct position and altitude
	Maintain the airspeed $\pm$ 10 knots, altitude $\pm$ 100 ft
Final Approach	Maintain approach speed of 90 + 5 knots
	Promptly react to landing gear failure by maintaining bearing $\pm$ 10 deg and bank angle $\pm$ 5 deg
	Avoid troubleshooting by checking circuit breakers and reporting non flight essential information when requested to ATC
Missed Approach / Go Around	Perform the following sequence of actions: - Pitch Up -Full Power -Flaps to TO position -Check VSI for positive rate -Flaps up
	Don't let the speed drop below 92 knots (best rate of climb speed)
	Disregard or delay the reply to non flight critical requests from ATC
Downwind	Maintain circuit altitude and a downwind speed of 100 $\pm$ 5 knots
	Assess circuit breakers, cycle landing gear at the end of Downwind
	Report downwind and respond to ATC emergency requests

Table 4.4: criteria for assessing the flight performance of participants in the Landing Gear scenario (LG) [10], [11]

A key aspect of the performance assessment relates on investigating whether differences exist between the experimental and control group in the ability of prioritising aviating over troubleshooting in conditions of high workload and when maintaining a stable and safe flight path is crucial for a safe landing. To draw meaningful conclusions from the experiment without obtaining ambiguous results which could be open to multiple interpretations, it is of utmost importance to clearly establish an Observable Behaviour which undoubtfuly indicate an appropriate application of the ABC technique and specifically of the *"Aviate"* step.

From the experimental point of view, a suitable strategy to achieve this goal is to induce pilots into troubleshooting in the most crucial phase of the scenario, namely after the failure of the landing gear and when a go around maneuver should be performed. To this purpose ATC requests or suggestions related to the issue being experienced are made to pilot, which, following the fundamental airmanship principle of ANC, are expected to be neglected or addressed at a later stage.

The use of checklists to respond to the specifc issue will also be made available to the pilot during the scenario: part of the checklist consists in steps involving an immediate response to the failure, these being the vital ones to guarantee a safe flight. Another part of the checklist involves troubleshooting and shall not therefore be performed until a stable flight path is ensured.

4.4.4. LANDING GEAR FAILURE SCENARIO: HYPOTHESES

Based on what discussed up to this point, the following hypotheses can be formulated concerning the flight performance and subjective workload of the experimental and control group:

- Experimental and control group are expected to experience the same subjective workload and not to show significant performance differences while flying the holding pattern: as mentioned, no relevant events happen at this stage of the scenario and being the procedure not yet applicable, there is no reason to assume that any difference in performance or in subjective workload could exist between the two groups.

- The experimental and control group are expected to show similar flight performance and similar subjective workload when compared to the control group during the missed approach and go around: the reason behind this hypothesis follows from the consideration that a landing gear failure is not expected to generate a remarkable surprise. As a matter of fact, since all the participants are well trained airline pilots, it is expected that all of them will recognise the problem and aviate accordingly, without significant differences in workload management. On the other hand this hypothesis also implies that the experimental group will not perform worse than the control group and will produce results within the standards defined in Table 4.4.

- The experimental group is expected to have a faster correct response to the landing gear failure when compared to the control group: specifically reminding pilots to aviate the aircraft is supposed to support pilots in preventing a delayed response in situations when ensuring a safe flight path and guaranteeing a correct energy management is of utmost importance for the safety of the flight and hence it is expected to result in more responsive actions from the experimental group when compared to the control group.

- The experimental group is expected to experience a significantly lower workload and significantly better flight performance on the second approach when compared to the control group: this hypothesis follows from the expected benefits of the "Breath" and "Check" steps: muscle relaxation and deep breathing followed by a methodical call out of the basic instrument readings should result in lower stress levels hence reducing the perceived workload with a consequent positive impact on performance.

- The experimental group is expected to experience an overall lower subjective workload when compared to the control group: once more, the reason lies in the supposed benefits that the pilot would gain by applying the procedure. A premise to this hypothesis is that the training previous to the assessment should be effective in stimulating a sense of self efficacy. That is, the training itself is expected to make pilots confident that the presented technique will support them in coping with startle and surprise effects.

4.4.5. Problem Solving related scenario: performance criteria

The performance criteria for this scenario aim at assessing the problem solving and decision making skills of the pilot on a complex secondary problem, as well as the pilot ability to follow the planned route.

Such performance criteria are therefore summarized in Table 4.5, for the categories of Problem Solving and Aviating respectively.

Performance Category	Performance Criterion \Dependent Variable
Problem Solving	Recognise and acknowledge that low fuel warning and reduction in RPM are not related to each other
	Total time taken to correctly diagnose the problem
	Pilot disregards spurious fuel warning
Aviating	Overall deviation from the nominal course in NM
	The pilot does not violate Schipol CTR, Schipol TMA nor the TRA located at Almerestad

Table 4.5: Performance criteria for the problem solving related scenario

4.4.6. PROBLEM SOLVING RELATED SCENARIO: HYPOTHESES

The following hypotheses are formulated based on the performance criteria previously defined:

- The overall deviation from the nominal route in nautical miles is expected to be significantly reduced for the experimental group than for the control group: this hypothesis follows from the key goal of supporting pilots in adhering to the right priorities that the ABC technique is based upon. The slowly increasing asymmetric thrust introduced in the scenario will not only lead to a potential upset condition, but it will also cause a significant deviation from the nominal route shown in Figure 4.4, with the consequent risk of violating the busy Schipol CTR airspace or the Almerestad TRA. Correctly applying the ABC steps will hopefully make sure that the participants will firstly leave the secondary low fuel warning problem on the background therefore guaranteeing a safe flight at all times.

- The experimental group is expected to take significantly less time to correctly diagnose the problem compared to the control group: this scenario introduces a more subtle surprise than the previous landing gear scenario, therefore possibly leading pilots to a more effortful re framing process. As discussed in section 3.2, applying a stress management technique and performing an unbiased assessment of the situation as prescribed by the *"Breath"* and *"Check"* steps respectively is supposed to support pilots in regaining situation awareness and establishing a suitable frame, therefore enhancing the chances of correctly diagnosing the problem.

4.5. SCENARIO CONTINGENCY PLAN

The newly developed scenarios discussed in the previous sections involve some variability, namely pilots could act differently from the expected story line as depicted in subsection 4.3.2 and subsection 4.3.3 : in order to minimise such variability and ensuring a consistent performance assessment among the participants, it is hence important to consider ahead the possible occurences that could cause a discrepancy with the planned sequence of events in these scenarios and define a contingency plan accordingly.

To this purpose a risk management approach is considered: firstly all possible risks are identified and causes and and effects of each are commented upon. Secondly, the importance of the specific risk is quantified in terms of likelihood of occurence and impact as shown in Table 4.7. Lastly, a mitigation measure is proposed and discussed for all the entries in Table 4.6.

Table 4.6 lists the possible risks and related causes and effects that the researcher could think of. It is important to remark that a project is never risk free, and although effort was invested in the risk identification process, during the experiment unexpected events could still occur that had not been thought of beforehand.

Risk Identifier	Scenario	Risk	Cause	Effect on Scenario	Effect on Experiment
A	Landing gear scenario	Landing gear is cycled during final approach	Too much time available on final	Safe landing on first approach	Subjective workload and performance will not be assessed on the remaining of the scenario
B	Landing gear scenario	Pilot forgets to extract landing gear on final	Distraction, others...	Unsafe Landing	Subjective workload and performance will not be assessed on the remaining of the scenario
C	Landing gear scenario	Pilot does not respond to WKLD mesurement	Distraction, others ...	[-]	Loss of WKLD measurement
D	Landing gear scenario	Pilot does not comply with ATC instructions	Distraction, pilot judgement, others...	Expected flight path is not followed	Performance assessment from participant not useful
E	Landing gear scenario	Pilot applies the full ABC procedure on final	Too much time available on final	[-]	Performance comparison becomes more difficult
F	Problem solving Scenario	Pilot does not notice the low fuel warning	Pilot is focused on the outside visuals	[-]	Surprise is not elicited
G	Problem solving Scenario	Landmarks not visible	Poor simulation layout	expected route is not followed	performance cannot be assessed

Table 4.6: Risk Identification Table

For each risk listed in Table 4.6, likelihood of occurence and impact are estimated in Table 4.7 on a scale from 1 to 4 , with 1 representing remote likelihood and negligible impact, and 4 indicating high likelihood and catastrophic impact respectively. The assigned likelihood and impact scores are briefly commented upon below.

Likelihood and Impact	Remote (1)	Unlikely (2)	Possible (3)	Likely (4)
Catastrophic (4)		D,A		
Critical (3)		F	E,B	
Marginal (2)		C,G		
Negligible (1)				

Table 4.7: Risk Assessment Matrix (RAM)

- **A** : After a brief consultation with a former military pilot from the Vliegclub Rotterdam, it is concluded that a pilot would normally not cycle the landing gear during final approach. In fact the standard procedure is to prepare the aircraft for landing when intercepting the glide slope, therefore lowering the gear and deploying flaps as required to set the aircraft in a stable configuration, with a rate of descent of approximately $500\frac{ft}{min}$. Formally a "$500ft$" call would follow to assess that the landing gear is extended and locked and that the aircraft is ready for a possible go-around. It seems therefore unlikely that pilots would spend precious time in fixing a landing gear problem during final approach rather than performing a go-around and evaluating the problem later on. Nonetheless, the occurrence of such event would have a catastrophic impact on the experiment as it would defeat the purpose of assessing the efficacy of the procedure within this scenario.

- **B** Despite the training, it has happened in past experiments that pilots do forget to lower the landing gear during the final approach: although unlikely, such event would have an important impact on the outcome of the scenario, as not only it would result in an unsafe landing, but it would also prevent the landing gear failure to happen, therefore compromising the key surprise that this scenario features...

- **C**: Similarly to the previous risk, it is unlikely but still possible that a pilot fails to respond to one of the workload measures, which would result in loosing a data point. The impact on the experiment is however only marginal, as this loss is compensated for by the numerous other data points collected from the other participants.

- **D**: Not complying with ATC instructions would be either result from distraction or from pilot judgement based on the consideration that compliance with the given instructions would compromise the safety of the flight. Within the experimental environment, such possibility is quite unlikely, but could have a critical impact on performance assessment in case the given instructions imply directives on the flight path to be followed.

- **E**: Once more, the possibility of the pilot applying the full ABC technique on the first approach attempt implies that the pilot has enough time to both stabilise the aircraft and assesses what the problem with the landing gear entails. This would possibly compromise performance comparison with those participants who delay the execution of the *"Breath"* and *"Check"* step to the downwind leg, as originally scripted.

- **F**: throughout the last scenario it could happen that, because the pilot is too focused on outside visual references, the intermittent triggering of the stall warning goes unnoticed. Even though the likelihood of this risk is remote as the low fuel warning indicator light is well visible when triggered, the consequence is remarkable as no surprise nor startle will be elicited.

- **G**: The last scenario requires the pilot to fly few legs based on visual references. If such references are not visible due to a possible lack of fidelity and/or quality in the simulation, the planned route cannot be flown, therefore defeating the purpose of the scenario.

From Table 4.7, risks **E** and **B** result to be the most important ones to take care of, followed in decreasing total score by risks **D**, **A** and risk **C** with a score of 4. For each risk, a mitigation measure is proposed in Table 4.8. The mitigation measures reported aims either at decreasing the likelihood of occurrence or at mitigating the impact of the specific risk. These strategies are briefly commented upon:

Risk Identifier	**Risk**	**Mitigate Likelihood**	**Mitigate Impact**
A	LG is cycled during final approach	Delay LG light, delay extension of LG, increase pilot workload	[-]
B	Pilot forgets to extract LG	[-]	ATC can report observation that LG is not extended
C	Pilot does not respond to workload measurement	No mitigation	No mitigation
D	Pilot does not comply with ATC instructions	No mitigation	No mitigation
E	Pilot applies the full ABC procedure on final	Increase pilot workload	[-]
F	Pilot does not notice the low fuel warning	[-]	ATC explicitly requests to report remaining fuel on board
G	Landmarks not visible	Re-plan the flight route	[-]

Table 4.8: Risk Mitigation Table

- **A**,**E**: In order to mitigate the likelihood of this risk, few strategies could be considered: for instance, the extension time of the landing gear could be increased, such for the full extension of the gear to take $10-15s$ more than the nominal deployment time. Similarly, the landing gear position lights in the cockpit could be programmed to lit few seconds later than they would normally do. However, it is also possible that these artefacts that are only meant to take time during final approach could mislead the pilot and give anticipatory cues on the imminent landing gear failure. A less artificial way of preventing the pilot to cycle the landing gear could instead involve further increasing workload: in particular, an additional unexpected event can be introduced, such as the failure of the Glide Slope antenna (assuming an IFR ILS appraoch is being performed). By doing so, the pilot is not only forced to spend additional focus on the approach charts to maintain the correct speed and rate of descent to follow the correct glide slope, but the standard operating procedures would also force him/her to execute the go around at a higher altitude than in a normal situation with a functioning glide slope antenna. The same strategy could be useful to mitigate the likelihood of risk **F**.

- **B**: In the unlucky event that the pilot forgets to extend the landing gear during the first final approach, the experimenter playing the role of ATC could intervene and report to the pilot the observation that " the landing gear has not been extended ".

- **C**, **D**: as it can be seen from Table 4.8, no mitigation is considered for these risks. Concerning risk **C**, the loss of one or few workload measurements is mitigated by the collection of numerous other data points from the other participants. When considering risk **D**, it is instead difficult to forecast what the non-compliance with ATC instructions will be and hence taking appropriate mitigating measures.

- **F**: in the unlikely case that the low fuel warning is not noticed, the intervention from the experimenter is required, who, by playing the role of the controller on duty will draw the attention of the pilot to the fuel indicators within the cockpit by explicitly requesting to report about the remaining fuel onboard.

- **G**: the impact of this risk can be mitigated by testing the scenario at the simulator and assessing therefore if each checkpoint along the route is clearly visible.

4.6. PARTICIPANTS

To allow for a meaningful comparison with the experiment carried out by van Middeldaar et al. [9], the experiment will be performed with the voluntary participation of about 20 airline pilots.

Furthermore, the involvement of pilots holding an ATPL rather than a PPL follows from multiple additional considerations. Firstly, a group of only airline pilots is more likely to feature homogeneity of experience compared to private pilots. In addition, the former have also undergone a training process which is more strict and standardised compared to the average training that an holder of a private pilot license goes through: this is because throughout their training private pilots might be followed by multiple instructors who can have very different backgrounds and might consequently have different training philosophies based on their experience.

Lastly, airline pilots have experience on multi-engine piston aircraft, which is not necessarily the case for private pilots, who most commonly are in possession of only a SEP rating. As all the scenarios will be flown on a small twin-prop, it follows that airline pilots are more suitable for the current experiment than private pilots.

4.7. APPARATUS

The experiment will be run on the SIMONA Reseach Simulator of the TU Delft Aerospace engineering faculty, which, for the purpose of this research, will implement a model of a Piper Seneca PA-34 developed by Muync and Hesse [65]. The following subsections will give brief introduction to the simulator itself and to the Piper Seneca PA-34 III aircraft.

4.7.1. SIMONA

The SIMONA research simulator is a full flight simulator (FFS) located at the Aerospace Engineering Faculty in TU Delft. Its systems are comparable to a level D flight simulator, even though the standards and specifications set by EASA do not allow for such qualification as the simulator is often reconfigured for research purposes, these mainly involving human in the loop experiments and validation of aerospace and automotive related concepts.

The motion base of the simulator features 6 hydraulic actuators which can therefore guarantee six degrees of freedom (see Figure 4.6a), allowing to stimulate the pilot vestibular system and providing the required inertial motion perception for performing manual and supervisory control tasks.

As regards the visual system, the pilot can experience a 180 x 40 field of view over a collimation mirror placed in front of the cockpit. Such collimation system involving three high resolution projectors allow for the image to be projected to infinity, therefore correcting for differences in image perception resulting from the displacement between the pilot and co-pilot stations. This set up enhances the realism of the in-flight simulation experience: on the other hand however, distortions in perception arise when flying close to objects since in such case the foreground objects would not anymore be placed at infinite distance from the observer. This could have an impact on correctly judging the height above the runway for example, which is especially important in all the scenarios involved in the experiment.

Another relevant factor impacting the simulator experience is the layout of the cockpit, this being not aircraft specific and resembling the one of a general airliner. It follows therefore that, although the flight dynamics have been modelled after the Piper Seneca PA-34, the participants in the experiment will have to get acquainted with a different internal layout than the one of the real system. Figure 4.6b shows the interior of the SIMONA flight simulator.

(a) External view of SIMONA Research Simulator

(b) Internal layout of SIMONA Research Simulator

Figure 4.6: SIMONA Research Simulator[1]

4.7.2. PIPER SENECA

The Piper Seneca PA-34 (see Figure 4.7) is a light twin engine aircraft produced by Piper Aircraft Company, firstly entering production in 1971 and developing in five different variants over the following thirty years [66]. The non linear flight dynamic model of the Piper Seneca II that is implemented in the SIMONA research simulator was developed by Muync and Hesse [65]. This model is considered particularly suitable for the experiment since it has been already used in previous research. Furthermore, the Piper Seneca is a rather simple aircraft which doesn't require a type rating and hence no special training is needed, therefore simplifying the training procedure of the subjects in the simulator.

Figure 4.7: Piper Seneca PA-34 [2]

[1] pictures retrieved from: `https://cs.lr.tudelft.nl/simona/facility/visual-display-system/` and `https://www.airtn.eu/downloads/airtn---simona.pdf`

[2] picture retrieved from: `https://cutteraviation.com/to-fix/piper-seneca-v-pa-34-220t/`

Table 4.9 reports few performance parameters that are of basic importance when flying the aircraft [12]:

Parameter	Speed [$KIAS$]	Description
V_{NE}	195	Never Exceed Speed
V_{NO}	163	Maximum Structural Cruising Speed
V_{Ma}	136	Design Maneuvering Speed
V_{FE}	107	Maximum Flaps Extended Speed
V_{LE}	129	Maximum Gear Extended Speed
V_{LO}	129	Maximum Landing Gear Extending Speed
V_{LO}	107	Maximum Landing Gear Retracting Speed
V_{Mc}	66	Air Minimum Control Speed
V_y	89	Best Rate of Climb Speed
V_x	79	Best Angle of Climb Speed
$V_{downwind}$	98	Speed on Downwind
V_{final}	83	Speed on Final
V_S (at MTOW = 2073Kg)	63	Stall Speed in Clean Configuration
V_S (at MTOW = 2073Kg)	60	Stall Speed in Landing Configuration (40 deg)

Table 4.9: Performance Speed Table of the Piper Seneca PA-34-II [12]

4.8. PROCEDURE

Figure 4.8 shows the procedure that each participant will follow throughout the experiment. For each activity, the related expected duration is reported in-between brackets. The total estimated duration of each experimental session is between 2.5 and 3 hours.

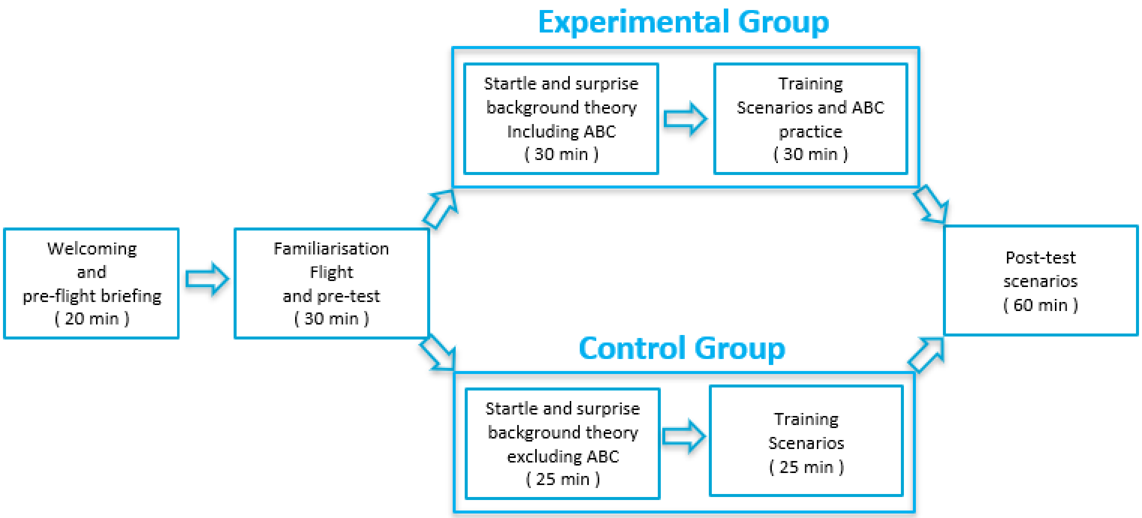


Figure 4.8: Experimental procedure

Before starting the experiment, participants are welcomed and are firstly requested to fill in a questionnaire about their flight experience. Next, they are briefed about the goal of the experiment and the aircraft model they will fly.

A familiarisation round on the simulator follows thereafter: at this stage of the experiment, pilots have the chance of getting acquainted with the model of the Piper Seneca and with the layout of the cockpit within the SIMONA simulator. No malfunctions or surprises are yet introduced and the participants are simply required to fly a standard circuit from take off to landing and to practice with the missed-approach manoeuvre. For the purpose of the familiarisation process, flying conditions are simulated to be optimal with wind being calm, ceiling above 5000 ft and unlimited visibility.

Contrary to the familiarisation round, the following pre-test scenario is characterised by an engine failure during take-off, a scenario that pilots should be very well familiar with from their training. The purpose of this pre-flight test is to assess the flying skills of the participants and to allow therefore to balance the experimental and control groups.

After completing the familiarisation and the pre-test, participants step out of the simulator and undergo a theoretical training session, where the concepts of startle and surprise are discussed and few relevant aircarft accidents are presented, therefore stressing the role of startle and surprise management in aviation. For the experimental group only, the ABC technique is then introduced in its steps, after which pilots from the control and the experimental group are invited to undertake a further training session in the simulator followed by a post test.

The training scenarios are taken from the experiment of van Middelaar et al. and involve 5 different simulations featuring a rudder malfunction, a RPM indicator failure and a right engine failure at take off respectively [9]. The postest is designed to address the goals discussed in section 4.1 and involves the scenarios previously introduced in section 4.3.

4.9. EXPERIMENTAL CONSTRAINTS AND LIMITATIONS

To allow for meaningful conclusions from the discussed experiment while still ensuring the simulations to be as realistic as possible a few constraints and limitations have to be imposed: firstly, the freedom within the scenarios is limited as pilots are constrained to fly a specific route and adhere to the requests of the controller. As an example, in the landing gear scenario, the pilot is instructed to fly a standard circuit after the missed approach following the landing gear failure to extend: in reality, the way such a distress is handled is very much dependent on the dynamics of the situation and most often the controller tries to accommodate the requests of the pilots. In this case, however, the interaction between the pilot and the controller will be more unilateral, namely requests will mainly be addressed from the controller to the pilots and not the other way around. Therefore the artefact of the former instructing a standard circuit pattern rather than asking for the pilots intentions allows to limit the action space of the participants and to make sure that their performance is comparable according to the established criteria.

For the same purpose, pilots in the last scenario will be instructed to fly the legs shown in Figure 4.4, even though VFR flying in a class G airspace would allow pilots to freely move within the boundary of this airspace (of course taking into account limitations imposed by the active NOTAMs).

In addition to what discussed so far, one limitation influencing the realism of the experiment concerns the layout of the simulator: as mentioned in subsection 4.7.1, the interior of the SIMONA does not resemble the cockpit of the real aircraft, even though the location of the instruments and controls will not significantly differ from the ones of the PA-34. On the other hand, the experiment has been designed not to be type specific: therefore while the lack of fidelity in reproducing the real cockpit environment could influence the overall simulation experience, the specific tasks that the participants are called to carry out will not be affected.

Lastly, all scenarios involve one pilot only rather than a standard crew of two people as it is the case in normal airline operations. While performing an experiment with a complete crew would surely lead to interesting results and conclusions on the topic of startle and surprise management on the flight deck, due to time limitations the possibility of assessing crew performance on the topic falls outside the scope of the project.

4.10. SUMMARY AND CONCLUSIONS

This chapter thoroughly discussed the experiment design. The main goal of the experiment is threefold, namely it aims at assessing whether the new technique allows to correctly identify the right priorities when flying, at assessing whether workload management in high workload conditions is significantly improved when implementing the ABC technique compared to the control group and at testing the efficacy of the procedure in supporting problem solving and decision making in relation to a secondary complex task while not compromising the primary task of aviating the aircraft.

To address these goals the MASS and FLAP scenarios have been implemented from van Middelaar et al. [9], while two additional new scenarios have been developed. The landing gear scenario focuses on workload management and on correctly addressing the right priorities at the right time (i.e aviating over trouble shooting): this simulation is based in Rotterdam and features radio communications coordinated by the TWR controller who is intentionally inducing the participant into troubleshooting a landing gear malfunction during final approach. The performance assessment is based on both workload measurements as well as on specific aviating criteria as specified in Table 4.4

The last scenario involves a more subtle surprise and requires the pilot to navigate based on visual references (VFR) from Rotterdam to Lelystad while experiencing a spurious fuel warning and slow decrease in one of the engine thrust, which could upset the aircraft up to an unrecoverable extent. Furthermore, the unfavourable weather conditions are likely to cause a drift from the nominal course if the pilot is not constantly aware of his/her position. Similarly to the landing gear scenario, the performance assessment criteria for this scenario are summarised in Table 4.5 and address dependent variables such as deviation from the nominal route and time taken to identify the system failure, therefore related to avaiting and problem diagnosis respectively.

Around twenty airline pilots will join the experiment: these are preferred over private pilots due to a higher expected homogeneity of experience. Furthermore, airline pilots have flown on multi-engine piston aircraft during their training, which is often not the case for PPL holders: as participants will be flying a model of a light twin aircraft in the SIMONA research simulator, the former seems to be the most appropriate group pf participants to the experiment.

5
CONCLUSIONS

This document provides a thorough literature review of startle and surprise, therefore discussing the influence that these have on human performance, and the consequent impact on the flight deck performance.

From such review it is firstly concluded that although startle and surprise are not rare in aviation and seldomly have major negative consequences on the outcome of a flight, these responses have the potential to compromise safety to an unrecoverable extent as the cases of AF447 and WCA708 have shown.

Previous research has developed dedicated training interventions to improve startle and surprise management on the flight deck. The core of the procedures briefed and practiced throughout these interventions consists in the implementation of stress management techniques aiming at coping with the physiological response from startle and surprise as well as at the unbiased collection of observations through all the senses, therefore allowing to regain situation awareness.

Overall, the reviewed training interventions were found to be useful to support pilots in recovering from startle and surprise events and the related techniques have generally been appreciated by the participants who seem to be willing to apply such techniques in real flight. On the other hand, a previous experiment conducted by van Middelaar et al. [9] concluded that the COOL procedure might have a distraction effect as aviating actions were observed to be delayed and to consequently exacerbate an already upset aircraft condition.

The current research aims therefore at addressing the following research question:

How can the COOL procedure be adapted and tested to improve task prioritization and workload management ?

A first step in answering such question involved the definition of the ABC mnemonic following the recommendations provided by Landman et al. [60]. While the "*Breath*" and "*Check*" steps of the mnemonic are comaprable to the "*Calm Down*" and "*Observe*" steps from the COOL technique, the "*Aviate*" step has been introduced to explicitly remind pilots that stabilising the flight path as much as possible is always the first priority before taking any other step in the recovery process from startle and surprise.

The so defined procedure will therefore be tested in an experiment involving about twenty airline pilots and four different test scenarios aiming at assessing whether the technique allows to identify the right priorities when flying (i.e. flying the aircraft first over troubleshooting), at assessing whether there exist a significant difference in workload management in high workload conditions when compared to the control group and at testing the efficacy of the procedure itself in supporting problem solving and decision making in relation to a secondary complex task.

Bibliography

[1] International Air Transport Association, *Loss of Control In-Flight Accident Analysis Report Edition 2019 Guidance Material and Best Practices* (2019).

[2] W. L. Martin, P. S. Murray, and P. R. Bates, *P11 The Effects of Startle on Pilots During Critical Events: A Case Study Analysis*, Tech. Rep.

[3] A. Landman, L. Groen, Eric, M. R. van Paassen, M., W. Bronkhorst, and M. Mulder, *Dealing With Unexpected Events on the Flight Deck: A Conceptual Model of Startle and Surprise,* Human Fcators , 1161 (2017).

[4] M. A. Staal, *Stress, Cognition, and Human Performance: A Literature Review and Conceptual Framework*, Tech. Rep.

[5] Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile, *Interim Report n° 3 - On the accident on 1st June 2009to the Airbus A330-203registered F-GZCPoperated by Air Franceflight AF 447 Rio de Janeiro - Paris*, Tech. Rep. (2009).

[6] A. Landman, *Managing startle and surprise in the cockpit,* (), 10.4233/uuid:b0382c6a-52af-42a5-b5bf-91368fd9c284.

[7] A. M. Shahsavarani, E. Azad, M. Abadi, and M. H. Kalkhoran, *International Journal of Medical Reviews Systematic Review International Journal of Medical Reviews*, Tech. Rep. 2 (2015).

[8] Civil Aviation Authority, *Flight-crew human factors handbook*, Tech. Rep.

[9] S. H. Van Middelaar, *Startle & Surprise Intervention Training Evaluating Training Effects on Pilot Performance in Off-Nominal Operations*, Tech. Rep. (2018).

[10] Piper Aircraft Corporation, *1999 Kg. SENECA III PA-34-220T PILOT'S OPERATING HANDBOOK,* (1981).

[11] Faa, *Airline Transport Pilot and Type Rating for Airplane Airman Certification Standards*, Tech. Rep. (2019).

[12] Piper Aircraft Company, *Pilot Operating Handbook Seneca II,* (1976).

[13] C. M. Belcastro, L. Groff, R. L. Newman, J. V. Foster, D. A. Crider, D. H. Klyde, and A. M. Huston, *Preliminary Analysis of Aircraft Loss of Control Accidents: Worst Case Precursor Combinations and Temporal Sequencing*, Tech. Rep.

[14] J. Reason, E. Hollnagel, and J. Paries, *REVISITING THE « SWISS CHEESE » MODEL OF ACCIDENTS*, Tech. Rep. (2006).

[15] Boeing Company, *STATISTICAL SUMMARY OF COMMERCIAL JET AIRPLANE ACCIDENTS*, Tech. Rep. (2021).

[16] Boeing Company, *Statistical Summary of Commercial Jet Airplane Accidents*, Tech. Rep. (2001).

[17] Boeing Company, *Statistical Summary of Commercial Jet Airplane Accidents*, Tech. Rep. (2002).

[18] A. Talone and F. Jentsch, *Evaluating startle, surprise, and distraction: An analysis of aircraft incident and accident reports*, Tech. Rep. (2015).

[19] S. M. Casner, R. W. Geven, and K. T. Williams, *The effectiveness of airline pilot training for abnormal events,* Human Factors **55**, 477 (2013).

[20] P. Van Oorschot, *Using Unpredictability and Variety in Pilot Training to Improve Performance in Surprise Situations*, Tech. Rep. (2017).

[21] J. Field, E. Boland, J. van Roij, J. Mohrmann, and J. Smeltink, *Startle Effect Management*, Tech. Rep.

[22] M. Gillen, *A Study Evaluating if Targeted Training for Startle Effect can Improve Pilot Reactions in Handling Unexpected Situations in a Flight Simulator*, Tech. Rep. (2016).

[23] J. Rivera, A. B. Talone, C. T. Boesser, F. Jentsch, and M. Yeh, *Startle and surprise on the flight deck: Similarities, differences, and prevalence,* in *Proceedings of the Human Factors and Ergonomics Society*, Vol. 2014-January (Human Factors an Ergonomics Society Inc., 2014) pp. 1047–1051.

[24] W. L. Martin, P. S. Murray, P. R. Bates, and P. S. Lee, *Fear-Potentiated Startle: A Review from an Aviation Perspective,* International Journal of Aviation Psychology **25**, 97 (2015).

[25] W.-U. Meyer, R. Reisenzein, and A. Schützwohl, *Motivation and Emotion*, Tech. Rep. 3 (1997).

[26] U. Neisser, *Perceiving, Anticipating, and Imagining,* , 92 (1978).

[27] G. Klein and D. A. Peluso, *A data-frame theory of sensemaking*, Tech. Rep. (2007).

[28] A. Landman, E. L. Groen, M. M. van Paassen, A. W. Bronkhorst, and M. Mulder, *The Influence of Surprise on Upset Recovery Performance in Airline Pilots,* International Journal of Aerospace Psychology **27**, 2 (2017).

[29] A. Rankin, R. Woltjer, and J. Field, *Sensemaking following surprise in the cockpit—a re-framing problem,* Cognition, Technology and Work **18**, 623 (2016).

[30] American Psychological Association, *What's the difference between stress and anxiety ?* , (2020).

[31] James Cook University, *Stress and anxiety*, Tech. Rep.

[32] National Educational Psychological Service (NEPS), *Managing Stress and Anxiety*, Tech. Rep.

[33] R. Lazarus, *Psychological stress and the coping process* (New York: McGraw Hill, 1966).

[34] A. Nieuwenhuys and R. R. Oudejans, *Anxiety and performance: perceptual-motor behavior in high-pressure contexts,* (2017).

[35] F. Ozel, *Time pressure and stress as a factor during emergency egress*, Tech. Rep.

[36] R. M. Yerkes and J. D. Dodson, *The relation of strength of stimulus to rapidity of habit-formation,* (1908).

[37] D. H. Lamb, *On the Distinction Between Physical and Psychological Stressors A Review of the Evidence,* Motivation and Emotion **3** (1979).

[38] P. Sivasubramanian, *Eustress Vs Distress-A Review,* International Journal of Research in Humanities & Soc. Sciences **4** (2016).

[39] G. B. Senanayake and C. Arambepola, *Understanding chronic stress: a narrative review of literature,* Journal of the College of Community Physicians of Sri Lanka **25**, 30 (2019).

[40] L. Burges, *What to know about General Adaptation Syndrome,* (2017).

[41] M. W. Eysenck, N. Derakshan, R. Santos, and M. G. Calvo, *Anxiety and cognitive performance: Attentional control theory,* (2007).

[42] A. Ohman, A. Flykt, F. Esteves, and K. Institute, *Emotion Drives Attention: Detecting the Snake in the Grass,* Journal of Experimental Psychology: General **130**, 466 (2001).

[43] A. Nieuwenhuys and R. D. Raǒul, *Effects of anxiety on handgun shooting behavior of police officers: A pilot study,* Anxiety, Stress and Coping **23**, 225 (2010).

[44] R. K. Dismukes, T. E. Goldsmith, and J. A. Kochan, *Effects of Acute Stress on Aircrew Performance: Literature Review and Analysis of Operational Aspects*, Tech. Rep. (2015).

[45] G. A. Klein, *A Recognition Primed Decision (RPD) Model of Rapid Decision Making Chapter 6 A Recognition-Primed Decision (RPD) Model of Rapid Decision Making**, Tech. Rep.

[46] Y. Bar-Haim, A. Kerem, D. Lamy, and D. Zakay, *When time slows down: The influence of threat on time perception in anxiety,* Cognition and Emotion **24**, 255 (2010).

[47] C. M. Inzana, J. E. Driskell, E. Salas, and J. H. Johnston, *Effects of Preparatory Information on Enhancing Performance Under Stress,* Journal of Applied Psychology **81**, 429 (1996).

[48] J. E. Driskell, C. Copper, and A. Moran, *Does Mental Practice Enhance Performance?* Journal of Applied Psychology **79**, 481 (1994).

[49] J. E. Driskell S Brian Mullen Craig Johnson R Sandy Hughes R and C. L. Batchelor, *DEVELOPMENT OF QUANTITATIVE SPECIFICATIONS FOR SIMULATING THE STRESS ENVIRONMENT*, Tech. Rep. (1992).

[50] N. R. T. G. HFM-292, *"Understanding and Reducing Skill Decay", final report (in prep),* Tech. Rep.

[51] J. A. Kochan, E. G. Breiter, and F. Jentsch, *SURPRISE AND UNEXPECTEDNESS IN FLYING: DATABASE REVIEWS AND ANALYSES*, Tech. Rep.

[52] SKYbrary, *Flight Control Laws,* (2018).

[53] Gobierno Bolivariano de Venezuela, *Expediente JIAAC-9-058-2005 i*, Tech. Rep. (2005).

[54] H. Soll, S. Proske, G. Hofinger, and G. Steinhardt, *Decision-Making Tools for Aeronautical Teams: FOR-DEC and Beyond,* Aviation Psychology and Applied Human Factors **6**, 101 (2016).

[55] H. Soll, S. Proske, G. Hofinger, and G. Steinhardt, *Decision-Making Tools for Aeronautical Teams: FOR-DEC and Beyond,* Aviation Psychology and Applied Human Factors **6**, 101 (2016).

[56] E. Jacobson, *Progressive Relaxation,* Source: The American Journal of Psychology **36**, 73 (1925).

[57] W. Martin, *Developing Startle and Surprise Training Interventions for Airline Training Programs,* .

[58] A. Landman, P. van Oorschot, M. M. van Paassen, E. L. Groen, A. W. Bronkhorst, and M. Mulder, *Training Pilots for Unexpected Events: A Simulator Study on the Advantage of Unpredictable and Variable Scenarios,* Human Factors **60**, 793 (2018).

[59] M. Piras and J. Crijnen, *Evidence Based Training-Data Generation and Analysis of Air-line Pilots Training Performance Internship Assignment*, Tech. Rep. (2021).

[60] A. Landman, *Managing startle and surprise in the cockpit,* (), 10.4233/uuid:b0382c6a-52af-42a5-b5bf-91368fd9c284.

[61] J. Burki-Cohen, *Technical Challenges of Upset Recovery Training: Simulating the Element of Surprise,* .

[62] National Transportation Safety Board, *Aircraft Accident Report - United Airlines, INC. McDonnel-Douglas, DC-8-61, N8082U,* (1978).

[63] Federal Aviation Administration, *Lockheed Model L-1011 Eastern Airlines Flight 401, N310EA,* (1972).

[64] F. Zijlstra, *Efficiency in Work Behaviour: a designed approach for modern tools,* .

[65] R. Muynck and M. Hesse, *The a priori simulator software package of the piper pa34 seneca iii,* (1990).

[66] Piper Aircraft Company, *Piper Seneca,* .

Part III

Appendices

A
LANDING GEAR SCENARIO ATC SCRIPT

(PRELIMINARY REPORT)

The following appendix reports the script of the ATC communication which will be implemented in the landing gear scenario. Comments on the storyline are in boldface, while the possible expected answers from the participants are in-between square brackets.

The following roles are involved in such scenario:

- **PH-DOC**: Medical Helicopter crossing the CTR (Life Liner)
- **PH-SPZ**: Student Pilot
- **PH-SVT**: Private Pilot
- **PH-TUD**: Pilot flying the simulation

The scenario starts with the Pilot already being cleared to enter the CTR and approaching MIKE. Right before reaching the waypoint, the pilot is instructed to perform a couple of orbits at MIKE to give way to the medical helicopter crossing the CTR. Meanwhile a private pilot (PH-SVT) is requesting take off clearance (see Figure A.1).

Figure A.1: Initial traffic involved in the simulation

PH-DOC: PH-DOC approaching Gouda 1000 ft, request to cross CTR in the direction Gouda – The Hague at 1000 ft for medical transportation.

TWR: PH-DOC cleared to cross the CTR, maintain 1000 [ft] or below, mind traffic approaching MIKE from the east, same altitude.

PH-DOC: Cleared to cross the CTR, will maintain 1000 [ft], traffic in sight PH-DOC.

TWR: PH-TUD make a right orbit at MIKE, maintain 1000 ft, mind the medical helicopter crossing from the South.

PH-TUD: [Wilco right orbit 1000 ft … looking for traffic PH-TUD].

PH-SVT: PVT at V4 ready for MIKE departure.

TWR: PVT clear for take off runway 24, the wind is 260, 15 knots.

PH-SVT: Clear for take off, runway 24, PVT.

TWR: … and PUD after orbit complete perform another right orbit, 1000ft.

PH-TUD: [Wilco right orbit 1000 ft PUD].

PH-DOC: PH-DOC approaching Delft 1000ft, request frequency change.

TWR: PH-DOC frequency change approved.

PH-DOC: Frequency change approved PH – DOC … bedankt !

Event: The low fuel warning is triggered during the second orbit. Probably the pilot will report the event. In any case, soon after this warning TWR clears the pilot for a direct approach to runway 24. At this stage a student pilot (PH-SPZ) asks for start up clearance (see Figure A.2).

Figure A.2: Further development of the simulation

PH-TUD: [PUD We have critically low fuel readings, request priority for landing].

TWR: PUD roger, direct final runway 24, clear to land.

PH-TUD: [Clear to land runway 24 PUD].

PAUSE FOR FEW SECONDS . . .

WKLD MEASUREMENT: RATE YOUR WORKLOAD IN THE PAST MINUTE ON A SCALE FROM 1 TO 10.

PH-SPZ: Rotterdam TWR, PHSPZ good morning.

TWR: PPZ, Rotterdam Tower go-ahead.

PH-SPZ: Robin DR-400 at the Vliegclub . . . local VFR flight. . . information LIMA is received. . . ehm correction information KILO, request start up PPZ.

TWR: PPZ information KILO is correct, start up approved, runway 24 is in use, QNH 1009, report when ready for taxi

.

PH-SPZ: Start up approved ehm. . . QNH. . . Ehm . . . Say again slowly for PPZ.

TWR: PPZ, Start Up Approved, runway 24 is in use, QNH 1009 [HPa], report when ready for taxi.

PH-SPZ: QNH 1009, runway 24, will report ready for taxi . . . PPZ.

Event: When the pilot lowers the landing gear, causing the landing gear failure, an unintentional transmission occurs from the student pilot, with the purpose of providing a distraction from the occurrence of the landing gear failure event.

PH-SPZ: Yeah, so do you think we should depart via ROMEO or via HOTEL ? I have never flown the HOTEL departure before, so I would like to proceed with HOTEL if you agree ...

The right (or left) main landing gear fails to extend: at this point the pilot is expected to counteract the yawing motion generated and subsequently to report the issue to the controller. Following the request of the pilot, the controller gives clearance for a go around and induces the pilot himself/herself into troubleshooting the problem while he/she is flying the final approach. As discussed, this artefact should have the purpose of distracting the pilot from flying the aircraft and makes potential differences between the aviating behaviour of the two groups more likely to be noticed.

PH-TUD: [PUD we are experiencing a malfunction with the landing gear, we will go around]

TWR: PUD, roger ... PUD How many POB ?

PH-TUD: [PUD 1 POB], ALTERNATIVELY [PUD, stand by ...]

TWR: PUD I fly the Piper Seneca myself... I have already experienced a similar issue in the past. Could you try to check if the circuit breakers are all in ? Or Maybe Check the Ammeter ?

PH -TUD: [Checking the circuit breakers .... They are all in ....], ALTERNATIVELY [PUD, stand by ...]

Some intereferences on the frequency occur

TWR: Roger ...

TWR: PUD, radio check on 118.205 ...

PH-TUD: [PUD, I read you : 5], ALTERNATIVELY [PUD, stand by ...]

TWR: PUD Roger...

TWR: PUD, after touch down, climb to 1000ft, right hand turn and join right hand downwind runway 24.

PH-TUD: [PUD wilco 1000 ft, right hand downwind runway 24]

The go around takes place at this point: when climbing after touch down the controller asks if the pilot needs emergency services next to the runway ....

TWR: PUD, do you want to declare an emergency and do you need further assistance ?.

PH-TUD: [PUD affirmative, we would like to declare an emergency, and we do require assistance], ALTERNATIVELY [PUD, stand by ...].

PAUSE FOR FEW SECONDS ....

WKLD MEASUREMENT: RATE YOUR WORKLOAD IN THE PAST MINUTE ON A SCALE FROM 1 TO 10.

PH – SPZ: Tower, PPZ at the Vliegclub, request taxi.

TWR: PPZ runway 24, V4.

PH-SPZ: runway 24, V4, PPZ.

When on downwind the pilot has time to stabilise the aircraft, assess the situation and prepare the aircraft for a second landing attempt, therefore recycling the landing gear. Once the flight path is under control, the pilot has also time to reply to the requests from the controller, if these have been delayed before...

Furthermore when on downwind there is still room for further requests from ATC which could interfere with flying the aircraft...

TWR: PUD report amount of fuel remaining.

PH -TUD: [Report fuel remaining], ALTERNATIVELY [PUD, stand by ...]

PH-SVT: PVT approaching MIKE, 1000 ft, request frequency change.

TWR: PVT, frequency change approved.

PH-SVT: Frequency change approved, PVT.

PH-SPZ: PPZ at V4 ready for departure.

TWR: PPZ hold short at V4, there is an aircraft on downwind with a landing gear issue... Expect late take off clearance.

PH-SPZ: Holding at V4 and roger for the traffic, PPZ.

TWR: PUD cleared to land runway 24, wind is 260, 10 kts.

PH-TUD: [Cleared to land, runway 24, PUD]

B
PERFORMANCE PLOTS

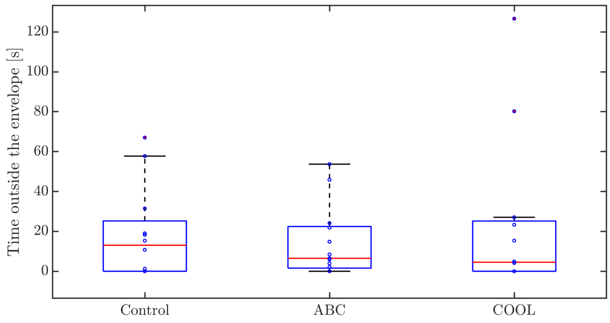


Figure B.1: Time outside of the envelope in relation to the pre-test scenario.

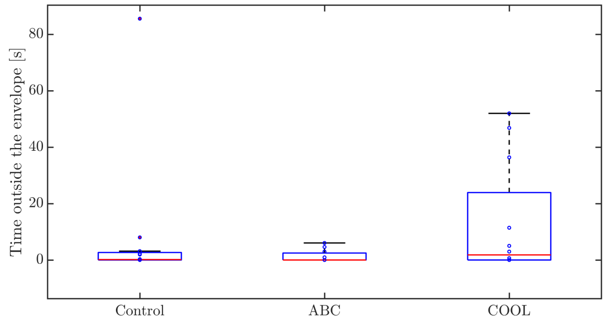


Figure B.2: Time outside of the envelope in relation to the flap asymmetry scenario.

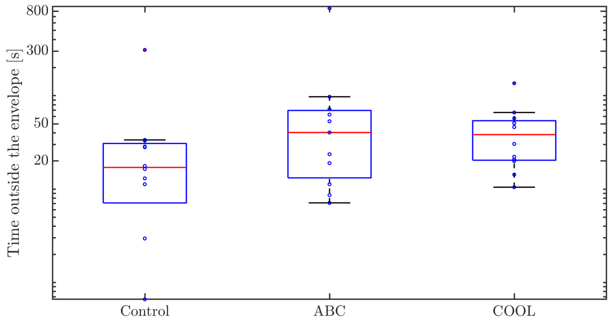


Figure B.3: Time outside of the envelope in relation to the mass shift scenario.

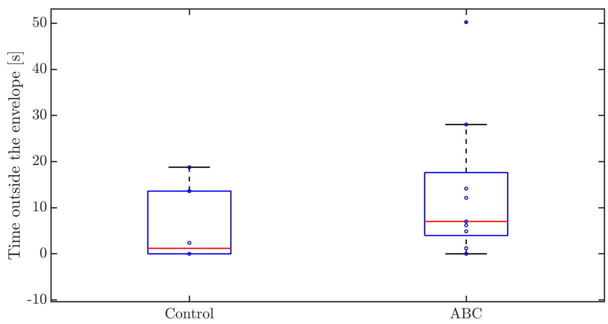


Figure B.4: Time outside of the envelope in relation to the landing gear failure scenario.

C
SUMMARY OF STATISTICAL RESULTS

Table C.1: Average RSME scores per scenario and related significance (within participants analysis ABC and Control).

	Flap ($\mu = 67.08$)	Gear ($\mu = 61.41$)
Mass ($\mu = 85.41$)	$p = 0.018$	$p = 0.034$

Table C.2: Average stress scores per scenario and related significance (within participants analysis ABC and Control groups)

	Flap ($\mu = 4.25$)	Gear ($\mu = 4.5$)
Mass ($\mu = 6.83$)	$p < 0.01$	$p < 0.01$

Table C.3: Average RSME scores per group and scenario and related significance (between participant analysis ABC, Control, and COOL groups).

	Flap	Mass	Gear
ABC	$\mu = 65.73$	$\mu = 79.18$	$\mu = 64.11$
COOL	$\mu = 62.33$	$\mu = 74.41$	N/A
Control	$\mu = 66.25$	$\mu = 81.25$	$\mu = 51.83$
	$p = 0.871$	$p = 0.647$	$p = 0.369$

Table C.4: Average Stress scores per group and scenario and related significance (between participant analysis ABC, Control, and COOL groups).

	Flap	Mass	Gear
ABC	$\mu = 0.066$	$\mu = 0.312$	$\mu = 4.66$
COOL	$\mu = 0.228$	$\mu = 0.247$	N/A
Control	$\mu = -0.601$	$\mu = -0.286$	$\mu = 3.83$
	$p = 0.563$	$p = 0.275$	$p = 0.529$

Table C.5: Mean rank of the rated usefulness scores given by the *ABC* and *COOL* groups over the flap asymmetry and the mass shift scenario with related significance.

	ABC	COOL
Flap	$\mu_{rank} = 10.91$	$\mu_{rank} = 13$
Mass	$\mu_{rank} = 14.36$	$\mu_{rank} = 9.83$
	$p = 0.449$	$p = 0.100$

D
RESULTS OF THE SHAPIRO-WILK AND LEVENE'S TESTS.

Table D.1: Results of the Shapiro-Wilk and Levine's test for normality and equality of variance in relation to H1.

	Shapiro-Wilk
Control	$p < 0.01$
ABC	$p < 0.01$
COOL	$p = 0.01$
Levine's test	$p = 0.052$

Table D.2: Results of the Shapiro-Wilk test for normality in relation to the hypothesis H2.

		Stress	**RSME**
Flap Asymmetry	Control	$p = 0.103$	$p = 0.863$
	ABC	$p = 0.070$	$p = 0.332$
Mass Shift	Control	$p = 0.73$	$p = 0.012$
	ABC	$p = 0.007$	$p = 0.053$
Gear Failure	Control	$p = 0.804$	$p = 0.131$
	ABC	$p = 0.228$	$p = 0.146$

Table D.3: Results of the Levene's test for homogeneity of variance in relation to hypothesis H2

	Stress	**RSME**
Flap Asymmetry	$p = 0.385$	$p = 0.018$
Mass Shift	$p = 0.006$	$p = 0.119$
Gear Failure	$p = 0.178$	$p = 0.004$

Table D.4: Results of the Shapiro-Wilk test for normality in relation to the hypothesis H3.

		Stress	**RSME**
Flap Asymmetry	ABC	$p = 0.319$	$p = 0.711$
	COOL	$p = 0.912$	$p = 1.000$
Mass Shift	ABC	$p = 0.354$	$p = 0.033$
	COOL	$p = 0.476$	$p = 0.911$

Table D.5: Results of the Shapiro-Wilk test for normality in relation to the hypothesis H3.

	Stress	**RSME**
Flap Asymmetry	$p = 0.274$	$p = 0.165$
Mass Shift	$p = 0.273$	$p = 0.569$

E
EXPERIMENT SCRIPT

Scenario	Location at start	Wind	Task	Failure	Failure timing
Familiarisation 1	RWY 18C	NIL	Fly std. circuit	-	-
Familiarisation 2	RWY 18C	090/10	Fly std. circuit	-	-
Pre-test	Approach to RWY 18C (2.7 NM, 675 ft)	090/10	Land on RWY 18C	Left engine failure	600 ft threshold
Training 1	RWY 18C	NIL	Fly std. circuit	-	-
Training 2	Approach to RWY 18C (2.7 NM, 675 ft)	090/19	Land on RWY 18C	rudder stuck at 0 deg during approach	300 ft threshold
Training 3	RWY 18C	160/4	Fly std. circuit	Right engine failure	90 Knots
Training 4	Approach to RWY 18C (2.7 NM, 675 ft)	270/19	Land on RWY 18C	rudder hardover (-15 deg)	15 s from start
Test 1	RWY 18C	270/12	Fly std. circuit	flap asymmetry	upon flap deployment
Test 2	RWY18C	270/13	Fly std. circuit	mass shift	100 ft
Test 3	2.7 NM east of MIKE (EHRD CTR)	260/8	Land on RWY 24	Right landing gear extension failure	upon gear deployment

Table E.1: Experiment script.

F

SAMPLE QUESTIONNAIRE (ABC GROUP)

- How **startled** were you by the event ? (with startle we mean a quick, brief physiological stress response.)

Not at all | Extremely

0	1	2	3	4	5	6	7	8	9	10

- How **surprised** were you by the event ? (with surprise we mean the extent to which the events mismatched with your expectations)

Not at all | Extremely

0	1	2	3	4	5	6	7	8	9	10

- How difficult was it to understand what happened ?

Not at all | Extremely

0	1	2	3	4	5	6	7	8	9	10

- How much stress or anxiety did you feel during the scenario ?

Not at all | Extremely

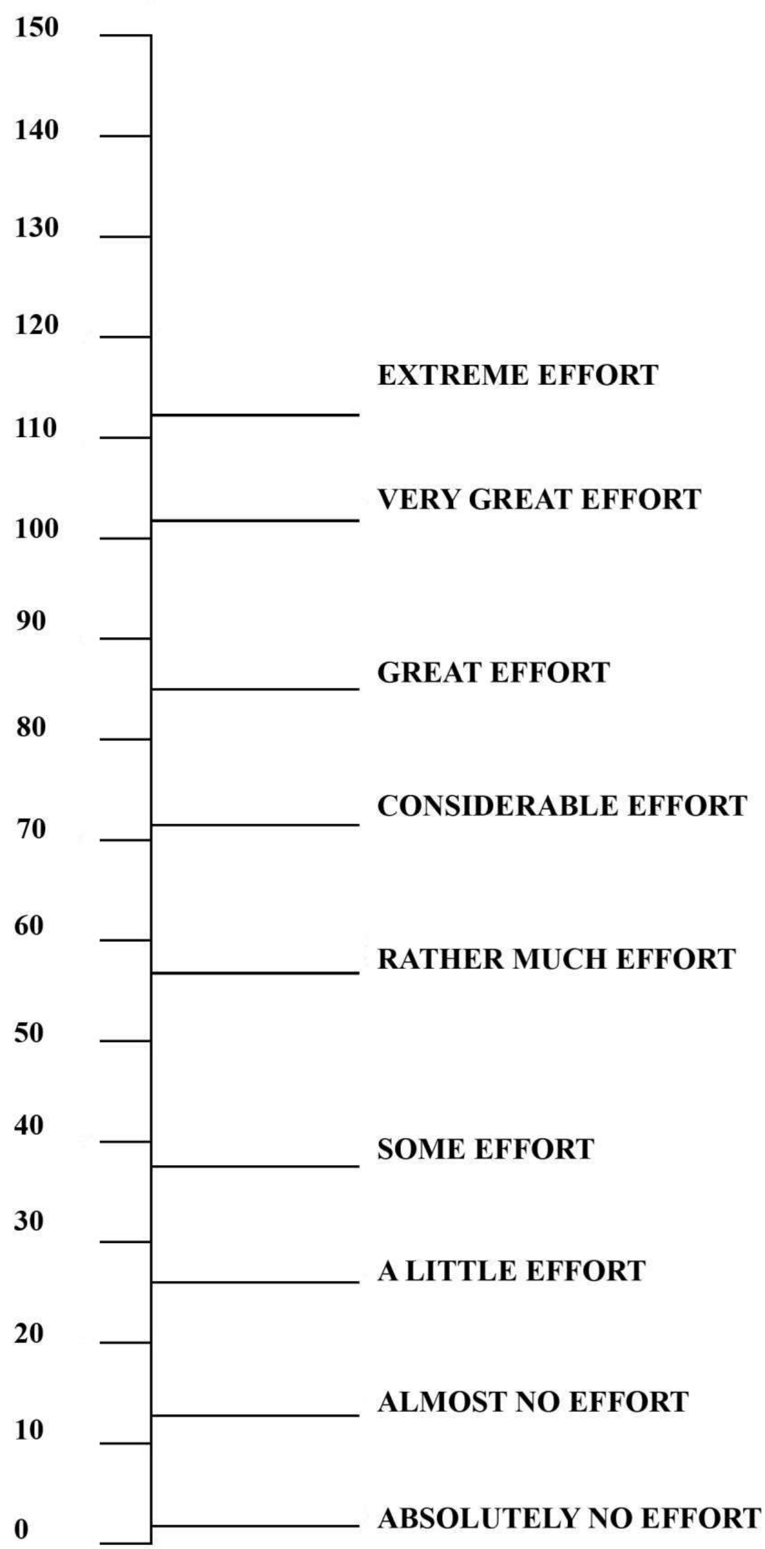
0	1	2	3	4	5	6	7	8	9	10

- If you applied the “ABC” intervention method, which aspects of such method did you use ?
    - **Aviate** (flight path control + energy management)
    - **Breath** (Sit upright, deep breathing, become aware of applied control forces)
    - **Check** (Scan Primary and Secondary instruments outloud)
- To what extent did the ABC intervention method help you ?

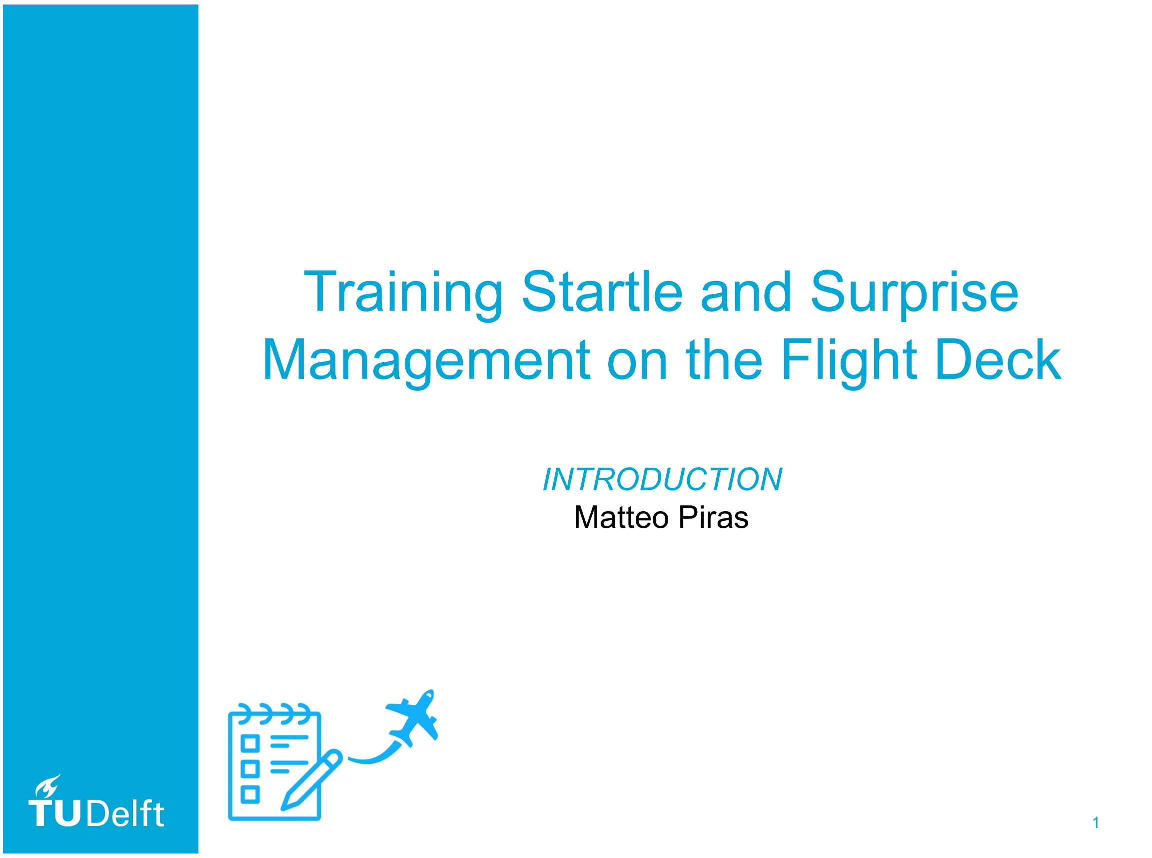
Not at all | Extremely

0	1	2	3	4	5	6	7	8	9	10

How much mental effort did the scenario require? (Please place a cross on the line.)



G
BRIEFING SLIDES



Experiment Schedule

- Familiarisation with the SIMONA simulator (30 min)
- Coffee break ! (10 min)
- Startle and surprise theory (20 min)
- Practice session (45 min)
- Test session (45 min)

TUDelft

Few remarks !!!

- Experiment consent form !
- Results of the experiment are anonymous: don't worry if you screw up !
- **Please do not interrupt the briefing ! Save your questions for later ☺**

Apparatus (1/3)

SIMONA Simulator

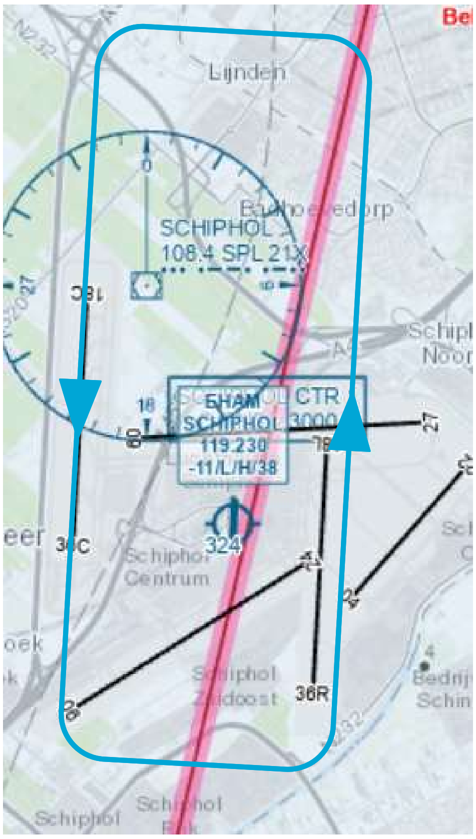
PA-34 Seneca III

TU Delft

Apparatus (2/3)

Apparatus (3/3)

FAMILIARISATION FLIGHT



- **RWY 18C EHAM – Left Hand @ 1000 ft**
- **Current METAR:**
  - *VRB01KT CAVOK*

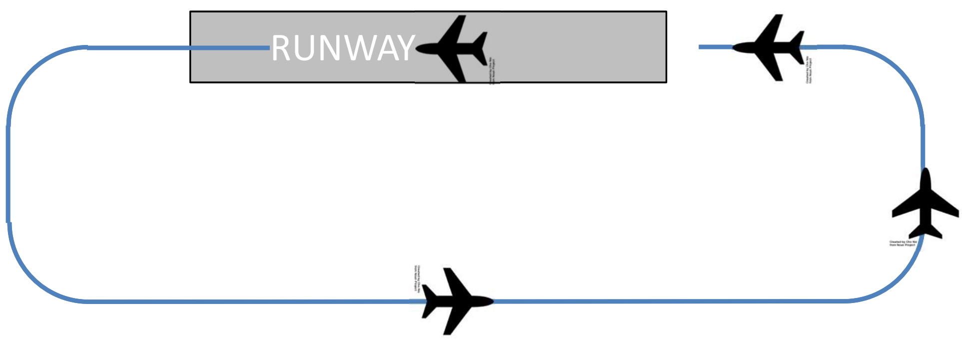
TUDelft

PA-34 Speed parameters

TAKE OFF
- V2 = 90 knots
- Gear = UP

TAKE OFF
- Vr = 80 knots
- Flaps = UP

FINAL
- V = 90 knots
- Flaps = 40 deg



BASE LEG
- V = 90 knots
- Gear = down
- Flaps =25 deg

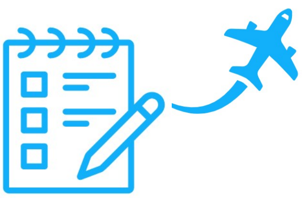
DOWNWIND @ 1000ft
- V = 115 knots

TU Delft

Training Startle and Surprise Management on the Flight Deck

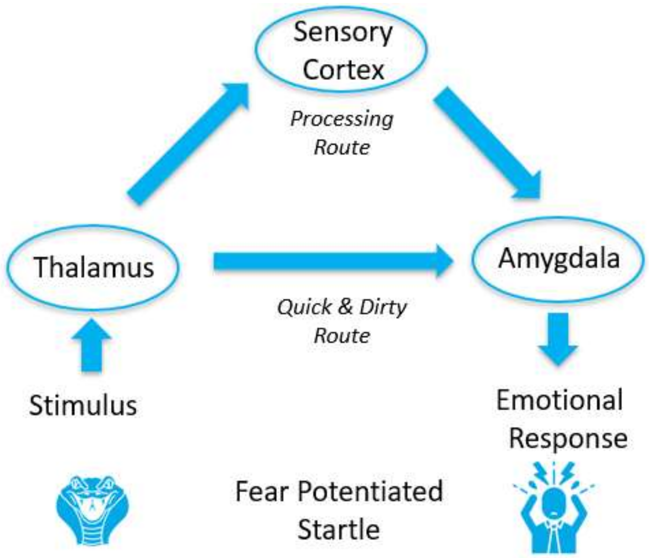
TRAINING BRIEFING

Matteo Piras



Introduction: startle and surprise

- **Startle**: *complex combination of physiological, emotional and cognitive reactions to a sudden stimulus*



- **Surprise**: *cognitive emotional response to unexpected events resulting from a mismatch between active frame and factual reality*

How to cope with startle and surprise ?

Mnemonic aid procedures !

- Currently mnemonics aids are often used in aviation to support decision making …
- Similar procedures have been developed to support pilots in recovering from startle and surprise events

Procedure

GOALS

- Supporting pilots in prioritizing the correct sequence of tasks
- Recognizing and controlling the physiological and psychological reactions resulting from startle and surprise events
- Supporting the reframing process

- Aviate
- Breathe
- Check

TUDelft

ABC Procedure

- **Aviate**
  - Control flight path
  - Aim at straight and level flight
- **Breathe**
  - Sit up right
  - Relax your arms, legs, shoulders
  - Deeply inhale and hold your breath for 2-3 seconds
  - Exhale and focus on the control forces
- **Check:**
  - Focus on your senses: do you feel, see, hear, smell anything unusual ?
  - Scan and call out the basic instruments

H

RAW DATA FROM THE ABC GROUP

Table H.1: Raw data in relation to the familiarization Scenario.

ID	RSME	Stress
1	65	4
3	70	3
4	67	6
5	38	1
8	38	0
10	80	6
11	67	4
12	43	2
13	70	2
16	50	1
20	68	3
21	43	2
25	48	1

Table H.2: Raw data from the questionnaires in relation to the pre-test Scenario.

ID	Startle	Surprise	Difficulty	Stress	RSME
1	7	6	6	6	102
3	9	8	7	7	100
4	8	7	6	7	72
5	7	9	6	8	108
8	6	4	2	3	57
10	7	8	6	8	70
11	4	2	3	6	95
12	3	6	6	4	65
13	4	4	2	2	87
16	3	1	1	2	56
20	4	9	5	7	102
21	7	6	0	4	78
25	4	2	2	4	57

Table H.3: Raw data from the questionnaires in relation to the flap asymmetry scenario.

ID	**Startle**	**Surprise**	**Difficulty**	**Stress**	**A**	**B**	**C**	**Usefulness**	**RSME**
1	4	3	5	3	1	0	0	2	66
3	7	7	2	2	1	1	1	4	57
4	7	7	9	6	1	1	1	7	85
5	5	8	9	4	1	1	1	8	78
8	6	6	5	4	1	1	1	6	50
10	6	7	6	6	1	1	1	7	70
11	3	6	4	6	1	1	1	6	57
12	4	6	4	4	1	1	1	7	84
13	2	6	7	2	1	1	1	7	70
16	4	3	9	4	1	1	1	8	70
20	6	5	3	2	1	1	1	7	43
21	7	7	4	5	1	0	1	5	78
25	2	2	7	3	1	0	0	6	39

Table H.4: Raw data from the questionnaires in relation to the mass shift scenario.

ID	**Startle**	**Surprise**	**Difficulty**	**Stress**	**A**	**B**	**C**	**Usefulness**	**RSME**
1	7	7	7	6	1	1	1	7	57
3	9	8	8	8	1	1	1	7	85
4	8	7	8	7	1	1	1	7	80
5	9	10	8	9	1	1	1	7	102
8	8	8	8	7	1	1	1	6	100
10	8	8	10	8	1	1	1	7	102
11	6	9	8	6	1	1	1	8	67
12	5	7	7	4	1	1	1	6	65
13	3	8	8	3	1	1	1	7	65
16	8	9	10	8	1	1	1	9	104
20	3	7	4	3	1	1	1	7	50
21	10	9	7	9	1	1	1	7	100
25	4	5	7	3	1	1	1	6	48

Table H.5: Raw data from the questionnaires in relation to the landing gear failure scenario.

ID	Startle	Surprise	Difficulty	Stress	A	B	C	Usefulness	RSME
1	-	-	5	5	1	0	0	4	80
3	5	5	9	3	0	0	0	0	40
4	7	8	9	7	1	1	1	8	90
5	2	7	4	7	1	1	1	8	72
8	-	-	-	-	0	0	0	0	38
10	5	7	5	6	1	1	1	7	65
11	4	6	0	7	1	1	1	5	85
12	3	7	7	3	1	1	1	5	55
13	0	3	0	1	1	1	1	7	89
16	2	1	1	3	1	1	1	5	41
20	2	1	2	1	0	0	0	0	35
21	8	8	2	7	1	1	1	7	75
25	1	0	1	1	1	1	1	5	20

Table H.6: Time Oudside the Envelope (TOE) in seconds for each test scenario.

ID	Pre-Test	Flap Asymmetry	Mass Shift
10	0	0	97.42
11	45.9	0	11.26
12	53.7	6.04	8.62
13	3.94	0	18.98
16	0	0	-
1	14.84	0.94	40.36
20	0	0	-
21	6.5	4.62	71.74
25	5.92	-	7.12
3	21.88	2.98	23.64
4	24.16	-	62.74
5	8.46	0	53.2
8	2.08	0	862.06

Table H.7: Number of times the safe envelope was exceeded for each test scenario.

ID	Pre-Test	Flap Asymmetry	Mass Shift
10	0	0	5
11	1	0	1
12	1	1	1
13	2	0	1
16	0	0	5
1	1	1	2
20	0	0	1
21	3	1	3
25	1	0	2
3	1	1	1
4	2	7	3
5	1	0	2
8	1	0	1

I

RAW DATA FROM THE CONTROL GROUP

Table I.1: Raw data in relation to the familiarization Scenario.

ID	RSME	Stress
2	68	5
6	28	1
7	38	3
9	65	1
14	60	2
15	96	6
17	73	0
18	50	3
19	50	4
22	57	2
23	45	2
24	38	1

Table I.2: Raw data from the questionnaires in relation to the pre-test scenario.

ID	Startle	Surprise	Difficulty	Stress	RSME
2	7	8	2	7	84
6	4	1	1	3	53
7	6	7	3	7	58
9	7	6	3	6	78
14	6	10	4	6	103
15	7	4	1	6	93
17	3	5	1	0	73
18	2	3	1	3	50
19	8	8	2	7	68
22	2	8	1	2	83
23	1	1	1	2	50
24	7	8	4	5	81

Table I.3: Raw data from the questionnaires in relation to the flap asymmetry scenario.

ID	Startle	Surprise	Difficulty	Stress	RSME
2	4	8	9	3	68
6	3	2	2	2	37
7	4	4	4	4	37
9	2	4	3	2	57
14	6	5	4	3	69
15	7	8	8	8	108
17	7	9	8	2	99
18	4	3	2	3	57
19	8	8	6	6	68
22	2	5	5	1	77
23	1	2	1	2	45
24	7	8	8	7	73

Table I.4: Raw data from the questionnaires in relation to the mass shift scenario.

ID	Startle	Surprise	Difficulty	Stress	RSME
2	7	7	6	6	80
6	6	5	2	3	75
7	8	8	8	8	104
9	4	5	5	3	85
14	8	7	6	6	78
15	9	9	9	8	110
17	6	9	9	3	85
18	4	4	6	4	67
19	9	8	8	7	75
22	4	6	8	3	94
23	5	3	6	4	57
24	8	8	5	6	65

Table I.5: Raw data from the questionnaires in relation to the landing gear failure scenario.

ID	Startle	Surprise	Difficulty	Stress	RSME
2	3	4	0	4	48
6	2	2	2	2	40
7	2	2	2	1	26
9	2	4	2	1	42
14	2	5	0	3	40
15	7	6	4	7	100
17	2	8	1	1	72
18	0	2	-	3	58
19	7	7	8	6	52
22	3	7	2	2	70
23	2	4	1	2	45
24	6	7	6	4	50

Table I.6: Time Oudside the Envelope (TOE) in seconds for each test scenario.

ID	Pre-Test	Flap Asymmetry	Mass Shift
14	57.74	0	33.6
15	31.46	85.58	0.66
17	0	0	11.26
18	0	2.12	28.08
19	1.24	8.02	33.04
22	10.76	0	2.96
23	19	0	28.54
24	67.02	0.3	0
2	0	1.94	13.02
6	0	0	17.7
7	18.3	0	309.06
9	15.32	3.14	16.42

Table I.7: Number of times the safe envelope was exceeded for each test scenario.

ID	**Pre-Test**	**Flap Asymmetry**	**Mass Shift**
14	1	0	3
15	1	4	1
17	0	0	1
18	0	1	1
19	1	1	2
22	1	0	2
23	2	0	1
24	1	1	0
2	0	2	2
6	0	0	1
7	1	0	1
9	1	1	2

J

RAW DATA FROM THE COOL GROUP

Table J.1: Raw data from the questionnaires in relation to the pre-test scenario.

Startle	Surprise	Difficulty	Stress	RSME
7	7	1	0.2	38
6	10	2	6.8	86
3	6	5	5.1	74
6	7	2	6.9	79
3	2	0	1.4	69
3	8	1	3.8	54
6	6	3	6.1	57
6	5	4	5.2	67
4	3	2	3.7	37
6	8	3	2	39
1	9	1	1.6	38
8	9	2	2.1	81

Table J.2: Raw data from the questionnaires in relation to the flap asymmetry Scenario.

Startle	Surprise	Difficulty	Stress	C	O	O	L	Usefulness	RSME
8	4	8	1.6	1	1	1	1	4	40
6	7	9	4.8	1	1	1	1	4	46
6	8	9	5.4	1	1	1	1	4	83
6	9	7	9.2	1	1	1	1	4	108
2	10	7	5.1	1	1	1	1	4	78
0	3	1	0.6	0	1	1	1	1	21
8	8	9	7.7	1	1	1	1	4	72
7	6	7	6	1	1	1	1	4	68
5	5	3	6.1	1	1	1	1	2	60
7	8	8	5.8	1	1	1	1	4	57
4	7	9	2.9	1	1	0	1	3	51
4	5	5	3	1	1	1	1	3	64

Table J.3: Raw data from the questionnaires in relation to the mass shift Scenario.

Startle	Surprise	Difficulty	Stress	C	O	O	L	Usefulness	RSME
9	9	7	4	1	1	1	1	3	54
5	7	7	6	1	1	0	0	3	80
6	9	7	5.5	1	1	1	1	4	77
8	9	10	9.3	1	1	1	1	4	89
8	8	10	7.5	1	1	1	1	4	106
3	7	8	2.8	0	1	0	1	2	85
7	7	6	7.3	1	1	1	1	3	77
9	8	8	7.4	1	1	1	1	4	93
6	6	8	7	0	1	1	1	1	75
6	8	8	6.8	1	1	1	1	3	52
4	8	8	3	1	1	1	1	3	38
8	8	8	5.8	1	1	1	1	3	67

Table J.4: Time Oudside the Envelope (TOE) in seconds for each test scenario

ID	Pre-Test	Flap Asymmetry	Mass Shift
1	0	0.64	19.98
11	27.06	0	30.46
14	0	0	50.54
15	4.16	0.64	14.3
18	0	0	10.48
20	15.36	11.44	66.08
21	80.22	36.4	135.8
24	4.88	0	50.72
3	126.66	2.98	57.24
5	23.34	52	46.14
7	0	5.02	22.2
8	0	46.84	20.76

Table J.5: Number of times the safe envelope was exceeded for each test scenario.

ID	Pre-Test	Flap Asymmetry	Mass Shift
1	0	1	3
11	1	0	2
14	0	0	4
15	1	1	1
18	0	0	2
20	2	1	2
21	1	3	2
24	2	0	4
3	5	1	3
5	1	5	2
7	0	1	2
8	0	4	1

K

PLOTS OF THE TIME SPENT OUTSIDE OF THE ENVELOPE FOR THE ABC GROUP

K.1. Pre-Test Scenario

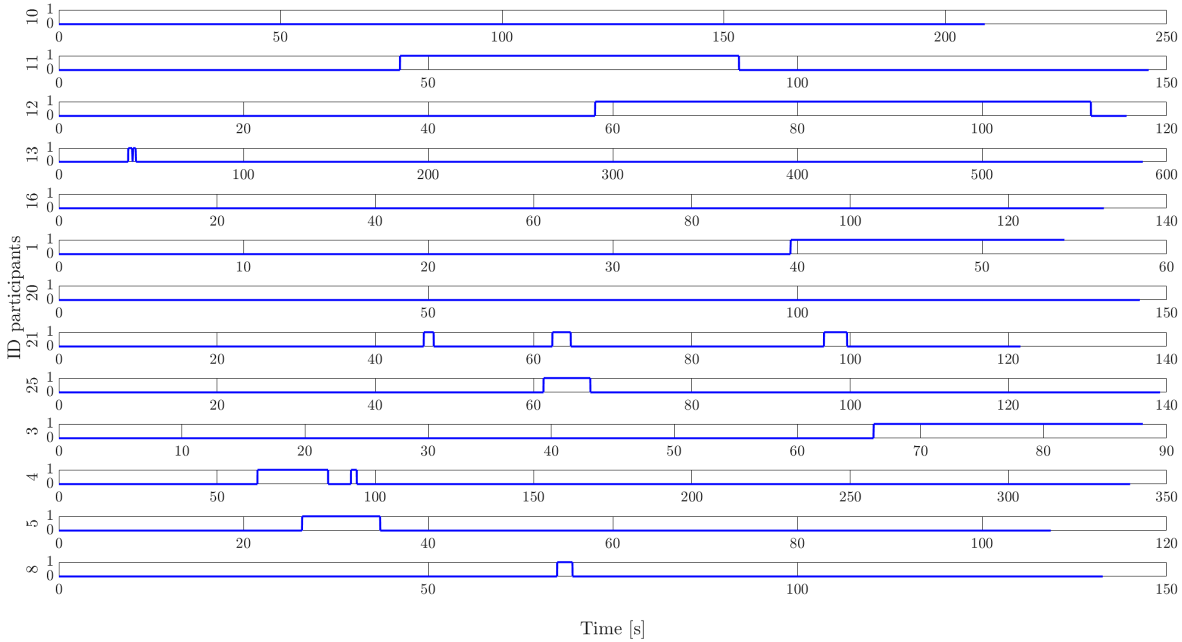


Figure K.1: Total time spent outside the envelope.

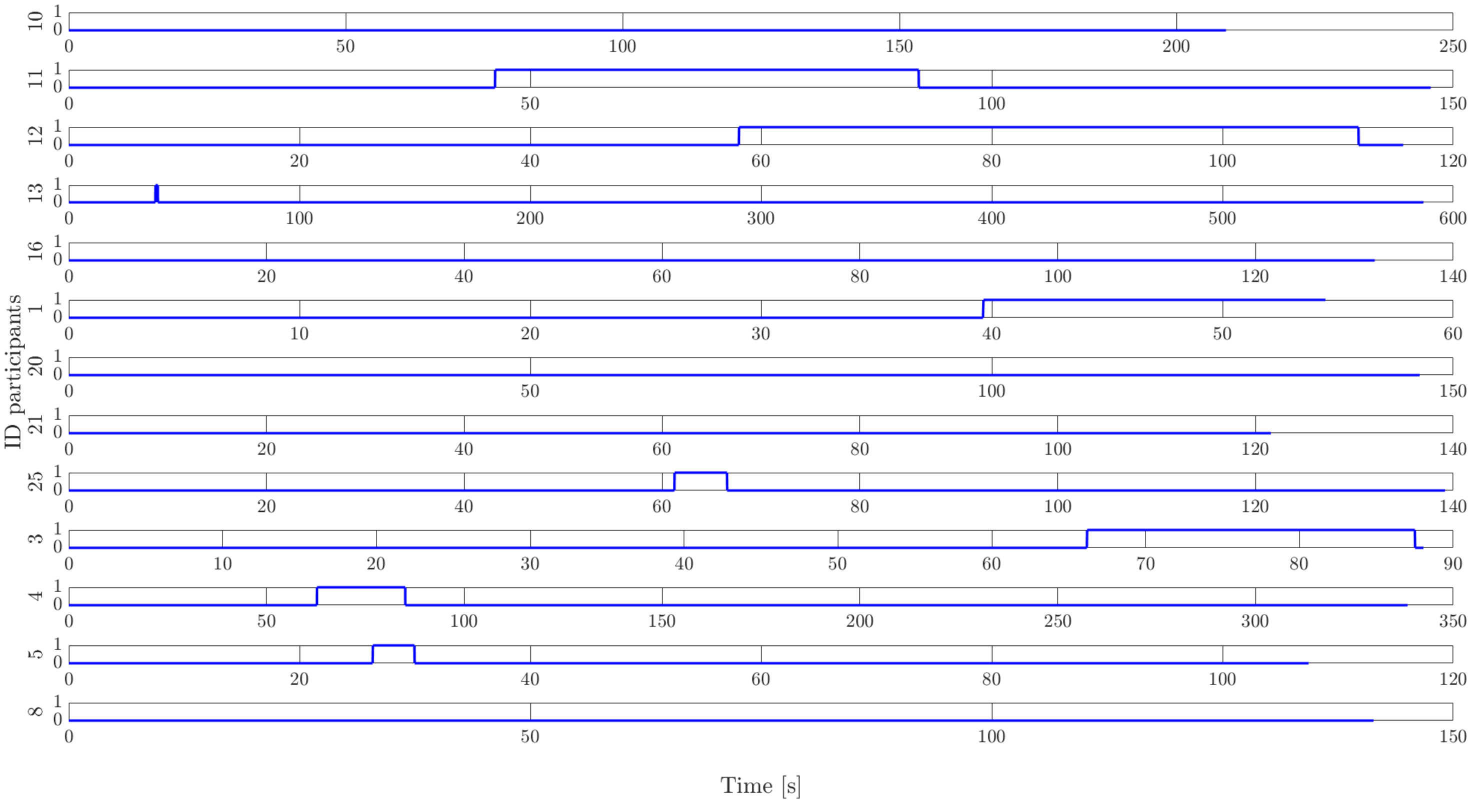


Figure K.2: Contribution of V_{TAS} to the total time spent outside of the envelope.

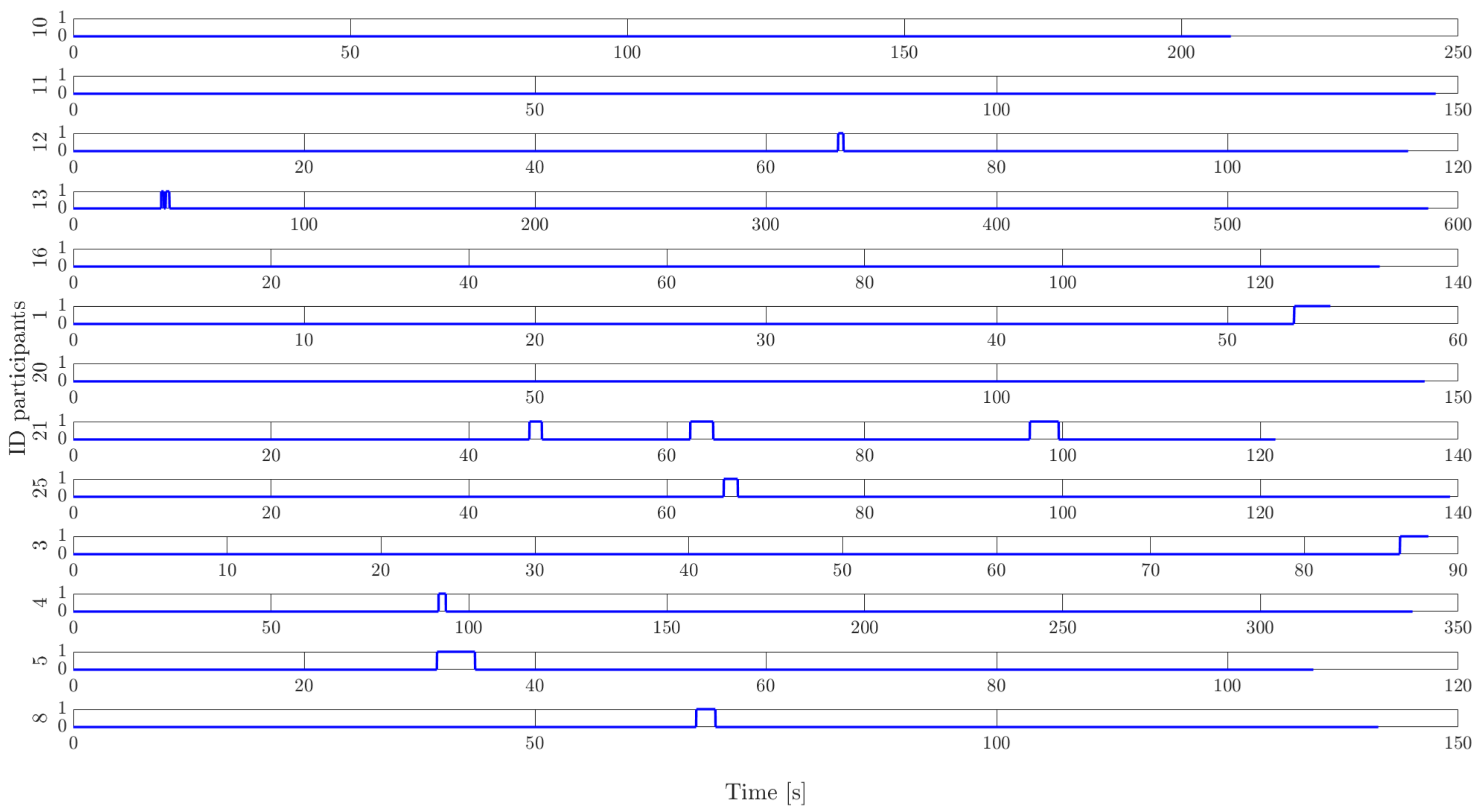


Figure K.3: Contribution of θ to the total time spent outside of the envelope.

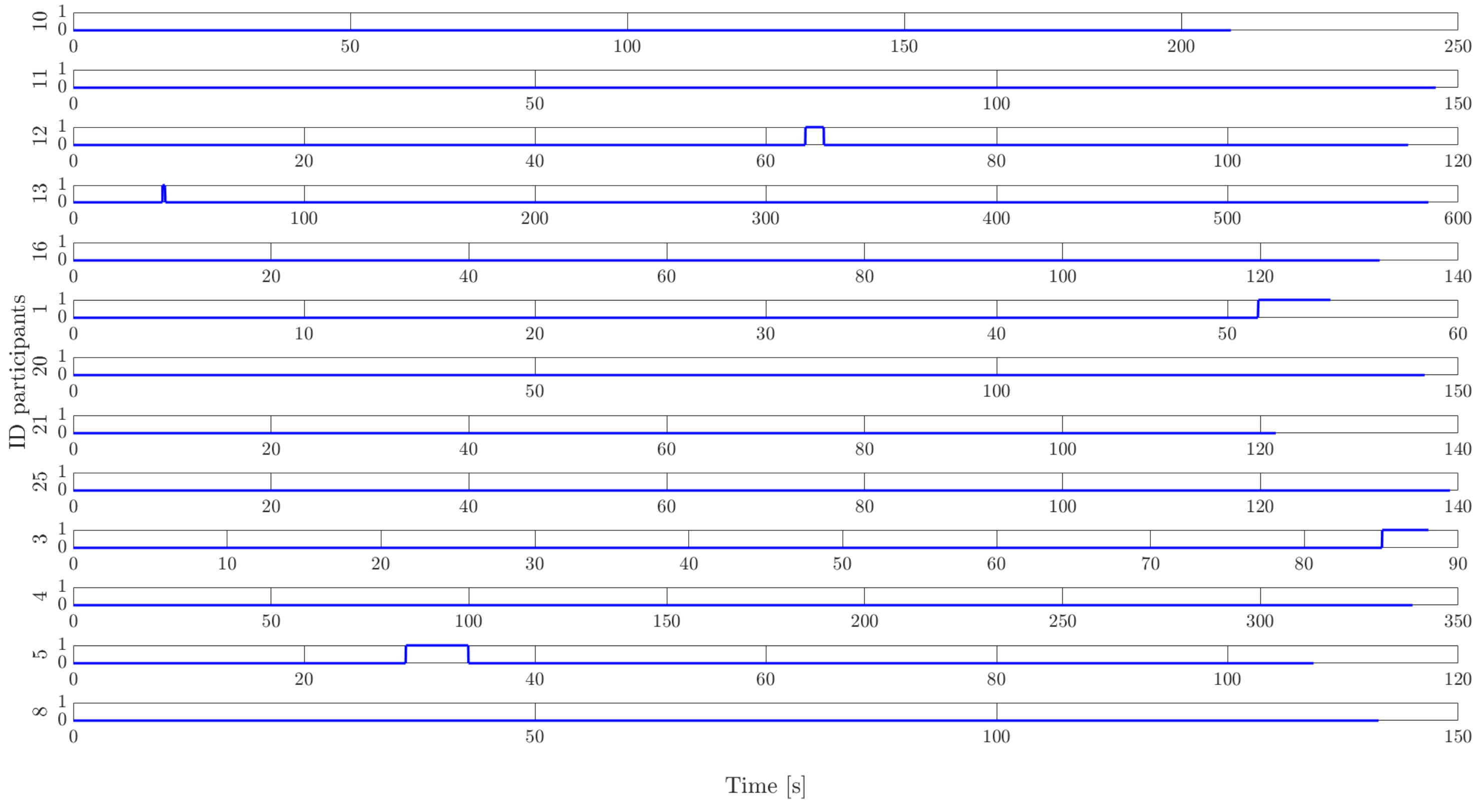


Figure K.4: Contribution of ϕ to the total time spent outside of the envelope.

K.2. Flap Asymmetry scenario

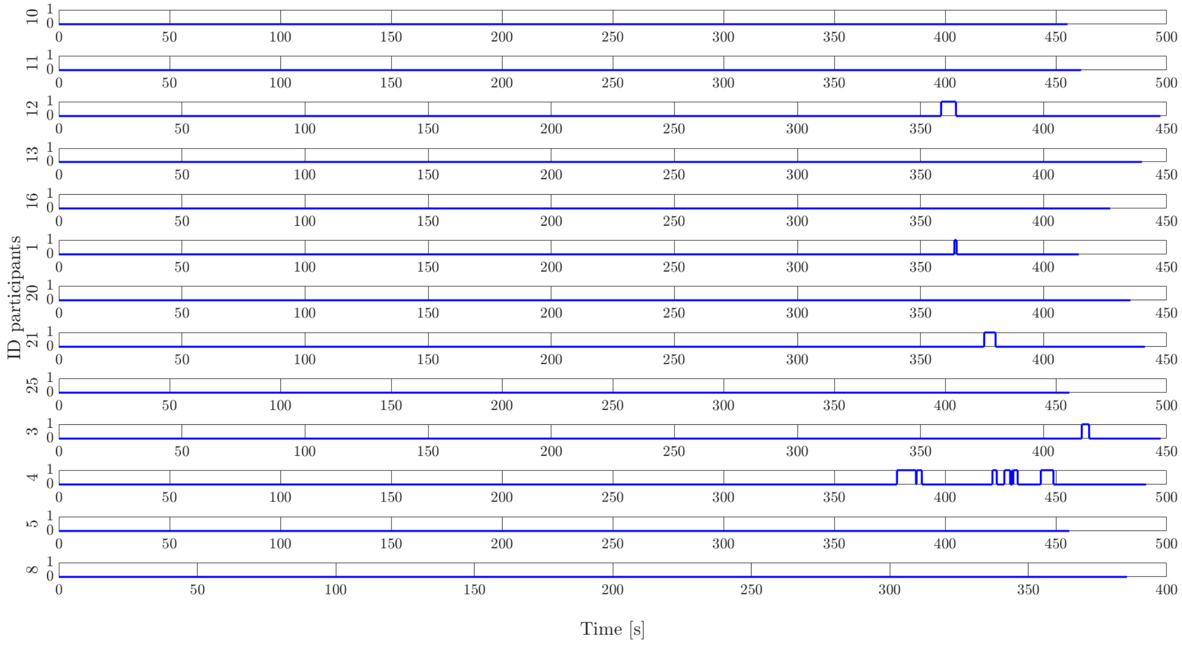


Figure K.5: Total time spent outside the envelope.

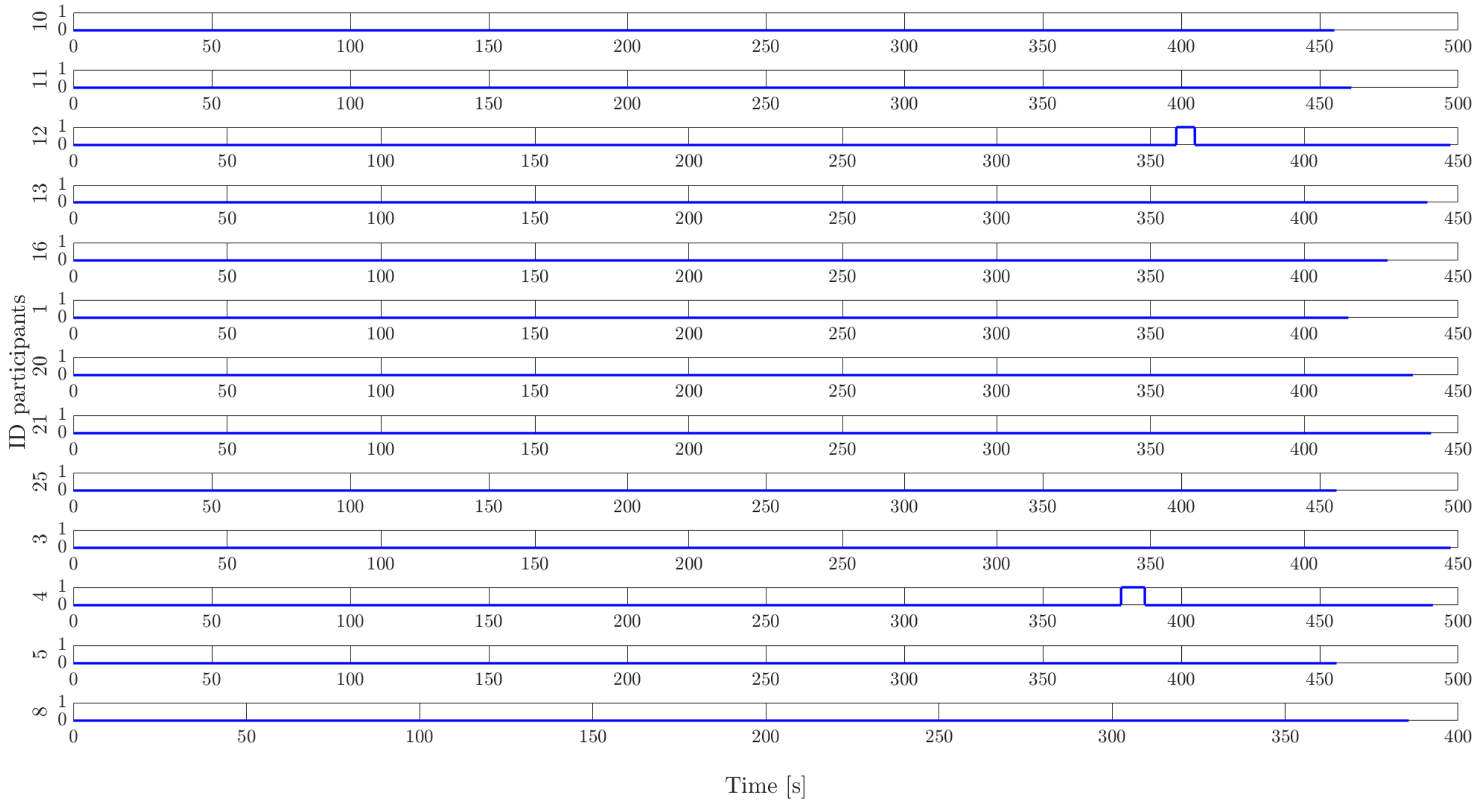


Figure K.6: Contribution of V_{TAS} to the total time spent outside of the envelope.

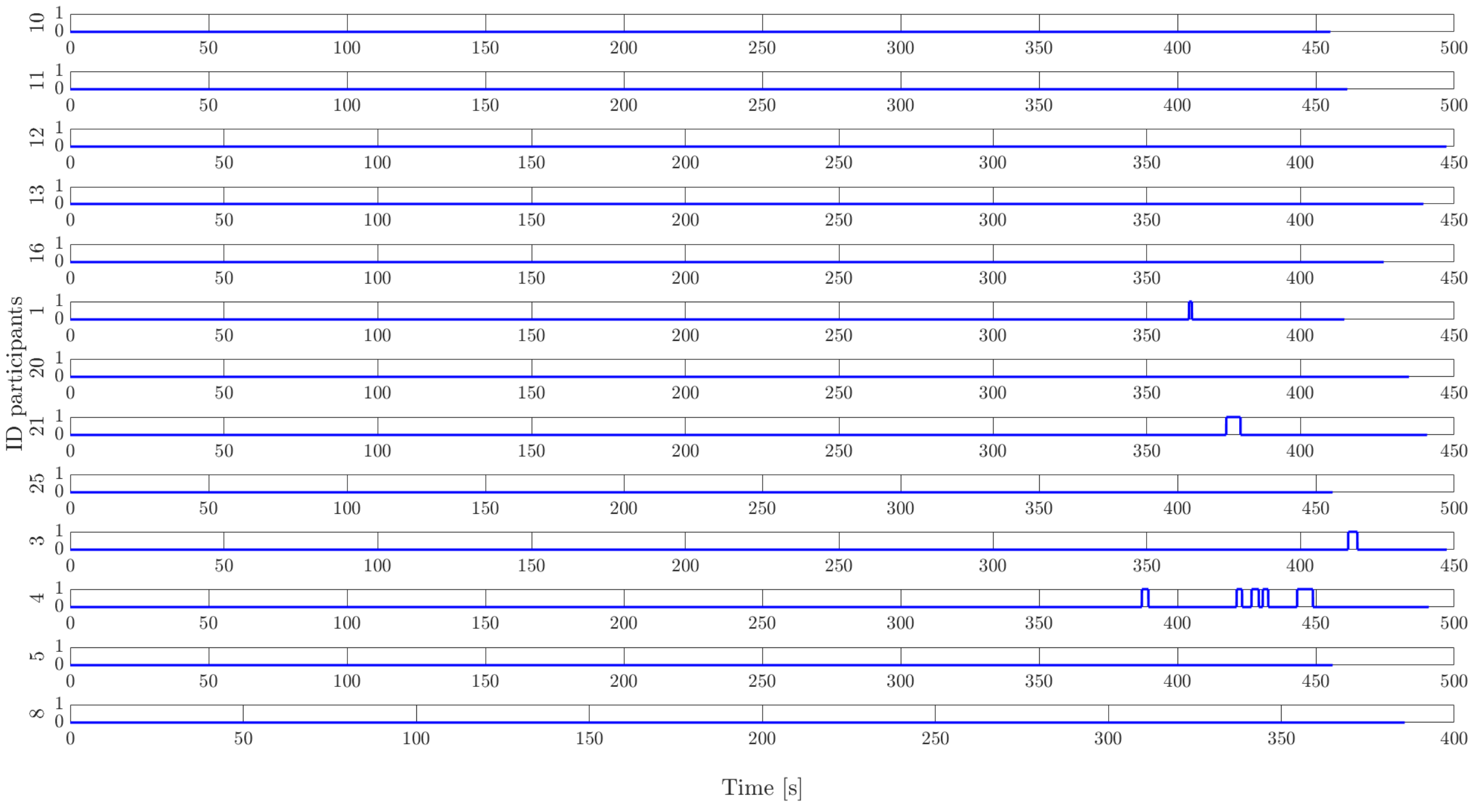


Figure K.7: Contribution of θ to the total time spent outside of the envelope.

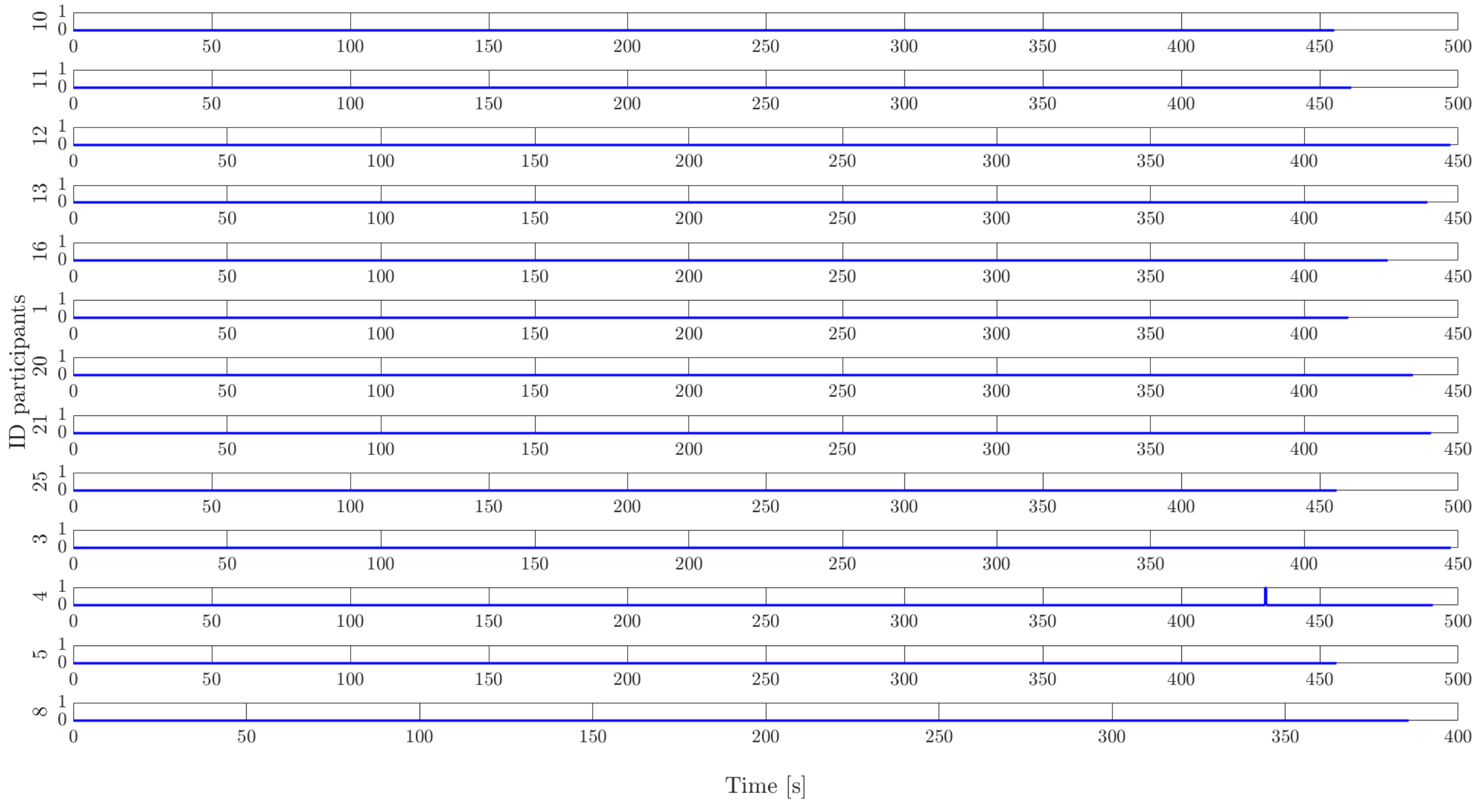


Figure K.8: Contribution of ϕ to the total time spent outside of the envelope.

K.3. Mass Shift scenario

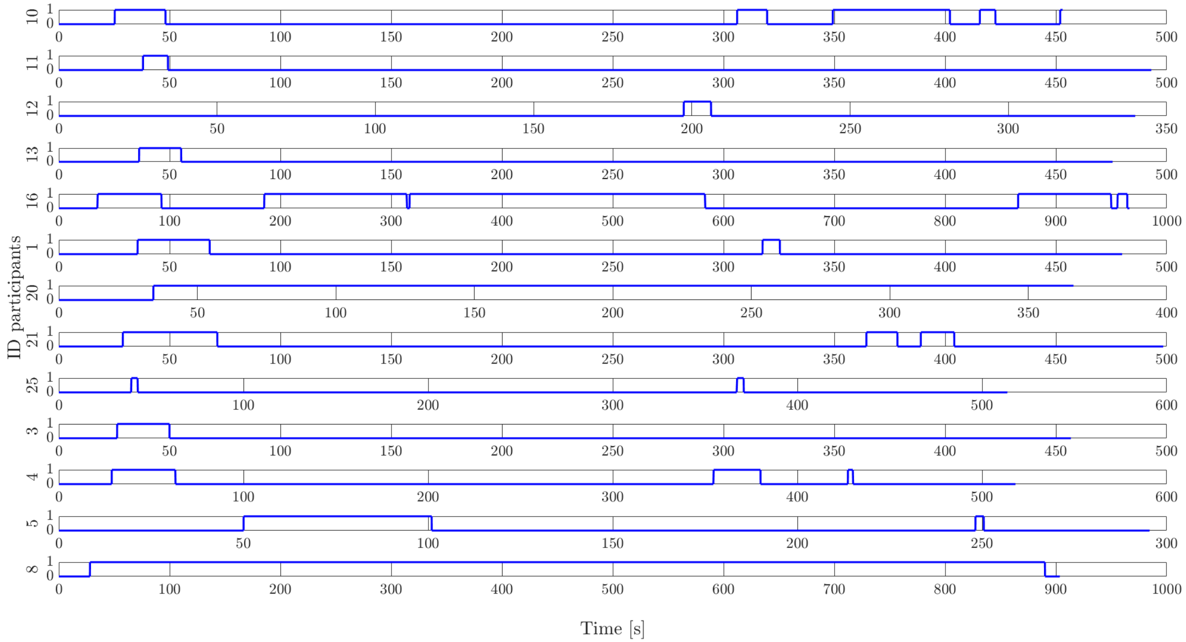


Figure K.9: Total time spent outside the envelope.

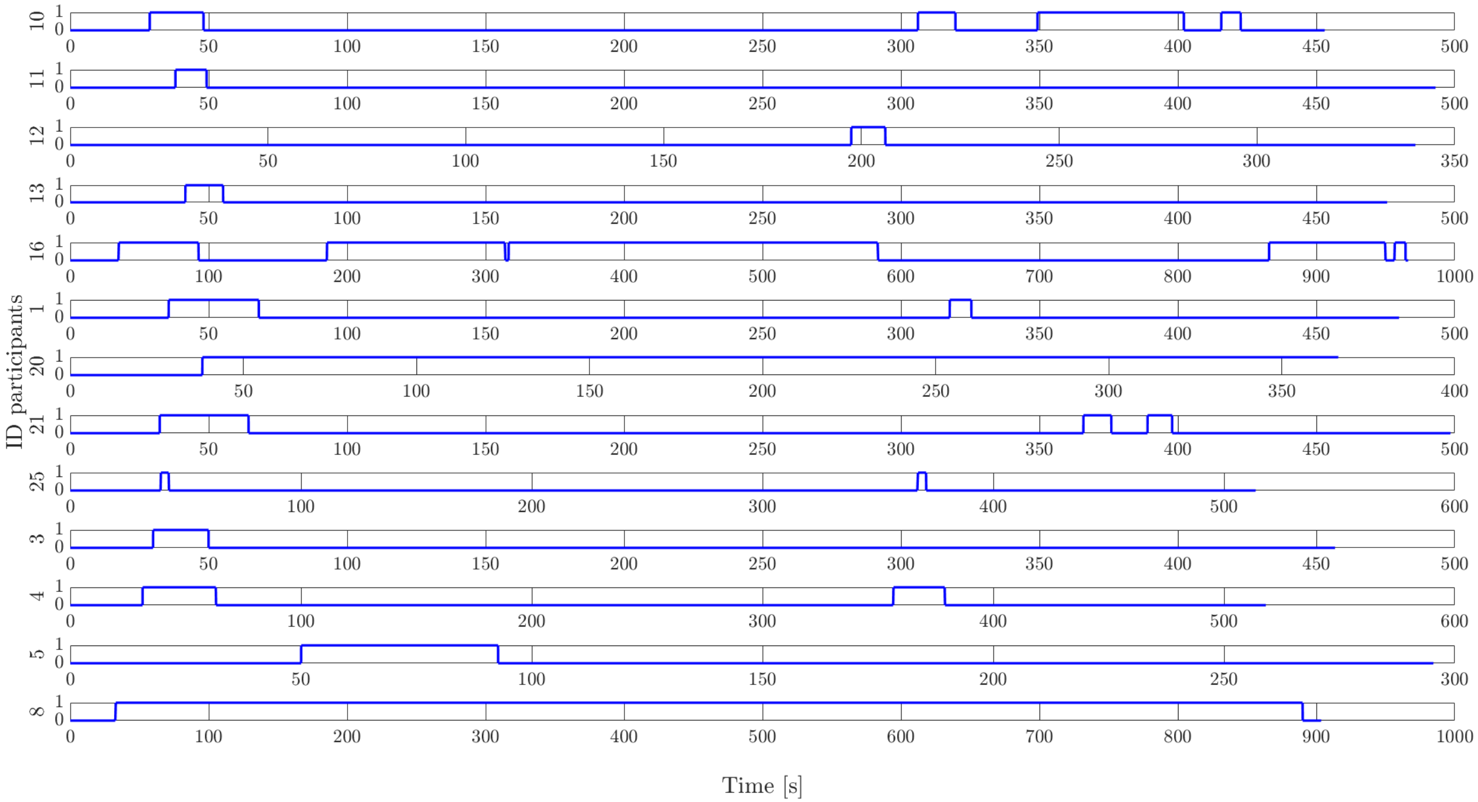


Figure K.10: Contribution of V_{TAS} to the total time spent outside of the envelope.

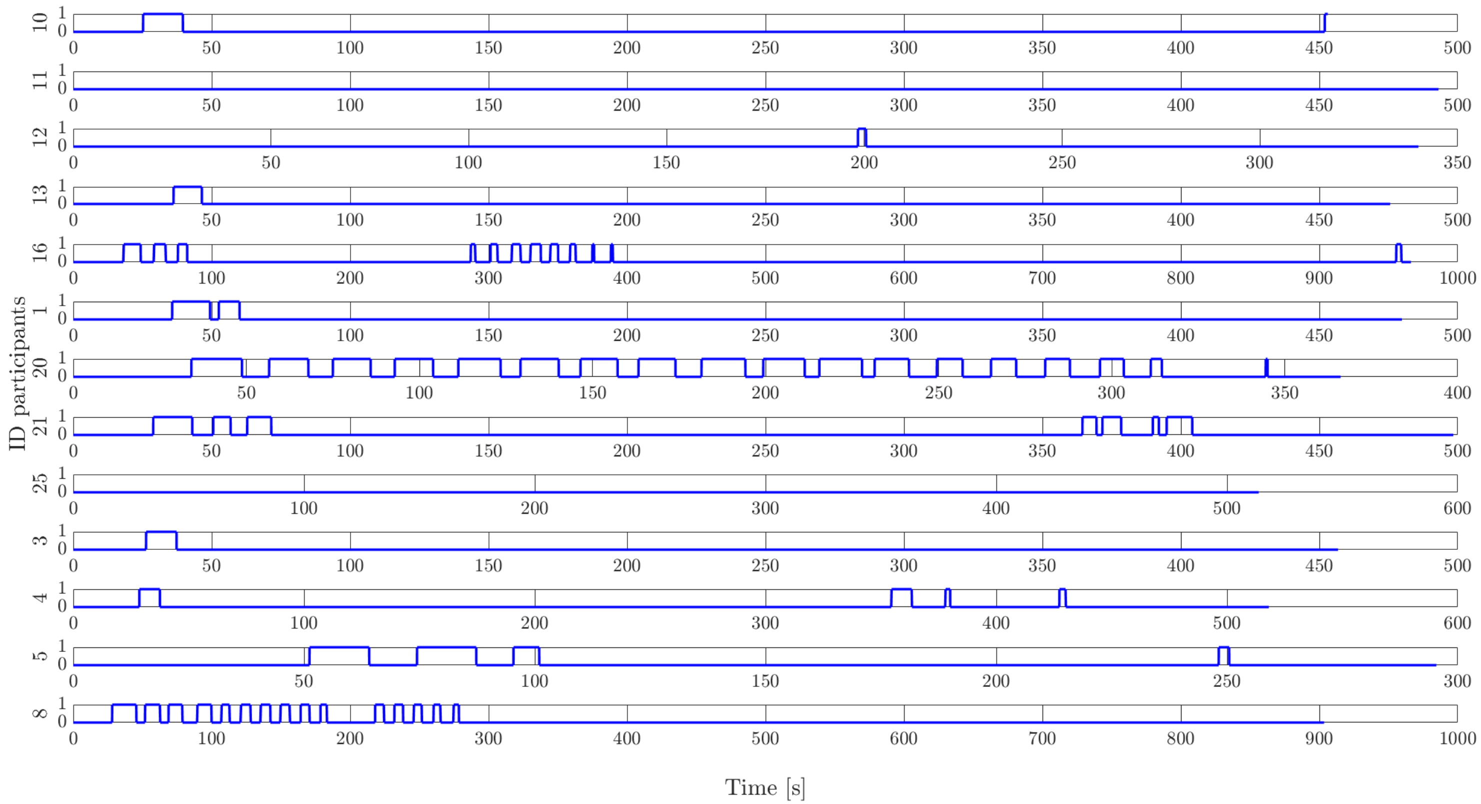


Figure K.11: Contribution of θ to the total time spent outside of the envelope.

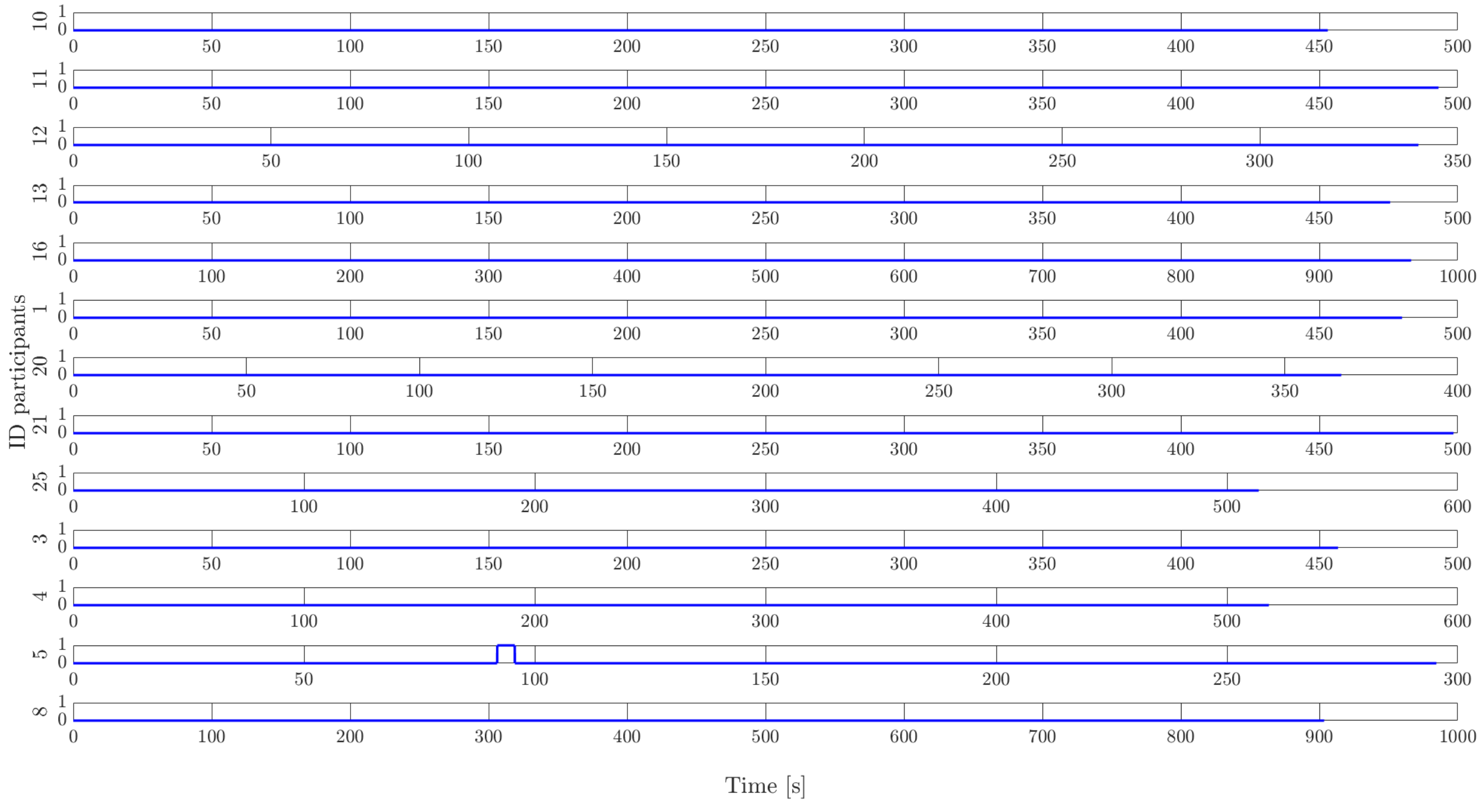


Figure K.12: Contribution of ϕ to the total time spent outside of the envelope.

K.4. OUTCOME OF EACH SCENARIO PER PARTICIPANT

Table K.1: Outcome of each scenario per participant.

ID Participant	Pre-test	Flap Asymmetry	Mass shift
1	Crashes	Lands	Lands
3	Crashes	Lands	Lands
4	Lands	Lands	Lands
5	Lands in a field	Lands	Lands
8	Lands	Lands	Lands
10	Lands	Lands	Crashes (RWY 18R)
11	Lands	Lands	Lands
12	Crashes	Lands	Lands
13	Lands	Lands	Lands
16	Lands	Lands	Scenario stops
20	Crashes	Lands	Scenario stops
21	Lands	Lands	Lands
25	Lands	Lands	Lands

L

Plots of the total time spent outside of the envelope for the control group

L.1. Pre-Test Scenario

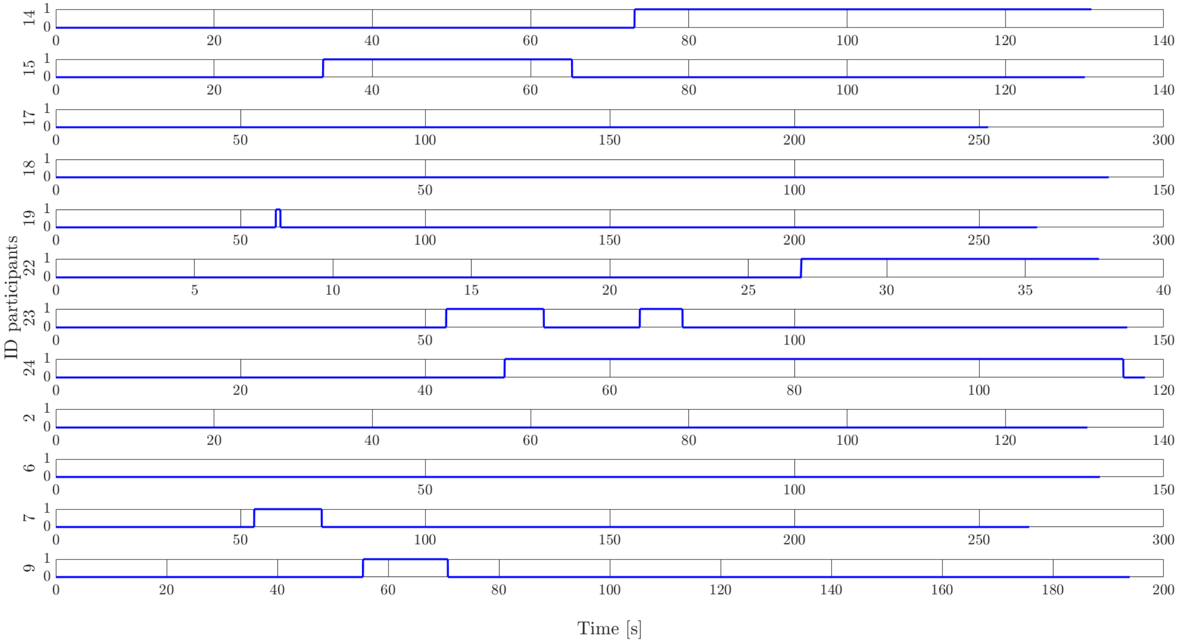


Figure L.1: Total time spent outside the envelope.

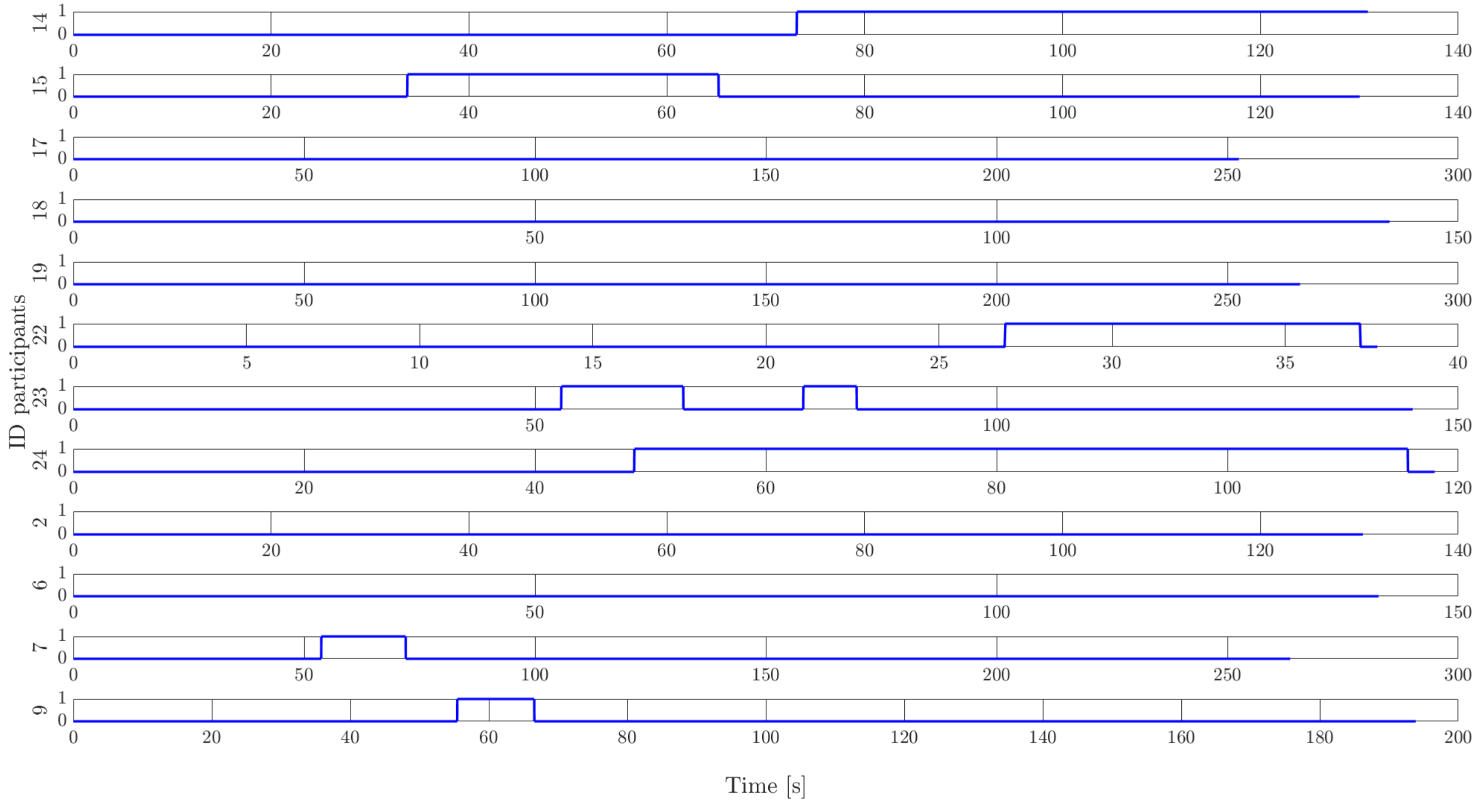


Figure L.2: Contribution of V_{TAS} to the total time spent outside of the envelope.

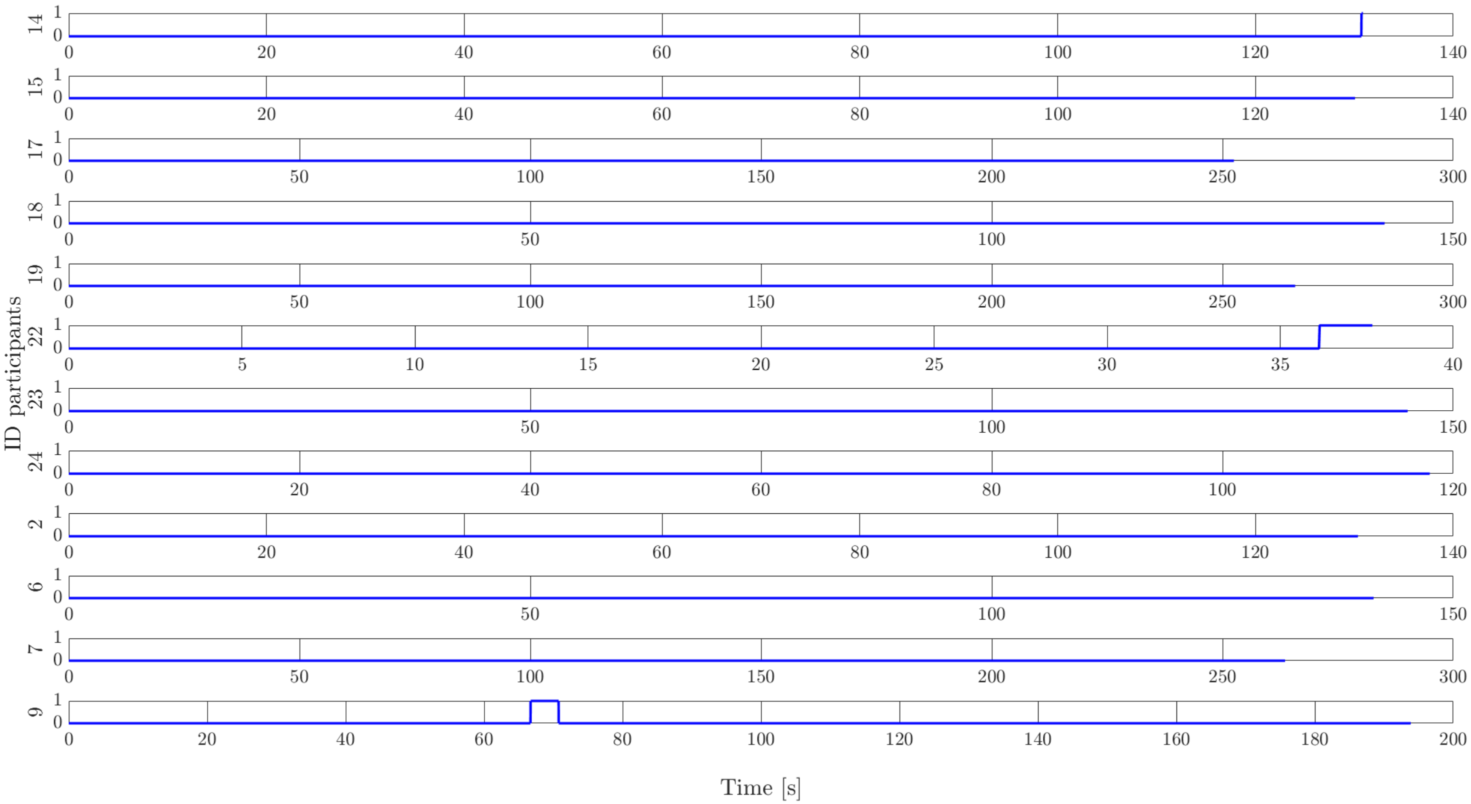


Figure L.3: Contribution of θ to the total time spent outside of the envelope.

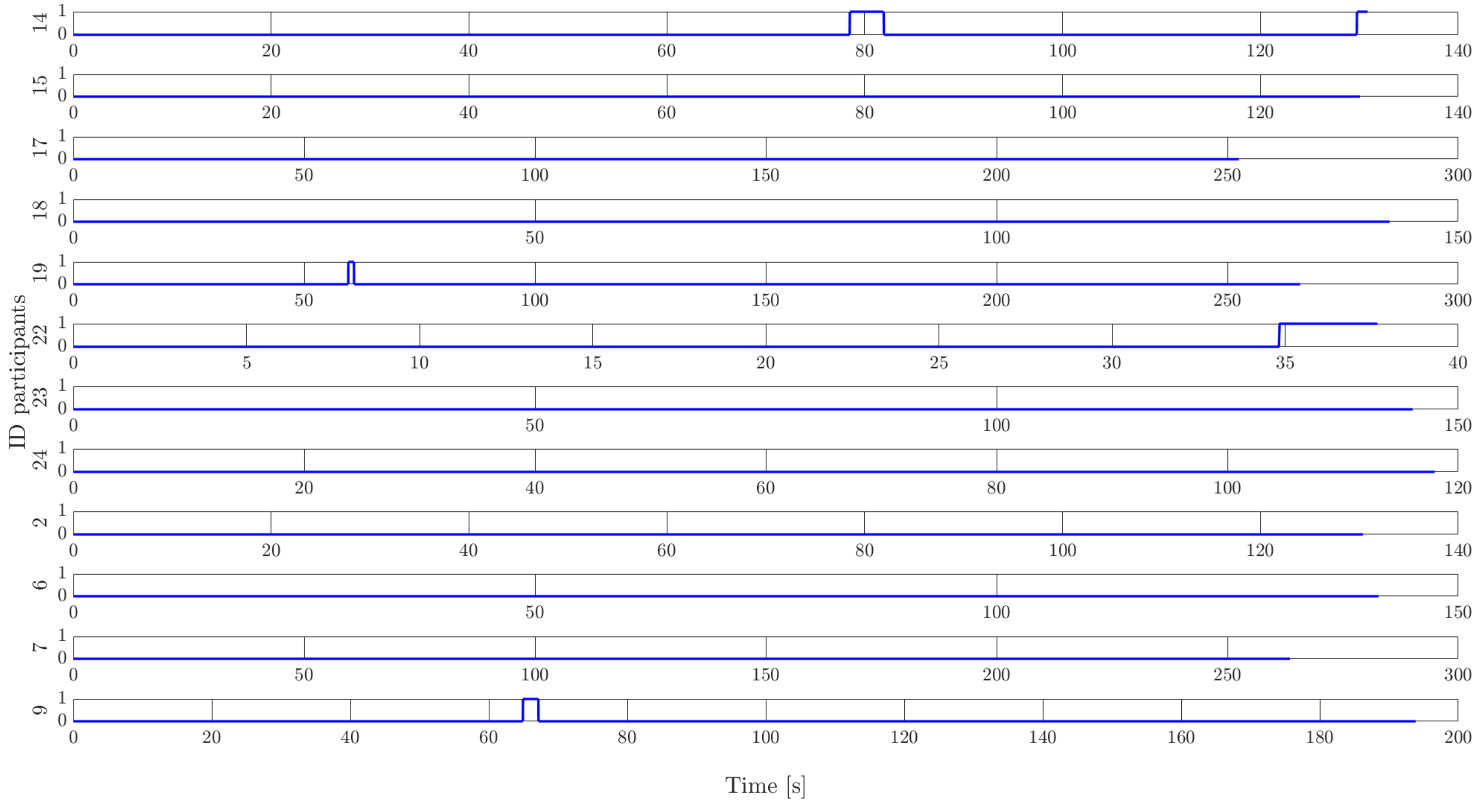


Figure L.4: Contribution of ϕ to the total time spent outside of the envelope.

L.2. Flap Asymmetry scenario

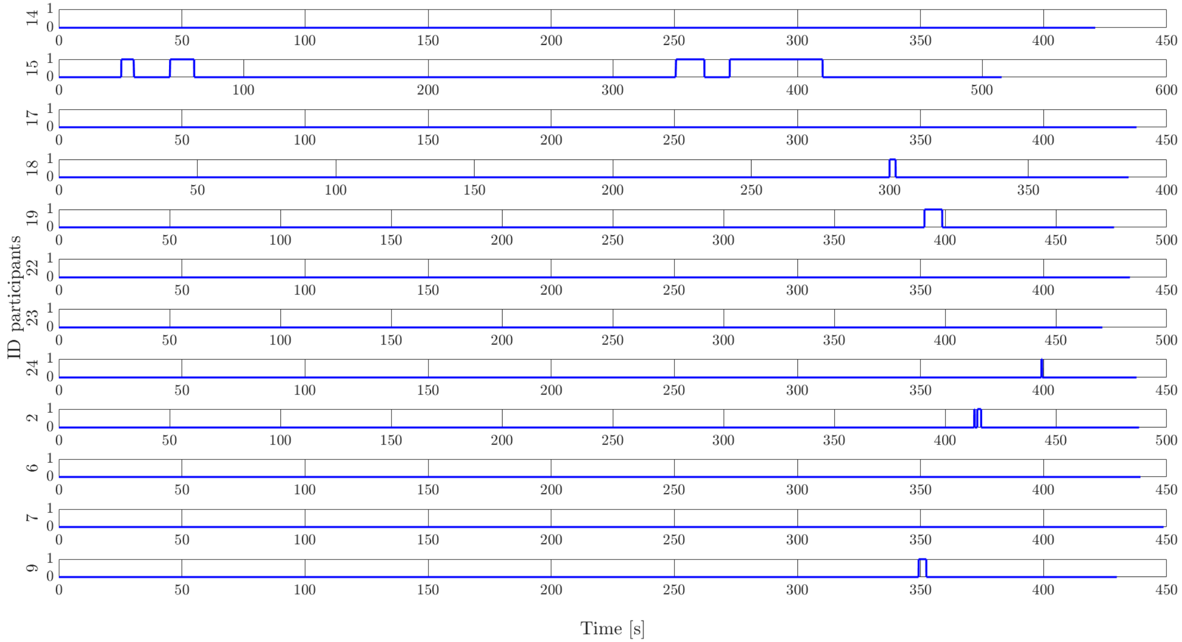


Figure L.5: Total time spent outside the envelope.

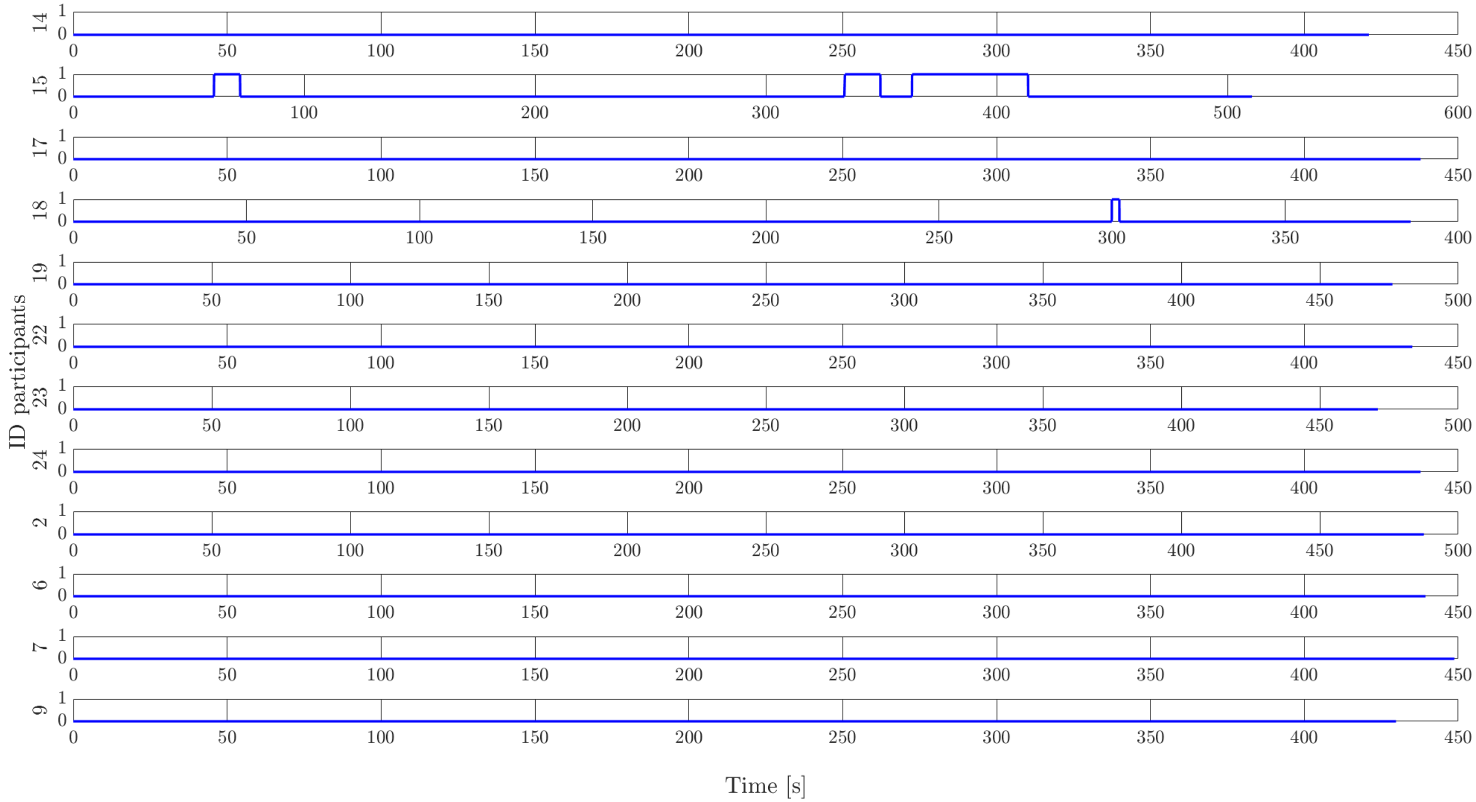


Figure L.6: Contribution of V_{TAS} to the total time spent outside of the envelope.

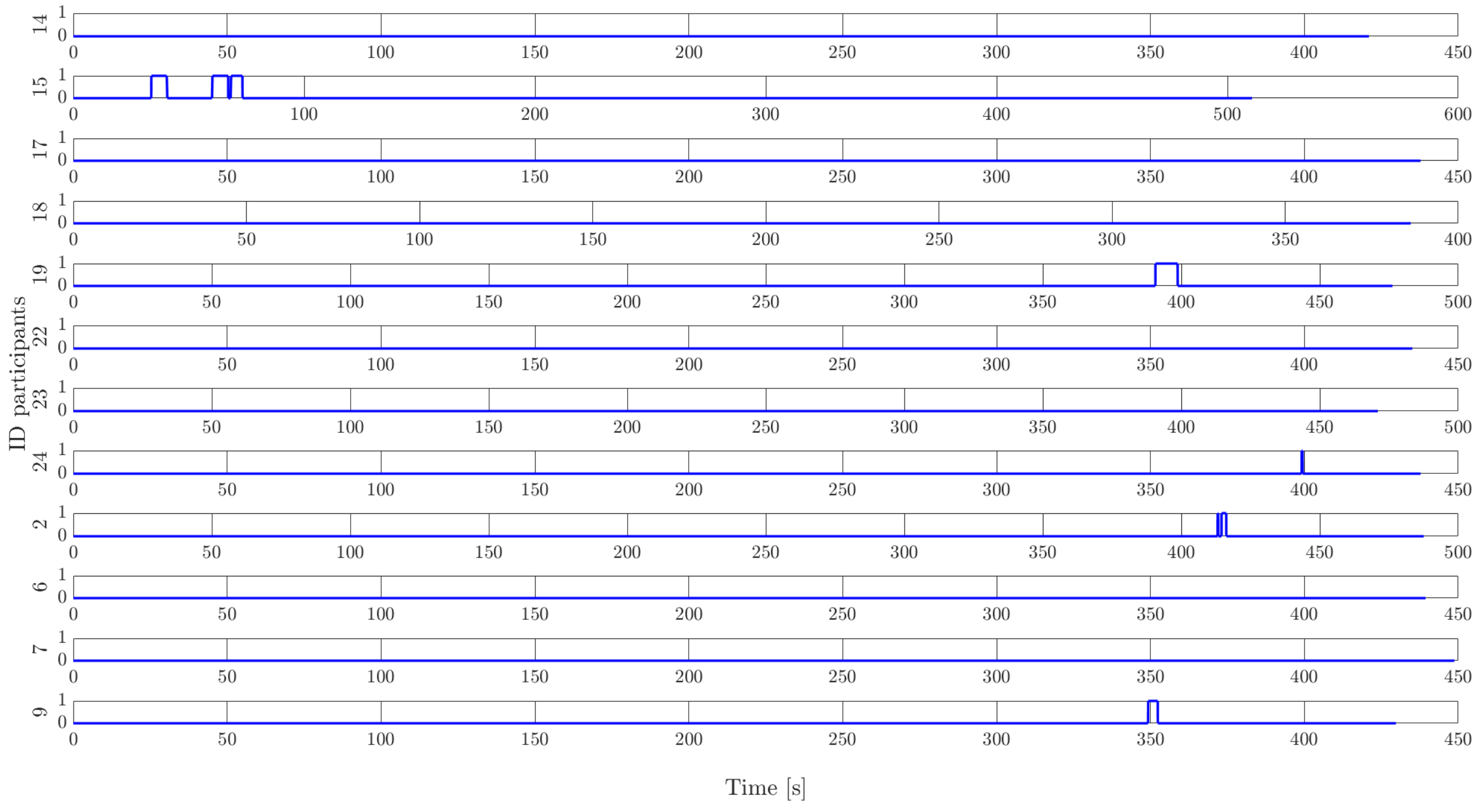


Figure L.7: Contribution of θ to the total time spent outside of the envelope.

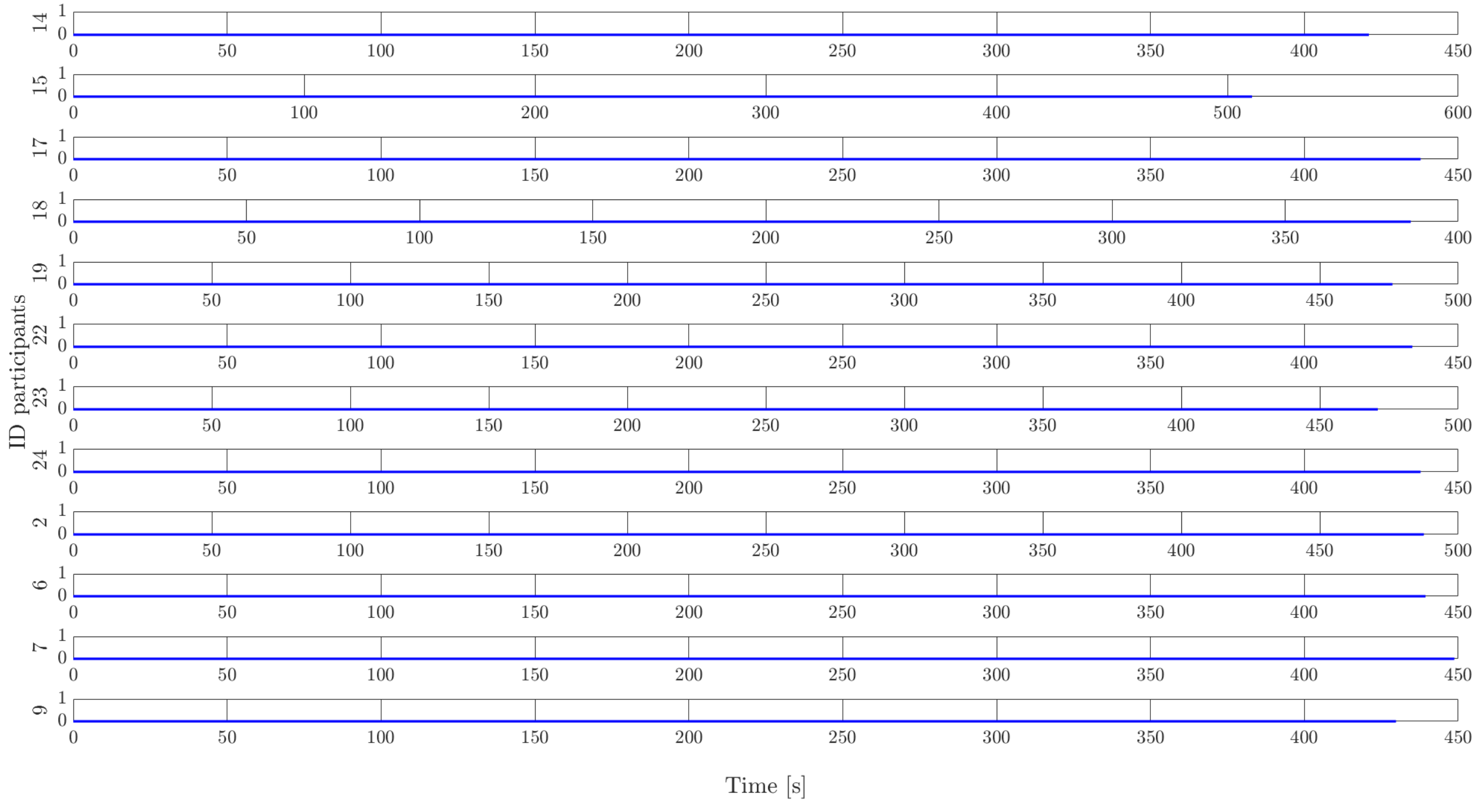


Figure L.8: Contribution of ϕ to the total time spent outside of the envelope.

L.3. MASS SHIFT SCENARIO

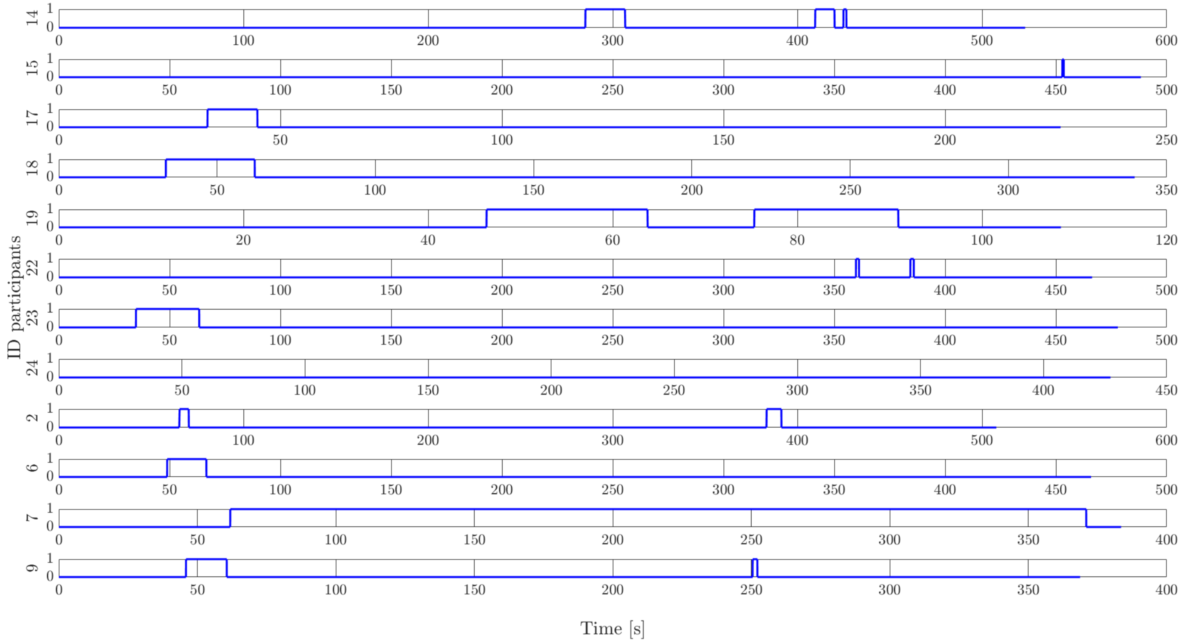


Figure L.9: Total time spent outside the envelope.

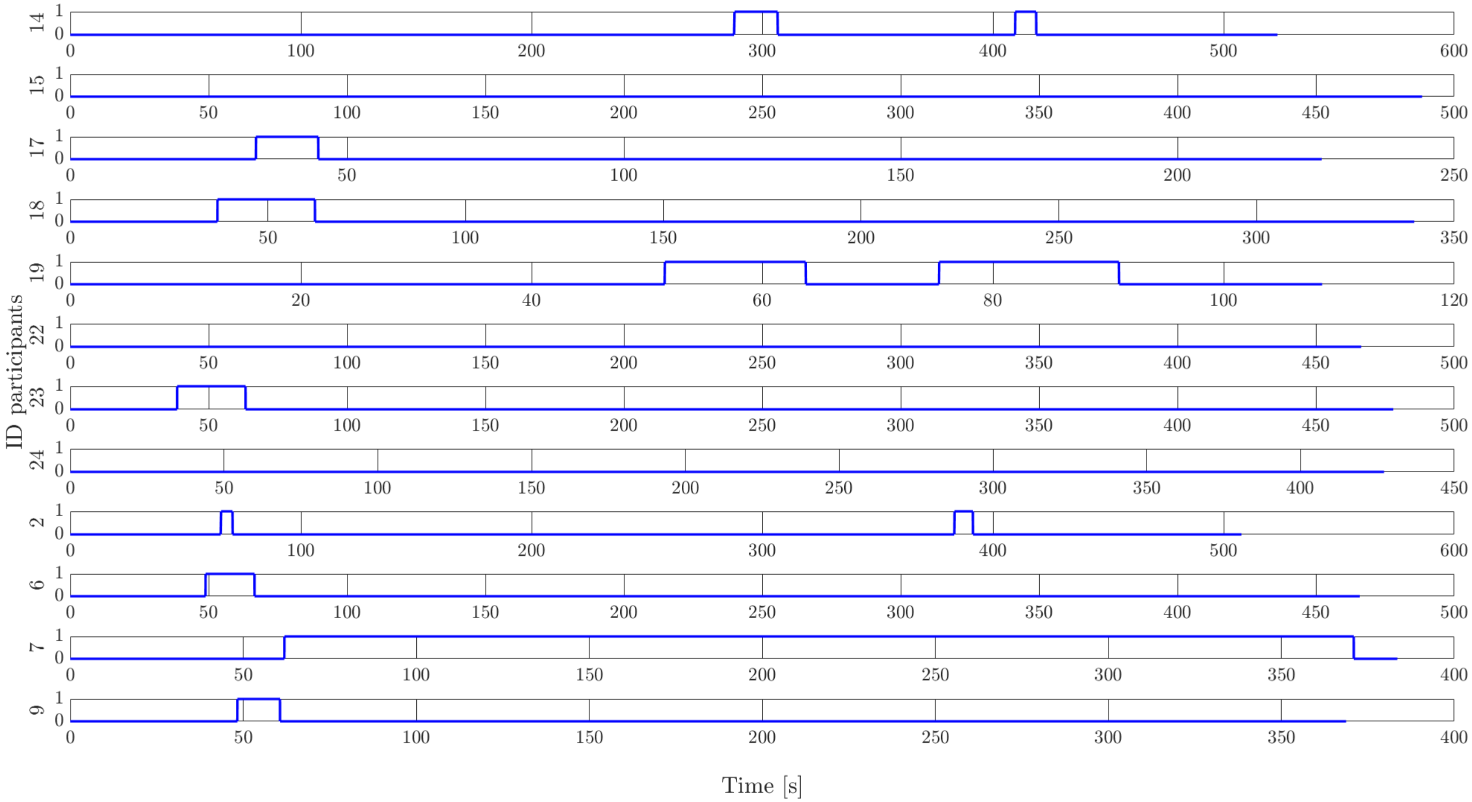


Figure L.10: Contribution of V_{TAS} to the total time spent outside of the envelope.

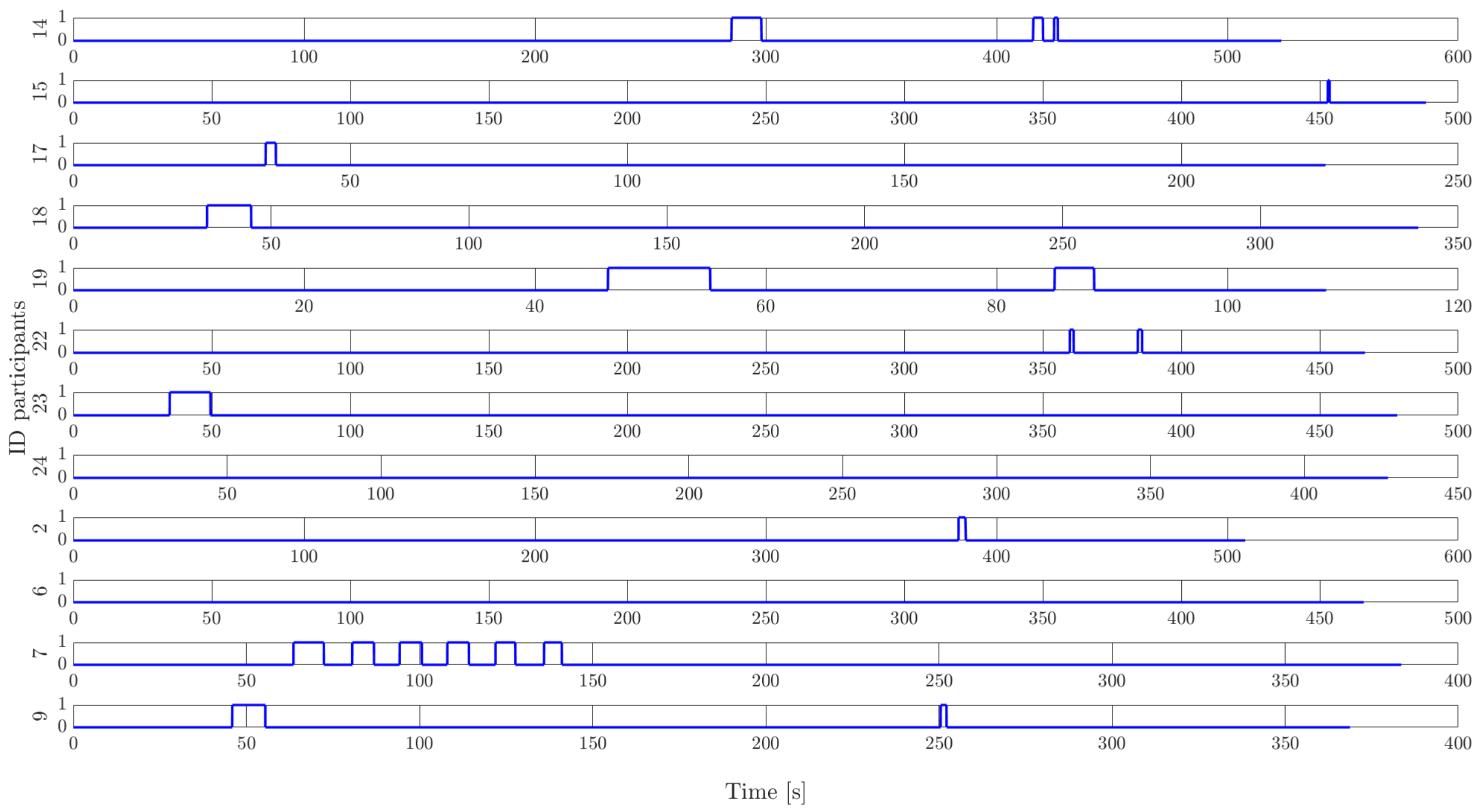


Figure L.11: Contribution of θ to the total time spent outside of the envelope.

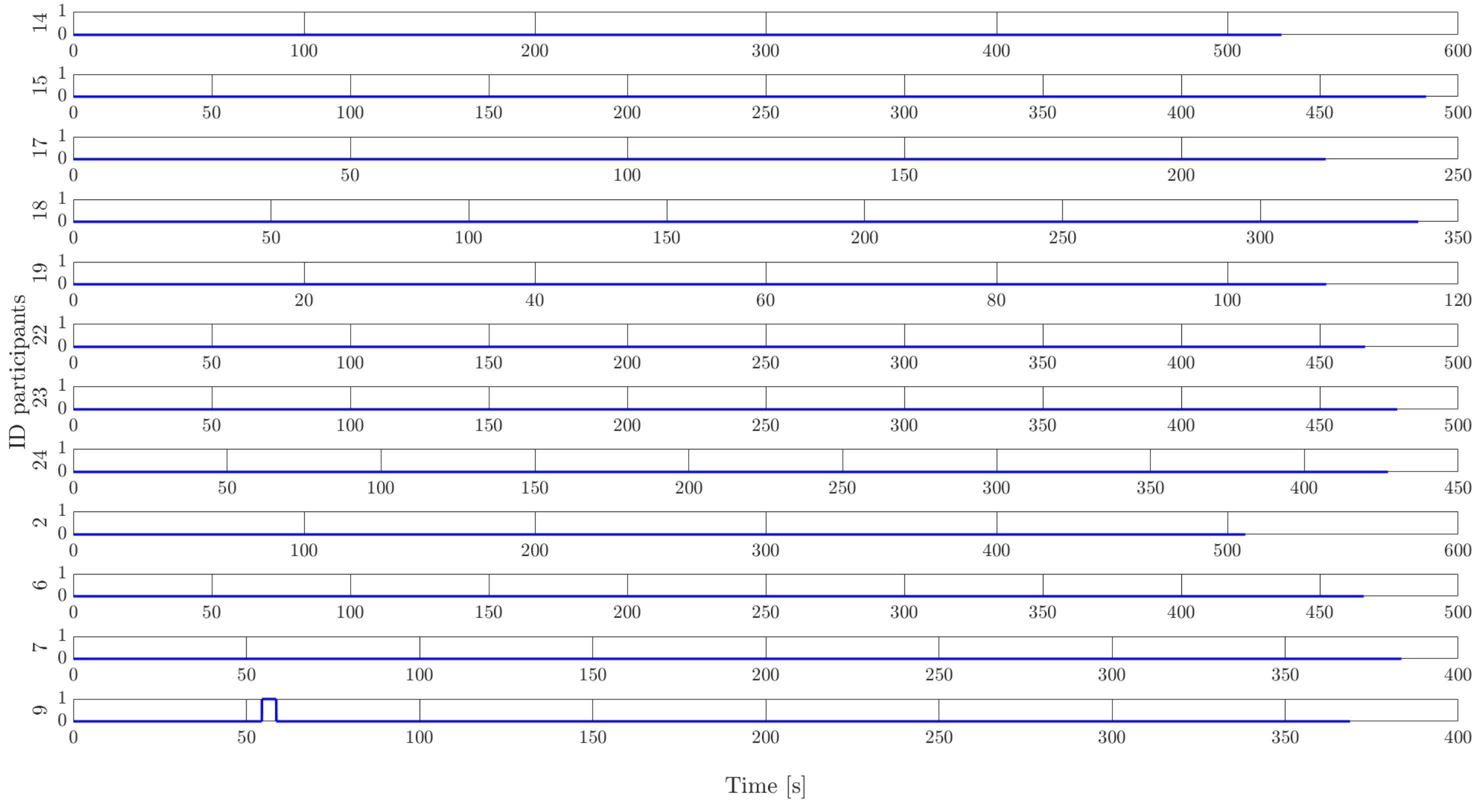


Figure L.12: Contribution of ϕ to the total time spent outside of the envelope.

L.4. OUTCOME FOR EACH SCENARIO PER PARTICIPANT

Table L.1: Outcome of each scenario per participant.

ID Participant	Pre-test	Flap Asymmetry	Mass shift
2	Lands	Lands	Lands
6	Lands	Lands	Lands
7	Lands	Lands	Lands (RWY27)
9	Crashes	Lands	Lands
14	Crashes	Lands	Lands
15	Lands	Lands	Lands
17	Lands	Lands	Lands
18	Lands	Lands	Lands
19	Lands (RWY18R)	Lands	Lands
22	Crashes	Lands	Lands
23	Lands	Lands	Lands
24	Crashes	Lands	Lands

M

PLOTS OF THE TOTAL TIME SPENT OUTSIDE OF THE ENVELOPE FOR THE *COOL* GROUP

M.1. Pre-Test Scenario

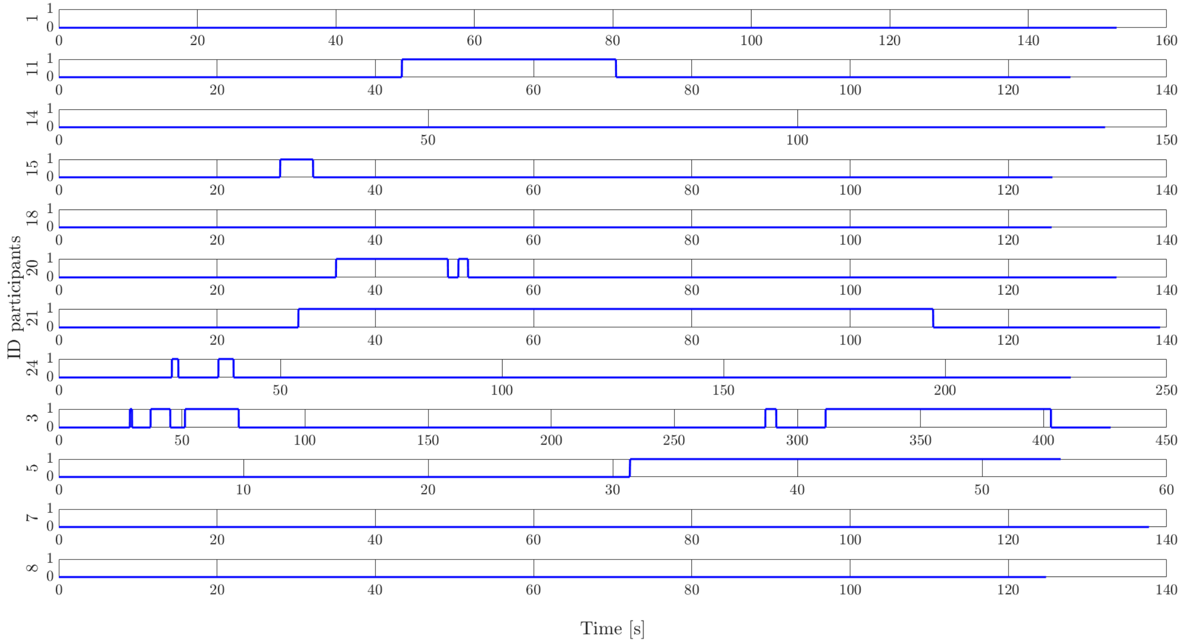


Figure M.1: Total time spent outside the envelope.

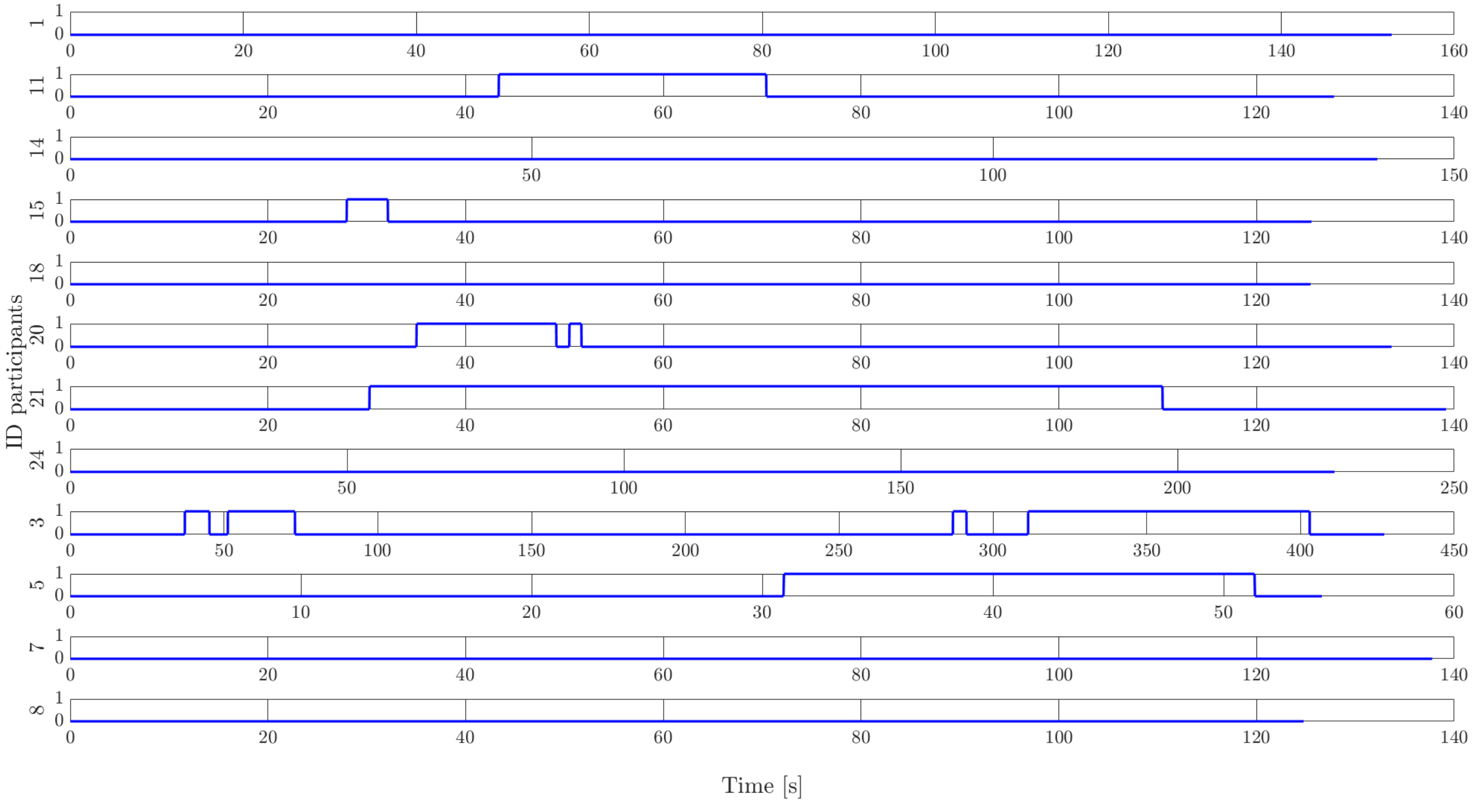


Figure M.2: Contribution of V_{TAS} to the total time spent outside of the envelope.

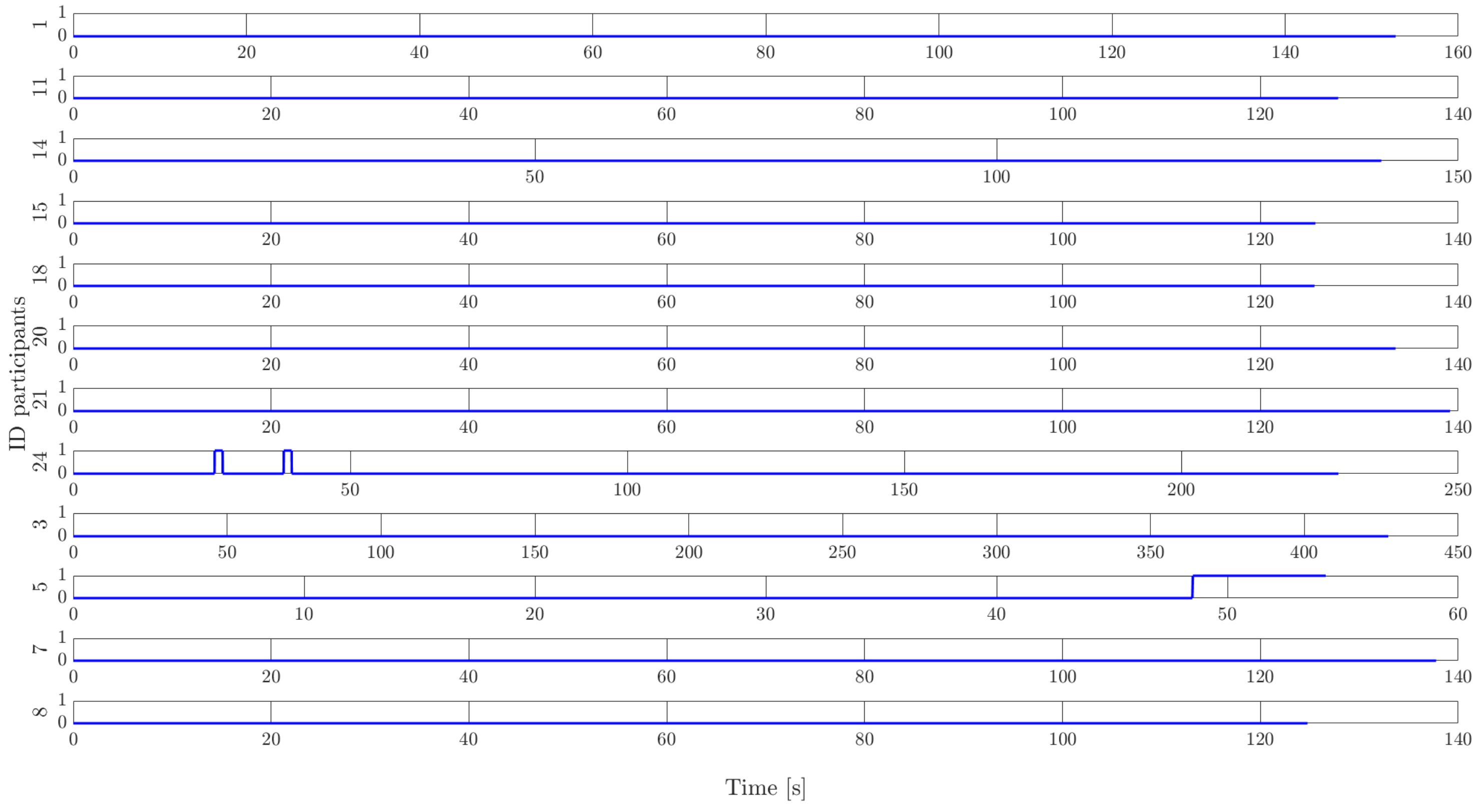


Figure M.3: Contribution of θ to the total time spent outside of the envelope.

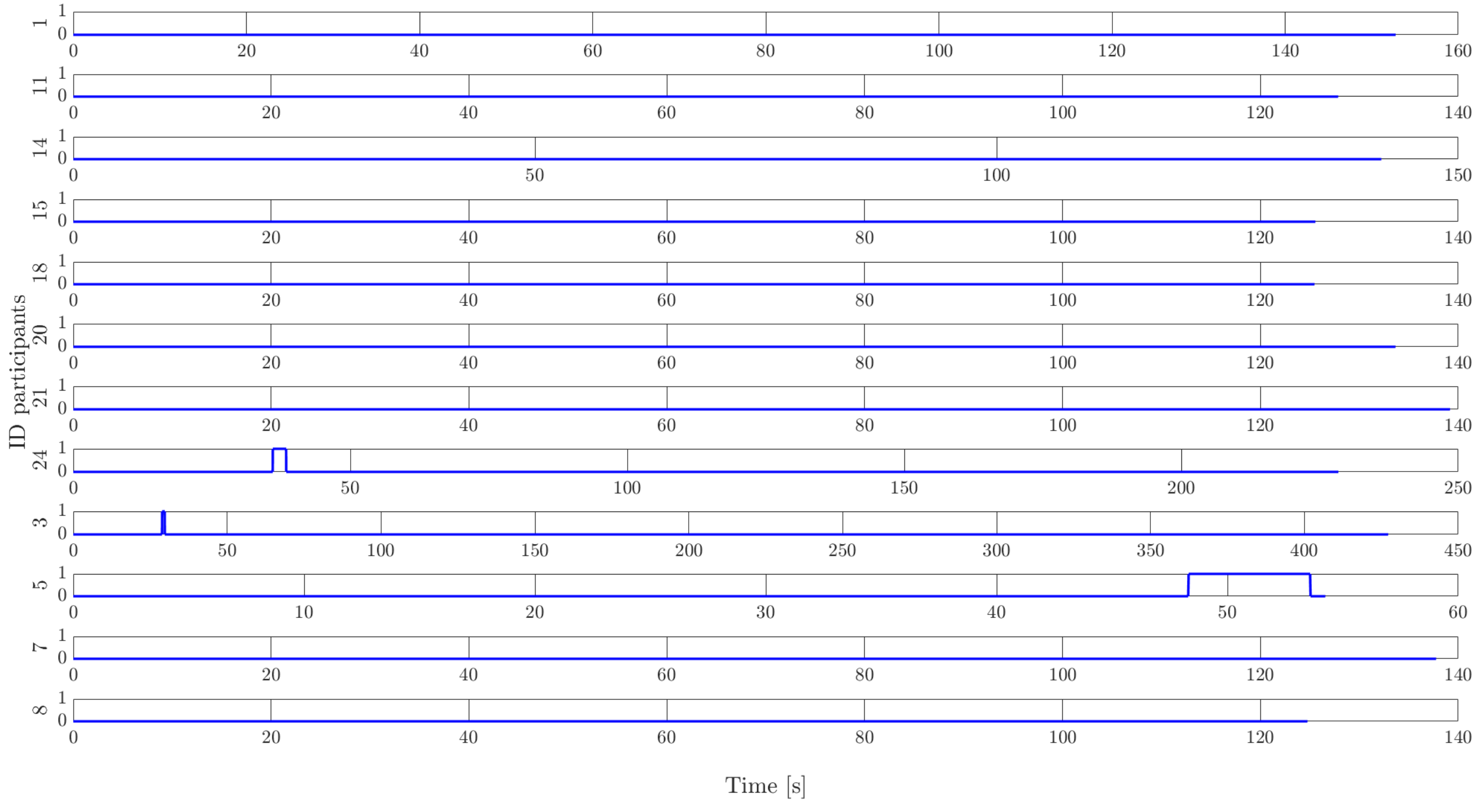


Figure M.4: Contribution of ϕ to the total time spent outside of the envelope.

M.2. Flap Asymmetry scenario

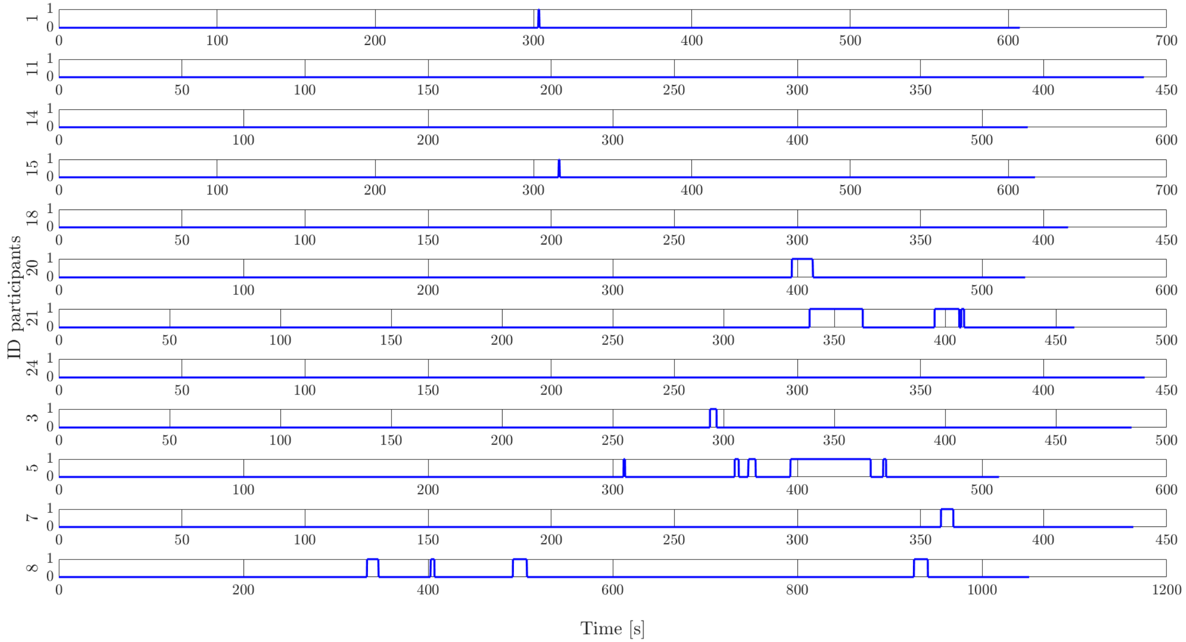


Figure M.5: Total time spent outside the envelope.

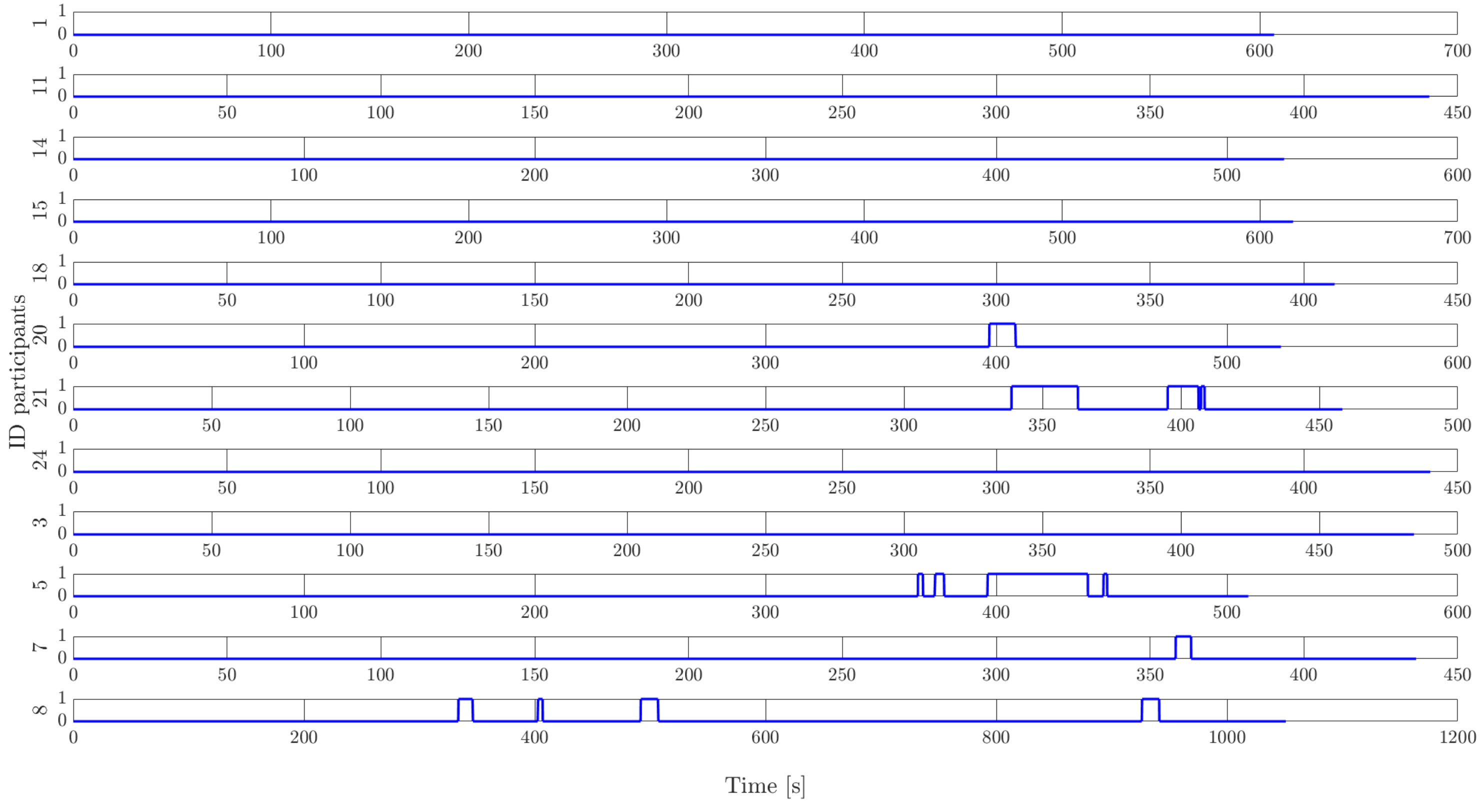


Figure M.6: Contribution of V_{TAS} to the total time spent outside of the envelope.

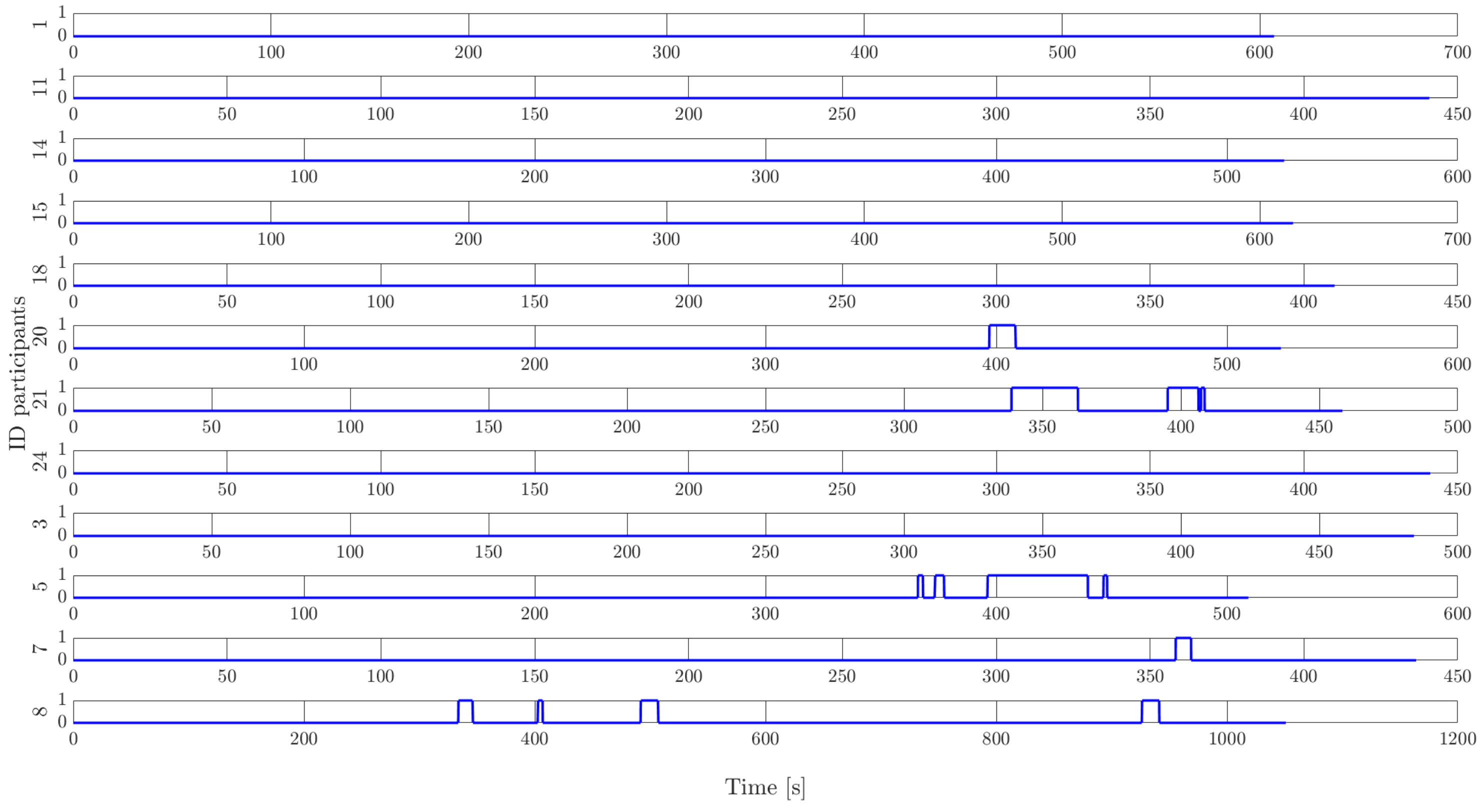


Figure M.7: Contribution of θ to the total time spent outside of the envelope.

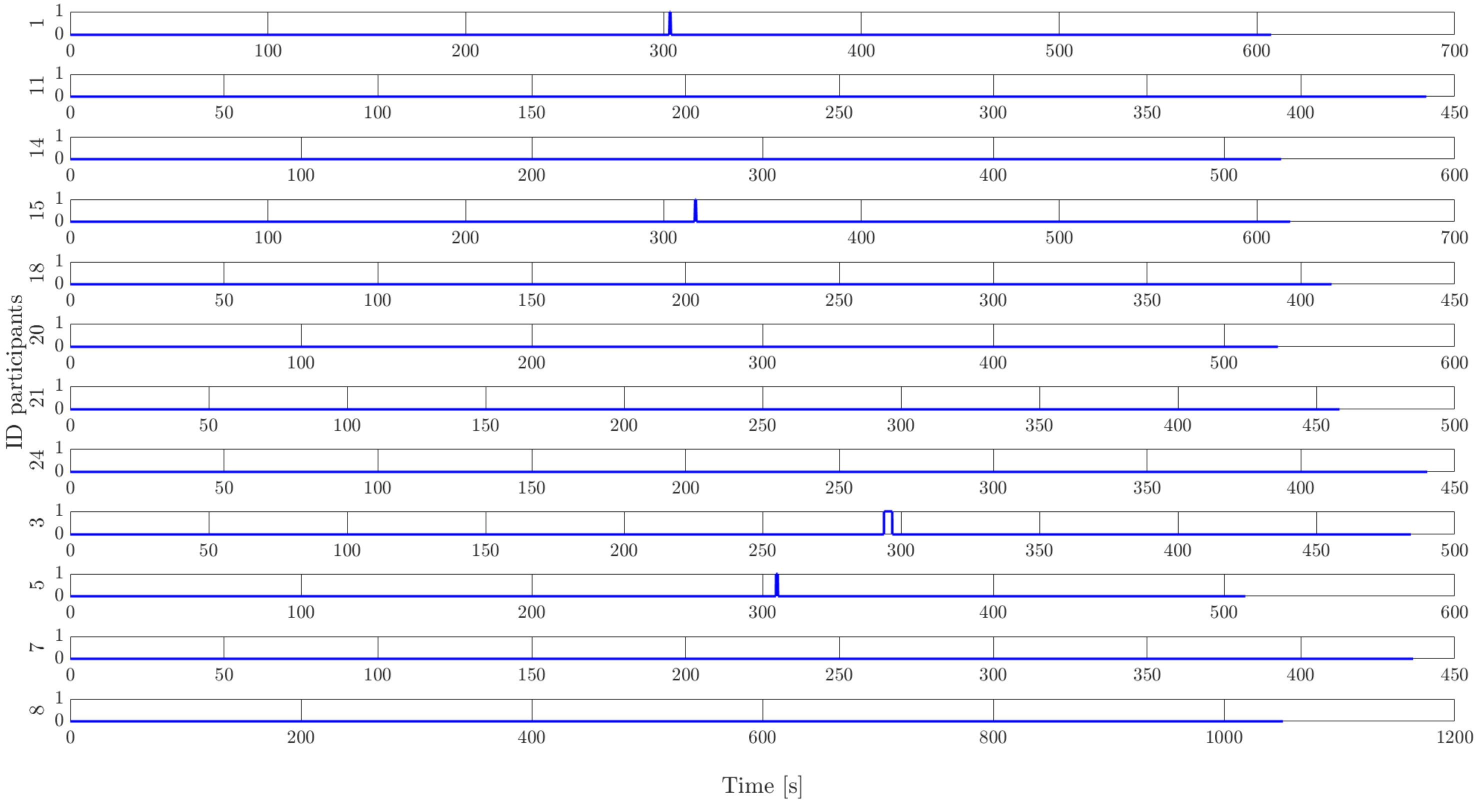


Figure M.8: Contribution of ϕ to the total time spent outside of the envelope.

M.3. Mass Shift scenario

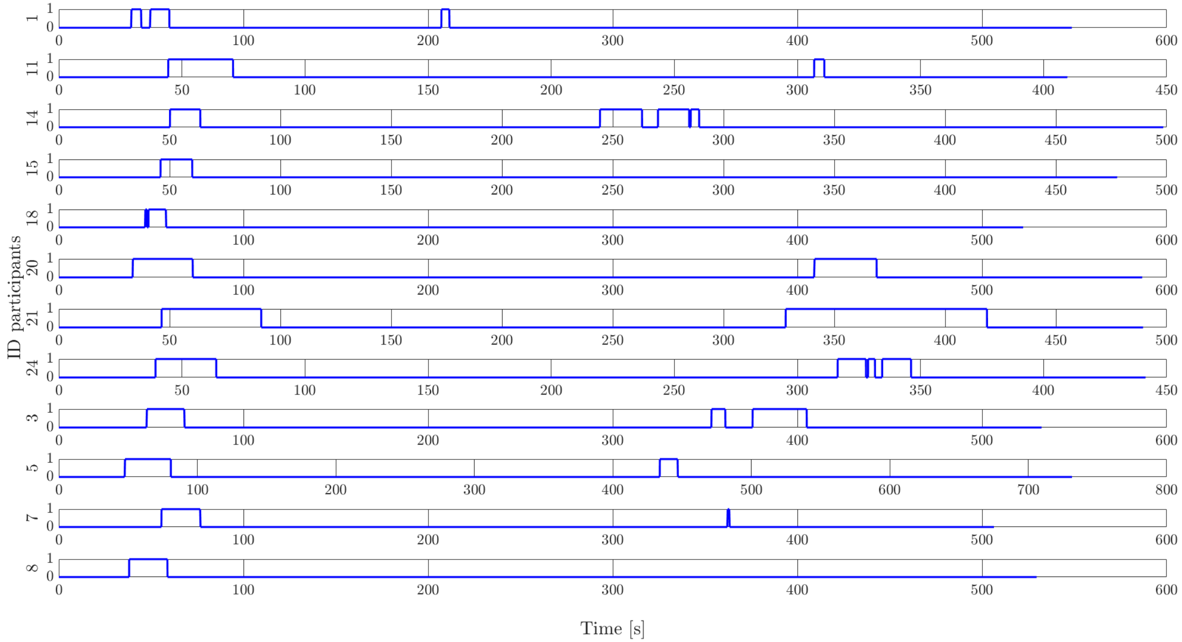


Figure M.9: Total time spent outside the envelope.

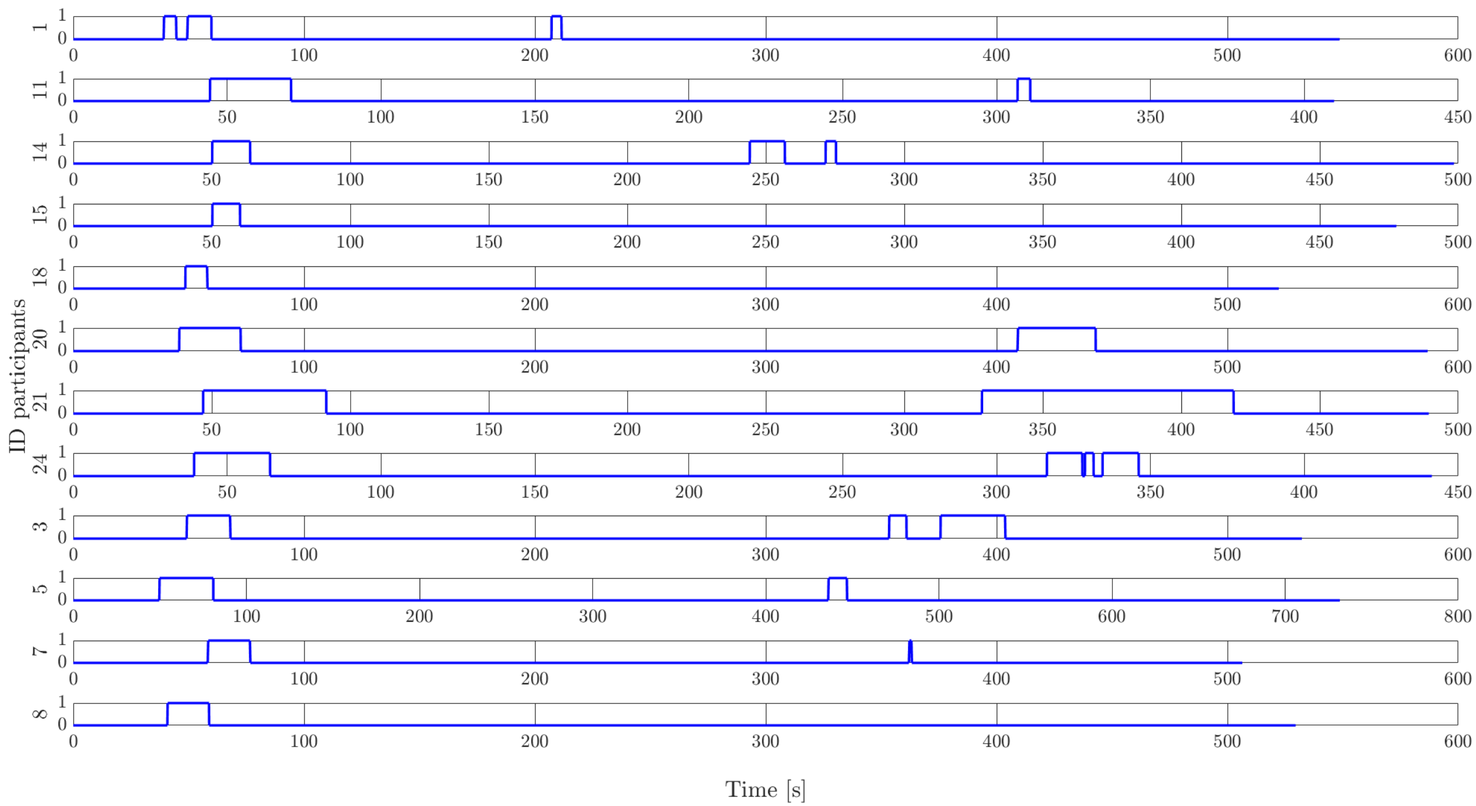


Figure M.10: Contribution of V_{TAS} to the total time spent outside of the envelope.

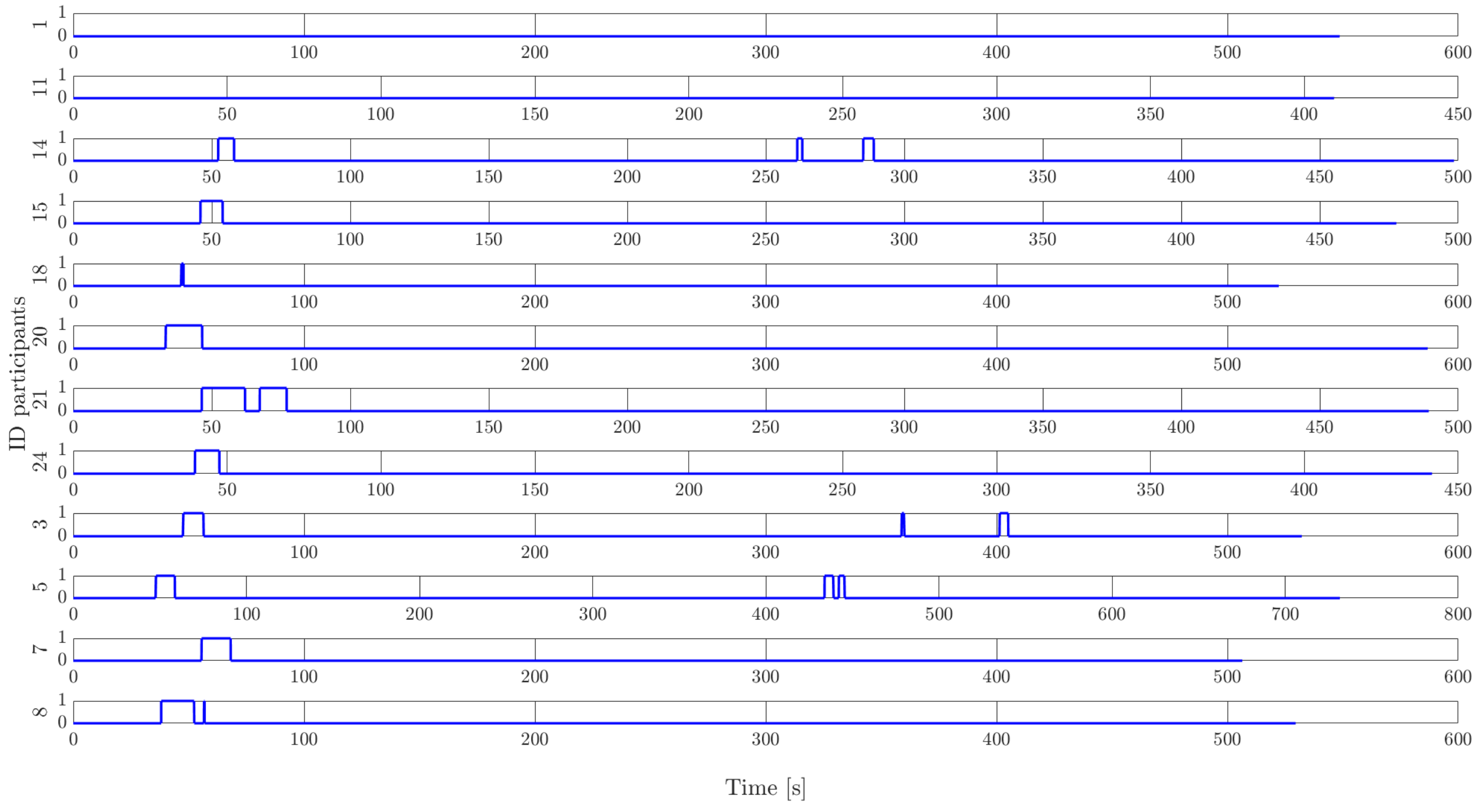


Figure M.11: Contribution of θ to the total time spent outside of the envelope.

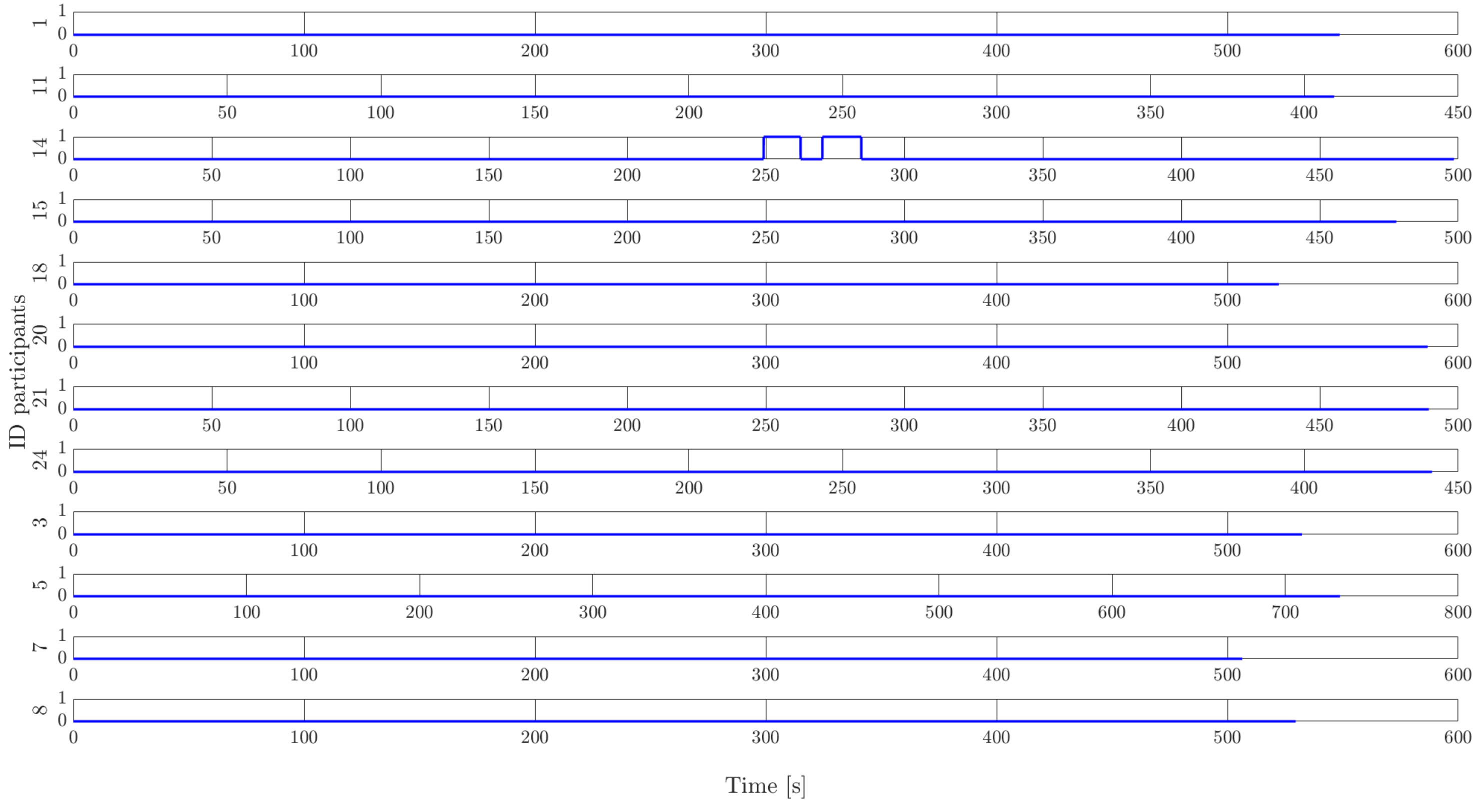


Figure M.12: Contribution of ϕ to the total time spent outside of the envelope.