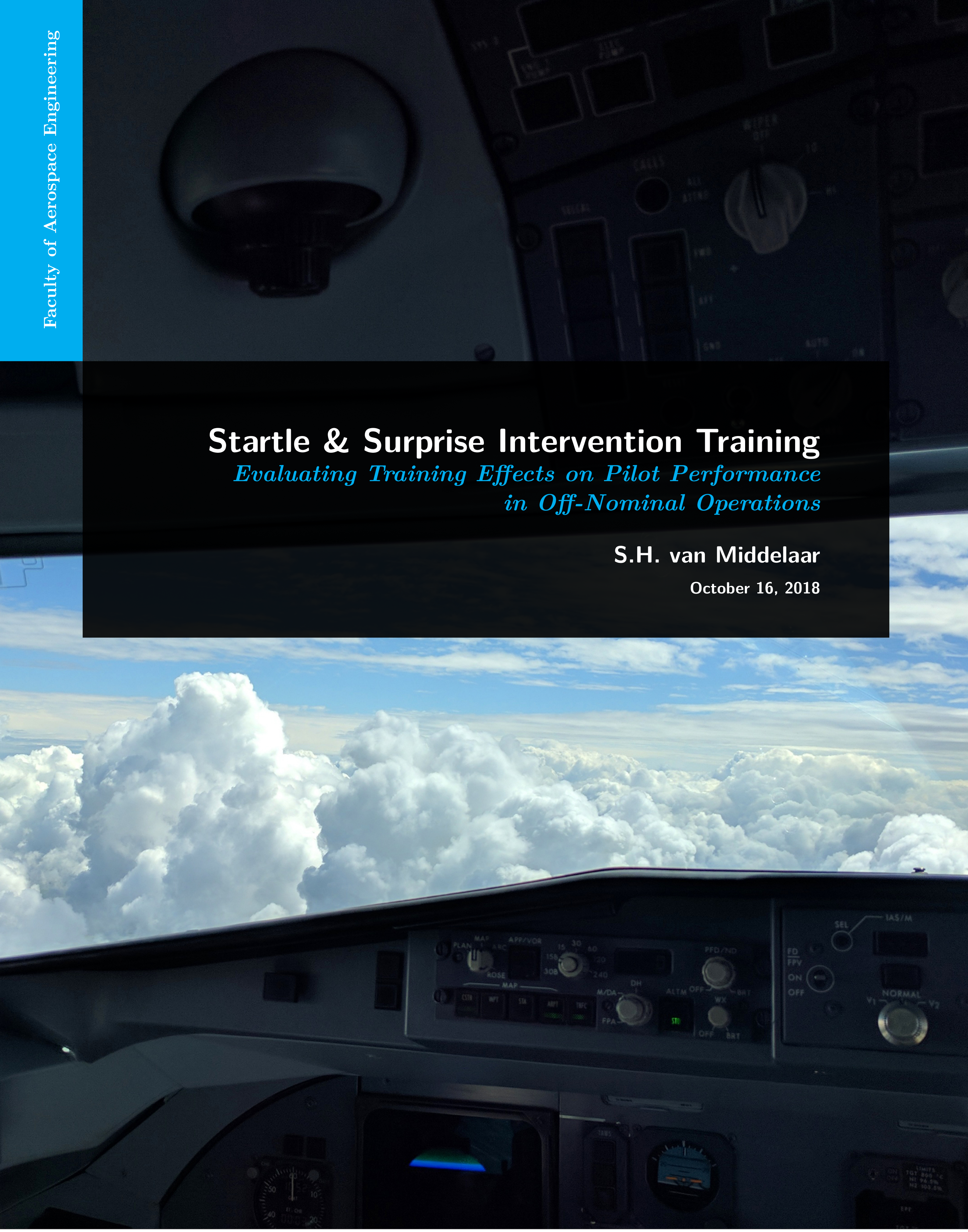
Faculty of Aerospace Engineering



Startle & Surprise Intervention Training

Evaluating Training Effects on Pilot Performance in Off-Nominal Operations

S.H. van Middelaar

October 16, 2018

Delft University of Technology

Challenge the future

Startle & Surprise Intervention Training

Evaluating Training Effects on Pilot Performance in Off-Nominal Operations

Master of Science Thesis

For obtaining the degree of Master of Science in Aerospace Engineering
at Delft University of Technology

S.H. van Middelaar

October 16, 2018

Faculty of Aerospace Engineering · Delft University of Technology

Delft University of Technology

Delft University Of Technology
Department Of
Control and Simulation

The undersigned hereby certify that they have read and recommend to the Faculty of Aerospace Engineering for acceptance a thesis entitled **“Startle & Surprise Intervention Training”** by **S.H. van Middelaar** in partial fulfillment of the requirements for the degree of **Master of Science**.

Dated: October 16, 2018

Readers:

dr. ir. M. M. van Paassen

H.M. Landman, MSc

dr. E. L. Groen

prof. dr. ir. M. Mulder

dr. O. A. Sharpanskykh

List of Figures

List of Tables

Contents

Preface

Before you lies my thesis titled *Startle & Surprise Intervention Training: Evaluating Training Effects on Pilot Performance in Off-Nominal Operations*, based on a human-in-the-loop experiment conducted among twenty-four Dutch airline pilots in the SIMONA Research Simulator. It has been written as part of the graduation requirements of the MSc Aerospace Engineering at Delft University of Technology.

I would like to express my gratitude towards the many people who helped me to finish this 'final project'. First of all, my daily supervisor Annemarie for her support and help during the past year. It was great fun working with you and I wish you all the best finishing your PhD. René, for his guidance and time spent on the DUECA modules to get it all up and running in the SIMONA. Olaf, for spending countless hours getting the SIMONA ready for each run and fixing either the simulator itself or my own coding mistakes. And not to mention both of them for installing this awesome surround sound system in the SIMONA, making my simulation way more startling and realistic compared to the headset mono sound previously used. Eric and Max, for their useful comments and observations. Dirk, for testing the experimental scenarios. Herman and Peter, for developing and expanding the Seneca III model. Alexei, for being able to join my graduation committee last minute.

Many thanks to all the pilots who participated so generously in my experiment, all just for a single bottle of wine! You have no idea how much of a relief it was to find enough participants, despite sending emails and messages to probably every Dutch pilot alive. I really appreciate all of your comments and suggestions.

Thank you Ronald, for explaining working with SPSS, DUECA and providing me with general advice on my thesis. Sorry for not including the serialist/holist theory in my thesis and denying you a precious reference!

After seven in years in Delft, including six years at the Aerospace Faculty, it would be quite foolish to not give a big 'shout-out' to my parents. Besides making these years in Delft possible, they have always encouraged me to follow my dreams (even when this meant spending time at the airfield flying gliders, instead of attending lectures - I'm so glad I can finally admit this now!). As puzzling as it seems, they did not even resist - well, maybe a little - when I revealed I wasn't exactly planning on finding a job just yet, but instead wanted to pursue my longtime goal of becoming a commercial pilot. Thank you guys for making it possible for me to start my ATPL training this November at the KLM Flight Academy. One day I will take you up in the air without the risk of landing in a crop field after, like what, twenty minutes?

See you around,
Sophie van Middelaar

Delft, October 16, 2018

Part I

Paper

Startle & Surprise Intervention Training: Evaluating Training Effects on Pilot Performance in Off-Nominal Operations

S.H. van Middelaar*

Supervisors: H.M. Landman[†], M.M. van Paassen[‡], E.L. Groen[§] and M. Mulder[¶]

Delft University of Technology, Delft, the Netherlands

Recent studies and accident investigations show the detrimental effects of startle and (automation) surprise on flight crew performance in terms of cognitive reasoning and sensemaking. Previous research conducted at Delft University of Technology shows a positive effect of training variability on crew performance in surprising tasks related to training, but in case of novel, unrelated surprising tasks, training methods remain to be investigated. Other recent studies and proposals suggest to use checklist-based training methods to prepare pilots for puzzling and surprising scenarios, but effects on flight crew performance have yet to be published. Therefore a simulator-based, between-subjects experiment was conducted to explore the effects of a four-item checklist-based training on pilot performance. Training elements focused on managing stress and enhancing situation awareness, as well as improving sensemaking and decision-making processes. It was expected that trained pilots would exhibit better overall performance in test scenarios involving off-nominal, surprising situations. Nevertheless, results indicated no significant difference in pilot response to an initial disturbance, but did show significant performance improvement in scenarios involving a second, ensuing upset compared to the baseline group. This suggests that checklist-based training methods offer a structured problem-solving approach, resulting in better response to consecutive disturbances in select scenarios.

I. Introduction

Nowadays commercial aviation has become one of the safest modes of transport, with the fatality rate per million flights dropping from 12 (1959) to 0.2 (2016) in the last decades.[1] In modern cockpit environments dials and gauges have been replaced by glass cockpits with high levels of automation and system reliability, resulting in a significant change in the role of the flight crew: its active, controlling nature has evolved into a more supervising and monitoring role.[2] Serious off-nominal conditions have become so rare that the current generation of pilots might not even encounter such a condition apart from during simulator training. They do however still occur and the unexpectedness of these malfunctions and automation surprises is thought to have influenced the outcome of several recent accidents, such as Turkish Airlines 1951 and Air France 447.[3,4]

While several startling scenarios, such as engine failures and turbulence, are trained frequently and Upset Prevention & Recovery Training (UPRT) will become mandatory in pilot training programs starting from 2019, general training to deal with startle and surprise is still largely omitted.[5] A recent TUD[a]

*MSc student, Department of Control & Simulation, Faculty of Aerospace Engineering, P.O. Box 5058, 2600GB Delft, The Netherlands; s.h.vanmiddelaar@student.tudelft.nl.

†PhD candidate, Department of Control & Simulation, Faculty of Aerospace Engineering, P.O. Box 5058, 2600GB Delft, The Netherlands; h.m.landman@tudelft.nl.

‡Associate Professor, Department of Control & Simulation, Faculty of Aerospace Engineering, P.O. Box 5058, 2600GB Delft, The Netherlands; m.m.vanpaassen@tudelft.nl. Member AIAA.

§Senior scientist, TNO Soesterberg, P.O. Box 23, 3769ZG Soesterberg, The Netherlands; eric.groen@tno.nl.

¶Professor, Department of Control & Simulation, Faculty of Aerospace Engineering, P.O. Box 5058, 2600GB Delft, The Netherlands; m.mulder@tudelft.nl. Associate Fellow AIAA.

[a]Delft University of Technology

study concluded that introducing variability in training helps pilots respond to surprising tasks, in case the surprising failure is related to the context of the prior training.[6] Alternatively, the industry itself proposes training general problem-solving strategies to aid pilots in novel, surprising tasks unrelated to prior training. Some of these checklists and decision-making models are already used in aviation (FORDEC, DECIDE, DODAR, etc.),[7–9] while extended and improved models (ROC, BAD),[10,11] aimed specifically at dealing with startle and surprise, are being proposed and developed. However, very little experimental research has been performed to evaluate the effectiveness of this type of training. This study will therefore focus on evaluating the effectiveness of such startle and surprise intervention training, ultimately to help develop a training method to aid pilots in unexpected tasks. The research objective is to find out how the particular startle and surprise intervention training used in this experiment affects pilot performance in surprising tasks unrelated to the practice scenarios used for training.

First of all, some background to the topic of will be provided by discussing the characteristics and effects of startle and surprise, as well as the content of the training given to experiment participants. Subsequently the experimental design will be discussed, followed by the results. Finally a discussion and conclusion is presented.

II. Startle & Surprise

II.A. Characteristics of Startle and Surprise

Startle is a fast, involuntary response, generated as a result of exposure to an intense and sudden stimulus.[12,13] It comes with a virtually instantaneous reflex involving muscle contraction in different parts of the face and body,[12,14] possibly interrupting motor performance for 0.1 to 3 seconds in simple tasks.[13] As the startle develops, it is accompanied by emotional, cognitive and psycho-physiological effects, including increase of both heart rate and blood pressure, hormone release[14] and the direction of the subject's attention towards the source of the startle stimulus.[13,15] When exposed to an intense startle, humans can demonstrate impaired response for up to 30 seconds[16] or, in extreme cases, even prolonged tonic immobility.[17,18] In case of a single, short-duration startle, the associated symptoms will fade quickly ('false alarm').[14]

In contrast to startle, *surprise* is a slower, more complex emotional and cognitive response, occurring when a difficult to explain reality does not match a subject's expectations.[3,19] Surprise can arise with or without a prior startle and has comparable physiological effects. However, the human surprise response develops over time instead of instantly, therefore being less defensive and dismaying compared to the startle response.[3,20] The physiological effects of both the startle and surprise response are believed to have a defensive function, interrupting current actions and drawing attention to the stimulus source.[13] This might help the subject to decide whether to fight, flee or investigate.[21]

Whether startle and surprise result in a fully developed stress response, depends on the presence of a persistent threat or stimulus.[14,16,22] In such case, stress arises from the perceived lack of being able to come up with a solution to the problem, caused by a mismatch between the level of environmental demands and the response capabilities of the subject.[23] The resulting anxiety inhibits goal-directed attention and increases stimulus-driven attention.[24] Combined with excessive physiological arousal and frustration due to the inability to determine the cause and solution to a problem, this can seriously impact subjects' troubleshooting and decision-making capabilities needed to solve the situation.[3,25]

In existing literature, several (conceptual) models have been presented to explain the human startle and surprise response. The common factor here involves issues related to *sensemaking*, described by Klein et al. as *"a motivated, continuous effort to understand connections (which can be among people, places, and events) in order to anticipate their trajectories and act effectively"* [26, p. 71]. In order to understand these connections, subjects start from a certain perspective, framework or viewpoint, called a *frame*. During the process of sensemaking, subjects can extend frames, but can also question frames, leading to rejecting and possibly replacing them (*reframing*). The concepts of sensemaking and frames are useful when developing a startle and surprise intervention training, since these provide clues on how pilots respond to surprising events. A clear example is confirmation bias, which according to sensemaking theory can be interpreted as subjects unsuccessfully extending a frame in order to fit the situation, instead of reframing and thereby replacing the frame with a new one.[27]

II.B. Startle and Surprise In Aviation

The effect of surprising tasks on flight crew performance, especially in modern cockpit environments, has been mentioned in studies since the late 20th century.[2] High levels of automation and obscured system principles add additional challenges to dealing with surprising and unexpected tasks, rendering pilots prone to loss of situation awareness as a result of discrepancies between frame and reality. A clear example in modern day cockpits is the *mode error*, when pilots mistakenly assume the autopilot to be in a certain mode.[28] A simulator-based experiment conducted by the FAA in 1994,[28] involving twenty airline pilots, showed that most problems between pilots and automation occur in off-nominal, time-critical situations such as autopilot disengagement during approach, aborted takeoff and loss of glideslope during final approach. Lack of knowledge about the automation led to surprises when the aircraft did not react as expected.

Although the 1994 FAA experiment was almost 25 years ago and pilot training has improved considerably since, recent studies still show the same problems. In 2012, a simulator-based experiment by NASA[29] revealed issues regarding the predictable, repetitive nature of recurrent pilot training. In frequently trained scenarios, such as low-level wind shear, aerodynamic stalls and engine failures, presented in the routine way as seen during training, pilots responded correctly and with little variability. However, when presented in an *unexpected* way, the same scenarios resulted in significantly worse pilot performance. In the latter case, the majority of pilots exhibited signs of miscategorization, procedural missteps and/or states of confusion. A similar experiment by researchers at the Griffith University Aerospace Strategic Study Centre[30] showed that during a low-visibility approach, the introduction of a startling stimulus just above decision height led to significantly delayed or even dangerously unstable performance in about one-third of the participating pilots. Moreover, most participants acknowledged physiological sensations due to the startling stimulus, such as adrenaline rush and increased heart rate. More than half of the subjects also reported feeling confused or indecisive for mixed duration, similar to the participants in the NASA experiment . A previous experiment by Landman et al.,[31] in which pilots were subjected to aerodynamic stalls in both anticipated and surprise conditions, shows similar results with pilots performing significantly worse in surprise conditions compared to anticipated conditions.

Apart from experimental results, the aviation industry itself also acknowledges the need for startle and surprise intervention training. In the past few decades, the introduction of advanced autopilots has dramatically reduced pilot hands-on flying time. Moreover, thanks to the high reliability of aircraft systems, off-nominal flying conditions have become very rare. When a malfunction or unintended autopilot command does occur, the level of surprise can be significant. Survey results among pilots confirm this and raise flags over issues involving managing unexpected situations in highly automated cockpit environments.[32] A recent survey among 200 Dutch commercial pilots indicated that pilots experience automation surprise about three times per year. Nearly two-thirds of the pilots mentioned system malfunction, interpreted by the respective authors as lack of knowledge about the systems, as one of the main causes of the perceived levels of surprise.[33] Furthermore, in several (recent) accidents such as Turkish Airlines 1951, Air France 447, Colgan Air 3407, Pinnacle Airlines 3701 and West Caribbean Airlines 708, startle and surprise are thought to have impaired the crews problem-solving capabilities.[4,14,34] Negative effects attributed to startle and/or surprise in these cases included degraded information processing, disrupted reasoning, loss of situation awareness and inappropriate control responses.

II.C. Dealing with Startle and Surprise

Existing literature, accident reviews and recent studies clearly show the detrimental effects of startle and surprise on flight crews' troubleshooting and decision-making capabilities. Fortunately, information on dealing with startle and surprise is also available, both from aviation itself as well as from other domains, such as police departments and health care.

When being confronted with a persisting threatening stimulus, the activation of general stress-related systems within the body can result in both cognitive and psychomotor impairment.[14] In stressful environments, quick relaxation techniques are therefore part of training to help subjects overcome or at least minimize these initial detrimental effects. In the military, practices to cope with acute stress include muscle relaxation, breathing exercises and cognitive methods such as positive self-talk.[35] In general, breathing exercises and muscle relaxation techniques are well-known methods for reducing stress-related issues.[36]

Another issue arising from startle and surprise is the direction of the subjects' attention towards the stimulus – which is helpful in certain life-threatening situations, but poses a risk in critical and complex work environments such as the flight deck. In aviation, tunnel vision has been a suspected factor in many incidents and accidents such as China Airlines Flight 006, Birgenair 301 or more recently, Air France 447.[4,37] In all cases, crews had all information necessary to figure out their situation, but nevertheless focused on only part of the available information. It is therefore of great importance to prevent confirmation bias and loss of situation awareness, by helping subjects to enforce goal-directed or 'top-down' attention.[15,24] In aviation, this is usually accomplished by incorporating instrument scanning patterns, checklists and decision-making strategies (e.g., *'Aviate,Navigate,Communicate'*) in both initial and recurrent training.

Apart from the current decision-making strategies, new training methods specifically aimed at startle and surprise are being developed and proposed as well. Examples include *'Relax, Observe, Confirm'* (ROC) by the Netherlands Aerospace Centre[11] and *'Breathe, Analyze, Decide'* (BAD) by the University of Southern Queensland.[10] A recent experiment by the Netherlands Aerospace Centre showed that such techniques can be trained relatively well, with 70% of 44 participants applying the full ROC method during training. Moreover, a majority indicated that they liked the training, regarded it as helpful and intended to apply the training in daily operations.[11,38] However, the particular experiment focused solely on application of the method, instead of exploring effects on pilot performance as well.

III. Method

In order to evaluate the effects of startle and surprise intervention training on pilot performance, a checklist-based training (*'COOL'*) was developed based on elements from existing stress management and problem-solving techniques. A simulator-based, between-subjects experiment was conducted involving a baseline (control) and experimental group, in which the received training served as the main independent variable. Figure 1 shows an overview of the different experimental phases.

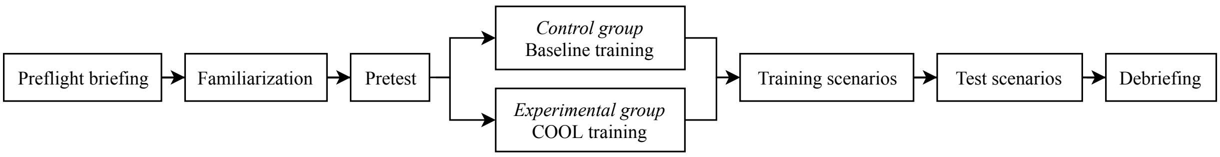


Figure 1. Flowchart showing the different phases of the experiment.

III.A. Participants

A total of twenty-four Dutch, currently employed commercial pilots from various airlines participated in the experiment. This study complied with the tenets of the Declaration of Helsinki and informed consent was obtained from each participant. Participants were randomly assigned to one of the two groups ($2 \times n = 12$), unless the balance, in terms of experience, age, etc., tended to be distorted. An overview is presented in Table 1 and Table 2. No significant differences between the two groups existed in terms of age (t-test, $t(22) = -0.435,\ p = 0.668$), employment years (Mann-Whitney, $U = 65,000,\ p = 0.707$) and flight hours on large, high-performance aircraft (t-test, $t(22) = -0.160,\ p = 0.874$).

Participants' trait anxiety was evaluated using the State-Trait Anxiety Inventory (STAI) test.[39] The results from the STAI questionnaire filled out by the participants before commencing the experiment showed no significant difference (Mann-Whitney, $U = 57,000,\ p = 0.384$) in anxiety ratings between the two groups.

Table 1. Participant comparison - age, employment years, flight hours and STAI score.

	Control Mean (SD)	*Experimental* Mean (SD)
Age	39.6 (11.7)	37.4 (12.7)
Years employed	14.7 (10.9)	13.5 (10.8)
Flight hours[b]	7544 (5851)	7172 (5549)
STAI score	24.9 (4.27)	28.9 (12.3)

Table 2. Participant comparison - gender, rank.

	Control No.	*Experimental* No.
Gender, Male	11	12
Gender, Female	1	0
Rank, CPT	6	4
Rank, FO	5	6
Rank, SO	1	2

[b]On high-performance, commercial jets and/or turboprops such as Boeing, Airbus, Embraer, Bombardier and Fokker aircraft

Due to inaccurate and/or missing information regarding flight experience on small aircraft (SEP/MEP[c]) provided by the participants, it was not possible to determine the exact difference in flight hours in this aircraft category between the control and experimental group. However, most participants had only flown these types of aircraft during their initial training and did not hold an active SEP/MEP rating anymore. Three participants in the control group logged 500 or more SEP hours and one participant in the experimental group logged more than 50 MEP hours.

III.B. Apparatus

The simulator-based experiment was conducted using the SIMONA (Simulation, Motion and Navigation) research simulator, or SRS, at the Faculty of Aerospace Engineering at Delft University of Technology. The SRS is a six-degrees-of-freedom full-motion simulator with a hydraulic hexapod motion system, that can be configured as basically any fixed- or rotary wing aircraft. The simulator has a collimated 180 degrees horizontal by 40 degrees vertical field of view (FOV) and features DLP projectors capable of projecting high-resolution computer generated images, rendered by FlightGear flight simulator software[d]. A 5.1 surround sound system was installed to enhance fidelity.

The Piper PA-34 Seneca III, a light twin-engine propeller aircraft, served as the aircraft model used throughout the experiment. The corresponding software model is a non-linear, six-degrees-of-freedom model developed by R. de Muynck and M. van Hesse (1990), which has been adapted for use in SIMONA experiments by H.J. Koolstra.[40] The same model was used in the preceding experiment by Landman et al., but some improvements were made to the model in order to facilitate new failure modes. Furthermore, some of the original mono audio was replaced by files from a stereo audio package developed by Carenado for Microsoft Flight Simulator 2004[e]. The use of a relatively basic aircraft model such as the Piper Seneca ensures relatively quick pilot familiarization, but also allows for interesting scenarios due to the twin-engine propeller configuration.

As the SIMONA simulator is a research simulator rather than a training simulator, the flight deck is modeled after a generic multi-crew cockpit. This means that both the cockpit layout and the instruments were not modeled after the Piper Seneca III. In order to minimize the familiarization time and to prevent participants from benefiting from relevant type experience, only the basic flight control and instruments were used in this experiment. This included a control column, pedals, throttle, gear lever, trim and three flap settings (0°, 25° and 40°). The (digital) instruments used were based on a Cessna Citation II and included a Primary Flight Display (PFD) and several other instruments such as a gear- and flap indicator, Exhaust Gas Temperature (EGT) display, RPM and torque indicator, fuel quantity and oil temperature/pressure displays. Except for a single test scenario all scenarios were flown within the vicinity of Schiphol Airport.

III.B.1. Flying Characteristics of the Piper Seneca III

The Piper Seneca is a twin-engine propeller aircraft frequently used as a trainer aircraft and sports aircraft. The overall handling of the model used in this experiment is very gentle and the model does not stall aggressively, although the pitch control can be considered to be quite sensitive. In case of a single-engine failure, the twin-engine configuration requires pilots to give full rudder input to compensate the asymmetric thrust. The aircraft is still able to climb comfortably in case of a single-engine failure, although in clean configuration (flaps and gear up) only. Furthermore, when flying single-engine in cruise, the Seneca model tends to dissipate its energy quickly, resulting in the aircraft quickly dropping below its minimal control speed of 80 kts for single-engine operations if the pitch attitude is not corrected for. Below 80 kts, the aileron and rudder authority is no longer sufficient to compensate the asymmetric thrust and the aircraft will roll towards the dead engine and pitch down, unless thrust of the remaining engine is reduced. In case of a loss of aileron authority the aircraft can be recovered by reducing thrust and/or adjusting the pitch attitude to pick up speed.

[c]Single or Multi Engine Piston
[d]https://home.flightgear.org/
[e]http://www.carenado.com/sitecarenado/

III.C. Checklist-Based Intervention Design - *'COOL'*

The startle and surprise intervention training used in this experiment is based on proven elements from existing stress management techniques, decision-making models and troubleshooting strategies used in both aviation and other industries. The goal of the intervention training is to aid pilots in surprising and possibly startling off-nominal situations by helping pilots to:

- Recognize and deal with the effects of the startle response ('de-startle')
- Become aware of own control inputs and aircraft response
- Prevent tunnel vision and/or confirmation bias
- (Re)gaining situation awareness
- Start troubleshooting and/or follow designated procedures

As many current aviation decision-making models aimed at assisting pilots in off-nominal situations (e.g., ROC/BAD, DECIDE/DESIDE, FORDEC) are based on checklists, this particular training was designed using the same principle. The training consists of four steps aimed at (1) stress management and/or relaxation, (2) enhancing situation awareness, (3) sensemaking / (re)framing and finally (4) decision-making. These four elements will be described next.

III.C.1. Stress Management / Relaxation

As part of the first training element, subjects are instructed to combine multiple breaths with muscle relaxation as described below:

- Slowly inhaling, holding breath for approximately three seconds, and slowly exhaling.
- Simultaneously sit upright, relax shoulders and be aware of applied stick and other control forces.

Almost any emergency situation allows the pilot to wait at least a few seconds before initiating a recovery procedure – the ones that require direct action, such as go arounds, onboard fire, uncommanded thrust reverser activation and explosive decompression are frequently trained. Participants are reminded to 'aviate first', but are also informed about the fact that 'pausing' for a few seconds can help to relax and, in some cases, even is essential to find out what is happening to the aircraft, as well as to prevent rushing into inappropriate actions. Shifting awareness to the applied stick and control forces helps pilots to become aware of their own (unintended) control inputs.

III.C.2. Enhancing Situation Awareness

The initial relaxation step is followed by the second step, which reminds subjects of the importance to 'mind the bigger picture'. In order to (re)gain situation awareness, subjects are instructed to:

- Call out the basic instrument readings: pitch, speed, bank angle, altitude and vertical speed.
- Check secondary instruments/indicators as necessary, such as throttle setting and engine indicators.
- Observe aircraft state, e.g., hard to keep level, continuously yawing to the right, etcetera.
- Be aware of own inputs, e.g., inadvertently pulling back on the stick.
- Focus on personal senses: what does one see, hear, smell, and feel?

Note that this step does not imply that the actual cause of the problem should be determined here, but should rather help the subject to (re)gain situation awareness and make sure any immediate threats are recognized (*'Aviate, Navigate, Communicate'*). Subjects are asked to vocalize their findings, both to force them to take a good look at their instruments, as well as to encourage Crew Resource Management (CRM) in a multi-crew cockpit.

III.C.3. Sensemaking and (Re)framing

The third training element focuses on sensemaking and (re)framing and encourages subjects to summarize their findings and observations from the previous step in order to formulate a plan of action. This does not require subjects to have identified the root cause of the problem, but rather should help them to link their observations together. Subjects are instructed to:

- Give an outline of the current situation, by stating observations that do not make sense – as well as the ones that do.
- Reflect on their observations: any missing information?

III.C.4. Decision-making

The fourth and final training element encourages pilots to formulate a plan of action and execute the steps necessary. Possible actions here could for example consist of following company troubleshoot procedures, running checklists or simply trying different control inputs. If the root cause of the problem has already been determined, the pilot can apply a possible solution here. However, while the method aims to help pilots in determining aircraft failure(s), its goal is not to ensure pilots find a solution to the presented problem, since every problem is unique and therefore requires a different approach. Rather, the checklist should help pilots to overcome the initial 'spooked' and possibly 'puzzled' feeling, in order to get to the troubleshooting itself. It is therefore expected that the subject will have to perform some more specific troubleshooting afterwards.

III.C.5. Mnemonic

Due to the relatively short training time and the fact that many existing aviation decision-making models, such as DECIDE, FORDEC and DODAR, are also based on acronyms, the same principle was used to introduce the training to participating pilots. In this case *'COOL'* (*'Calm down, Observe, Outline, Lead'*) was used as a mnemonic to remember the four elements of the training.

III.D. Experimental Procedure and Scenarios

Before commencing the experiment, all participants received a preflight briefing. Subsequently they completed five familiarization flights to get acquainted with the aircraft and simulator, followed by a pretest scenario with an unannounced engine failure. Participants then received either the baseline theoretical training or the experimental theoretical training, followed by five simulator-based practice scenarios. Lastly, all participants completed four test scenarios. The experiment concluded with a post-experimental debriefing and discussion.

III.D.1. Preflight Briefing, Familiarization Flights and Pretest

During the preflight briefing participants were informed about the nature of the experiment, as well as the aircraft model (controls, instruments and flying characteristics such as the minimum control speed and basic pitch-power information) and airport procedures, including the traffic pattern. Participants were instructed to complete the given task in order to be able to compare performance, unless it would lead to a loss of control. This means that in some cases participants were required to continue their approach, when they would normally go around (for example during the pretest). Certain aircraft speeds and settings presented in this preflight briefing were available in the form of a checklist during the flights in the simulator, as shown in Table 3.

Table 3. Kneepad information available to all participants during the simulator sessions.

Takeoff	Flaps: UP	V_r: 80 kts	V_{mc} SE: 80 kts
After takeoff	Pitch: $\sim$ 13 deg	V_2: 92 kts	Gear: UP
Downwind	V_{dw}: 115 kts	Torque: $\sim$ 43 N·m	
Base	V_{app}: 90 kts	Flaps: 25	Gear: DOWN
Final	Flaps: LAND		

All participants then practiced two takeoff rolls and three familiarization flights in order to get used to the aircraft model and its characteristics. The three familiarization flights consisted of a takeoff, level-off at 1,000 ft, two left-hand turns, joining a left-hand traffic pattern at 1,000 ft and landing. This task is hereafter referred to as a 'regular pattern'. During the third traffic pattern, participants were instructed to reduce throttle to idle and stall the aircraft for a brief moment in order to introduce them to the sound of the stall warning and gear alert. After these familiarization flights all pilots stated that they felt comfortable flying the aircraft.

After completing the familiarization flights, participants were informed that, in order to compare participant performance, a precision landing in strong crosswind conditions had to be performed. During this pretest, an unannounced left engine failure occurred on final and pilots were instructed to continue landing. A summary of the familiarization scenarios and pretest is given in Table 4.

Table 4. Familiarization scenarios and pretest for both groups.

No.	Task / Description	Runway	Wind
1	Takeoff roll	EHAM 18C	None
2	Takeoff roll	EHAM 18C	090 10KT
3	Regular pattern	EHAM 18C	None
4	Regular pattern	EHAM 18C	090 10KT
5	Regular pattern with gear alert and stall warning demo	EHAM 18C	090 10KT
Pretest	Approach and landing with left engine failure	EHAM 18C	090 10KT

III.D.2. Theoretical Training and Practice Sessions

After finishing the familiarization flights and pretest, the experimental group was trained in applying the *'COOL'* method by means of a training based on the principles of Stress Exposure Training (SET). Literature indicates the positive effects of this type of training on human performance in stressful situations, in which participants are (a) first informed about the stress environment and its consequences, then (b) taught skills aiding them in maintaining effective performance in such an environment and finally (c) offered to practice under safe, simulated conditions.[41] Both the control and experimental group received the first part (a) of the training, which consisted of background information such as the characteristics of startle and surprise, the relevance to aviation and some accident case analyses. The control group then continued directly to the third part (c) of the training in the simulator, while the experimental group first received the second part (b) of the training, introducing them to the *'COOL'* method. Afterwards, the experimental group also completed the third part (c) of the training. An overview is presented in Figure 2.

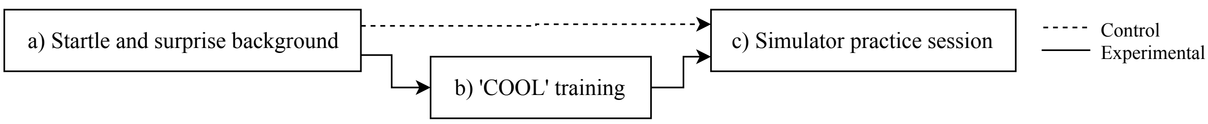


Figure 2. Theoretical and practical training phases for both groups.

In the third, simulator-based part (c) of the training, participants completed several scenarios involving off-nominal conditions. During these practice sessions, participants were informed that these scenarios were merely meant as training scenarios and therefore performance would not be monitored. While the control group was simply instructed to fly the scenarios and complete the given task, the experimental group was instructed to complete the task while also implementing and practicing the *'COOL'* training. An overview of the training scenarios is given in Table 5.

The first training scenario consisted of a regular takeoff, pattern and landing, like the familiarization flight, with no failures occurring. During this first training flight the experimental group was asked several times to practice applying the *'COOL'* method inflight. During the second training flight participants were instructed to fly a normal approach and landing, with a rudder failure occurring during final. The third training scenario again consisted of a normal takeoff, pattern and landing with an RPM indicator failure (showing a randomly oscillating and subsequently disappearing RPM value) occurring shortly before leveling off. In the fourth scenario participants were tasked with a centerline flyby instead of a landing, with a rudder hardover occurring just before the runway. The fifth and final scenario consisted of a takeoff, pattern and landing, with a right engine failure occurring shortly after rotation, thereby triggering the gear alert as well.

Table 5. Training scenarios for both groups.

No.	Task	Failure	Runway	Wind
1	Regular pattern	None	EHAM 18C	None
2	Approach and landing	No rudder	EHAM 18C	090 19KT
3	Regular pattern	RPM indicator failure	EHAM 18C	090 15KT
4	Approach and centerline flyby	Rudder hardover	EHAM 18C	270 19KT
5	Regular pattern	Right engine failure during initial climb	EHAM 18C	160 04KT

III.D.3. Test Scenarios and Debriefing

After completing the training scenarios, participants directly continued to the four test scenarios. Participants were informed about the end of the training and start of the test scenarios and were aware that performance would be monitored. In contrast to the training scenarios, participants did not complete the test scenarios in the same order, but rather according to a Latin square distribution. In each scenario, some conditions were changed compared to the practice sessions (e.g., traffic pattern side, visibility and takeoff airport) in order to distract the participants and increase workload. An overview of the test scenarios is presented in Table 6.

Table 6. Test scenarios for both groups.

No.	Task	Failure	Runway	Wind	Visibility
1	Regular pattern	Flap asymmetry	EHAM 18C	270 12KT	Low
2	Takeoff, level-off at 2,000 ft and right-hand pattern	False stall warning at 1,500 ft	EHAM 18C	260 06KT	Moderate
3	Regular pattern	Mass shift during rotation	EHAM 18C	270 13KT	CAVOK[f]
4	Regular pattern	Faulty airspeed indicator	EHLE 05	340 14KT	CAVOK

In the first (1) test scenario participants were instructed to perform a regular takeoff, pattern and landing, but this time in low visibility conditions. During this scenario, the aircraft suffered from a flap asymmetry causing the flaps on one side of the aircraft to not extend more than 10°, resulting in a rolling and yawing moment. This was first noticeable when deploying flaps 25° on base leg. Despite this condition the aircraft could still be controlled relatively well with flaps 25°. However, deploying flaps 40° rendered the aircraft difficult to control.

In the second (2) test scenario participants were instructed to climb to 2,000 ft (instead of 1,000 ft) and fly a right-hand pattern for the first time. During the climb the aircraft suffered an audible bird strike hitting the angle of attack vane at 1,500 ft, resulting in a false stall warning and stick shaker activation.

During the third (3) test scenario an audible mass shift upon rotation caused the center of gravity to shift backwards, creating a violent pitch-up moment. After recovery, by either power reduction or rolling the aircraft to get the nose down, the aircraft could be flown normally in clean configuration. However, deploying flaps rendered the aircraft extremely difficult to control as a result of the balloon effect (pitch-up moment due to flap deployment).

The fourth (4) test scenario took place at a different airport (Lelystad Airport, EHLE), but participants were tasked with the same regular pattern. Upon rotation, the indicated airspeed began to drop with 1 kts/s,

[f]Clouds and Visibility OK

causing erroneous airspeed readings. In such a case of unreliable airspeed readings, pilot training dictates reverting to flying according to pitch-power settings, as listed in Table 3.

Afterwards, participants received a debriefing, which included information about the experimental design, an explanation of the failures in each scenario and a discussion on the usefulness and allure of the training.

III.E. Dependent Variables

Participant performance was assessed using scenario-specific variables as listed below. Pretest performance evaluation was based on whether the pilots recognized the engine failure and responded accordingly. Dependent variables included whether or not pilots dropped below the minimum control speed V_{mc} of 80 kts, time spent flying below V_{mc}, occurrence of loss of aileron authority and in the latter case, recovery from the resulting loss of control. Furthermore the maximum negative vertical speed during approach was examined. Based on the balancing of the groups, no significant differences between the two groups were expected.

For the first (1) scenario in Table 5, involving flap asymmetry, performance was examined based on participants' recovery from the first upset (in terms of aileron input and maximum roll angle), subsequent flap selection choices (either deploying more flaps after the first upset, maintaining the current setting or retracting them), moment of configuration (downwind, base leg or final), response to a possible second upset and problem identification. To compensate for pilots deploying flaps in the turn from downwind to base leg, the maximum roll angle was defined as the maximum roll angle reached minus the roll angle at the time of flap deployment. It was expected that the experimental group would exhibit better problem understanding, resulting in more favorable flap selection choices and lower roll angles.

For the second (2) scenario, involving a false stall, recognizing the stick shaker and stall warning, as well as applying correct stall recovery procedures or 'unloading' (measured in terms of pilot elevator input, load factor, vertical speed and pitch rate and angle) and eventually recognizing the false stall conditions, were used to determine pilot performance. It was expected that the experimental group would again exhibit better problem understanding and identification. Furthermore the scenario was designed to explore any possible negative effects of the *'COOL'* training, such as pilots in the experimental group applying the *'COOL'* method before performing stall recovery procedures.

Performance measures used for the third (3) scenario, involving a mass shift, included recovery from the first mass-induced upset (in terms of elevator input, maximum pitch angle and minimum airspeed), subsequent flap selection choices and in case of flap deployment, recovery from the second flap-induced upset, as well as problem identification. It was expected that pilots in the experimental group would again show better problem understanding and identification, resulting in more favorable flap selection choices and lower pitch angles.

Finally, for the fourth (4) scenario, involving a faulty airspeed indicator, performance was assessed in terms of time before reverting to pitch-power settings, indicated and true airspeed at the time of switching to pitch-power settings, maximum true airspeed reached and time flying above 50 kts indicated airspeed ('following the faulty indicator'), as well as problem identification. It was expected that pilots in the experimental group would revert to pitch-power settings more quickly compared to pilots in the control group.

After each scenario, participants were asked to indicate perceived levels of startle, surprise, anxiety, workload and understanding by filling out a questionnaire. Levels of startle, surprise and understanding were indicated by 0-10 Likert-type scales,[42] while anxiety was rated using a 0-10 Visual Analog Scale.[43] As a measure for perceived workload the Rating Scale Mental Effort (RSME) was used.[44] Furthermore, participants were asked whether they enjoyed and liked the received training by means of a separate questionnaire. Apart from the RSME and STAI, the scales used in these questionnaires are not validated and were only used to provide additional insights.

III.F. Data Analysis

Only results from the pretest and four test scenarios were gathered, since the training scenarios were used for practice only. Raw simulator data, including for example aircraft state and pilot inputs, were processed using custom MathWorks MATLAB scripts.

Statistical analysis was performed using IBM SPSS software. Normality of the performance data was assessed using the Shapiro-Wilk normality test. In case of normally distributed data, an independent samples t-test was used (equality of variances tested using Levene's test), while for non-normally distributed data the Mann-Whitney test was used. Subjective ratings were processed using a parametric t-test. While Likert scales technically provide ordinal data, it is common to use parametric tests such as the t-test to analyze Likert-type scales, since the t-test is considered to be sufficiently robust in these cases.[45]

Apart from the independent samples tests, a mixed ANOVA was performed for the flap asymmetry and mass shift scenarios. While the first and second upset in these scenarios are technically not repeated measurements, they can be regarded as consecutive upsets due to the nature of the scenario and aircraft failure. Despite most data meeting all ANOVA requirements, the maximum pitch angle data from the mass shift scenario violated both the Box's Test of Equality of Covariance Matrices and the Levene's Test of Equality of Error Variances. Nevertheless, an ANOVA was performed as the test is considered to be fairly robust to these violations.

IV. Results

The experimental results in terms of performance data for each scenario, as well as the results of the subjective ratings given by the participants, are presented here.

IV.A. Pretest

During the pretest, eight out of twelve pilots in both groups allowed the airspeed to drop below V_{mc}, see Table 7. In the control group, this led to four losses of aileron authority, while in the experimental group three pilots lost aileron authority. In the control group, two pilots did not recover from the loss of aileron authority, resulting in a loss of control and subsequent crash or an off-airport emergency landing, while in the experimental group only one pilot did not recover.

Table 7. Pretest results, engine failure during final approach: number of pilots dropping below minimum control speed V_{mc}, experiencing a loss of aileron authority and number of pilots not recovering.

	Control	*Experimental*
	No.	No.
Below V_{mc}	8	8
LoA	4	3
Crash	2	1

Table 8. Pretest results, engine failure during final approach: time spent flying below minimum control speed V_{mc} and maximimum (negative) vertical speed for both groups.

	Control	*Experimental*
	Mean (SD)	Mean (SD)
Time < V_{mc} (s)	28.0 (30.4)	26.5 (46.8)
Max. VS (ft/min)	-1650 (1794)	-1560 (1686)

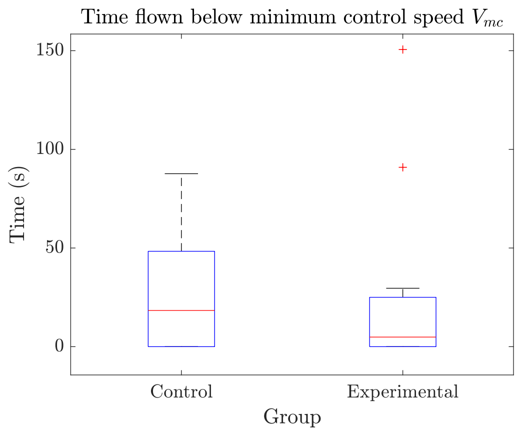


Figure 3. Pretest results: total time flown below V_{mc} during approach.

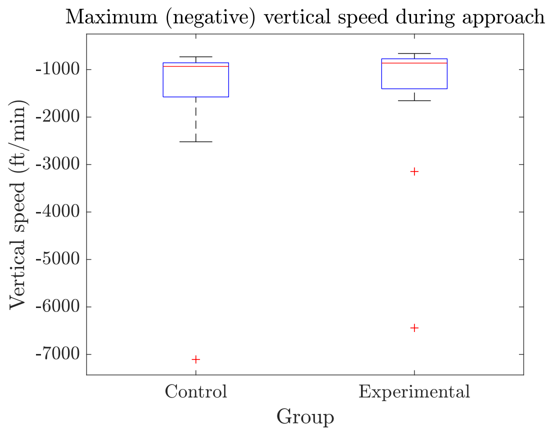


Figure 4. Pretest results: maximum (negative) vertical speed during approach.

Figure 3 shows the total time spent flying below the minimum control speed of 80 kts. The outlier of $t = 150$ s in the experimental group is due to one pilot recovering from a loss of control at very low altitude, requiring a considerable amount of time to climb and get back to the runway. Figure 4 shows the maximum (negative) vertical speed attained by the pilots of both groups during their approach. The three outliers of both groups correspond to the three losses of control. No significant differences were found between the two groups in terms of total time below V_{mc} (Mann-Whitney, $U = 64,000,\ p = 0.638$) and maximum negative vertical speed (Mann-Whitney, $U = 56,000,\ p = 0.356$), see Table 8. Overall performance of both groups therefore seems to be comparable.

IV.B. Flap asymmetry

For the scenario involving the flap asymmetry, Table 9 shows the performance results for both groups, while a visual representation of the participants' response is shown in Figure 7. In the experimental group, eleven out of twelve pilots indicated that they applied *'COOL'* completely, which could be confirmed by audio recordings for eight out of twelve pilots. Two pilots did not apply *'COOL'* until after the second upset caused by deploying flaps 40°, although one of these pilots did apply *'COOL'* prior to *any* flap deployment, because of perceived deteriorating visibility. After deploying flaps 25°, six pilots in the control group also deployed flaps 40° on final, compared to four pilots in the experimental group. One pilot in the control group reverted back to flaps 25° shortly after. Two pilots in the experimental group decided to retract the flaps and make a flaps-up landing. In both groups, seven pilots correctly identified the problem as a flap asymmetry. Other pilots stated aileron or rudder problems as possible causes.

Table 9. Flap asymmetry results: overview of results, including flap selection choices for each group.

	Control No.	*Experimental* No.
Applied *'COOL'*	N/A	11 (8*)
Identified problem	7	7
Landed F0	0	2
Landed F25	7	6
Deployed F40	6	4
Landed F40	5	4

* Confirmed by audio recording analysis

Table 10. Flap asymmetry results: maximum roll angle after first and second upset, as well as the difference between the first and second upset for both groups. A positive value indicates a performance improvement for the second upset.

	Control Mean (SD)	*Experimental* Mean (SD)
Max. ϕ (°), F25	-12.6 (5.88)	-16.6 (7.01)
Max. ϕ (°), F40	-17.7 (11.5)	-9.87 (4.77)
Max. ϕ (°), diff.	-5.72 (10.1)	8.82 (9.88)

Figure 5 and Table 10 show the maximum roll angles for both groups after deploying flaps 25° and for those who deployed flaps 40°. For the ten pilots deploying flaps 40°, the roll angle difference between the first upset (after deploying flaps 25°) and the second upset (after deploying flaps 40°) is shown in Figure 6. In this figure a positive value indicates a response improvement: the pilot countered the second upset (flaps 40°) more quickly than the first upset (flaps 25°). A negative value indicates that the second upset (flaps 40°) was larger than the first upset (flaps 25°).

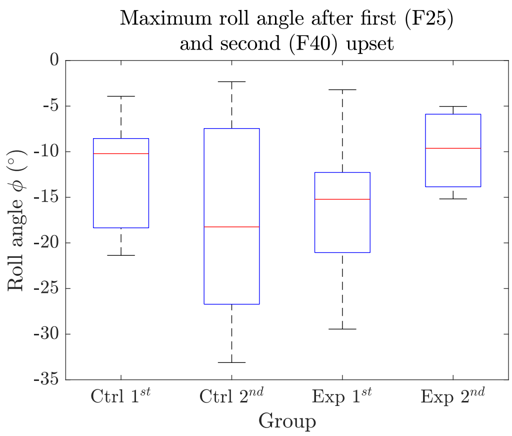


Figure 5. Flap asymmetry results: maximum roll angle ϕ after deploying flaps 25° and flaps 40°. For $F25°$ $n = 24$ ($C12$, $E12$), $F40°$ $n = 10$ ($C6$, $E4$).

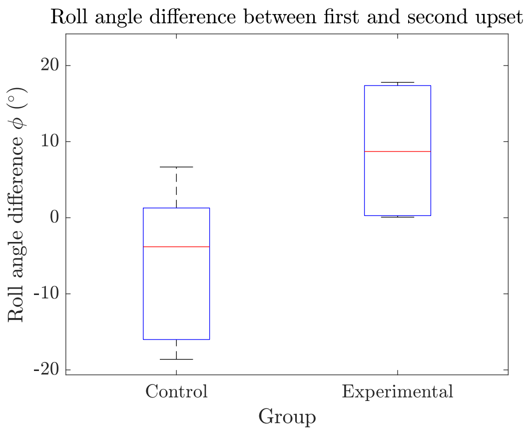


Figure 6. Flap asymmetry results: roll angle difference ϕ between first and second upset. Positive values indicate performance improvement. Total $n = 10$ ($C6$, $E4$).

For the first upset (flaps 25°), no significant difference in roll angle (t-test, $t(22) = -1.504,\ p = 0.147$) existed between the groups. When examining the pilots who selected flaps 40° ($n = 10$), the experimental group seems to perform better. However, the differences are non-significant, both in maximum roll angle after selecting flaps 40° (t-test, $t(8) = 1.268,\ p = 0.240$), as well as the roll angle difference between the first and second upset (Mann-Whitney, $U = 4,000,\ p = 0.088$).

When treating the first and second upset as repeated measurements, mixed ANOVA results again show no significant main effect ($F(1, 8) = 0.231,\ p = 0.644$) for upset moment (first/second) and similarly no significant main effect ($F(1, 8) = 0.021,\ p = 0.889$) for group (control/experimental). Also, no significant interaction ($F(1, 8) = 5.061,\ p = 0.055$) was found between upset moment and group. It therefore seems that both the t-test (independent samples) and mixed ANOVA (repeated measures) show similar results.

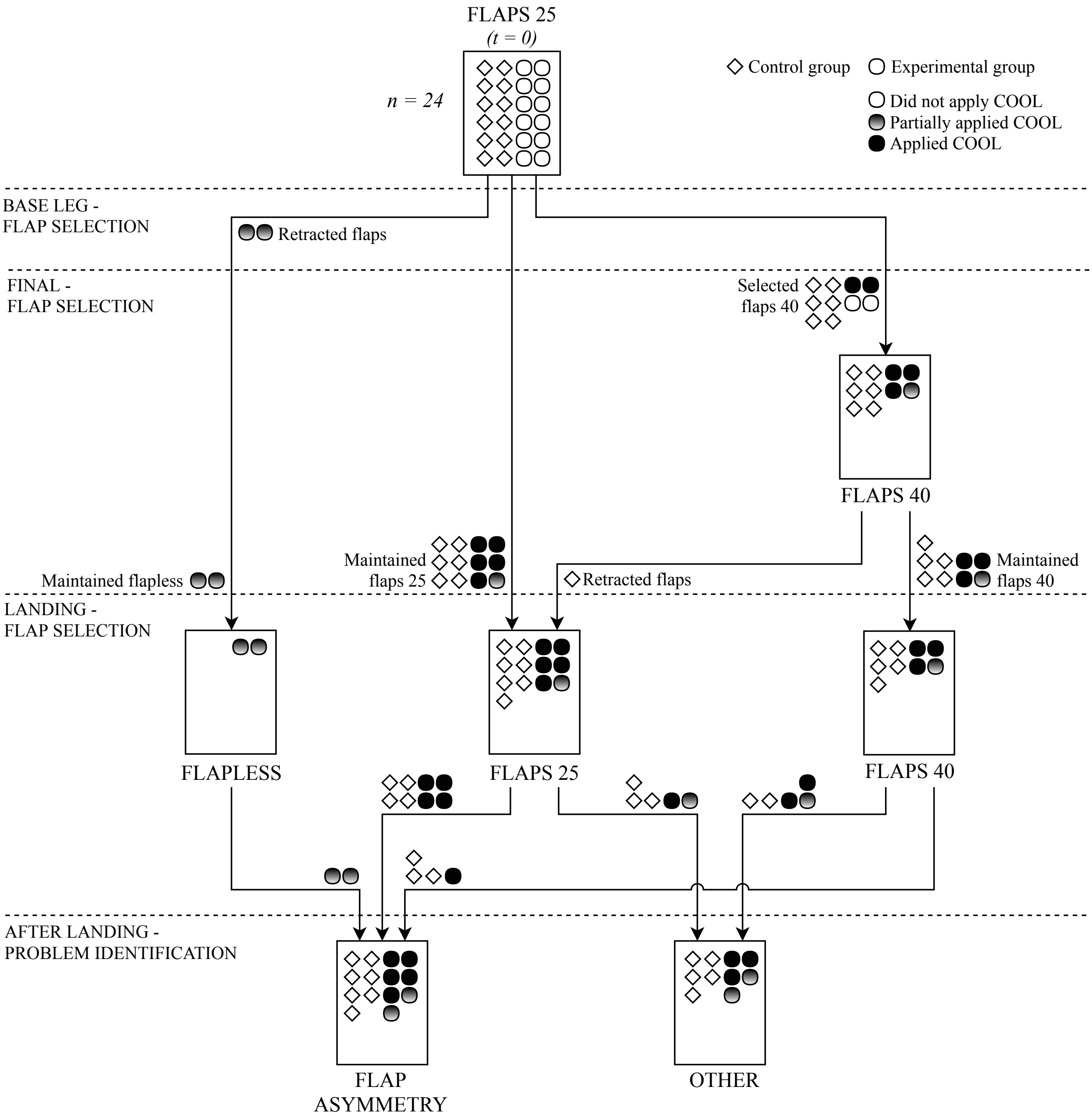


Figure 7. Flap asymmetry results: visualization showing the flap selection choices for both groups. Application of *'COOL'* is based on audio analysis.

IV.C. False stall warning

Table 11 shows an overview of participants' responses during the false stall warning scenario. In the experimental group, eleven out of twelve pilots indicated that they applied the *'COOL'* method, which could be confirmed by audio. Although all pilots were introduced to the stall warning and stick shaker during the familiarization flights, only nine pilots in the control group and eleven pilots in the experimental group recognized the stall warning and stick shaker during the scenario. Of these pilots, eight pilots in the control group and ten pilots in the experimental group correctly identified the problem as being a false stall warning and stick shaker afterwards. Other possible aircraft problems reported by the participants included structural damage, aerodynamic vibrations and control problems, which might be influenced due to the relatively high frequency of the stick shaker used in the simulation.

Table 11. False stall results: table showing an overview of the results for both groups.

	Control No.	*Experimental* No.
Applied *'COOL'*	N/A	11 (11*)
Identified false stall	8	10
Unload	8	8
Recognized stick shaker	9	11

* Confirmed by audio recording analysis

Pilot performance was also examined in terms of pilot elevator input during the first ten seconds after stick shaker activation, as well as the load factor, pitch rate, pitch angle and vertical speed. Both groups showed comparable responses to the false stick shaker, with eight pilots in each group unloading noticeably after activation of the stall warning and stick shaker. No significant differences in terms of load factor, pitch rate, etc. were found.

IV.D. Mass shift

An overview of the results for the mass shift scenario is shown in Table 12. A visual representation is shown in Figure 8. Nine out of twelve pilots in the experimental group indicated that they applied *'COOL'* completely, which could be confirmed by audio. Of these nine pilots, three pilots applied *'COOL'* for a second time after the flap-induced upset. None of the pilots identified the problem correctly – all pilots either identified the cause of the pitch controllability as an elevator problem, or as a pitch trim problem.

Table 12. Mass shift results: overview of results, including flap selection choices for each group.

	Control No.	*Exp.* No.
Applied *'COOL'*	N/A	9 (9*)
Identified problem	0	0
Never deployed flaps	1	1
Early configuration (downwind)	1	2
Deployed F25 (base)	10	9
Retracted flaps	8	6
Landed F0	10	9
Landed F25	1	3
Deployed F40	1	0
Crash	1	0

* Confirmed by audio recording analysis

Table 13. Mass shift results: maximum pitch angle and minimum airspeed after first and second upset, as well as the differences between the first and second upset for both groups. A positive value indicates a performance improvement for the second upset.

	Control Mean (SD)	*Experimental* Mean (SD)
Max. θ (°), 1^{st}	21.5 (3.67)	26.9 (6.14)
Max. θ (°), 2^{nd}	19.4 (15.1)	12.9 (3.37)
Max. θ (°), diff.	1.75 (17.4)	15.2 (7.33)
Min. IAS (kts), 1^{st}	70.8 (8.03)	60.9 (9.81)
Min. IAS (kts), 2^{nd}	62.9 (17.3)	70.9 (8.14)
Min. IAS (kts), diff.	-8.06 (22.3)	12.0 (13.4)

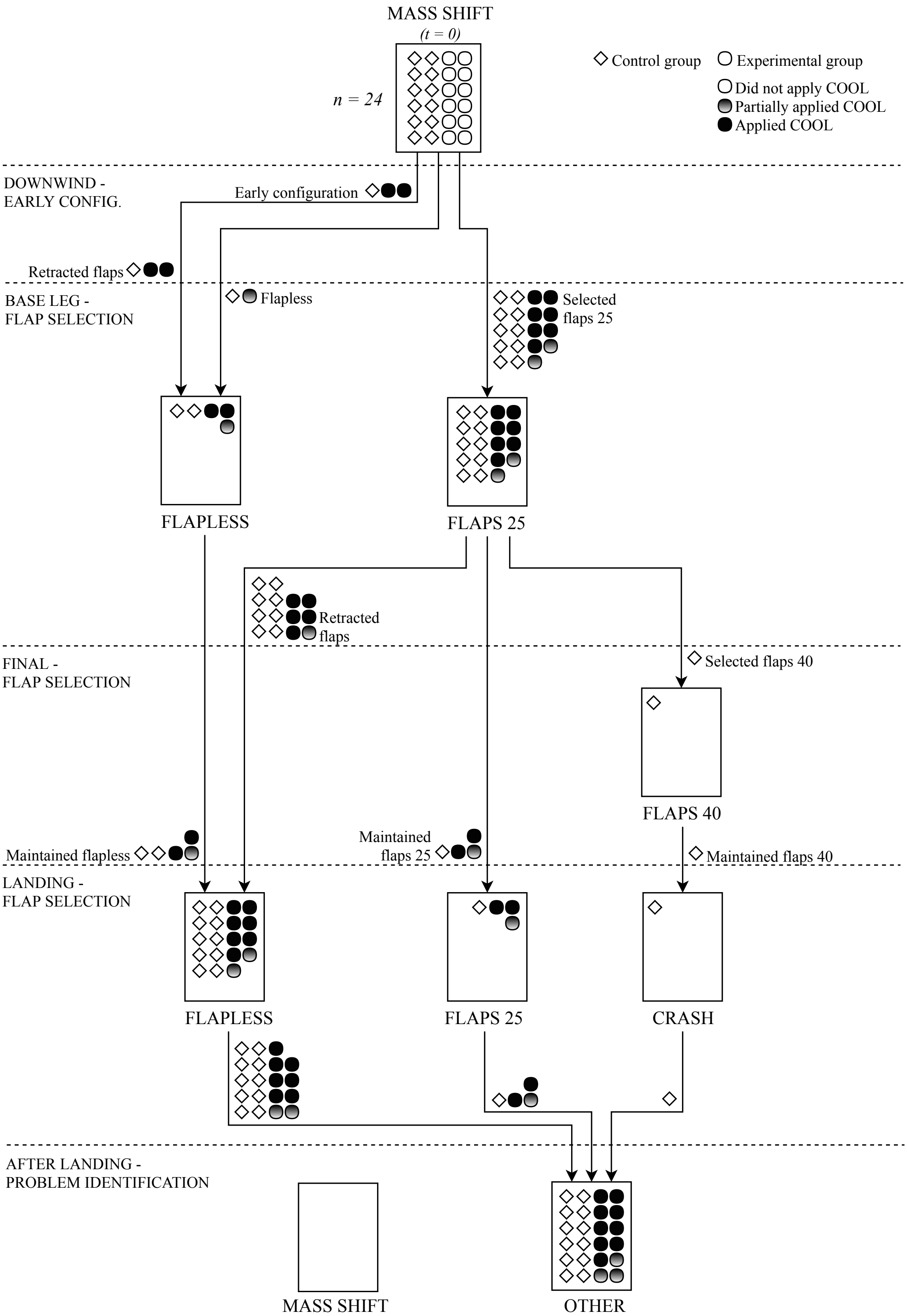


Figure 8. Mass shift results: visualization showing the flap selection choices for both groups. Application of *'COOL'* is based on audio analysis.

One pilot in each group decided to perform a flapless landing after the first upset, due to the expected excessive pitch-up in case of flap deployment. One pilot in the control group and two pilots in the experimental group instead decided to configure for landing early (above 1,000 ft above ground level), after which they immediately retracted their flaps. Furthermore, ten pilots in the control group deployed flaps 25° on base leg – eight pilots then retracted the flaps, one pilot maintained flaps 25° and one pilot deployed flaps 40° on final, resulting in a loss of control and crash. In the experimental group, nine pilots deployed flaps 25° on base leg – six pilots then retracted the flaps, while three pilots maintained this flap setting.

Figure 9 and Figure 10 show the maximum pitch angle and minimum airspeed for both groups after the first upset (due to the mass shift) and second upset (due to flap deployment, only for pilots who deployed flaps 25°). As for the scenario involving the flap asymmetry, one can compare the differences in maximum pitch angle and minimum airspeed between the first and second upset. The results are shown in Figure 11 and Figure 12.

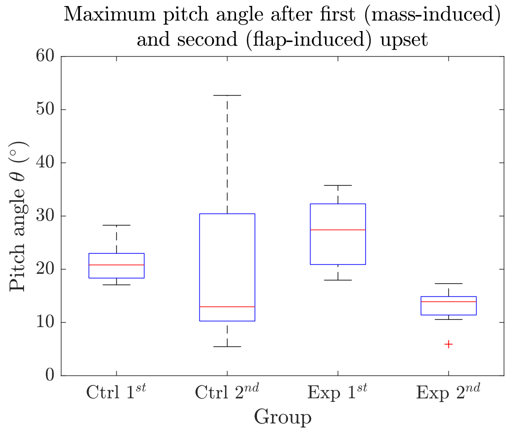


Figure 9. Mass shift results: max. pitch angle θ after the first (mass-induced) and second (flap-induced) upset. For the first upset $n = 24$ ($C12$, $E12$), second upset $n = 19$ ($C10$, $E9$).

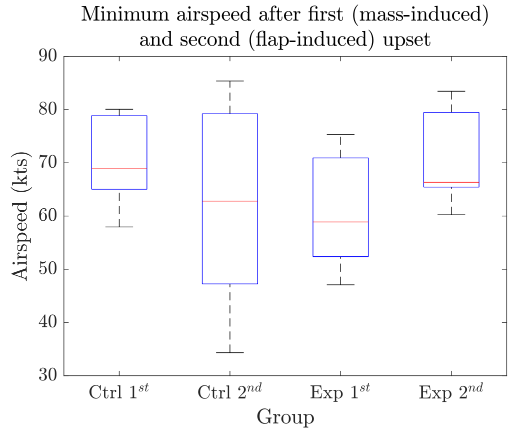


Figure 10. Mass shift: min. airspeed after the first and second upset. For the first upset $n = 24$ ($C12$, $E12$), second upset $n = 19$ ($C10$, $E9$).

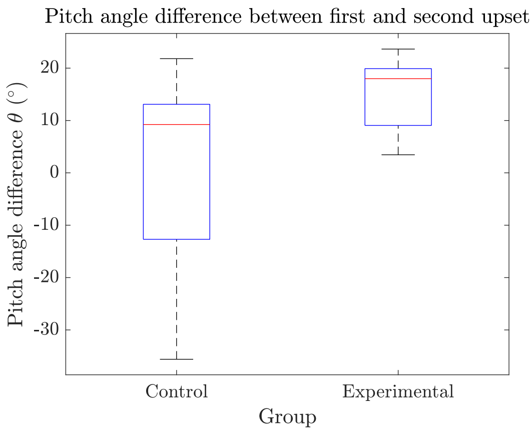


Figure 11. Mass shift results: pitch angle difference θ between the first and second upset. Positive values indicate performance improvement. Total $n = 19$ ($C10$, $E9$).

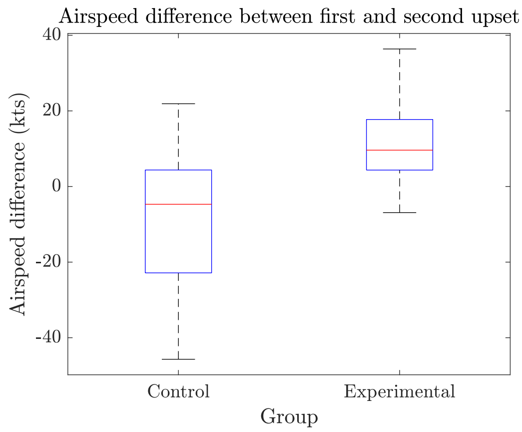


Figure 12. Mass shift results: airspeed difference between the first and second upset. Positive values indicate performance improvement. Total $n = 19$ ($C10$, $E9$).

For the first upset due to the mass shift, significant differences in maximum pitch angle (t-test, $t(22) = 2.616,\ p = 0.017$) and minimum airspeed (t-test, $t(22) = -2.709,\ p = 0.013$) exist in favor of the control group. However, for the second upset due to flap deployment ($n = 19$), no significant differences between the groups are present anymore, both for the maximum pitch angle (Mann-Whitney, $U = 42,000,\ p = 0.806$) and minimum airspeed (t-test, $t(17) = 1.260.\ p = 0.225$). Significant differences in favor of the experimental group were found for both the maximum pitch angle difference (t-test, $t(17) = 2.222,\ p = 0.046$) and the minimum airspeed difference (t-test, $t(17) = 2.345,\ p = 0.031$) between the first and second upset. An overview of mean values and standard deviations is presented in Table 13.

When the first and second upset are treated as repeated measurements, like in the flap asymmetry scenario, mixed ANOVA results show some significant effects. For the maximum pitch angle data, the results show a significant main effect ($F(1,17) = 7.271,\ p = 0.015$) for upset moment (first/second), but no significant main effect ($F(1,17) = 0.009,\ p = 0.927$) for group (control/experimental). Also, a significant interaction ($F(1,17) = 4.569,\ p = 0.047$) was found between upset moment and group. Both significant effects are however open to discussion, as for a violation of the Levene's Test it is usually recommended to lower the critical alpha level from 0.05 to, for example, 0.025 or 0.01. Post-hoc tests on the interaction effect revealed that the experimental group attained significantly higher pitch angles ($\Delta = 6.930,\ p = 0.006$) during the first upset compared to the control group. However, the difference between both groups during the second upset turned out to be non-significant ($\Delta = -6.480,\ p = 0.225$). When looking at the differences between the first and second upset, the experimental group showed significant performance improvement ($\Delta = -15.16,\ p = 0.004$), while the control group did not ($\Delta = -1.753,\ p = 0.690$).

For the minimum airspeed data, mixed ANOVA results show no significant main effect ($F(1,17) = 0.213,\ p = 0.650$) for upset moment (first/second) and similarly no significant main effect ($F(1,17) = 0.484,\ p = 0.496$) for group (control/experimental). However, a significant interaction ($F(1,17) = 5.498,\ p = 0.031$) was found between upset moment and group. Post-hoc tests on the interaction effect revealed that, like the pitch angles, the experimental group performed worse with significantly lower airspeeds ($\Delta = -12.10,\ p = 0.005$) during the first upset compared to the control group. However, the difference between both groups during the second upset turned out to be non-significant ($\Delta = 7.960,\ p = 0.225$). When looking at the differences between the first and second upset, no significant difference was found for both the experimental ($\Delta = 12.01,\ p = 0.070$) and control group ($\Delta = -8.056,\ p = 0.189$). It therefore seems that both the t-test (independent samples) and mixed ANOVA (repeated measures) show similar results.

IV.E. Faulty airspeed indicator

In the faulty airspeed indicator scenario, all pilots correctly identified the problem as an unreliable airspeed reading and all pilots landed the aircraft using pitch-power settings. Both groups showed comparable responses to the faulty airspeed reading and no significant differences could be determined. All pilots in the experimental group indicated that they applied the *'COOL'* method, which could be confirmed by audio for ten out of twelve pilots.

Table 14. Faulty airspeed indicator results: table showing an overview of the results for both groups.

	Control	*Experimental*
	No.	No.
Applied *'COOL'*	N/A	12 (10*)
Identified unreliable airspeed	12	12
Switched to pitch-power	12	12

* Confirmed by audio recording analysis

IV.F. Subjective ratings

The results of the Likert scale questionnaires show participants rating the scenario involving the false stall warning and stick shaker as being the most startling scenario, while the scenario involving the faulty airspeed indicator was rated as being the least startling. The mass shift scenario was rated to be the most surprising scenario, as well as the scenario that was the most difficult to understand. All pilots indicated that the scenarios used were realistic and convincing.

When comparing startle, surprise, understanding, anxiety, training helpfulness and RSME ratings of both groups for each scenario, most differences turned out to be non-significant. However, for the pretest the control group reported significantly lower understanding ratings compared to the experimental group (t-test, $t(22)$=-2.200, p =0.039). Looking at the test scenario results, the experimental group indicated significantly higher surprise ratings (t-test, $t(22)$=2.229, p =0.036) for the false stall scenario, as well as significantly lower understanding ratings (t-test, $t(22)$=2.899, p =0.011) for the mass shift scenario. Averaging all four test scenarios, results show significantly (t-test, $t(22)$=2.417, p =0.024) lower problem understanding ratings for the experimental group compared to the control group. Also, when comparing the pretest and post-test understanding ratings for each group, a significant interaction (ANOVA, $F(1, 22) = 14.743,\ p = 0.001$) was found between moment (pretest versus post-test) and group.

Participants in the experimental group were generally positive about the tested method, acknowledged both by the given ratings as well as the post-experimental debriefing and discussion. For each test scenario, participants in the experimental group were asked to give a rating on a 1-5 scale indicating how much the *'COOL'* method helped them to deal with the scenario. A mode of 4 out of 5 combined with comments from the debriefing show that pilots liked the method and found it to be helpful. Comments by pilots included that the method has a 'natural flow', counters tunnel vision and fits within a multi-crew cockpit environment. The first two steps of the method (*'Calm down'* and *'Observe'*) were regarded as being the most helpful, while (*'Outline'* and *'Lead'*) were considered to be steps any pilot would perform naturally, regardless of specific training.

Only two participants rated the *'COOL'* method as insufficiently helpful after at least one of the test scenarios, with a rating of either 1 or 2 on a 1-5 scale. One of these two participants indicated that while the method could work in a relatively basic environment such as a single pilot sports aircraft, the method was regarded to be distracting and less helpful in complex, multi-crew airliner cockpits. The other participant indicated that due to the nature of the scenarios, the situation became too time-critical to apply the complete *'COOL'* method. However, the same participant also stated that the method might be helpful in less time-critical situations, such as problems arising during the cruise phase.

All participants in the experimental group indicated that they applied the *'COOL'* method to each test scenario, although in some cases this could not always be confirmed by means of an audible response during the scenario. The results are shown in Table 15. The mass shift scenario, rated as being both the most surprising and the most difficult scenario to understand, shows the most incomplete applications of the *'COOL'* method. Furthermore, results from the flap asymmetry and mass shift scenario show some participants 'picking up' the method, despite skipping the first *'Calm down'* step.

Table 15. Table showing the number of participants in the experimental group applying the four elements of *'COOL'*, as stated by participants themselves.

	No. of pilots *'Calm down'*	*'Observe'*	*'Outline'*	*'Lead'*	Complete *'COOL'*
Scenario					
Flap asymmetry	11	12	12	12	11
False stall	12	12	12	11	11
Mass shift	10	12	10	11	9
Faulty ASI	12	12	12	12	12

V. Discussion

When interpreting the results, it is somewhat surprising to see significant performance differences between both groups in the mass shift and flap asymmetry scenarios, while practically no differences are seen in the other two scenarios. A possible explanation here is the fact that only the mass shift and flap asymmetry scenarios involve a second upset. It suggests that while a method such as *'COOL'* might not help to counter an initial upset or interruption, it possibly *does* help pilots to analyze this disruption more consciously, in this particular case resulting in a better response to the second upset. Audio analysis results imply the same, with most pilots first recovering from an initial upset and subsequently applying the *'COOL'* method. On the contrary, the control group responds (significantly) worse to the second upset in both scenarios. Nevertheless, in both scenarios the response of the experimental group to the *initial* upset is (significantly) worse compared to the control group, implying that the application of *'COOL'* in some cases might delay or inhibit initial pilot response.

Surprisingly, pilots in the experimental group rated the scenarios significantly more difficult to understand compared to the control group. This is most likely a direct result from the two groups having different task expectations. While the control group was simply instructed to fly the scenarios, the experimental group was simultaneously tasked with applying the *'COOL'* training. It is possible that pilots in the experimental group therefore experienced higher pressure levels, since the training explicitly requires them to formulate a problem statement in the third step (*'Outline'*). This not only requires a better problem understanding, but also a more conscious approach compared to the control group. Or, as one pilot in the control group stated, 'as long as I get the aircraft safely on the ground I'm happy; it is up to the technicians to find out what is wrong'.

Despite lower understanding ratings, pilots seemed to appreciate the *'COOL'* method. Also, pilots in the experimental group applied the *'COOL'* method relatively well given the short training time, although it seems that in some scenarios, application of the method was deemed unnecessary due to pilots recognizing a certain failure early on. During the post-experimental discussion, pilots expressed that a structured problem-solving strategy is regarded helpful in startling and surprising scenarios. As acute stress management (*'Calm down'*) and enhancing situation awareness (*'Observe'*) were valued most and the other two steps (*'Outline'* and *'Lead'*) follow naturally from the first two steps, the latter two could therefore be deemed redundant. In such case, the method might be improved by focusing on optimizing the first two steps. As the significant effects result from scenarios with a consecutive upset, a recommendation for a future experiment would be to include more of these scenarios. Other recommendations include expanding the group size and exploring training retention in a follow-up experiment, since the experiment duration was only three hours. Finally, since the experiment was based on single-pilot operations only, the application of the training within a multi-crew cockpit needs to be explored as well.

As with any human-in-the-loop experiment involving pilots, it is difficult to draw strong conclusions from the results, both due to the small sample size as well as the confounding factors inherent to an experiment involving airline pilots. These factors include for example varying skill levels (flight hours, type ratings, employment years), airline operations (regional airlines versus legacy carriers) and training (flight training school, airline recurrent training), rendering it extremely difficult to arrange a large, homogeneous group. A noticeable example in this experiment is the response of pilots to the nose-high pitch upset(s) during the mass shift scenario. The recovery method applied varied with airline and flight school and included, for example, recovery by throttle reduction or rolling the aircraft.

VI. Conclusion

The results from the simulator-based experiment show that in off-nominal conditions unrelated to training, a checklist-based method such as *'COOL'* might benefit pilot performance. While not improving pilot response to an initial disturbance or upset, the results show significant performance improvement in case of a second, consecutive disturbance. This indicates that a structured and conscious approach to problem analysis benefits pilot performance in case of consecutive upsets.

Subjective ratings show that pilots appreciate the training method and participants expressed that a structured problem-solving strategy is regarded helpful in startling and surprising scenarios. The first two steps, focusing on acute stress management and enhancing situation awareness, were being valued most. Also, pilots applied the training relatively well despite the short training time.

Regarding the implementation and application of the method, previous research by Landman et al.[6] showed that simulator training, presented in a unpredictable and variable way, significantly improves pilot performance in scenarios related to training. Therefore introducing unpredictability and variability in simulator-based emergency training, combined with a checklist-based method such as *'COOL'*, could help to train pilots for off-nominal operations, both related to training as well as completely novel scenarios unrelated to training. Recommendations for a follow-up experiment include expanding the group size, exploring long term training retention and the application of the training method in multi-crew cockpits, as well as including more scenarios involving a consecutive upset and/or a deteriorating aircraft condition.

References

[1]Airbus, *A Statistical Analysis of Commercial Aviation Accidents 1958-2016*, 2017. Available at: https://flightsafety.org/wp-content/uploads/2017/07/Airbus-Commercial-Aviation-Accidents-1958-2016-14Jun17-1.pdf

[2]Wiener, E.L., Curry, R.E., *Flight-Deck Automation: Promises and Problems*, Ergonomics, vol. 23, no. 10, p. 995-1011, 1980. DOI: 10.1080/00140138008924809

[3]Landman, H.M., van Paassen, M.M., Groen, E.L., Bronkhorst, A.W., Mulder, M., *Dealing With Unexpected Events on the Flight Deck: A Conceptual Model of Startle and Surprise*, Human Factors, vol 59, no. 8, p. 1161 1172, 2017. DOI: 10.1177/0018720817723428

[4]Martin, W.L., Murray, P.S., Bates, P.R., *The Effects of Startle on Pilots During Critical Events: A Case Study Analysis*, 30th EAAP Conference: Aviation Psychology & Applied Human Factors - Working Towards Zero Accidents, 2012. Available at: https://research-repository.griffith.edu.au/handle/10072/54072

[5]Schroeder, J., *Research and Technology in Support of Upset Prevention and Recovery Training*, AIAA Modeling and Simulation Technologies Conference, 2012. DOI: 10.2514/6.2012-4567

[6]Landman, H.M., van Oorschot, P., van Paassen, M.M., Groen, E.L., Bronkhorst, A.W., Mulder, M., *Training Pilots for Unexpected Events: A Simulator Study on the Advantage of Unpredictable and Variable Scenarios*, Human Factors, vol. 60, no. 6, p. 793-805, 2018. DOI: 10.1177/0018720818779928

[7]Federal Aviation Administration, *Aeronautical Decision Making*. Available at: https://www.faa.gov/regulations_policies/handbooks_manuals/aviation/phak/media/04_phak_ch2.pdf

[8]Schutte, P.C., Trujillo, A.C., *Flight Crew Task Management in Non-Normal Situations*, Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 40, no. 4, p. 244-248, 1996. DOI: 10.1177/154193129604000422

[9]Li, W-C., Harris, D., Hsu, Y-L., Wang, T., *Understanding Pilots' Cognitive Processes for Making In-flight Decisions under Stress*, Proceedings of the 42nd Annual International Seminar of the International Society of Air Safety Investigators, Salt Lake City, 2011. Available at: https://dspace.lib.cranfield.ac.uk/handle/1826/8148

[10]Martin, W.L., *Developing Startle and Surprise Training Interventions for Airline Training Programs*, Presented at PACDEFF 2016, 2016. Available at: http://pacdeff.com/wp-content/uploads/2017/08/PACDEFF-FC-Forum-Presentation-on-Startle.pdf

[11]Boland, E., *Managing Startle & Surprise*, Presented at PACDEFF 2016, 2016. Available at: http://pacdeff.com/wp-content/uploads/2016/11/PACDEFF-Startle-Surprise-Management.pdf

[12]Koch, M., *The Neurobiology of Startle*, Progress in Neurobiology, vol. 59, no. 2, p. 107-128, 1999. DOI: 10.1016/s0301-0082(98)00098-7

[13]Rivera, J., Talone, A.B., Boesser, C.T., Jentsch, F., Yeh, M., *Startle and Surprise on the Flight Deck*, Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 58, no. 1, p. 1047-1051, 2014. DOI: 10.1177/1541931214581219

[14]Martin, W.L., Murray, P.S., Bates, P.R., Lee, P.S.Y., *Fear-Potentiated Startle: A Review from an Aviation Perspective*, The International Journal of Aviation Psychology, vol. 25, no. 2, p. 97-107, 2015. DOI: 10.1080/10508414.2015.1128293

[15]Nieuwenhuys, A., Oudejans, R.D., *Anxiety and Performance: Perceptual-Motor Behavior in High-Pressure Contexts*, Current Opinion in Psychology, vol. 16, p. 28-33, 2017. DOI: 10.1016/j.copsyc.2017.03.019

[16]Thackray, R.I., Touchstone, R.M., *Rate of Recovery and Subsequent Radar Monitoring Performance Following a Simulated Emergency Involving Startle*, FAA Civil Aeromedical Institute Report, no. FAA-AM-83-13, 1983. Available at: https://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/1980s/media/am83-13.pdf

[17]Bracha, H.S., *Freeze, Flight, Fight, Fright, Faint: Adaptationist Perspectives on the Acute Stress Response Spectrum*, CNS Spectrums, vol. 9, no. 9, p. 679-685, 2004. DOI: 10.1017/s1092852900001954

[18]Schmidt, N.B., Richey, J.A., Zvolensky, M.J., Maner, J.K., *Exploring Human Freeze Responses to a Threat Stressor*, Journal of Behavior Therapy and Experimental Psychiatry, vol. 39, no. 3, p. 292-304, 2008. DOI: 10.1016/j.jbtep.2007.08.002

[19]Foster, M.I., Keane, M.T., *Why Some Surprises Are More Surprising than Others: Surprise as a Metacognitive Sense of Explanatory Difficulty*, Cognitive Psychology, vol. 81, p. 74-116, 2015. DOI: 10.1016/j.cogpsych.2015.08.004

[20]Bradley, M.M., *Natural Selective Attention: Orienting and Emotion*, Psychophysiology, vol. 46, no. 1, p. 1-11, 2009. DOI: 10.1111/j.1469-8986.2008.00702.x

[21]Hagemann, T., Levenson, R.W., Gross, J.J., *Expressive Suppression During an Acoustic Startle*, Psychophysiology, vol. 43, no. 1, p. 104-112, 2006. DOI: 10.1111/j.1469-8986.2006.00382.x

[22]Civil Aviation Authority, *Flight Crew Human Factors Handbook*, 2014. Available at: http://publicapps.caa.co.uk/docs/33/CAP\%20737\%20DEC16.pdf

[23]Wheaton, B., *Stress, Personal Coping Resources and Psychiatric Symptoms: An Investigation of Interactive Models*, Journal of Health and Social Behavior, vol. 24, no. 3, p. 208-229, 1983. DOI: 10.2307/2136572

[24]Eysenck, M.W., Derakshan, N., Santos, R., Calvo, M.G., *Anxiety and cognitive performance: attentional control theory*, Emotion, vol. 7, no. 2, p.336-353, 2007. DOI: 10.1037/1528-3542.7.2.336

[25]Burki-Cohen, J., *Technical Challenges of Upset Recovery Training: Simulating the Element of Surprise*, AIAA Modeling and Simulation Technologies Conference, 2010. DOI: 10.2514/6.2010-8008

[26]Klein, G., Moon, B., Hoffman, R.R., *Making Sense of Sensemaking 1: Alternative Perspectives*, IEEE Intelligent Systems, vol. 21, no. 4, p. 70-73, 2006. DOI: 10.1109/mis.2006.75

[27]Klein, G., Moon, B., Hoffman, R.R., *Making Sense of Sensemaking 2: A Macrocognitive Model*, IEEE Intelligent Systems, vol. 21, no. 5, p. 88-92, 2006. DOI: 10.1109/mis.2006.100

[28]Sarter, N.B., Woods, D.D., *How in the World Did We Ever Get into That Mode? Mode Error and Awareness in Supervisory Control*, Human Factors, vol. 37, no. 1, p. 5-19, 1995. DOI: 10.1518/001872095779049516

[29]Casner, S.M., Geven, R.W., Williams, K.T., *The Effectiveness of Airline Pilot Training for Abnormal Events*, Human Factors, vol. 55, no. 3, p. 477-485, 2012. DOI: 10.1177/0018720812466893

[30]Martin, W.L., Murray, P.S., Bates, P.R., Lee, P.S.Y., *A Flight Simulator Study of the Impairment Effects of Startle on Pilots During Unexpected Critical Events*, Aviation Psychology and Applied Human Factors, vol. 6, no. 1, p. 24-32, 2016. DOI: 10.1027/2192-0923/a000092

[31]Landman, H.M., Groen, E.L., van Paassen, M.M., Bronkhorst, A.W., Mulder, M., *The Influence of Surprise on Upset Recovery Performance in Airline Pilots*, The International Journal of Aerospace Psychology, vol. 27, no. 1-2, 2017. DOI: 10.1080/10508414.2017.1365610

[32]Tenney, Y.J., Rogers, W.H., Pew, R.W., *Pilot Opinions on Cockpit Automation Issues*, The International Journal of Aviation Psychology, vol. 8, no. 2, p. 103-120, 1998. DOI: 10.1207/s15327108ijap0802_2

[33]de Boer, R., Dekker, S., *Models of Automation Surprise: Results of a Field Survey in Aviation*, Safety, vol. 3, no. 3, p. 20-30. DOI: 10.3390/safety3030020

[34]Dutch Safety Board, *Crashed During Approach, Boeing 737-800, near Amsterdam Schiphol Airport, 25 February 2009*, 2010. Available at: https://www.onderzoeksraad.nl/uploads/items-docs/1748/Rapport_TA_ENG_web.pdf

[35]United States Marine Corps, *Combat Stress Handbook*, 2000. Available at: http://www.au.af.mil/au/awc/awcgate/usmc/mcrp611c.pdf

[36]Benson, H., *The Relaxation Response*, Mind Body Medicine: How to Use Your Mind for Better Health, p. 125-149, Consumers Reports Books, 1993. Available at: http://www.upegroup.com/uploads/1/1/0/4/11040867/relaxation_response.pdf

[37]National Transportation Safety Board, *Aircraft Accident Report: China Airlines Boeing 747-SP, N4522V, 300 Nautical Miles Northwest of San Francisco, California, February 19, 1985*, 1986. Available at: https://www.ntsb.gov/investigations/AccidentReports/Reports/AAR8603.pdf

[38]Filluci, J.B., *Relax, Observe, Confirm*, Online. Available at: http://bluetoad.com/display_article.php?id=3104860&view=501561

[39]Julian, L.J., *Measures of anxiety: StateTrait Anxiety Inventory (STAI), Beck Anxiety Inventory (BAI), and Hospital Anxiety and Depression ScaleAnxiety (HADSA)*, Arthritis Care & Research, vol. 63, no. S11, p. S467-S472, American College of Rheumatology, 2011. DOI: https://doi.org/10.1002/acr.20561

[40]De Muynck, R., Van Hesse, M., *The a priori simulator software package of the Piper PA34 Seneca III (technical report)*, Delft University of Technology, 1990.

[41]Driskell, J.E., Johnston, J.H., Salas, E., *Does Stress Training Generalize to Novel Settings*, Human Factors, vol. 43, no. 1, p. 99-110, 2001. DOI: 10.1518/001872001775992471

[42]Likert, R., *A Technique for the Measurement of Attitudes*, Archives of Psychology, no. 140, p. 1-55, 1932.

[43]Freyd, M., *The Graphic Rating Scale*, Journal of Educational Psychology, no. 14, p 83-102, 1932. DOI: 10.1037%2Fh0074329

[44]Zijlstra, F.R.H., *Efficiency in work behaviour: A design approach for modern tools*, Doctoral thesis, Delft University of Technology, 1993. Available at: https://repository.tudelft.nl/islandora/object/uuid:d97a028b-c3dc-4930-b2ab-a7877993a17f

[45]Norman, G., *Likert scales, levels of measurement and the "laws" of statistics*, Advances in Health Sciences Educuation, vol 15, no. 5, p. 625-632, 2010. DOI: 10.1007/s10459-010-9222-y

Part II

Book of Appendices

Appendix A

Preflight Briefing

The following appendix contains the preflight briefing notes and Powerpoint presentation (Dutch only).

OPENINGSBRIEFING (+ PPT)

Algemeen

- Onderdeel van afstuderen Sophie en promotie Annemarie.
- Wordt verplicht in UPRT training, maar nog weinig trainingen beschikbaar.
- Daarnaast weinig experimentele resultaten die het effect van training ondersteunen.
- Testen, onderzoeken en vergelijken van verschillende trainingen gericht op het omgaan met startle en surprise.
- Straks eerst familiarisatie, dan uit de sim: theorie, pauze, daarna oefenscenarios, tenslotte testscenario's.

Piper Seneca

- Twin-engine piston, gebaseerd op de Seneca III.
- Controls: control column voor ailerons en elevator, pedalen voor je rudder, throttle, flaps, gear en pitch trim. Geen rudder trim. Digitaal display uitleggen (incl. torque).

SIMONA

- Flight deck niet natuurgetrouw vanwege functie van simulator.
- Flaps landing / 40 deg = flaps 30 deg op console.

- Problemen met simulatie op de grond, daarom taxiën oefenen.
- Slip indicator op PFD nogal gevoelig.
- Bij takeoff full power nodig, geen remmen.

Verklaring

- Formulier informed consent.
- Resultaten anoniem verwerkt.
- Misselijkheid, control forces, schrikken.
- Altijd mogelijk om af te breken.

Preflight briefing

- Splits Annemarie (technisch) en Sophie (scenario).
- Houd je aan gegeven opdracht (bijv. maak landing af, vlieg linkerhand circuit). Zo kunnen wij mensen goed vergelijken. Als crash onvermijdelijk, dan noodlanding.
- Aanpassing opdracht via headset.
- Binnen de opdracht enige vrijheid, bijv. qua snelheid en configuratie.
- Doe altijd je best om zo netjes/veilig mogelijk te vliegen.
- Bij landingen altijd richten op de centerline en blokken naast PAPI.
- Geef een callout bij afwijkingen of gekke dingen.
- Start, approach en kneepad doornemen.

Circuit

- Vliegen van/naar Schiphol 18C.
- Circuit standaard linksom, 1000 ft en 115 kts.
- Indraaipunt t.h.v. einde Polderbaan.
- Touchdown op blokken naast PAPI (die niet helemaal klopt).
- Geen ATC.

Startle & Surprise Experiment

Openingsbriefing

SIMONA, TU Delft

Mei 2018

Achtergrond

- MSc thesis Sophie
- PhD onderzoek Annemarie

Motivatie Onderzoek

- Vanaf 2019 verplicht door EASA/FAA
- Weinig experimentele resultaten
- Onderzoeken en vergelijken van trainingsmethoden

TU Delft

Programma

- Familiarisatie in SIMONA
- Theorie
- Pauze
- Oefensessie in SIMONA
- Experiment

TU Delft

Verklaring

- Toestemmingsverklaring
- Anonieme resultaten
- Misselijkheid, schrikken
- Mogelijkheid tot stoppen

TU Delft

Model

Piper Seneca III

TU Delft

Flight Controls

- Control column + pedalen
- Throttle
- Flaps en retractable gear
- Pitch trim
- Digitaal display met PFD

SIMONA

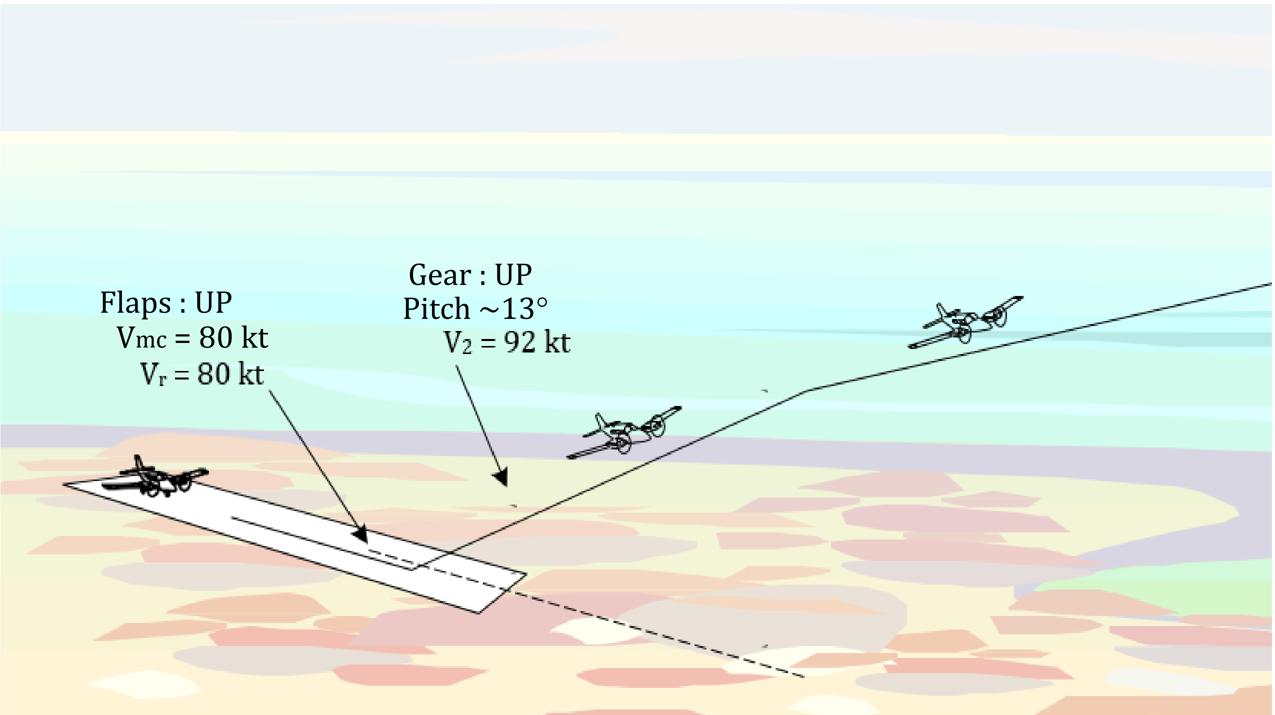
- Flight deck is niet natuurgetrouw
- Flaps 40° = 30° op hardware
- Problemen met simulatie op de grond
- Slip indicator in PFD erg gevoelig

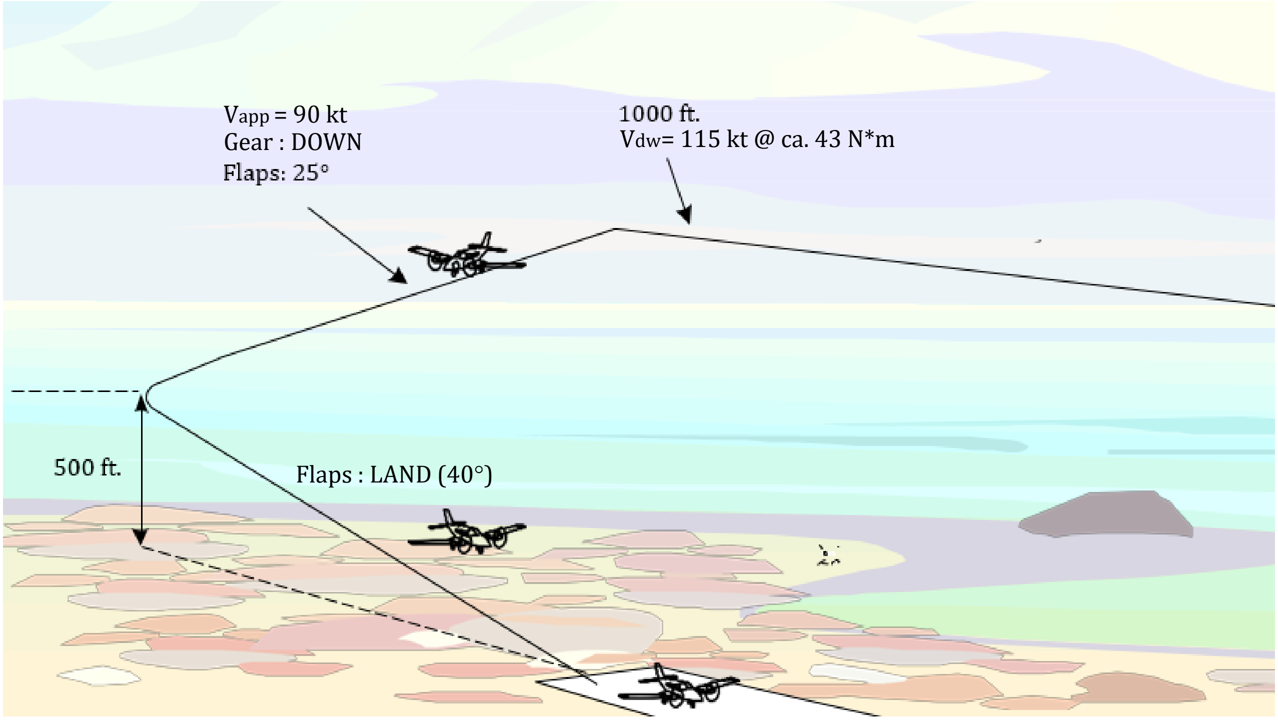
TU Delft

Preflight Briefing

- Houd je aan de gegeven opdracht
- Binnen de opdracht enige vrijheid
- Doe altijd je best
- Geef altijd een callout bij afwijkingen

TU Delft

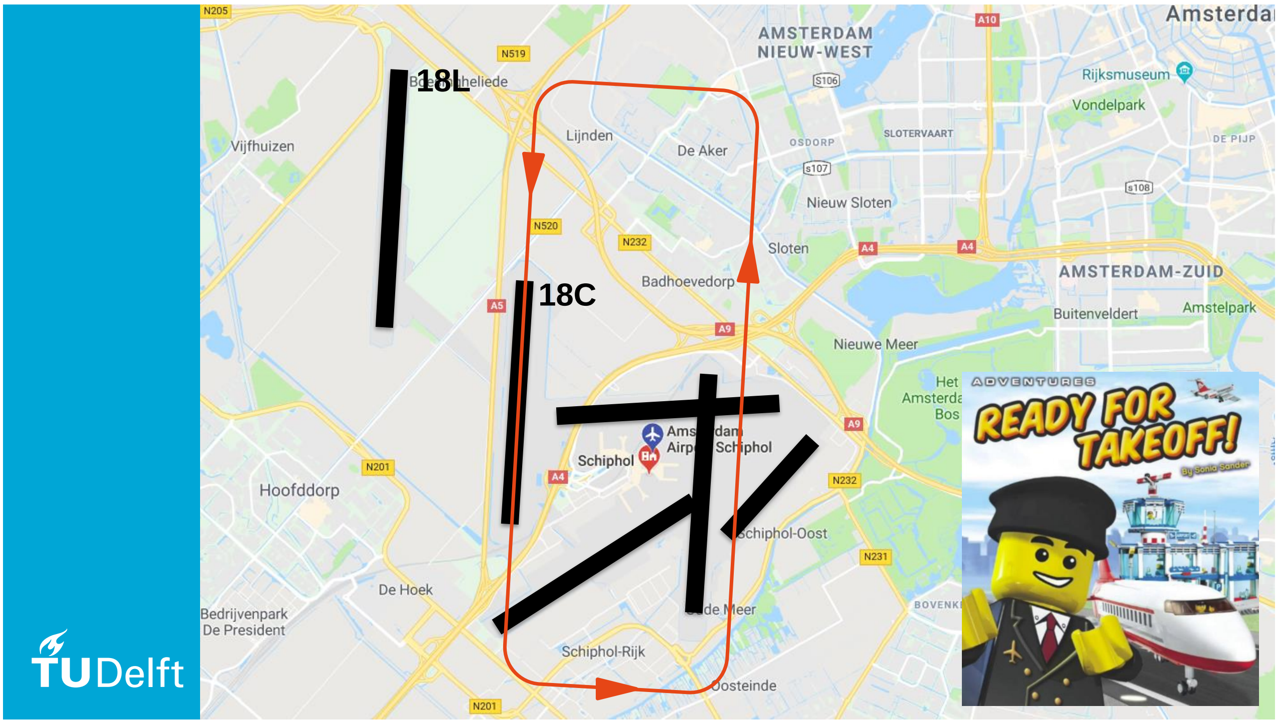




Circuit

- Schiphol 18C
- Circuit linksom, 1000 ft en 115 kts
- Touchdown op blokken naast PAPI
- Geen ATC

TU Delft


18L
18C
Vijfhuizen
Lijnden
De Aker
Badhoevedorp
Sloten
Nieuw Sloten
AMSTERDAM NIEUW-WEST
OSDORP
SLOTERVAART
Amsterda
Rijksmuseum
Vondelpark
DE PIJP
AMSTERDAM-ZUID
Buitenveldert
Amstelpark
Nieuwe Meer
Het Amsterda Bos
Schiphol
Amsterdam Airport Schiphol
Schiphol-Oost
Hoofddorp
De Hoek
Bedrijvenpark De President
Schiphol-Rijk
Oosteinde
N205
N519
N520
N232
A5
A4
A9
A10
N201
N231
ADVENTURES
READY FOR TAKEOFF!
TU Delft


TU Delft
Safety Briefing
SIMONA
TU Delft

Appendix B

Training Briefing

The following appendix contains the training briefing notes and Powerpoint presentations (Dutch only). Both groups received the first training briefing, while only the experimental group received the second one.

TRAININGSBRIEFING (+ PPT)

Trainingsbriefing #1

- Relevantie startle en surprise (FAA/EASA verplicht in UPRT vanaf 2019).
- CI006: crew verrast door flameout, gefocust op probleem, autopilot kan toestel niet op snelheid houden, door vertraging kan het rolmoment door de asymmetrische thrust niet gecounterd worden door de ailerons. Loss of control, verwarring leidt tot de verkeerde conclusie dat ADI's kapot zijn. Na 30,000 ft. uit de wolken en herstel.
- WCA708: De-icing systeem verlaagt service ceiling, crew kan snelheid niet houden. Toestel stallt, door verstoorde luchtstroom over de motoren lage EPR values. Crew denkt dat het een motorprobleem is. Stall leidt tot loss of control.
- Rol automatisering in startle en surprise: minder vliegtijd, weinig noodgevallen, automation surprise.
- Verschil tussen startle en surprise.
- Startle: schrikreflex, korte stressreactie.
- Voorbeeld startle: bird strike, blikseminslag.
- Surprise: verbazing, verwachtingen stroken niet met realiteit. Mentale plaatje aanpassen.
- Voorbeeld surprise: automation surprise, landing op verkeerde veld.

- Combinatie startle en surprise kan ervoor zorgen dat je ‘vastloopt’, wat invloed heeft op je troubleshooting en denkvermogen.
- Voorbeeld freeze: wegvallen communicatie, tegen elkaar insturen.

Trainingsbriefing #2

- Het is mogelijk om te trainen tegen startle en surprise en om te herstellen van ‘vastlopen’.
- Voorbeeld CAA i.c.m. workload (eisen omgeving versus capaciteiten).
- Methode testen die elementen uit bestaande en toegepaste methodes combineert.
- Doel van de methode is om startle en stress te managen en om je op weg te helpen met troubleshooting.
- Acronym: Keep it COOL!
- Methode tijdens het vliegen (Aviate-Navigate-Communicate).
- Eerste stap is ‘calm down’ – gericht op het managen van de eerste schrik:
  - Diep ademhalen, adem even vasthouden (2-3 sec), en rustig uitademen.
  - Tegelijkertijd rechtop gaan zitten, schouders naar achteren.
  - Wees je bewust van applied control forces en inputs.
- Tweede stap is ‘observe’ – gericht op het tegengaan van tunnelvisie en ‘het grote plaatje’:
  - Callout van basic six (pitch, speed, bank, altitude, vertical speed, koers).
  - Checken van secondary controls (throttle etc.).
  - Andere signalen? Wat hoor/zie/ruik/voel je?
- Derde stap is ‘outline’:
  - Wat klopt er wel, wat klopt er niet?
  - Probeer een ‘plan of action’ te verzinnen.
- Laatste stap is ‘lead’:
  - Voer je actieplan uit (bijv. configuratie aanpassen, vermoedens checken, etc.).
- We gaan nu enkele oefensessies in de SIMONA doen.
- Droog oefenen in SIMONA.


TU Delft
Startle & Surprise Experiment
Trainingsbriefing #1
SIMONA, TU Delft
Mei 2018
TU Delft


Case
CI 006
N4522V
TU Delft

Case

WCA 708

Ongevallen & Incidenten

- Air France 447
- Turkish Airlines 1951
- China Airlines 006
- West Caribbean Airways 708

Rol van de vlieger

Van actief vliegen naar ‘supervisor’

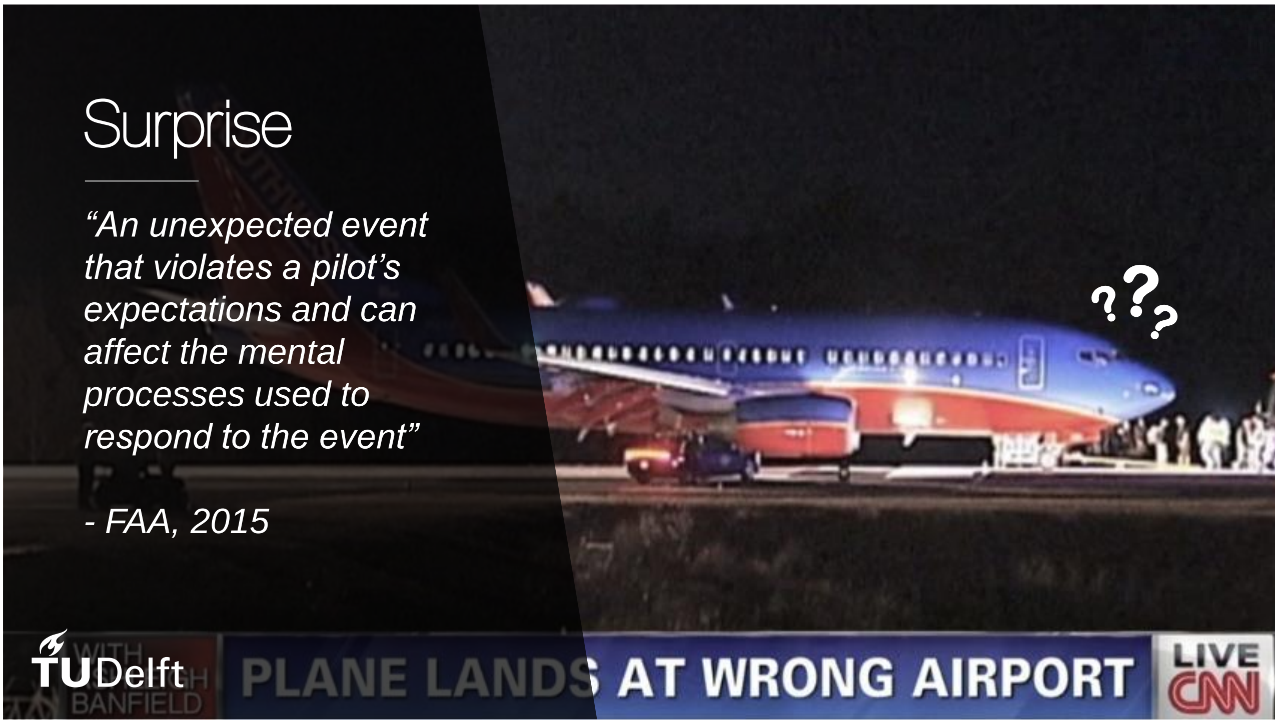
TU Delft

Startle, surprise en automatisering

- Automatisering zorgt voor complexe en uiterst betrouwbare systemen
- Minder ‘hands-on’ vliegtijd
- Noodgevallen komen nog maar weinig voor

TU Delft


Startle
“An uncontrollable, automatic muscle reflex, raised heart rate, blood pressure, etc., elicited by exposure to a sudden, intense event that violates a pilot’s expectations”
- FAA, 2015
TU Delft


Surprise
“An unexpected event that violates a pilot’s expectations and can affect the mental processes used to respond to the event”
- FAA, 2015
???
PLANE LANDS AT WRONG AIRPORT
LIVE CNN
TU Delft


Freeze
Wanneer een
oplossing niet
voor de hand ligt


SIMONA
Oefenen met
verrassende
situaties


TU Delft
Startle & Surprise Experiment
Trainingsbriefing #2
SIMONA, TU Delft
Mei 2018
TU Delft


INTERVENTION
Interventie
Trainen tegen 'vastlopen'
TU Delft

Interventie

"[During an IFR training flight] the increasing pressure (levels of concentration and air traffic input) eventually became too much. Loss of situational awareness and orientation was sudden with no obvious warning precursor. Despite staring at the instruments, nothing made sense (...) The reassuring words 'I have control' were sufficient to completely negate the workload pressure. As quickly as the mental picture was lost, it was regained."

- Civil Aviation Authority, 2016

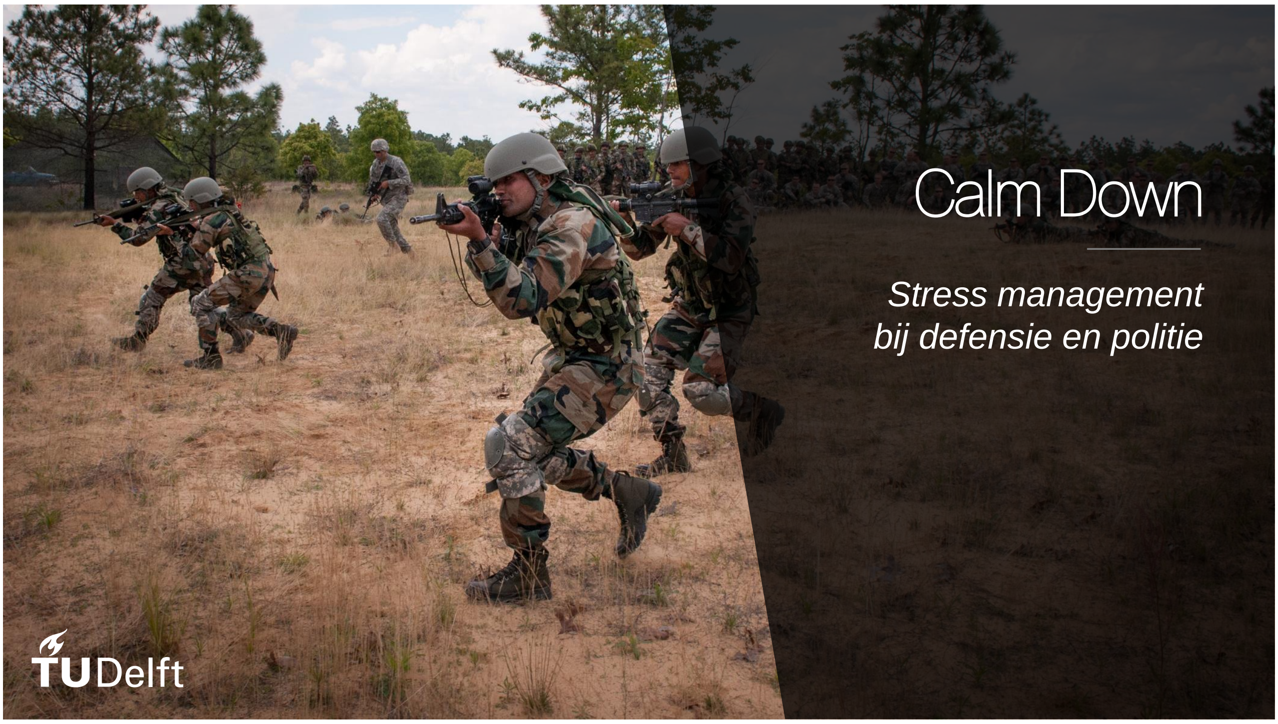
TU Delft

Omgaan met startle & surprise

- Training testen die bestaande methoden en technieken combineert
- Counteren van eerste schrik/stress
- Op weg helpen met troubleshooting

TU Delft


THIS IS FINE.
Problem?
Keep It COOL!
TU Delft


Calm Down
Stress management
bij defensie en politie
TU Delft

Calm Down

- Inademen, 2 sec vasthouden, uitademen
- Rechtop zitten, schouders naar achteren
- Bewust zijn van uitgeoefende control forces

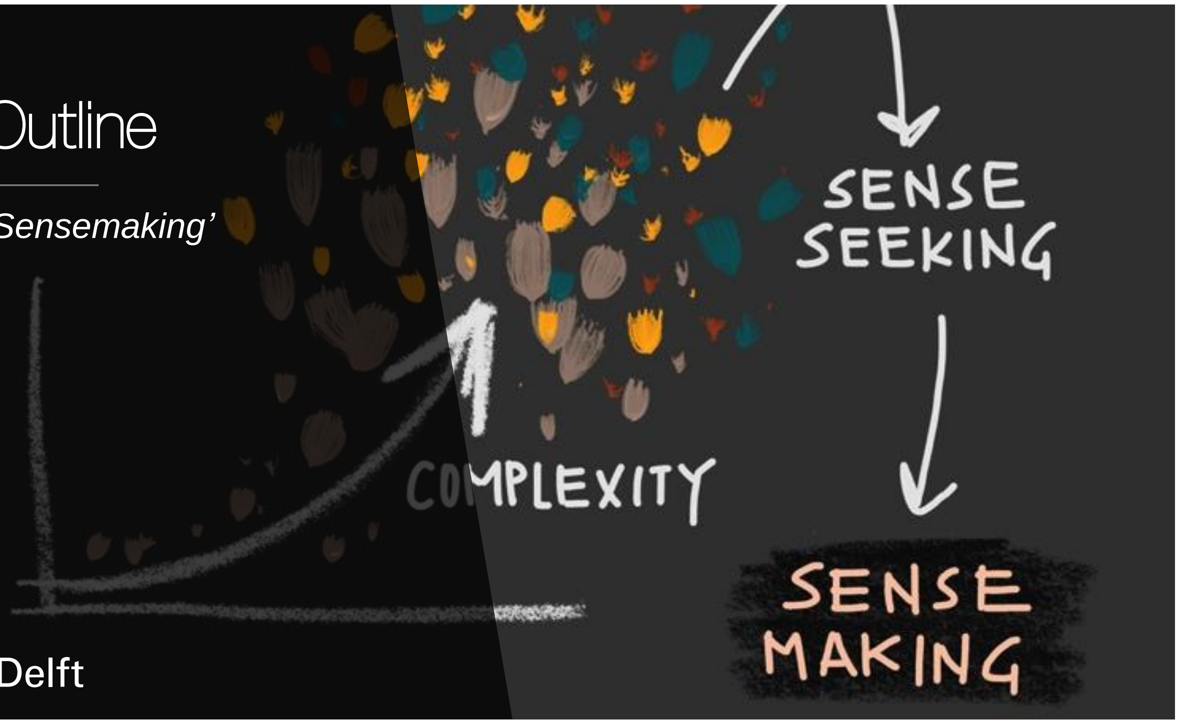
TU Delft

Observe

- Callout van basic six: pitch, speed, bank, altitude, vertical speed en heading
- Check secondary controls (throttle, etc.)
- Zintuigen: wat hoor/zie/voel/ruik je?

Outline

'Sensemaking'



Outline

- Op basis van je observaties
- Wat klopt er wel, wat klopt er niet?
- Probeer met een actieplan te komen

TU Delft

Lead

‘Decision-making’

TU Delft

Lead

- Actie ondernemen
- Troubleshooting

TU Delft

Appendix C

Piper Seneca III Kneepad

The following checklist was available to the participants during the simulator runs.

Takeoff	
Flaps	UP
V_r	80 kts
V_{mc} SE	80 kts
After takeoff	
Pitch	13 deg
V_2	92 kts
Gear	UP
Downwind	
V_{dw}	115 kts
Torque	43 Nm
Base leg	
V_{app}	90 kts
Flaps	25
Gear	DOWN
Final	
Flaps	LAND

Appendix D

Experimental Script

The following appendix contains the experimental script (Dutch only). Table D-2 shows the familiarization flights script, while table D-3 and table D-4 show the script for the training and test scenarios. Table D-1 shows the test scenario order for each participant (same for control and experimental group).

Table D-1: Test scenario order for each participant

Participant	1st	2nd	3rd	4th
1	701 (#1)	702 (#2)	501 (#4)	705 (#3)
2	702 (#2)	705 (#3)	701 (#1)	501 (#4)
3	705 (#3)	501 (#4)	702 (#2)	701 (#1)
4	501 (#4)	701 (#1)	705 (#3)	702 (#2)
5	701 (#1)	702 (#2)	501 (#4)	705 (#3)
6	702 (#2)	705 (#3)	701 (#1)	501 (#4)
7	705 (#3)	501 (#4)	702 (#2)	701 (#1)
8	501 (#4)	701 (#1)	705 (#3)	702 (#2)
9	701 (#1)	702 (#2)	501 (#4)	705 (#3)
10	702 (#2)	705 (#3)	701 (#1)	501 (#4)
11	705 (#3)	501 (#4)	702 (#2)	701 (#1)
12	501 (#4)	701 (#1)	705 (#3)	702 (#2)

Table D-2: Familiarization flights

Scenario ID	Omschrijving	Preflight	Wind	Tijdens vlucht / after failure	After landing
101 Fam. 1	1e keer takeoff oefenen	Je mag nu een takeoff roll oefenen. We zetten de simulatie stop op 400 ft.	Windstil	*400 ft*: top, we zetten je weer terug op de baan.	
102 Fam. 2	2e keer takeoff oefenen, nu met zijwind	Je mag nu een takeoff roll oefenen. We zetten de simulatie stop op 400 ft.	Zero niner zero at one zero	*400 ft*: top, we zetten je weer terug op de baan.	
101 Fam. 1	1e keer circuit oefenen	Takeoff, aflevelen op 1000 ft. Bocht linksom, aansluiten op std. circuit.	Windstil		Top, goed gedaan. Klaar voor de volgende?
102 Fam. 2	2e keer circuit oefenen	Takeoff, aflevelen op 1000 ft. Bocht linksom, aansluiten op std. circuit.	Zero niner zero at one zero		Top, goed gedaan. Klaar voor de volgende?
102 Fam. 2	2e keer circuit oefenen	Takeoff, aflevelen op 1000 ft. Bocht linksom, aansluiten op std. circuit.	Zero niner zero at one zero	*Downwind*: demonstratie stall, gear alert	Top, goed gedaan. Klaar voor de volgende?
423 Pretest	Engine failure tijdens approach	Land op de baan voor je. Je begint met gear down en flaps 25. We letten op je prestaties, **precisielanding** op blokken en centerline maken.	Zero niner zero at one zero		Dat was gemeen van ons, maar goed gedaan. Je mag de questionnaire over de pretest invullen

Table D-3: Training scenarios

Scenario ID	Omschrijving	Preflight	Wind	Tijdens vlucht / after failure	After landing
101 Fam. 1	1e keer COOL oefenen	Takeoff, aflevelen op 1000 ft. Bocht linksom, aansluiten op std. circuit. Wij zullen je af en toe vragen om COOL toe te passen.	Windstil	1x COOL tijdens takeoff, 1x cruise, 2x downwind, 1x base leg, 1x final	Top, COOL onder de knie? Nu door naar trainingsscenario's met verrassende elementen.
201	No rudder landing	Land op de baan voor je. Je begint met gear down en flaps 25. COOL toepassen bij afwijkingen.	Zero niner zero at one niner	*After rudder failure*: probeer je opdracht te voltooien	Goed gedaan, wat was het probleem? Feedback op COOL.
703	Engine indicator failure	Takeoff, aflevelen op 1000 ft. Bocht linksom, aansluiten op std. circuit. COOL toepassen bij afwijkingen.	Zero niner zero at one five	*After RPM indicator failure*: Check, doe wat je nodig vindt om je opdracht af te maken.	Goed gedaan, wat was het probleem? Feedback op COOL.
411	Rudder hardover	Flyby op 200 ft met 100 kts, boven de centerline. Gear down, flaps 25. COOL toepassen bij afwijkingen.	Two seven zero at one niner	*After rudder hardover*: probeer je flyby te voltooien.	Goed gedaan, wat was het probleem? Feedback op COOL.
704	Engine failure tijdens takeoff	Takeoff, aflevelen op 1000 ft. Bocht linksom, aansluiten op std. circuit. COOL toepassen bij afwijkingen.	One six zero at four	*After engine failure*: probeer door te klimmen naar 500 ft en de opdracht af te maken.	Goed gedaan, wat was het probleem? Feedback op COOL. Questionnaire training invullen.

Table D-4: Test scenarios

Scenario ID	Omschrijving	Preflight	Wind	Tijdens vlucht / after failure	After landing
701 (#1)	Flap asymmetry	Takeoff, aflevelen op 1000 ft. Bocht linksom, aansluiten op std. circuit. Probeer een **precisielanding** op de centerline en blokken te maken.	Two seven zero at one two		Goed gedaan, wat was het probleem? Questionnaire scenario 701 invullen.
702 (#2)	False stall	Takeoff, aflevelen op **2000 ft**. Bocht **rechtsom**, aansluiten op **rechterhandcircuit**. Hoogte eraf vliegen tot 1000 ft. op downwind. Probeer een **precisielanding** op centerline en blokken te maken.	Two six zero at six	*Stopzetten na herstel stall, bij stabiele snelheid rond 115 kts.*	Goed gedaan, wat was het probleem? Questionnaire scenario 702 invullen.
705 (#3)	Mass shift	Takeoff, aflevelen op 1000 ft. Bocht linksom, aansluiten op std. circuit. Probeer een **precisielanding** op centerline en blokken te maken.	Two seven zero at one three		Goed gedaan, wat was het probleem? Questionnaire scenario 705 invullen.
501 (#4)	ASI failure	Nieuw veld, Lelystad 05. Takeoff, aflevelen op 1000 ft. Bocht linksom, aansluiten op std. circuit. Probeer een **precisielanding** op de centerline en naast de PAPI te maken.	Three four zero at one four		Goed gedaan, wat was het probleem? Questionnaire scenario 501 invullen.

Appendix E

Questionnaires

The following pages contain the questionnaires used during the experiment. The first questionnaire is the preflight briefing questionnaire, followed by the scenario and training questionnaires.

Informatie deelnemer experiment Startle & Surprise

Naam:..

Leeftijd:.....................

Ervaring met vliegtuigtypen + vlieguren:

Type	Uren

Bijzondere vliegervaring (bijv. aerobatics, zweefvliegen):..................................

....

Aantal jaren werkzaam als vlieger:....................................

Huidige functie:..

Werkgever:...

Indien u reiskostenvergoeding (kilometervergoeding) wenst:

Adres vertrek:..

....

Rekeningnummer:..

Handtekening:..

Is er bij u in de training bijzondere aandacht besteed aan startle en surprise? Zo ja, op welke wijze?

....

....

....

....

STAI - zelfbeoordelingsvragenlijst

De volgende vragenlijst dient om inzicht te krijgen in de deelnemersgroepen. Lees iedere uitspraak door en omcirkel dan één cijfer om aan te geven hoe u zich ***in het algemeen*** voelt.

	bijna nooit	soms	vaak	bijna altijd
Ik voel me prettig.	1	2	3	4
Ik voel me nerveus en onrustig.	1	2	3	4
Ik voel me tevreden.	1	2	3	4
Ik kan een tegenslag maar moeilijk verwerken.	1	2	3	4
Ik voel me in vrijwel alles tekort schieten.	1	2	3	4
Ik voel me uitgerust.	1	2	3	4
Ik voel me rustig en beheerst.	1	2	3	4
Ik voel dat de moeilijkheden zich opstapelen zodat ik er niet meer tegenop kan.	1	2	3	4
Ik pieker teveel over dingen die niet zo belangrijk zijn.	1	2	3	4
Ik ben gelukkig.	1	2	3	4
Ik word geplaagd door storende gedachten.	1	2	3	4
Ik heb gebrek aan zelfvertrouwen.	1	2	3	4
Ik voel me veilig.	1	2	3	4
Ik voel me op mijn gemak.	1	2	3	4
Ik ben gelijkmatig van stemming.	1	2	3	4
Ik ben tevreden.	1	2	3	4
Er zijn gedachten die ik maar moeilijk los kan laten.	1	2	3	4
Ik neem teleurstellingen zo zwaar op dat ik ze niet van me af kan zetten.	1	2	3	4
Ik ben een rustig iemand.	1	2	3	4
Ik raak helemaal gespannen en in beroering als ik denk aan mijn zorgen de laatste tijd	1	2	3	4

Scenario #XXX

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 had happened?

Not at all | Extremely

0	1	2	3	4	5	6	7	8	9	10

How much tension or anxiety did you feel during the scenario? (please place a cross on the line).

Extremely little ◆————————————◆ Maximum

To what extend did your previous experience or training (before the experiment) help you to deal with the issue?

Not at all	Somewhat	Moderately	Much	Very much

If you applied the "COOL" intervention method: (Only shown for exp. group)

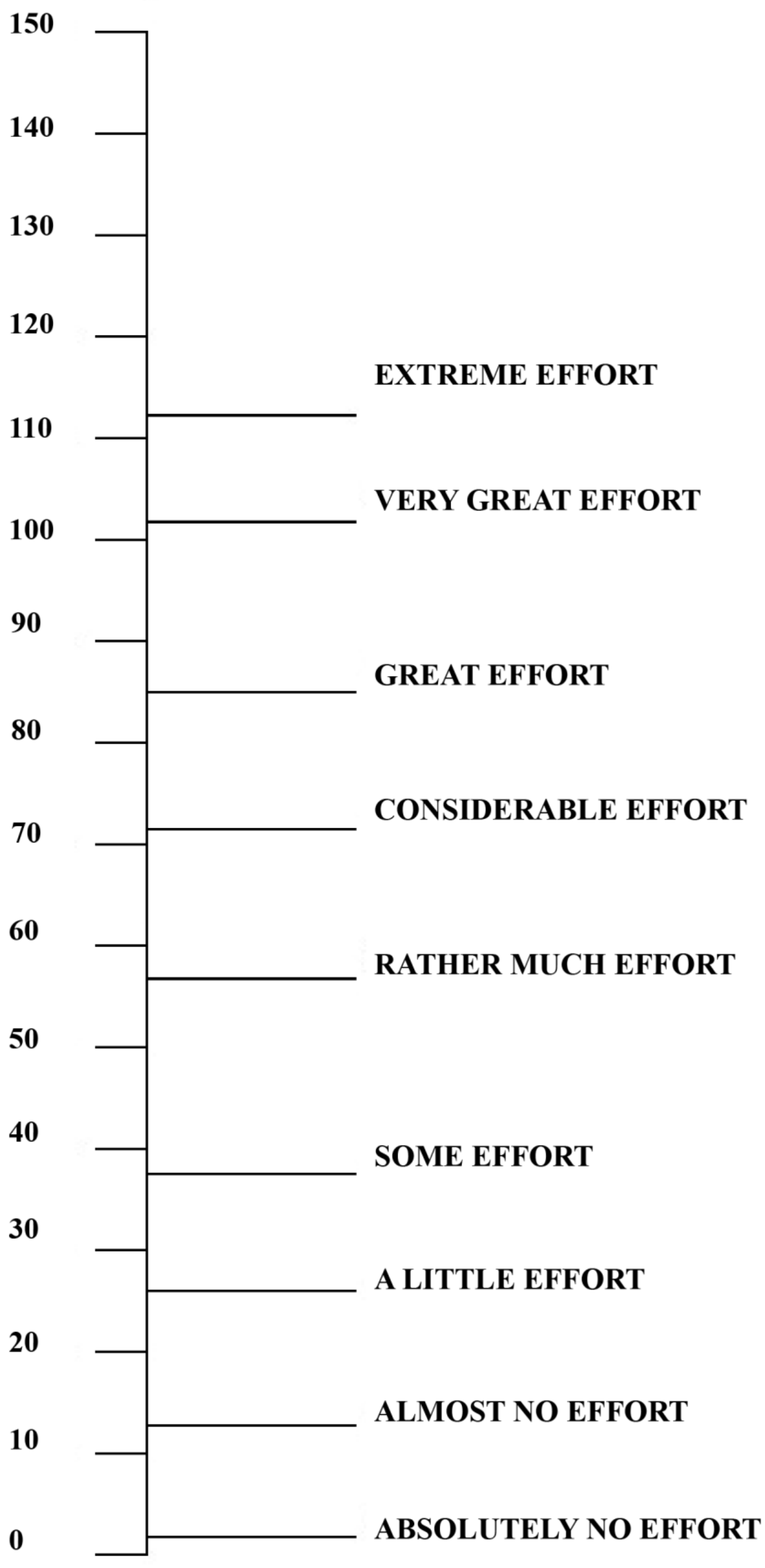
Which aspects of the method did you use?

☐ Calm down (breathe deeply, relax muscles)
☐ Observe (scan information out loud)
☐ Outline (formulate the issue)
☐ Lead (decide consciously to take an action)

To what extend did the COOL intervention method help you?

Not at all	Somewhat	Moderately	Much	Very much

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



Training

	Not at all true			Somewhat true			Very true
This training was fun to do.	1	2	3	4	5	6	7
I would describe this training as very interesting.	1	2	3	4	5	6	7
This training did not hold my attention at all.	1	2	3	4	5	6	7
I thought this training was quite enjoyable.	1	2	3	4	5	6	7
While I was doing this training, I was thinking about how much I enjoyed it.	1	2	3	4	5	6	7
I thought this was a boring training.	1	2	3	4	5	6	7
I enjoyed doing this training very much	1	2	3	4	5	6	7

Appendix F

Questionnaire Results

All questionnaire results have been processed by H.M. Landman. Table F-1 shows the balancing of the groups. The first rows contain participant data (age, experience, etc.), followed by the STAI score and enjoyment of the experimental training. Furthermore, additional ratings, experience and ranks are shown for each group.

The other tables show the questionnaire results, including startle, surprise and confusion (understanding) ratings for each group and scenario. Training ratings show how much previous training (before the experiment) helped the participants to fly the particular scenario on a 1-5 scale.

The final table shows the helpfulness of *'COOL'* per scenario, as indicated by pilots in the experimental group.

Table F-1: Balancing of groups

	Experimental group Mean (SD)	Control group Mean (SD)	t	p
Age	37.4 (12.7)	39.6 (11.7)	0.44	0.668
Hours SEP	228 (93)	372 (434)	1.13	0.282
Hours MEP	36.9 (21.9)	20.8 (8.6)	2.38	0.027
Hours Large	7172 (5549)	7544 (5851)	0.16	0.874
Years employed	13.5 (10.8)	14.7 (10.9)	0.26	0.801
STAI (20-80)	28.9 (12.3)	24.9 (4.3)	1.07	0.298
Enjoyment training	39.4 (5.6)	44.3 (3.6)	2.57	0.018

	No. of pilots	No. of pilots
> 50 hrs MEP	1	0
Aerobatics	2	4
Startle training	2	4
Glider rating	4	3
UPRT	4	4
Instructor	4	3
Rank, CPT	4	6
Rank, FO	6	5
Rank, SO	2	1

Table F-2: Pretest: questionnaire results

	Experimental group Mean (SD)	Control group Mean (SD)	p
Startle (0-10)	4.92 (2.07)	5.75 (2.30)	0.361
Surprise (0-10)	6.67 (2.43)	6.58 (2.84)	0.939
Confusion (0-10)	2.17 (1.40)	4.00 (2.52)	0.039
Anxiety (0-10)	3.74 (2.28)	4.14 (2.42)	0.681
Training (1-5)	3.92 (1.31)	3.83 (1.03)	0.864
RSME (0-80)	59.9 (18.6)	59.6 (23.3)	0.985

Table F-3: Flap asymmetry: questionnaire results

	Experimental group Mean (SD)	Control group Mean (SD)	p
Startle (0-10)	5.25 (2.42)	5.33 (2.35)	0.933
Surprise (0-10)	6.67 (2.10)	5.83 (2.79)	0.418
Confusion (0-10)	6.83 (2.59)	4.83 (2.98)	0.093
Anxiety (0-10)	4.85 (2.47)	3.90 (1.59)	0.274
Training (1-5)	3.33 (1.37)	3.25 (0.87)	0.860
RSME (0-80)	62.3 (22.4)	51.2 (18.1)	0.193

Table F-4: False stall: questionnaire results

	Experimental group Mean (SD)	Control group Mean (SD)	p
Startle (0-10)	6.75 (1.42)	7.25 (2.53)	0.556
Surprise (0-10)	7.33 (1.23)	5.42 (2.71)	0.036
Confusion (0-10)	5.83 (2.55)	5.33 (2.35)	0.622
Anxiety (0-10)	4.98 (2.07)	3.40 (2.25)	0.086
Training (1-5)	3.42 (1.08)	3.33 (1.44)	0.874
RSME (0-80)	58.0 (18.0)	45.0 (18.0)	0.091

Table F-5: Mass shift: questionnaire results

	Experimental group Mean (SD)	Control group Mean (SD)	p
Startle (0-10)	6.58 (1.93)	6.08 (2.61)	0.599
Surprise (0-10)	7.83 (0.94)	6.58 (3.09)	0.203
Confusion (0-10)	7.92 (1.17)	5.63 (2.48)	0.011
Anxiety (0-10)	6.03 (1.95)	5.00 (2.14)	0.233
Training (1-5)	3.50 (1.31)	3.83 (1.12)	0.510
RSME (0-80)	74.4 (19.1)	65.3 (26.7)	0.344

Table F-6: Unreliable airspeed: questionnaire results

	Experimental group Mean (SD)	Control group Mean (SD)	p
Startle (0-10)	4.75 (2.30)	4.75 (2.80)	>0.999
Surprise (0-10)	7.08 (1.83)	5.42 (2.68)	0.089
Confusion (0-10)	5.33 (2.81)	4.75 (2.34)	0.586
Anxiety (0-10)	4.26 (2.12)	3.23 (2.32)	0.267
Training (1-5)	3.83 (1.12)	4.42 (0.67)	0.134
RSME (0-80)	59.7 (21.5)	50.6 (20.0)	0.296

Table F-7: Test scenarios: average of questionnaire results

	Experimental group Mean (SD)	Control group Mean (SD)	p
Startle (0-10)	5.83 (1.60)	5.85 (2.13)	0.979
Surprise (0-10)	7.23 (1.06)	5.81 (2.55)	0.089
Confusion (0-10)	6.48 (1.19)	5.14 (1.51)	0.024
Anxiety (0-10)	5.03 (1.19)	3.88 (1.51)	0.115
Training (1-5)	3.52 (1.08)	3.71 (0.75)	0.626
RSME (0-80)	63.6 (15.9)	53.0 (14.7)	0.104

Table F-8: Test scenarios versus pretest: difference in questionnaire results (2x2 ANOVA)

	Experimental group Mean (SD)	Control group Mean (SD)	p
Startle (posttest-pretest)	0.92 (0.60)	0.10 (0.60)	0.350
Surprise (posttest-pretest)	0.56 (0.82)	-0.77 (0.36)	0.260
Confusion (posttest-pretest)	4.31 (0.59)	1.14 (0.59)	0.001
Anxiety (posttest-pretest)	1.29 (0.03)	0.26 (0.56)	0.063
RSME (posttest-pretest)	3.69 (5.51)	-6.58 (5.51)	0.201

Table F-9: Test scenarios: helpfulness of COOL on a 1-5 scale

	No. of pilots rating *'COOL'*:				
	1	2	3	4	5
Flap asymmetry	1	1	2	8	0
False stall	0	1	2	8	1
Mass shift	1	1	6	4	0
Unreliable airspeed	0	2	4	6	0

Appendix G

MATLAB Performance Plots

Pretest Scenario

Figure G-1: Pretest: Loss of Control (LoC) data for control (left) and experimental (right) group

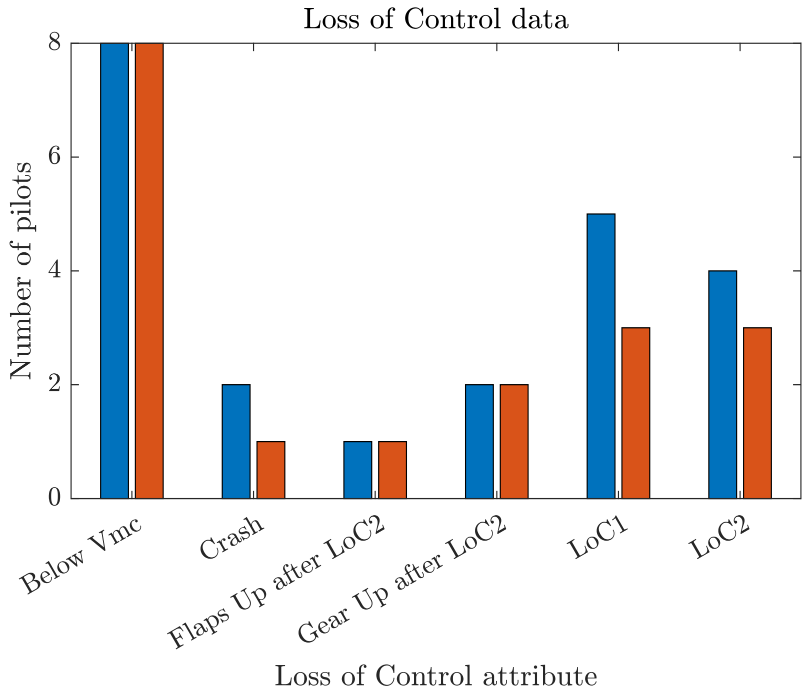


Figure G-2: Pretest: time spent below minimum control speed V_{mc}

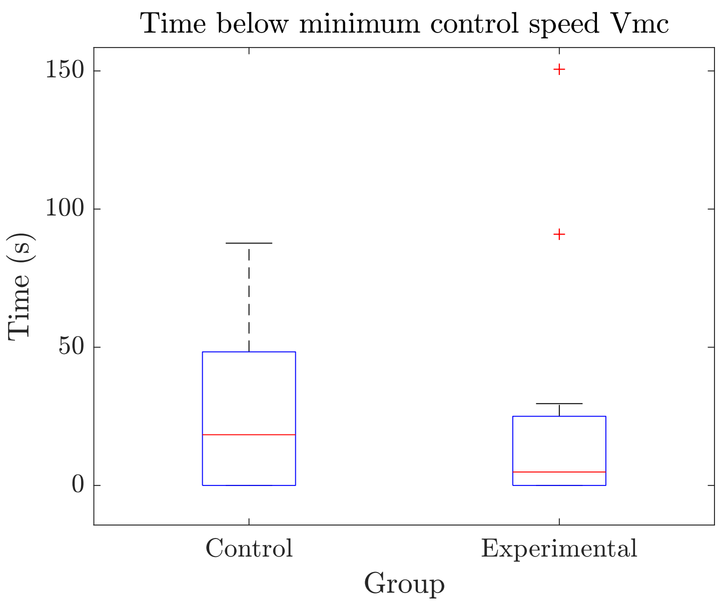


Figure G-3: Pretest: torque ratio for pilots dropping below V_{mc}

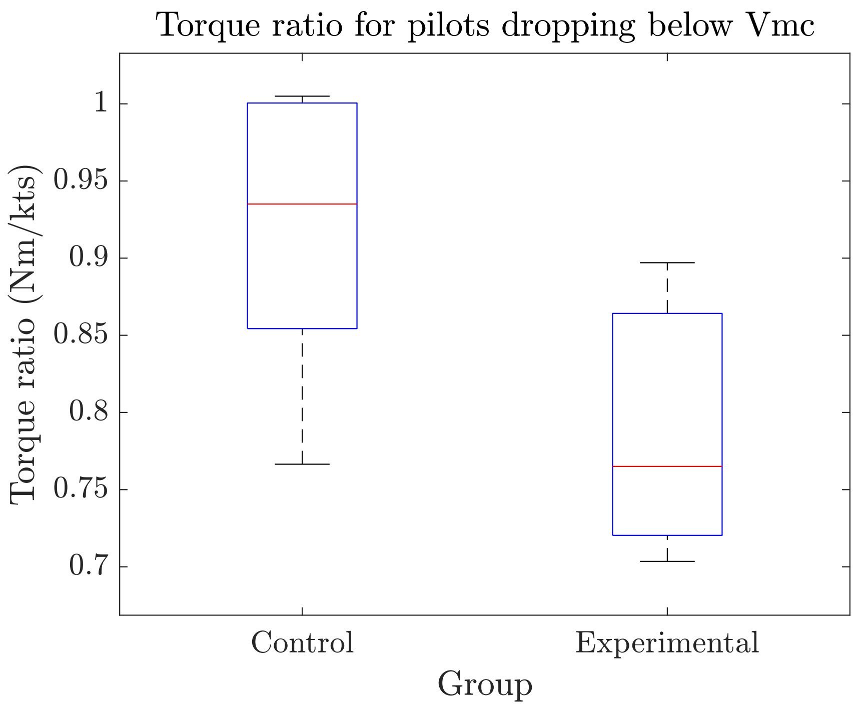
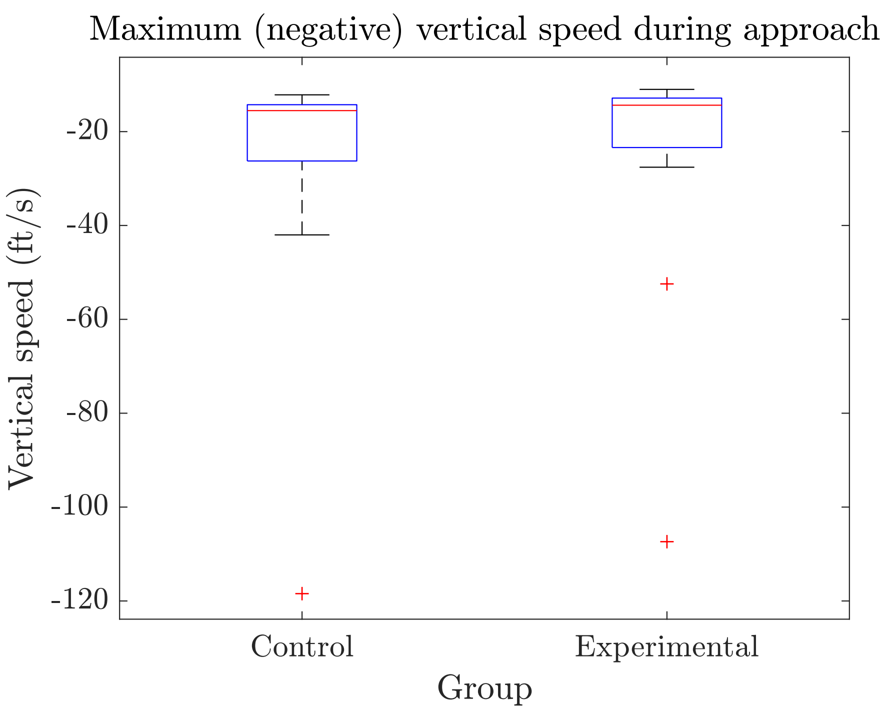


Figure G-4: Pretest: maximum (negative) vertical speed during approach



Flap Asymmetry Scenario

Figure G-5: Flap asymmetry: highest flap setting used for control (left) and experimental (right) group

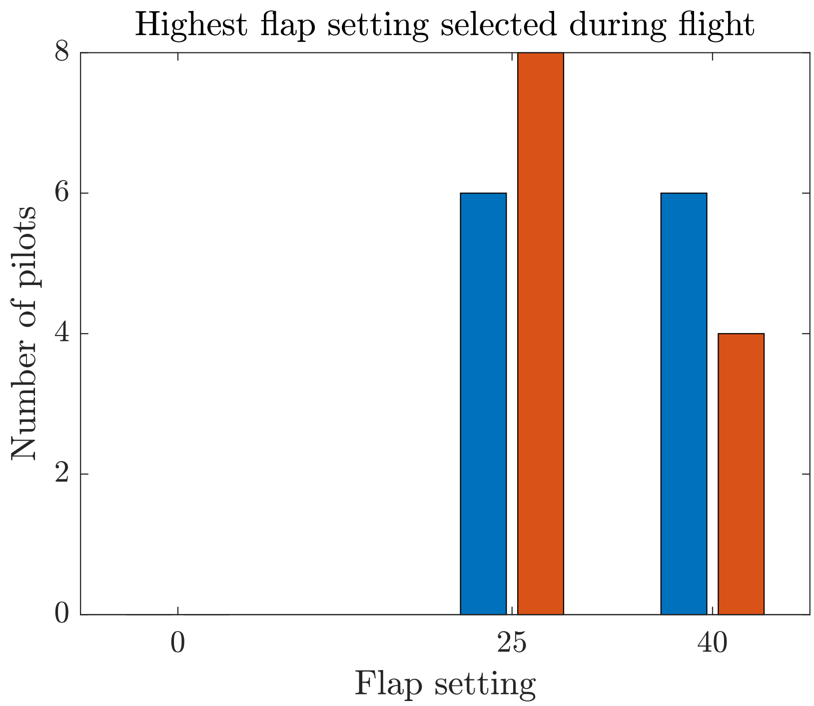


Figure G-6: Flap asymmetry: flap setting used for landing for control (left) and experimental (right) group

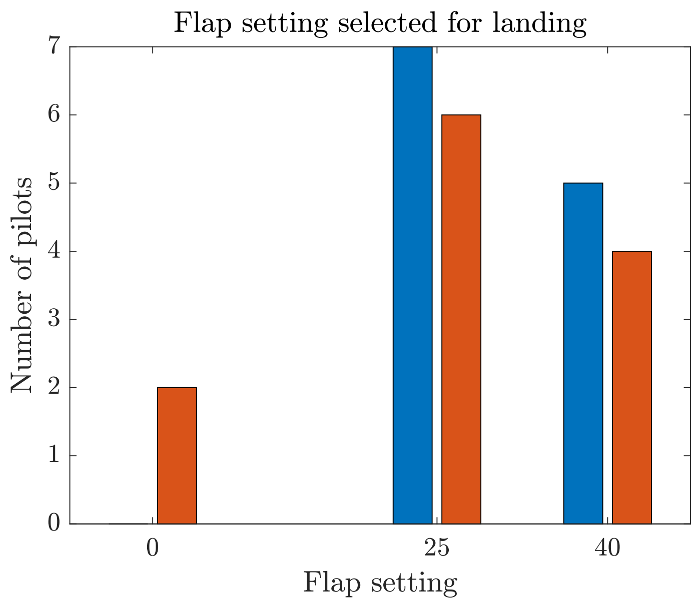


Figure G-7: Flap asymmetry: maximum roll angle attained after first (flaps 25) upset

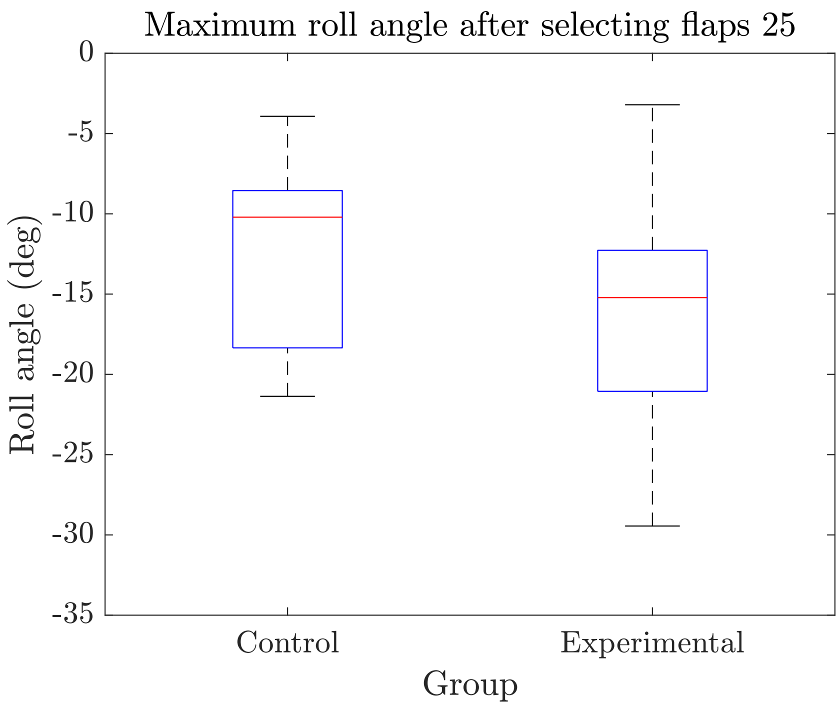


Figure G-8: Flap asymmetry: maximum roll angle attained after second (flaps 40) upset

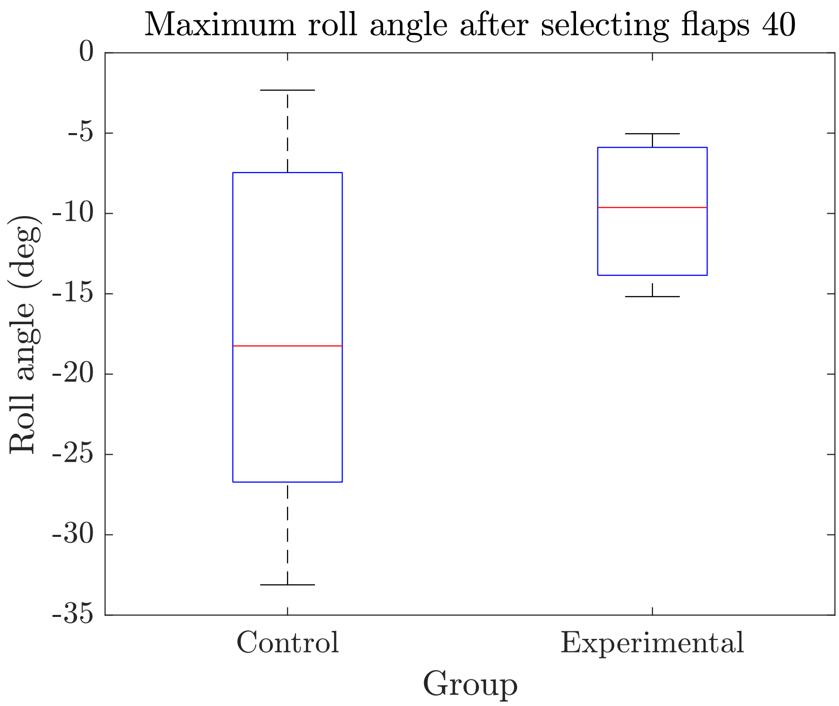
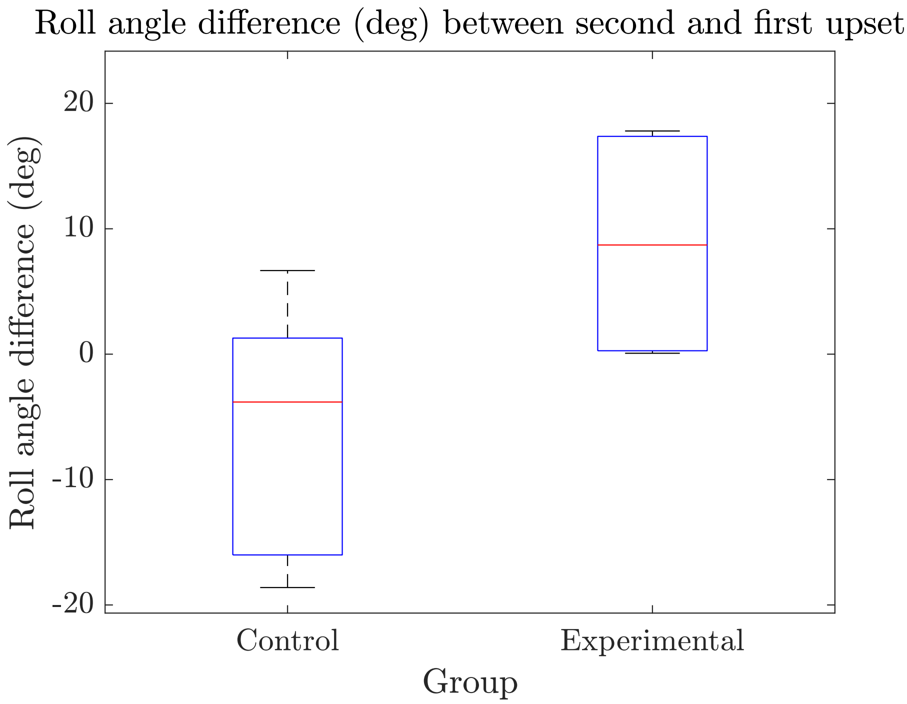


Figure G-9: Flap asymmetry: roll angle difference between first and second upset. Positive values indicate a performance improvement



False Stall Scenario

Figure G-10: False stall: maximum elevator deflection after stick shaker activation

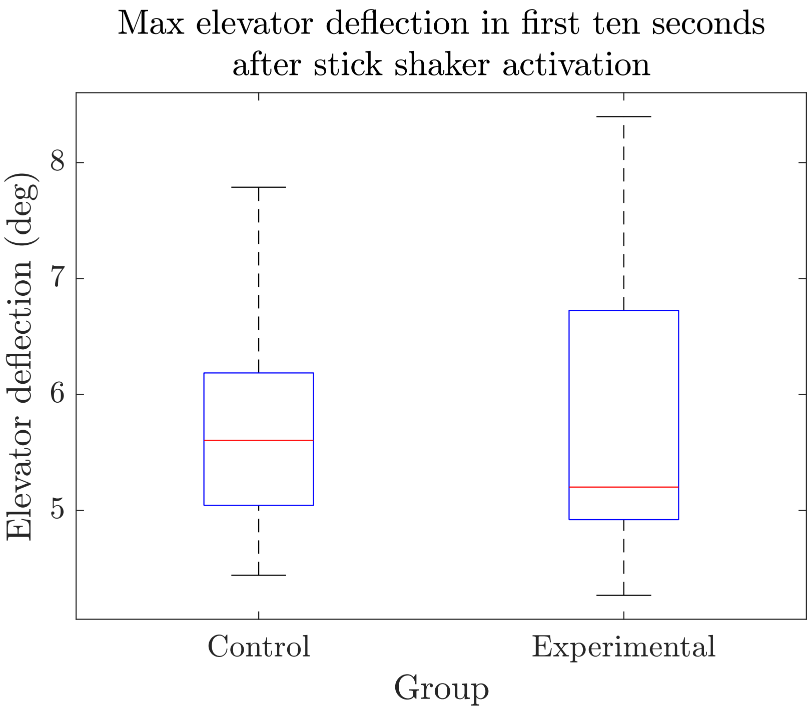


Figure G-11: False stall: minimum load factor after stick shaker activation

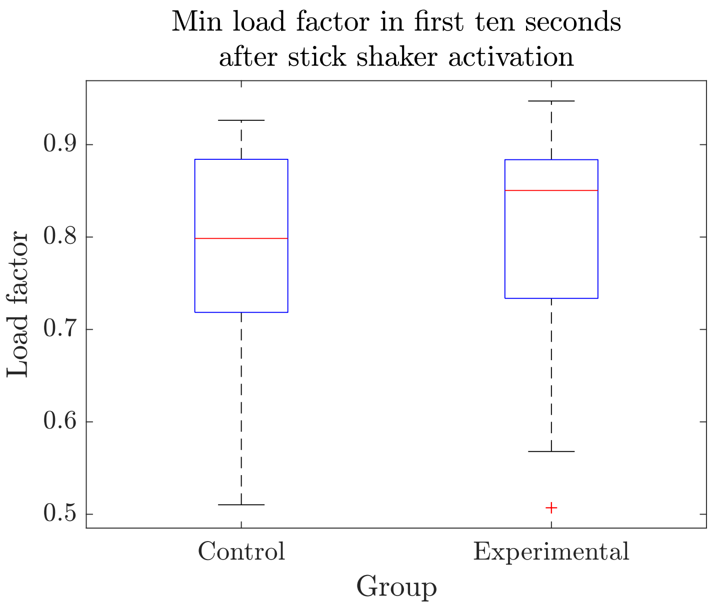


Figure G-12: False stall: minimum pitch angle after stick shaker activation

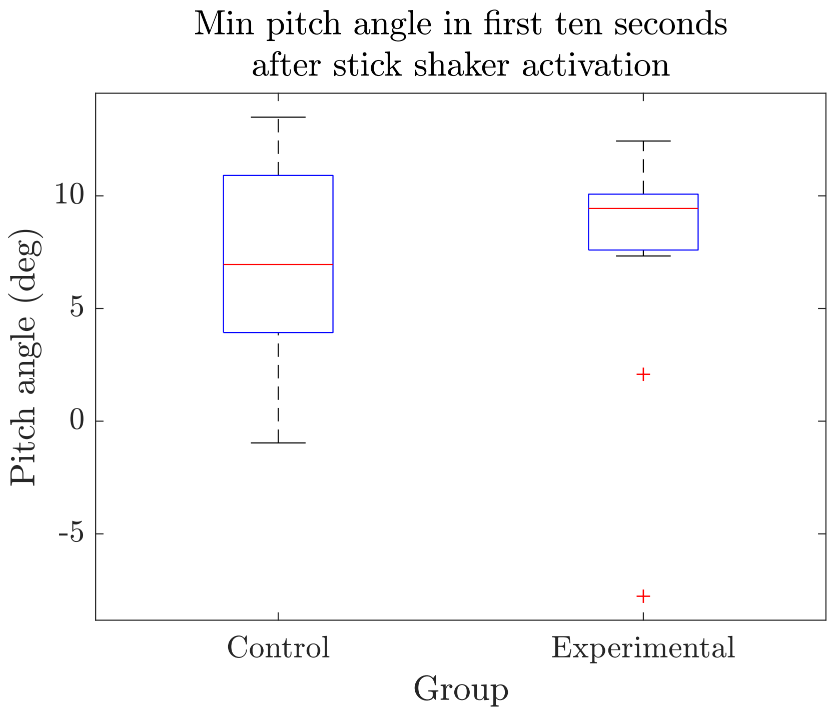


Figure G-13: False stall: minimum pitch rate after stick shaker activation

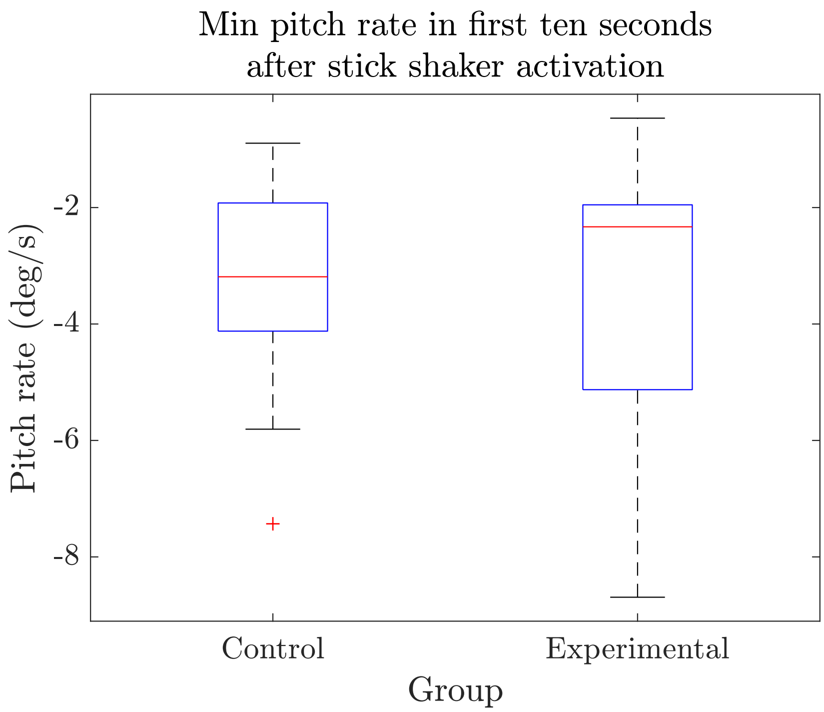


Figure G-14: False stall: minimum vertical speed after stick shaker activation

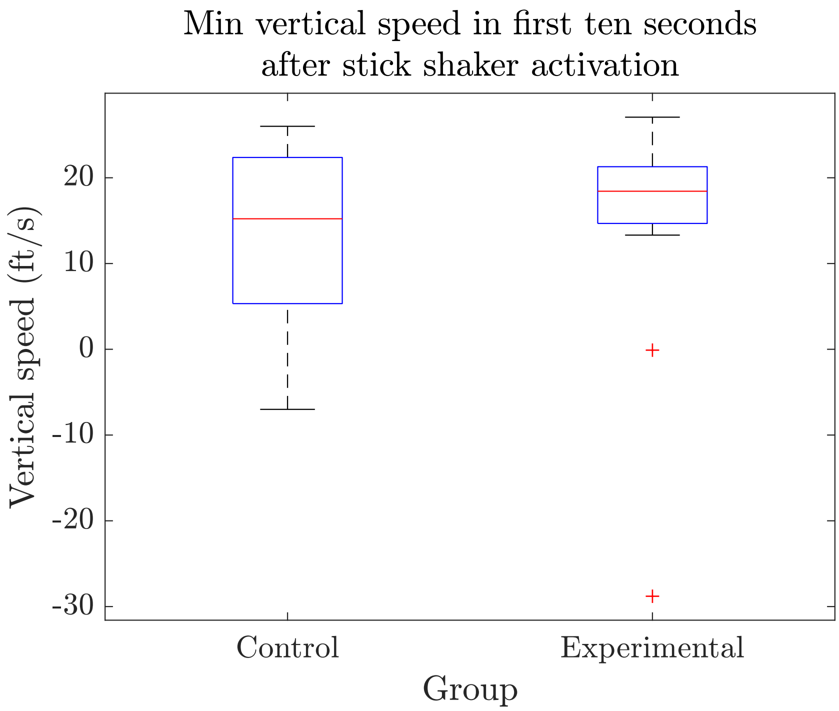
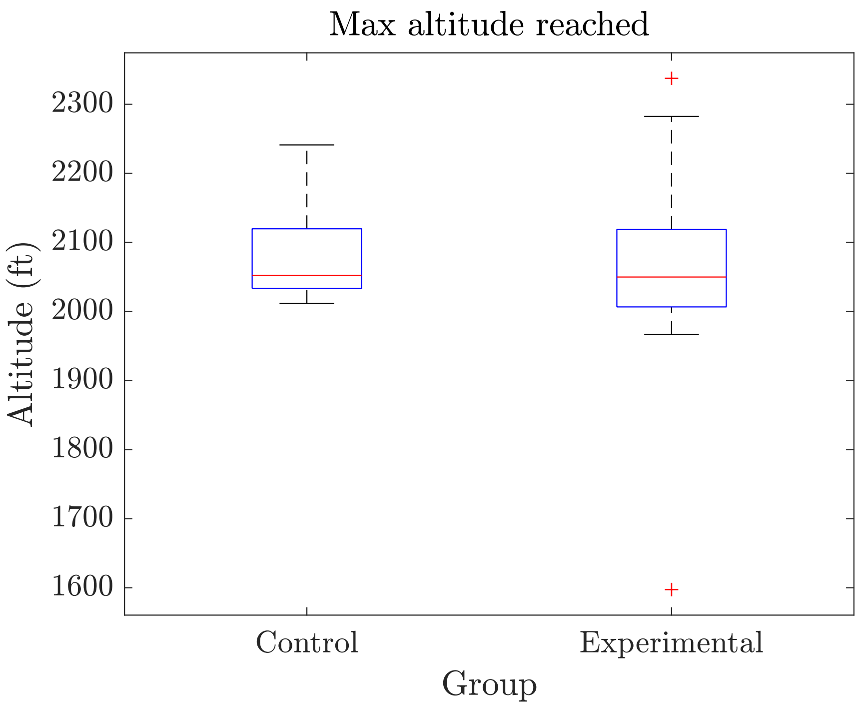


Figure G-15: False stall: maximum altitude



Faulty Airspeed Indicator Scenario

Figure G-16: Unreliable airspeed: time flown above 50 kts IAS

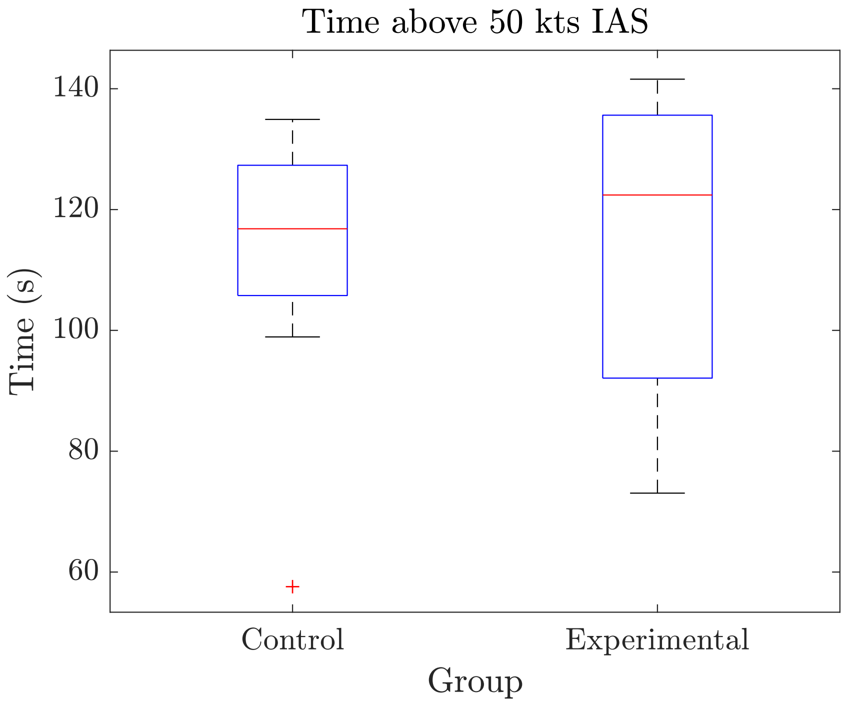


Figure G-17: Unreliable airspeed: time from rotation until below 60 kts IAS

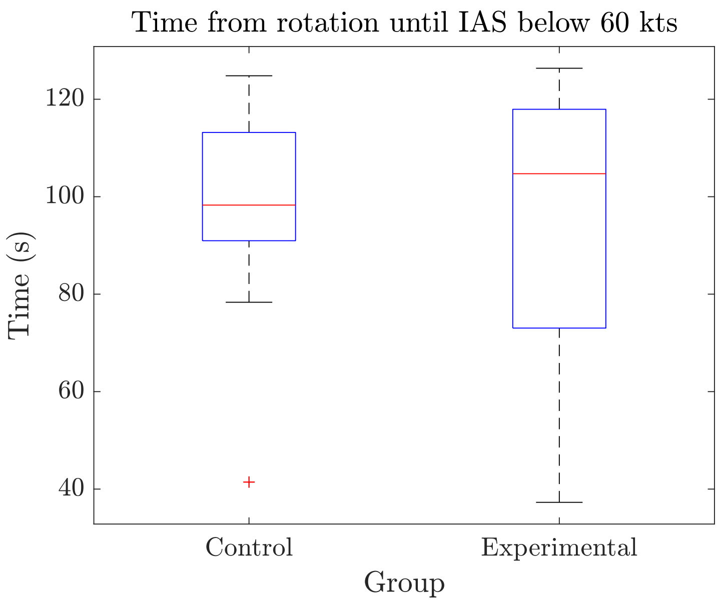


Figure G-18: Unreliable airspeed: time before selecting pitch-power

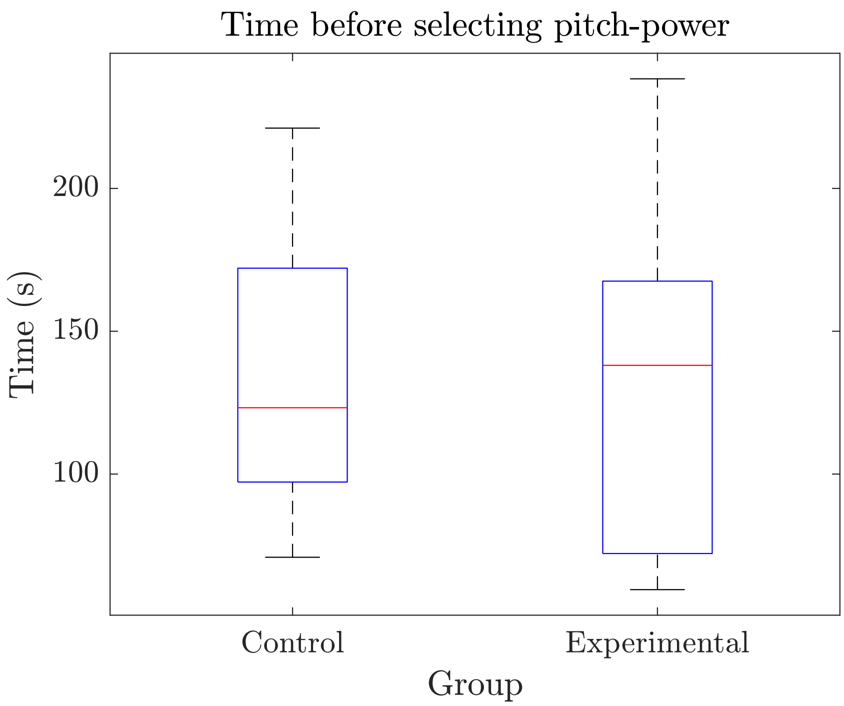


Figure G-19: Unreliable airspeed: IAS when selecting pitch-power

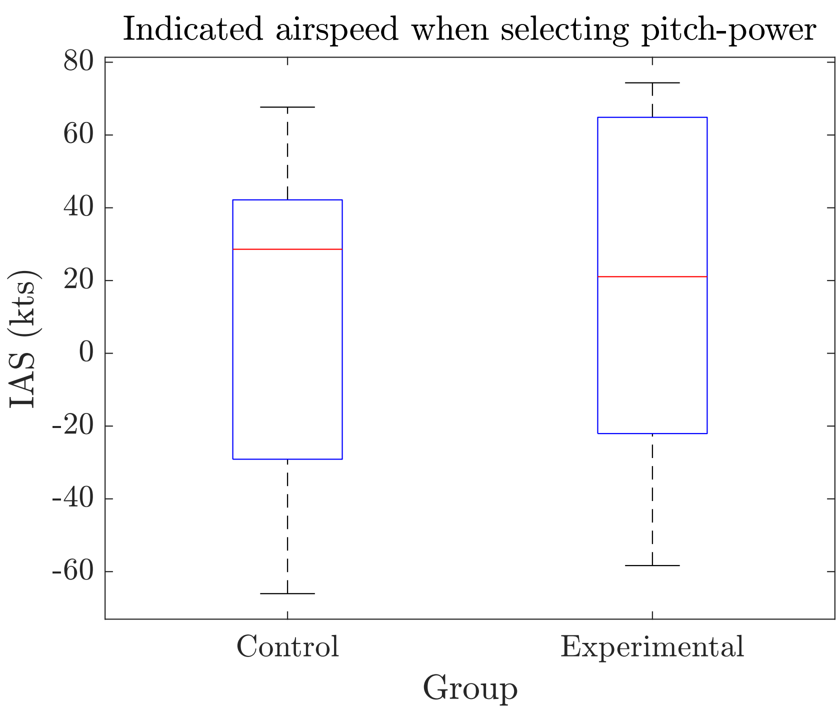


Figure G-20: Unreliable airspeed: TAS when selecting pitch-power

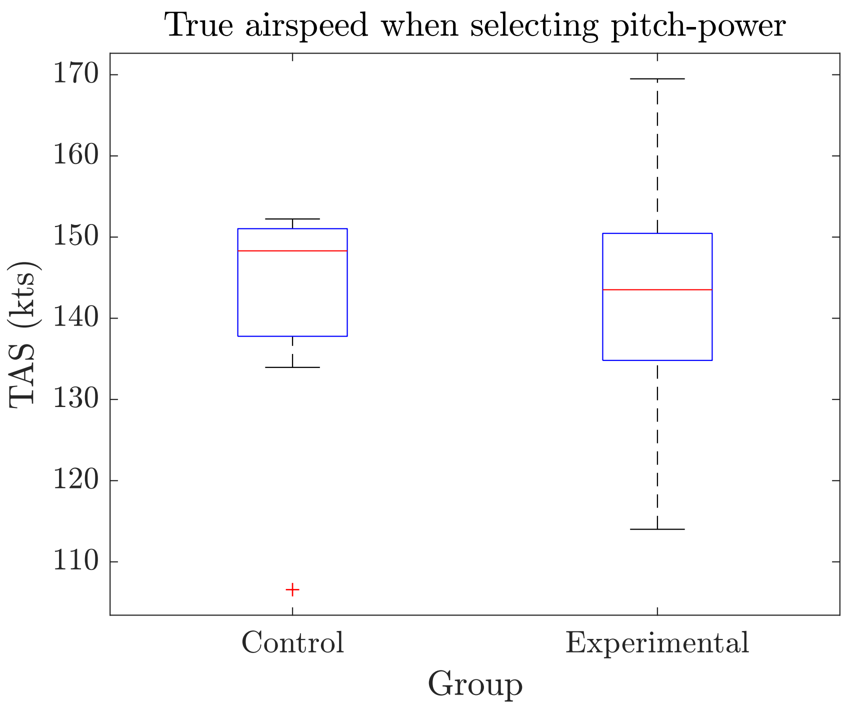


Figure G-21: Unreliable airspeed: maximum TAS

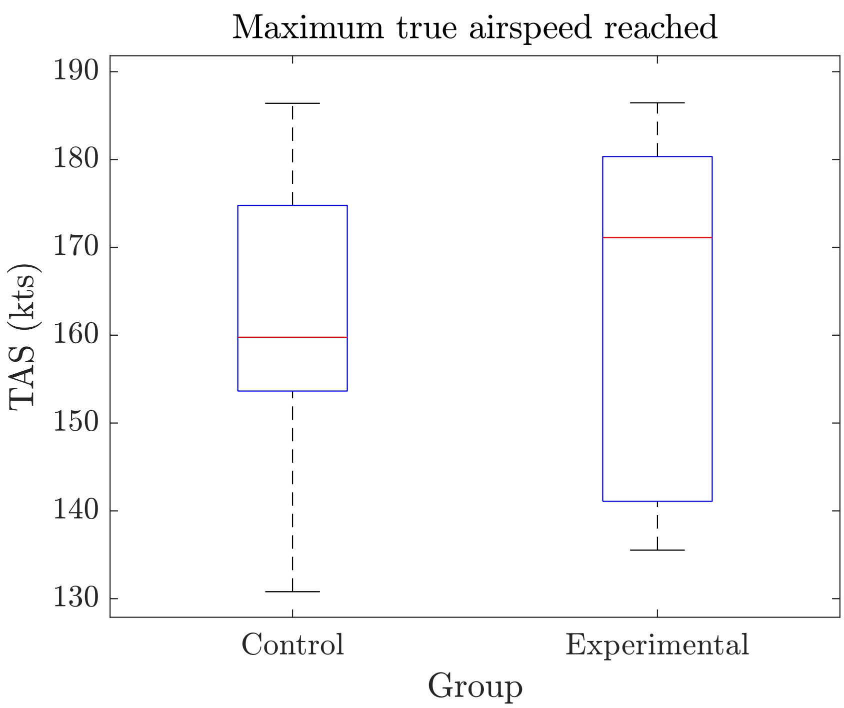
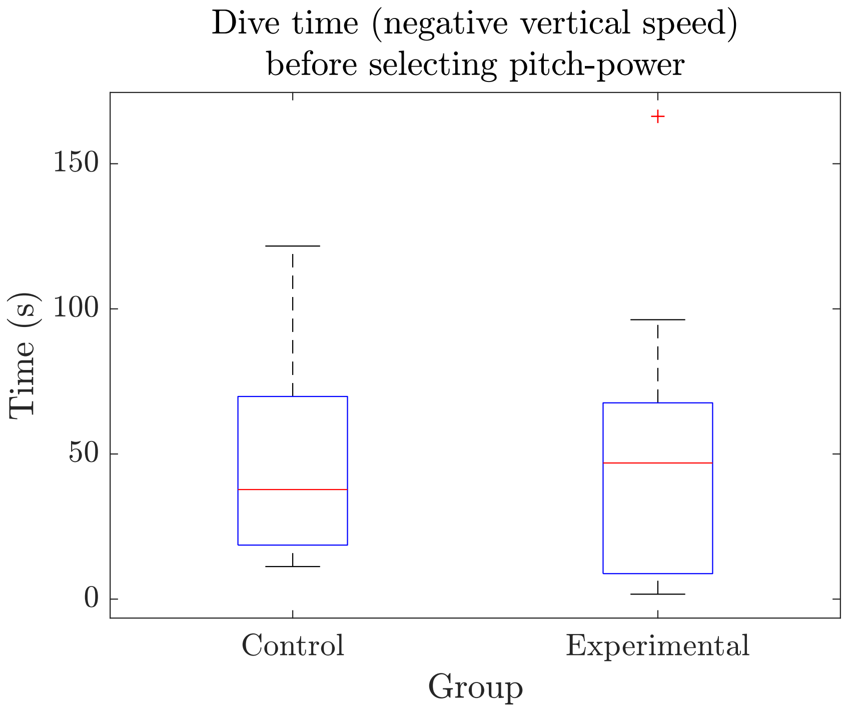


Figure G-22: Unreliable airspeed: time flown with a negative vertical speed before selecting pitch-power



Mass Shift Scenario

Figure G-23: Mass shift: results including crash, early configuration (>1000 ft), flap setting and power reduction (recovery from nose-high attitude) information for control (left) and experimental (right) group

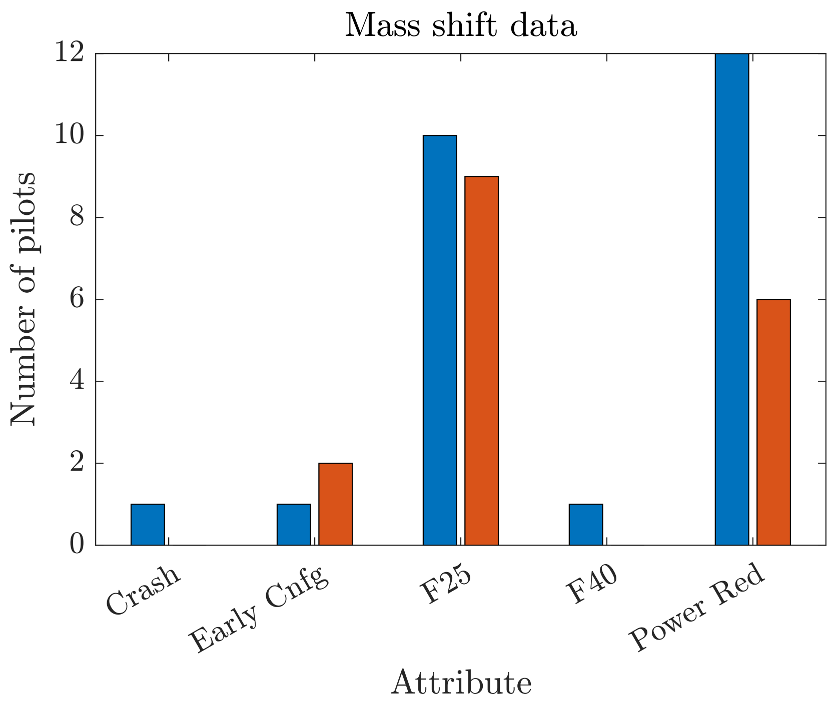


Figure G-24: Mass shift: flap setting used for landing for control (left) and experimental (right) group

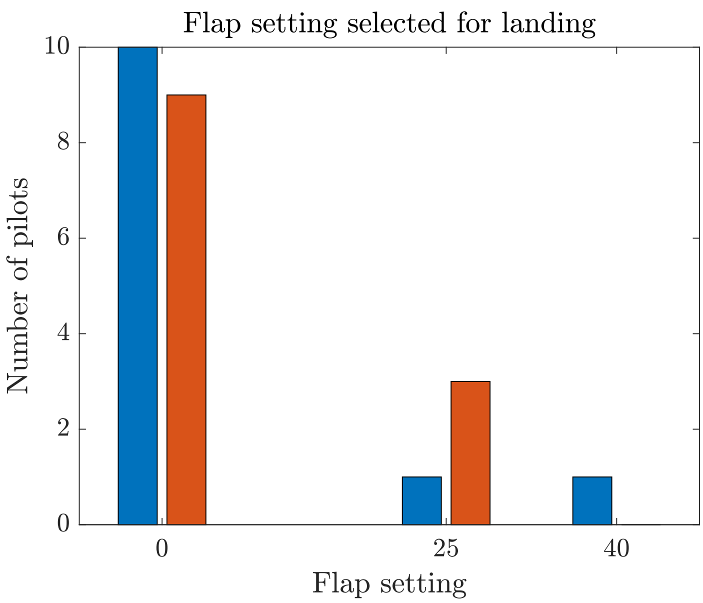


Figure G-25: Mass shift: maximum pitch angle attained after first, mass-induced upset

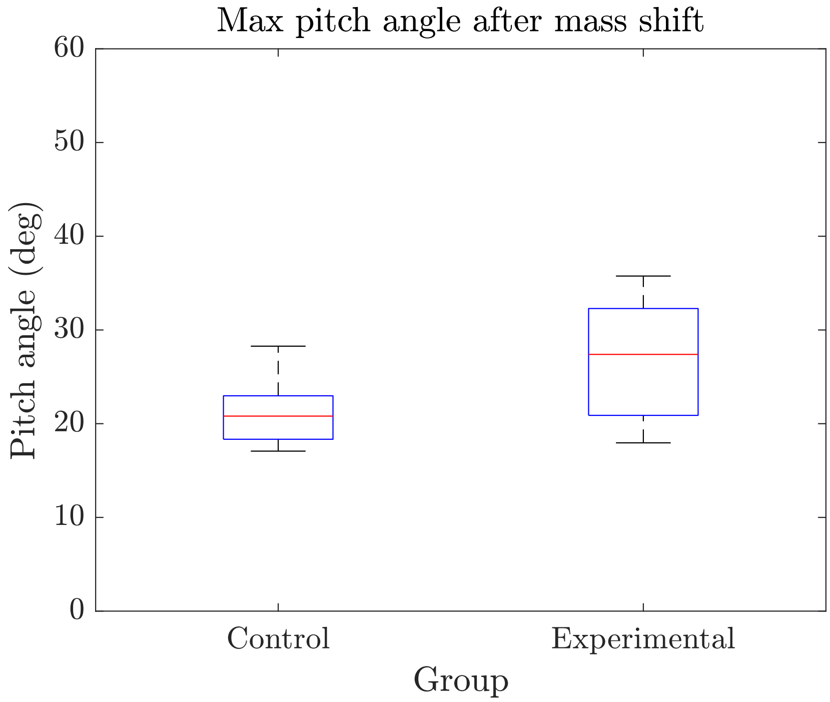


Figure G-26: Mass shift: maximum pitch angle attained after second, flap-induced upset

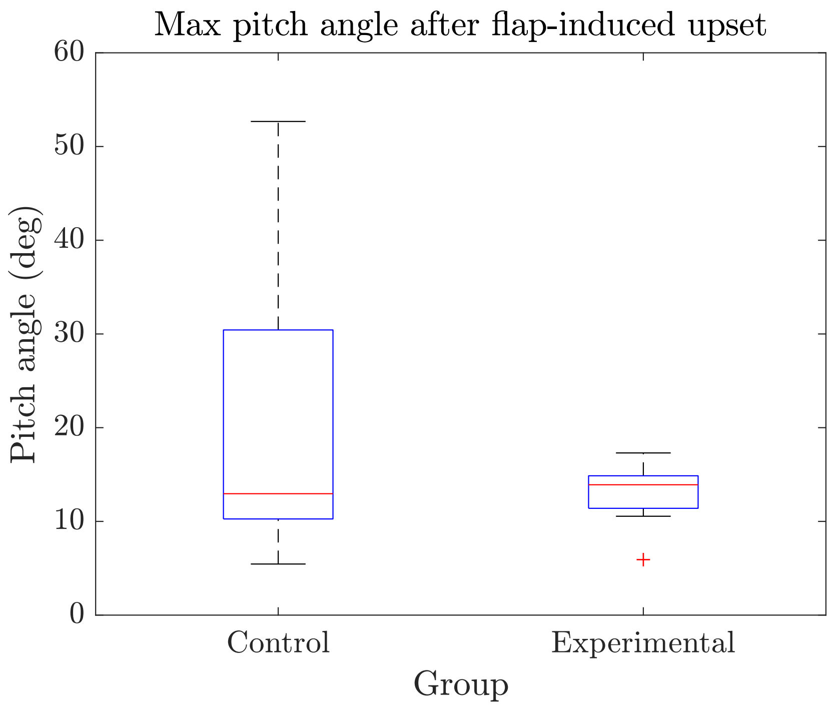


Figure G-27: Mass shift: pitch angle difference between first and second upset. Positive values indicate performance improvement

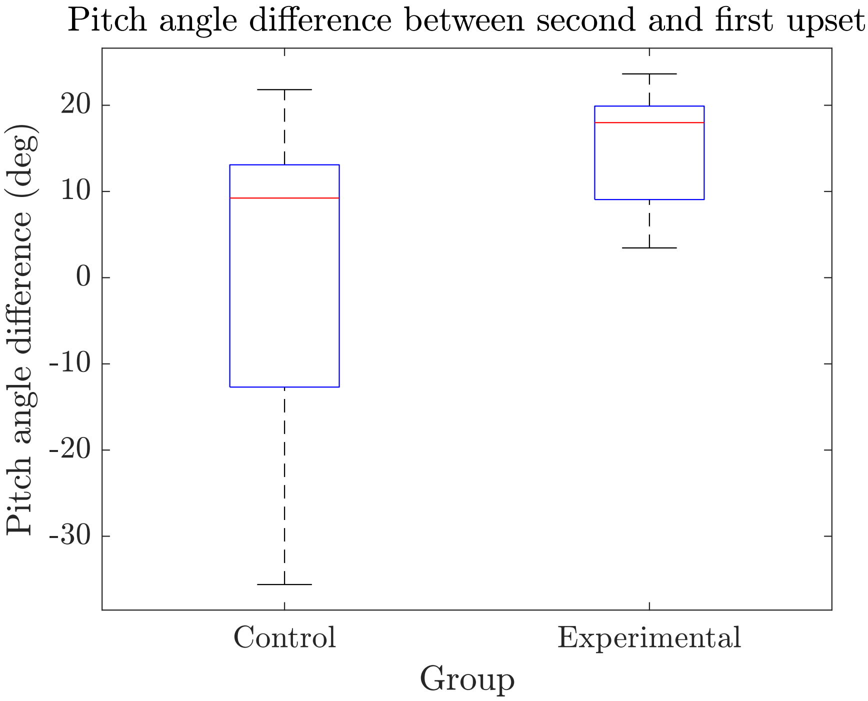


Figure G-28: Mass shift: minimum airspeed after first, mass-induced upset

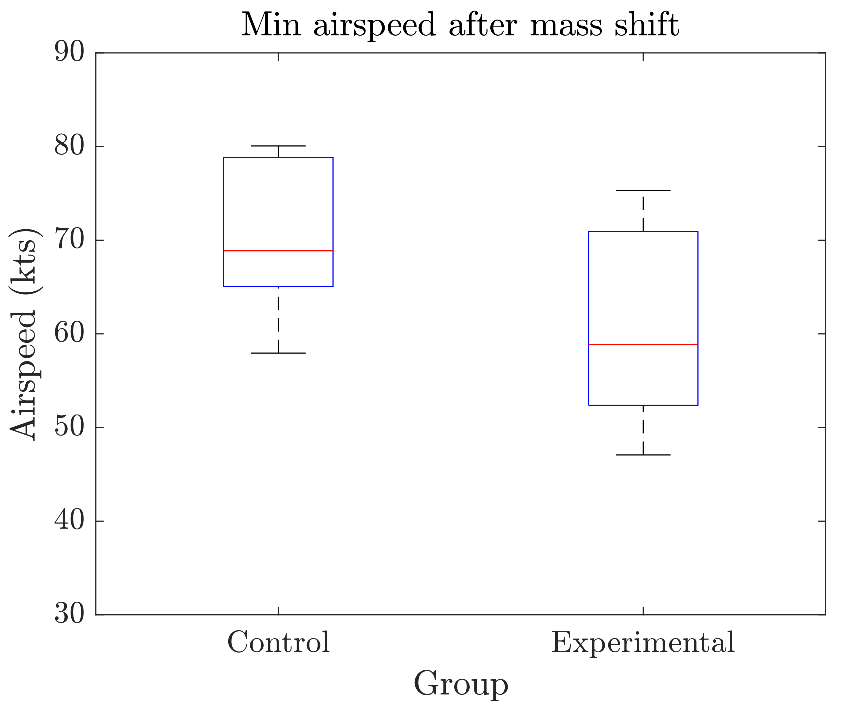


Figure G-29: Mass shift: minimum airspeed after second, flap-induced upset

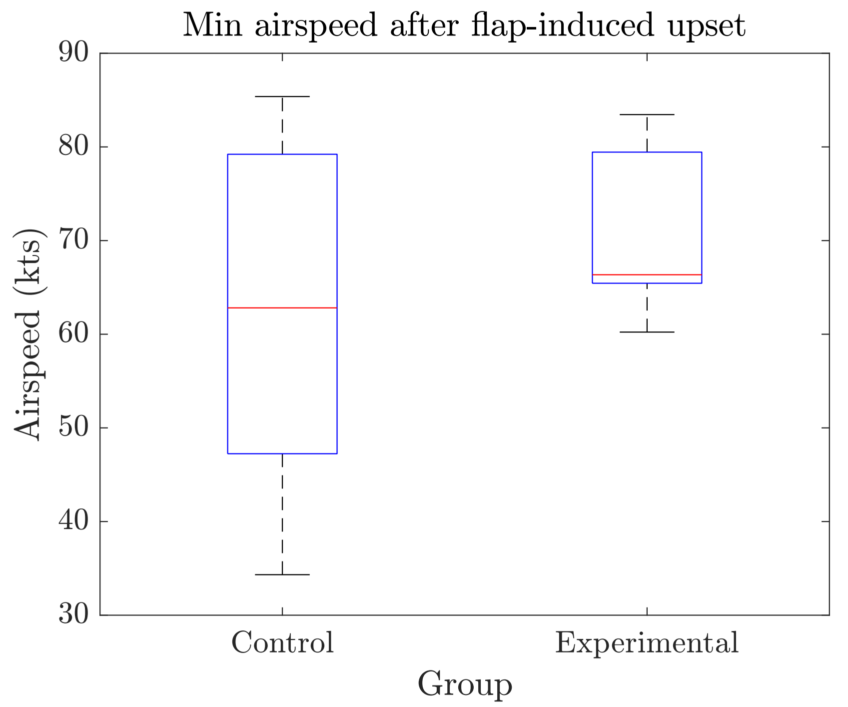
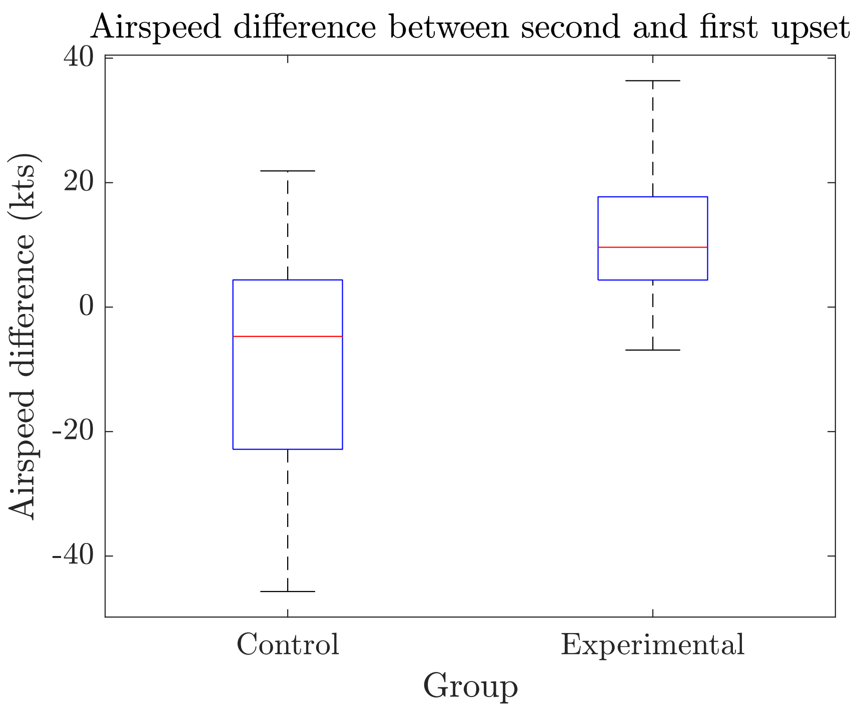


Figure G-30: Mass shift: airspeed difference between first and second upset. Positive values indicate performance improvement



Appendix H

MATLAB Scripts

This appendix contains the (most important) MATLAB scripts written to analyze the SIMONA output data, with one script per scenario. These scripts show the chosen variables and limits used in the performance analysis. All other MATLAB scripts (including the ones used for creating the performance plots) are available upon request.

```
% Written by SvM as part of the SenecaStartle project
% Analysis of scenario 423 - pretest
clear all;
clc;

%% Initialize variables
base_dir = 'C:\Users\sophi\Documents\SenecaStartle\Data\';
base_file = '_F423_Pretest.mat';
tt_under_vmc = [];
loss_of_control_vs_roll = [];
loss_of_control_maxail = [];
crash = [];
flap_start = [];
flap_final = [];
gear_up = [];
torque_ratio = [];
vs_max = [];

%% Analyze log file of each participant

for i=1:24
    % Go to folder of participant
    search_dir = [base_dir,num2str(i)];
    cd (search_dir);

    % Load participant log
    filename = [num2str(i),base_file];
    load (filename);

    %% Check time under minimum control speed Vmc

    % Count time under Vmc, above 15 ft altitude and save data
    t_under_vmc = length(find(vtas<80 & he>15)) / frequency;
    tt_under_vmc = [tt_under_vmc; t_under_vmc];

    %% Check throttle when flying under Vmc

    % Extract throttle settings if under Vmc
    torq_R_filtered = torq_R((find(vtas<80 & he>15)));
    % Extract airspeed if under Vmc
    vtas_filtered = vtas((find(vtas<80 & he>15)));
    % Calculate throttle ratio
    torq_ratio = (torq_R_filtered*1000) ./ vtas_filtered;
    if isempty(torq_ratio);
        % If not under Vmc, set ratio to zero
        torq_ratio = 0;
    end
```

```
    % Save data
    torque_ratio = [torque_ratio; max(torq_ratio)];

    %% Check for LoC by checking vertical speed and roll angle

    % Check for vertical speed < -20 ft/s
    loc_vs = he_dot(he_dot<-20);
    % Check for roll angle > 20 degrees
    loc_roll = phi(phi<-20 | phi>20);
    % If vs and roll conditions satisfied, count as LoC
    loc_roll_vs = ~isempty(loc_vs) & ~isempty(loc_roll);
    % Save data
    loss_of_control_vs_roll = [loss_of_control_vs_roll; loc_roll_vs
       ];

    %% Check for LoC by checking roll acceleration with max aileron

    % Find points where aileron deflection is max
    loc_ail = find(d_ail<-45 | d_ail>45);
    % Find roll acceleration values for max aileron
    loc_pdot = p_dot(loc_ail);
    % If signs of d_ail and p_dot are opposite, count as LoC
    loc_maxail = ~isequal(sign(d_ail(loc_ail)), sign(loc_pdot));
    % Create matrix containing LoC data for all participants
    loss_of_control_maxail = [loss_of_control_maxail; loc_maxail];

    %% Check if recovered or not

    if he_dot(length(he_dot)) < -10
        crash_vs = 1;
        crash = [crash; crash_vs];
    else
        crash_vs = 0;
        crash = [crash; crash_vs];
    end

    %% Check flap setting after 16 seconds (time of engine failure)

    flap_start = [flap_start; flap(frequency*16)];

    %% Check final flap, gear up, max. vertical speed and save

    flap_final = [flap_final; flap(length(t)-5)];
    gear_up = [gear_up; ismember(0,gear)];
    vs_max = [vs_max; min(he_dot)];

end
```

```
% Save .mat output file in main directory
cd (base_dir);
save analysis_423_pretest.mat tt_under_vmc loss_of_control_vs_roll
   loss_of_control_maxail crash flap_start flap_final gear_up
   torque_ratio vs_max
```

```
% Written by SvM as part of the SenecaStartle project
% Analysis of scenario 701 - flap asymmetry
clear all;
clc;

%% Initialize variables
base_dir = 'C:\Users\sophi\Documents\SenecaStartle\Data\';
base_file = '_E701_Flaps.mat';
max_phi_f25 = [];
max_phi_f40 = [];
diff_phi = [];
flap_highest = [];
flap_final = [];

%% Analyze log file of each participant

for i=1:24;
    % Go to folder of participant
    search_dir = [base_dir,num2str(i)];
    cd (search_dir);

    % Load participant log
    filename = [num2str(i),base_file];
    load (filename);

    %% Check maximum roll angle after selecting flaps 25

    % Find time stamps of flaps > 10 deg (start of asymmetry)
    flaps_25 = find(flap>10);
    % Find instances of flaps 25 at least one minute into the flight
    flaps_25 = flaps_25(flaps_25>(60*frequency));
    % Extract max roll angle first 20 seconds after selecting flaps 25
    extract_roll_25 = min(phi(flaps_25(1):(flaps_25+frequency*20)));
    % Compensate for selecting flaps in a turn
    max_phi_f25_comp = extract_roll_25 - phi(flaps_25(1));
    % Check maximum roll angle after selecting flaps 25 and save
    max_phi_f25 = [max_phi_f25; max_phi_f25_comp];

    %% Check for flaps 40 and save

    flap_highest = [flap_highest; max(flap)];

    %% If flaps 40, check maximum roll angle for flaps 40
```

```
    % Find first time stamp of flaps > 25 deg (selecting flaps 40)
    flaps_40 = find(flap>25,1);
    if isempty(flaps_40)
        % Set to NaN if no flaps 40
        max_phi_f40_comp = NaN;
    else
        % Compensate for selecting flaps in a turn
        max_phi_f40_comp = min(phi(flaps_40:end)) - phi(flaps_40(1)
            );
    end
    % Check maximum roll angle after selecting flaps 40 and save
    max_phi_f40 = [max_phi_f40; max_phi_f40_comp];

    %% Check final flap setting and save

    flap_final = [flap_final; flap(length(t)-5)];

    %% Calculate roll angle differences between second and first
       flap upset

    if max_phi_f40 ~= NaN
        diff_phi = [diff_phi; (max_phi_f40_comp - max_phi_f25_comp)
            ];
    else
        diff_phi = [diff_phi; NaN];
    end

end

% Save .mat output file in main directory
cd (base_dir);
save analysis_701_flaps.mat max_phi_f25 max_phi_f40 diff_phi
   flap_highest flap_final
```

```
% Written by SvM as part of the SenecaStartle project
% Analysis of scenario 702 - false stall
clear all;
clc;

%% Initialize variables
base_dir = 'C:\Users\sophi\Documents\SenecaStartle\Data\';
base_file = '_E702_Stall.mat';
unload_de = [];
unload_q = [];
unload_vs = [];
unload_n = [];
unload_theta = [];
h_max = [];
h_min = [];

%% Analyze log file of each participant

for i=1:24;
    % Go to folder of participant
    search_dir = [base_dir,num2str(i)];
    cd (search_dir);

    % Load participant log
    filename = [num2str(i),base_file];
    load (filename);

    %% Check for unload (stall recovery)

    % Find time stamp of stick shaker activation at 1,500 ft
    stick_shaker_time = find(he>1499 & he<1501,1);
    % Extract first ten seconds after stick shaker
    ten_sec =  stick_shaker_time:(stick_shaker_time+(10*frequency))
       ;

    % Elevator deflection
    unload_de = [unload_de; max(de(ten_sec))];
    % Pitch rate
    unload_q = [unload_q; min(qc(ten_sec))];
    % Vertical speed
    unload_vs = [unload_vs; min(he_dot(ten_sec))];
    % Load factor
    unload_n = [unload_n; min(n(ten_sec))];
    % Pitch angle
    unload_theta = [unload_theta; min(theta(ten_sec))];

    %% Check max altitude
```

```
    % Max altitude
    h_max = [h_max; max(he(stick_shaker_time:end))];
    % Min altitude after stick shaker
    h_min = [h_min; min(he(stick_shaker_time:end))];

end

% Save .mat output file in main directory
cd (base_dir);
save analysis_702_stall.mat unload_de unload_q unload_vs unload_n
   unload_theta h_max h_min
```

```
% Written by SvM as part of the SenecaStartle project
% Analysis of scenario 705 - mass shift
clear all;
clc;

%% Initialize variables
base_dir = 'C:\Users\sophi\Documents\SenecaStartle\Data\';
max_elevator = [];
max_theta = [];
min_vtas = [];
power_reduction = [];
max_elevator_2 = [];
max_theta_2 = [];
min_vtas_2 = [];
selected_f25 = [];
selected_f40 = [];
early_config = [];
land_flap = [];
crash = [];
diff_theta = [];
diff_vtas = [];

%% Analyze log file of each participant

for i=1:24
    % Go to folder of participant
    search_dir = [base_dir,num2str(i)];
    cd (search_dir);

    % Load participant log
    base_file = '_E705_Mass.mat';
    filename = [num2str(i),base_file];

    % If SIMONA stopped during run, check retry
    if exist(filename, 'file')
        load (filename);
    else
        base_file = '_E705_Mass2.mat';
        filename = [num2str(i),base_file];
        load (filename);
    end

    %% Check max/min values for some parameters during first mass
       shift upset;

    % Max elevator deflection in first 50 seconds
    max_elevator = [max_elevator; max(de(1:(50*frequency)))];
```

```
% Max pitch angle in first 75 seconds
max_theta = [max_theta; max(theta(1:(75*frequency)))];
% Min airspeed in first 75 seconds (excluding takeoff)
rotate = find(vtas>80);
min_vtas = [min_vtas; min(vtas(rotate(1):(75*frequency)))];

%% Check for throttle reduction as upset recovery procedure

power_r = find((Pla_left < 0.7) & (Pla_right < 0.7));
interval = (20*frequency):(21*frequency);
if ismember(1,(ismember(interval,power_r)))
    power_red = 0;
else
    power_red = 1;
end
power_reduction = [power_reduction; power_red];

%% Check for crash and/or landing

if he_dot(length(he_dot)) < -10
    crash_vs = 1;
    crash = [crash; crash_vs];
else
    crash_vs = 0;
    crash = [crash; crash_vs];
end

% Check for height < 10 ft
landing = find(he<10 & t>100);

%% Check flap selection

% Check for any flap selection
flap_25 = find(flap>1);
if isempty(flap_25)
    selected_flaps_25 = 0;
    early_cfg = 0;
% Exclude early configuration tryouts
elseif he(flap_25(1)) > 1000 || t(flap_25(1)) < 100
    selected_flaps_25 = 0;
    early_cfg = 1;
else
    selected_flaps_25 = 1;
    early_cfg = 0;
end
selected_f25 = [selected_f25; selected_flaps_25];
early_config = [early_config; early_cfg];
```

```
% Check for flaps 40
flap_40 = find(flap>26);
if isempty(flap_40)
    selected_flaps_40 = 0;
% Exclude early configuration tryouts
elseif he(flap_40(1)) > 900 || t(flap_40(1)) < 100
    selected_flaps_40 = 0;
else
    selected_flaps_40 = 1;
end
selected_f40 = [selected_f40; selected_flaps_40];

% Check flap setting used during landing
land_flap = [land_flap; flap(length(flap))];

%% Check for response to flap selection

if selected_flaps_25 == 1 && crash_vs == 0
    % Max elevator deflection in first 30 seconds after flap
        selection
    max_elevator_2 = [max_elevator_2; max(de(flap_25(1):(
        flap_25+30*frequency)))];
    % Max pitch angle between flap 25 selection and landing
    max_theta_2 = [max_theta_2; max(theta(flap_25(1):landing(1)
        ))];
    % Min airspeed between flap 25 selection and landing
    min_vtas_2 = [min_vtas_2; min(vtas(flap_25(1):landing(1)))
        ];
elseif crash_vs == 1
    % Max elevator deflection until crash
    max_elevator_2 = [max_elevator_2; max(de(flap_25(1):end))];
    % Max pitch angle between flap 25 selection and crash
    max_theta_2 = [max_theta_2; max(theta(flap_25(1):end))];
    % Min airspeed between flap 25 selection and crash
    min_vtas_2 = [min_vtas_2; min(vtas(flap_25(1):end))];
else
    % In case flaps 25 not selected
    max_elevator_2 = [max_elevator_2; NaN];
    max_theta_2 = [max_theta_2; NaN];
    min_vtas_2 = [min_vtas_2; NaN];
end

% Calculate max pitch angle difference between second and first
    upset
diff_theta = max_theta - max_theta_2;
```

```
    % Calculate min airspeed difference between second and first
       upset
    diff_vtas = min_vtas_2 - min_vtas;

end

% Save .mat output file in main directory
cd (base_dir);
save analysis_705_mass.mat max_elevator max_theta min_vtas
   power_reduction max_elevator_2 max_theta_2 min_vtas_2
   selected_f25 selected_f40 early_config land_flap crash
   diff_theta diff_vtas
```

```
% Written by SvM as part of the SenecaStartle project
% Analysis of scenario 501 - unreliable airspeed
clear all;
clc;

%% Initialize variables
base_dir = 'C:\Users\sophi\Documents\SenecaStartle\Data\';
base_file = '_E501_ASI.mat';
vtas_max = [];
asi_max = [];
time_above_50kts_ias = [];
time_to_60 = [];
time_to_pitch_power = [];
ias_pp = [];
vtas_pp = [];
dive_time = [];

%% Analyze log file of each participant

for i=1:24
    % Go to folder of participant
    search_dir = [base_dir,num2str(i)];
    cd (search_dir);

    % Load participant log
    filename = [num2str(i),base_file];
    load (filename);

    %% Check max airspeed and save

    vtas_max = [vtas_max; max(vtas)];

    %% Calculate indicated airspeed and save max value

    ias = vtas + offset_v_value;
    asi_max = [asi_max; max(ias)];

    %% Calculate time above 50 kts IAS and save

    t_above_50 = length(find(ias>50)) / frequency;
    time_above_50kts_ias = [time_above_50kts_ias; t_above_50];

    %% Check time from rotation until below 60 kts and save

    % Find takeoff time and save
    t_to = find(he>5);
    [t_mias_val, t_mias_index] = max(ias);
```

```
    % Check when below 60 kts and calculate time
    t_u60 = find(ias<60);
    t_u60 = t_u60(t_u60>t_mias_index);
    tt_b60 = (t_u60(1)-t_to(1)) / frequency;
    % Save data
    time_to_60 = [time_to_60; tt_b60];

    %% Check when reverting to pitch-power settings

    % Reduce thrust to <50 Nm, pitch positive and above 5 ft
    set_pp = find(torq_L<0.050 & torq_R<0.050 & theta>0 & he>5);
    % Takeoff
    takeoff = find(he>5);
    % Time to throttle pullback
    t2pp = (set_pp(1)-takeoff(1)) / frequency;
    % Save time to pitch power
    time_to_pitch_power = [time_to_pitch_power; t2pp];

    % IAS when reverting to pitch-power
    ias_pp = [ias_pp; ias(set_pp(1))];
    % True airspeed when reverting to pitch-power
    vtas_pp = [vtas_pp; vtas(set_pp(1))];

    %% Check 'dive time' before selecting pitch-power

    % Extract vertical speed up to selecting pitch-power and save
    extract_vs = he_dot(1:set_pp(1));
    dive_time = [dive_time; (length(extract_vs(extract_vs<0)) /
       frequency)];

end

% Save .mat output file in main directory
cd (base_dir);
save analysis_501_ASI.mat vtas_max asi_max time_above_50kts_ias
   time_to_60 time_to_pitch_power ias_pp vtas_pp dive_time
```

Appendix I

Digital Data

All data gathered during this experiment is available in digital form upon request and was handed in after graduation. This includes raw simulator data, audio logs and transcripts, MATLAB scripts, plots, SPSS files, Powerpoint presentations, briefings, questionnaires and DUECA files.

The DUECA project is available in the C&S repository as SenecaStartle and contains README files on how to use the project.