Facutly of Aerospace Engineering

Spatial Disorientation in a Hexapod Simulator

Evaluating the Effect of Expectation and Display Perception on Control Reversals for Experienced and Novice Pilots

Annemarie van den Hoed

February, 2020

Spatial Disorientation in a Hexapod Simulator

Evaluating the Effect of Expectation and Display Perception
on Control Reversals for Experienced and Novice Pilots

Master of Science Thesis

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

A. van den Hoed

February, 2020

Faculty of Aerospace Engineering · Delft University of Technology

Delft University of Technology
Control and Simulation Section

The undersigned hereby certify that they have read and recommend to the Faculty of Aerospace Engineering for acceptance a thesis entitled "**Spatial Disorientation in a Hexapod Simulator**" by **A. van den Hoed** in partial fulfillment of the requirements for the degree of **Master of Science**.

Dated: February, 2020

Readers:

dr. H.M. Landman

prof.dr. E.L. Groen

prof.dr.ir. M. Mulder

dr. M.A. Mitici

List of Figures

List of Tables

Contents

I

Paper

Spatial Disorientation Flight in a Hexapod Simulator Causes Pilots to Make Roll Reversal Errors

A. van den Hoed *
TU Delft, Faculty of Aerospace Engineering, the Netherlands

Supervised by: H.M. Landman, † O. Stroosma, ‡ M.M. van Paassen, § E.L. Groen, ¶ and M. Mulder ‖
TU Delft, Faculty of Aerospace Engineering, the Netherlands
and TNO, location Soesterberg, the Netherlands

In previous studies, pilots made roll reversal errors (RREs) when responding to a 'moving-horizon' type attitude indicator (AI). It was argued that it was the ambiguity of this display leading to RREs. Later, using non-pilots, it was found that RREs were in many cases caused by expectation-induced misperception of the AI bank angle, which can arise from a pilot experiencing spatial disorientation. The current study evaluated the role of expectation in the control strategy of professional pilots using a hexapod simulator.

First, an effective man-in-the-loop spatial disorientation scenario was developed, based on the 'leans' illusion. Here, the expectation of ten experienced and eight novice pilots was manipulated with the simulator motion without the AI, after which they had to turn the aircraft back level when the AI was shown again. An error was made in 16.7% of the runs where the bank angle on the AI was opposite to the earlier given physical motion cues, which is 2.9 times more compared to a baseline condition in which no motion, and thus no manipulated expectation, was present. In a third condition where a motion-induced expectation of the bank angle was present, but a level AI was shown later, no errors were made. It was also found that experienced pilots made slightly more errors than novice pilots and that their reaction time was on average half a second higher.

It was concluded that an initially induced expectation about the bank angle does not directly influence the control strategy of pilots, but that it does make them more vulnerable to misinterpretations of the AI, leading to RREs being made. This should be taken into account when developing new displays and technologies, as it should at all times guide pilots in their decision-making, minimizing the chance of misinterpretations.

List of Abbreviations

AI	Attitude Indicator
DLP	Digital Light Processing
FOV	Field of View
$LOC-I$	Loss-of-Control In-flight
$MATB$	Multi Attribute Test Battery
$NASA$	National Aeronautics and Space Administration
RRE	Roll Reversal Error
SRS	Simona Research Simulator
TLX	Task Load Index

---

[*] MSc. Student, Control and Simulation Section, Faculty of Aerospace Engineering, TU Delft, P.O. Box 5058, 2600 GB Delft, the Netherlands; a.vandenhoed@student.tudelft.nl.

[†] Scientist, Human Interface Department, TNO, P.O. Box 23, 3769 ZG Soesterberg, the Netherlands; h.m.landman@tudelft.nl.

[‡] Researcher Flight Simulator, Control and Simulation Section, Faculty of Aerospace Engineering, TU Delft, P.O. Box 5058, 2600 GB Delft, the Netherlands; o.stroosma@tudelft.nl.

[§] Associate Professor, Control and Simulation Section, Faculty of Aerospace Engineering, TU Delft, P.O. Box 5058, 2600 GB Delft, the Netherlands; m.m.vanpaassen@tudelft.nl.

[¶] Senior Scientist, Human Interface Department, TNO, P.O. Box 23, 3769 ZG Soesterberg, the Netherlands; eric.groen@tno.nl.

[‖] Professor, Control and Simulation Section, Faculty of Aerospace Engineering, TU Delft, P.O. Box 5058, 2600 GB Delft, the Netherlands; m.mulder@tudelft.nl.

I. Introduction

Aviation has advanced rapidly in the use of automation over the past decades. In modern aircraft, pilots are proactive decision-makers, their main task being to monitor that the automatic systems are operating correctly [1]. This has led to people believing that automation has fully replaced the pilot, but in fact it has mainly been used to enhance human capabilities [2]. In this man-machine system, pilots are still expected to take over whenever automation fails, which reflects the irony that the more advanced a man-machine system is, the more crucial the contribution of this human operator becomes [3]. Recent airplane accidents involving erroneous inputs from pilots have raised questions concerning the current form of the collaboration between pilot and (automated) vehicle [4].

For example, in Kenya flight KQA507 on May 4, 2007, the problem started with confusion in using the autopilot. During this night flight, the airplane showed to have a slight tendency to roll to the right, which was manually counteracted by the captain. However, due to high workload and improper instrument scanning, an increasing roll to the right was left unnoticed by the crew. Only when the bank angle aural warning sounded the pilots were informed about the potential dangers due to this increasing bank angle. The captain's intuitive reaction was to make an input to the right based on the assumption that he had over-corrected the autopilot before. At this moment, the pilot expressed surprise when referring to the attitude indicator (AI) since it now showed he was making the bank angle even steeper rather than recovering from it. Multiple corrective roll inputs to the left and right were done, but they were all in vain. The airplane crashed after loss of control, leaving no survivors [5].

In the Kenya flight accident, the pilot exerted a roll input opposite to the required direction, which is referred to as a roll reversal error (RRE). It has been argued that these errors occur due to a misinterpretation of the attitude indicator, originated in the ambiguity of the conventional 'inside-out' or 'moving-horizon' display. Here, the airplane symbol remains fixed on the display and the horizon rotates into the opposite direction of the control inputs [6]. Considering the design principle of the moving part, when controlling the moving part of the display, people expect it to move in the same direction as their control input [7]. Consequently, this ambiguity might cause the operator to control the horizon symbol as if it was the aircraft symbol, which is also known as a horizon control reversal (HCR) and will lead to a RRE [7]. In in-flight experiments, where subjects had to recover from steep turn upsets using the AI, non-pilots made 21.9% RREs, compared to around 4.9% of errors made by novice pilots and 1.5-4.7% by experienced pilots [8–10]. Non-pilots showed a RRE incidence of 19.3% in fixed-based simulators and 15-20% when using a motion-based simulator [11, 12]. Also here, professional pilots performed better, with a 4.5-8.7% error rate [10, 11, 13].

In all experiments mentioned above, the goal was to evaluate the 'inside-out' display and investigate whether it was the ambiguity leading to the errors made. For these studies, pilots' expectations about the bank angle were either minimized or balanced between the conditions. A few recent studies have tested the effect of expectations on errors. Landman et al.(2019)[14] showed in a fixed-based simulator experiment that RREs were actually in many cases caused by an expectation-induced misperception of the AI bank angle. This expectation was manipulated with a flying task starting with only the outside vision enabled. In a turn, this outside vision disappeared and shortly after a moving-horizon type AI was shown. Participants had to roll the wings level using the indication on the AI. In cases where an incongruent AI was shown, there was a 75% RRE incidence, compared to 9.8% in the baseline conditions, where the expectation matched the information given on the attitude indicator [14]. This indicated that the expectations induced by the outside visuals highly affected the inputs made and often outweighed the information shown on the instrument.

In-flight, an incorrect expectation about the aircraft attitude most often has loss of attitude awareness, or more commonly referred to as spatial disorientation, as its cause. It refers to the inability of a pilot to correctly interpret the aircraft attitude relative to the earth, or other points of reference, and is according to accident reports one of the major contributing factors in eventual occurrences of loss-of-control in-flight (LOC-I) [15]. Results from an in-flight experiment with non-pilots showed an increase from 5% RREs in the baseline condition to 63% in conditions where the expectation about the aircraft attitude, manipulated through applying vestibular cues when being blindfolded, mismatched the AI indication shown when the blindfold could be removed [6].

An overview of the above-mentioned researches into RREs is shown in Table 1. The two studies looking into misleading cues and expectations showed a significant effect of expectation on the occurrence of RREs. However, both

Table 1 **Overview of RRE incidence in previous research.**

Objective	Participants	Apparatus	RREs [%]	Comments	Research
Ambiguity of AI	Non-pilots	In-flight	21.9		Roscoe et al. (1975)
		Fixed-based simulator	19.3		Müller et al. (2018)
		Motion-based simulator	15-20		Ince et al. (1975)
	Pilots	In-flight	4.7	Experienced	Beringer et al. (1975)
			1.5	Experienced	Hasbrook et al. (1973)
			4.9	Novice	
		Fixed-based simulator	5.1		Beringer et al. (1975)
			4.5-8.7		Müller et al. (2018)
			8.3		Singer et al. (2008)
Misleading cues/ Expectation	Non-pilots	In-flight	75	Expectation induced misinterpretation	Landman et al. (2019)
			9.8	Baseline (solely misinterpretation)	
		Fixed-based simulator	63	Expectation induced misinterpretation	Landman et al. (2020)
			5	Baseline (solely misinterpretation)	

have been performed with non-pilots. It can be argued that professional pilots receive appropriate training in both the use of the attitude indicator as in how they should cope with spatial disorientation, not to make the same errors as non-pilots did. A recent study by Lewkowicz et al.(2020)[16] using spatial disorientation scenarios in a motion-based simulator found that there is no significant difference in coping with spatial disorientation flight between pilots and non-pilots. However, from this study no effects were reported about the influence of expectation on the control strategy.

Consequently, the main objective for this study is to investigate the role of expectation on the occurrence of RREs for pilots specifically. Based on flight hours as the most efficient criteria to compare expertise, two groups of professional pilots are distinguished for the experiment. Their expectation is manipulated by simulating spatial disorientation flight in an experiment using a hexapod simulator. Although promising results in the use of a hexapod simulator to simulate spatial disorientation flight were reported by Lewkowicz et al.(2020)[16] and Klyde et al.(2016)[17], this type of simulator has not often been used in trying to simulate spatial disorientation flight, especially not by shaping a pilots' expectation through the given motion cues. Therefore, a second objective of this study became to develop a man-in-the-loop spatial disorientation scenario for a hexapod simulator, allowing for the evaluation of the effect of expectation on pilot performance.

II. Background Information

A. Pilot Perception

To better understand why roll reversal errors occur, the 'conceptual model of pilot perception' can be used, which is shown in Figure 1. This is a simplified version of the 'conceptual model of startle and surprise' presented in a study by Landman et al. (2017)[18]. The full model with an explanation on how it has been modified for this study can be found in Appendix A.

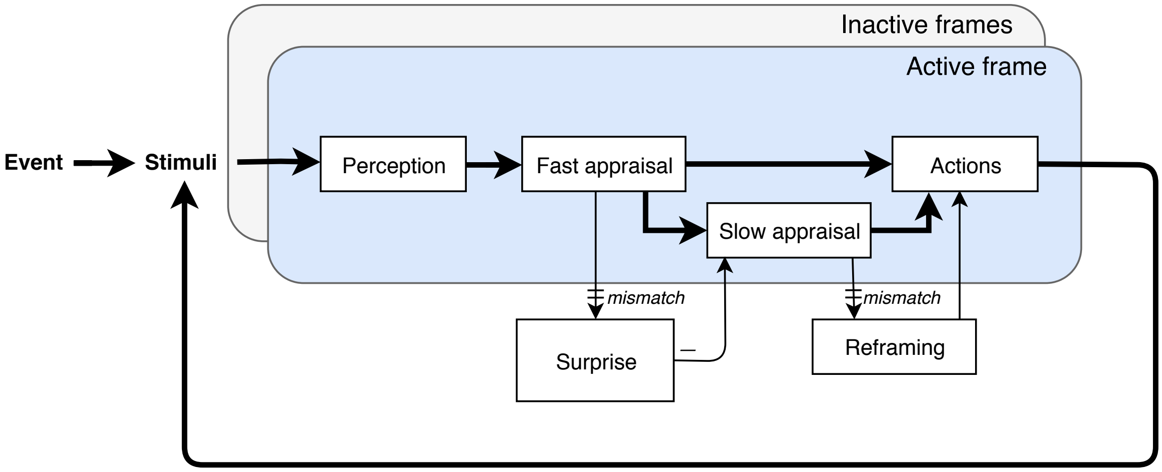


Fig. 1 Conceptual model of pilot perception [18].

The bold lines in the model show the perceptual cycle model from Neisser (1976)[19], where a person perceives stimuli, interprets those and assesses the information. Based on the data, an action or control strategy is chosen, which again leads to new data or stimuli. There are two types of cognitive systems to organize the judgment of stimuli [20]. In this model, these two types of appraisal are categorized as being fast or slow; fast meaning that it is done based on intuition and involves a highly automatic reaction, whereas slow appraisal indicates reasoning which evolves into an exploring and knowledge-based processing which takes more time and effort [21].

Furthermore, the model is based on the notion that perception of data and actions that follow are conceptualized as being guided by 'frames', which are mental knowledge structures that were previously learned. More precise, frames denote an explanatory structure or concept that defines entities by describing their relationship to other entities and are created based on previous experiences and knowledge [22]. Pilots have a number of frames based on their earlier understanding of a situation or concept which are stored in memory [18]. The data, or in the model 'stimuli', identify the relevant frame and vice versa the frame determines what data will be noticed. Consequently, as said by Beach (1997)[23], "*the frame may be in error, but until feedback or some other form of information makes the error evident, the frame is the foundation for understanding the situation and for deciding what to do about it*" (p.24). Someone might experience confusion or surprise whenever an inadequacy or mismatch between the existing frame and the perception of relevant data is found [22].

Expectations are known to guide people to a certain frame and therefore affect the way information is interpreted [24]. As an example, as found by Delk and Fillenbaum (1965)[25], "*an object is incorrectly judged as being more deeply red if it has the shape of a heart than if it has the shape of a square*". In this example, Bayesian decision theory shows that people would use prior knowledge or beliefs to interpret the information given [26]. A mismatch between what is expected and what has been observed triggers a reframing process in which a person 'fills the gap' and tries to structure perceived data to come up with a plausible understanding of the situation [24].

B. Spatial Disorientation

Related to aviation, a pilots' expectation can be shaped by experiencing spatial disorientation, which refers to the inability to correctly interpret the aircraft attitude to the earth or other points of reference [15]. In other words, pilots are put in a certain frame regarding the aircraft attitude based on visual or vestibular cues, which might mismatch with the real aircraft attitude.

Bank illusions are the most prevalent form of spatial disorientation and can lead to roll reversal errors being made [27]. A schematic illustration of one of these illusions, the 'leans', is shown in Figure 2. Like most spatial illusions, it happens in the absence of outside visual cues or with false outside visual cues (such as a sloping terrain interpreted as level, or a sloping cloud field) in combination with insufficient instrument scanning. Subfigure 2.1 shows an airplane in level flight seen from behind with a pilot perceiving the correct attitude relative to the earth (indicated by the red dotted

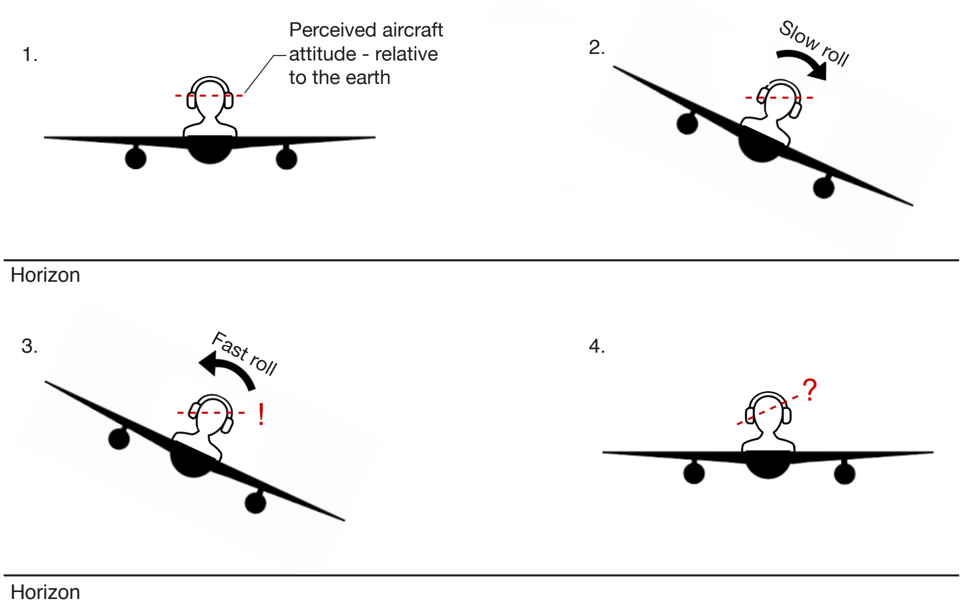
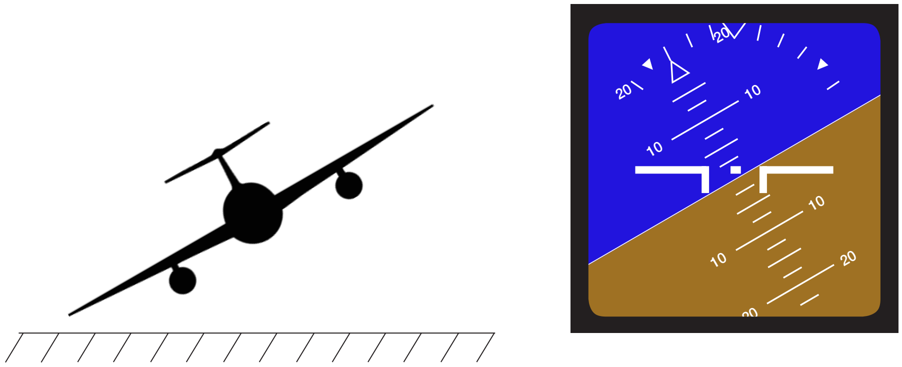


Fig. 2 Illustrating the leans illusion, in four phases.

line). In Subfigure 2.2, the airplane makes a slow turn to the right. The high-frequency dynamics of the semi-circular canal cause for low roll rates or prolonged rolling motions not to be detected, leaving the pilot to believe the aircraft is still flying straight [28]. Followed by a suprathreshold correction, as shown in Subfigure 2.3, the pilot will be left with the illusion of turning to the opposite direction as illustrated in Subfigure 2.4 [29]. This disorientation illusion is referred to as the ‘leans’ because the pilot feels compelled to lean using the perceived vertical plane after the abrupt correction as a reference. A control input done to counteract this turning sensation will not give the result the pilot expects, leading to an erroneous input or RRE [6, 13].

The wrong expectation of the aircraft attitude due to the leans illusion can increase the probability that the attitude indicator would be misinterpreted when being in a banked angle, as this expectation now coincides with the horizon indication as illustrated in Figure 3. The airplane on the left (seen from behind) indicates the aircraft attitude as perceived in Subfigure 2.4. When referring to the AI showing the actual aircraft attitude, the pilot might read it to believe that the information matches, leading to a horizon control reversal.



(a) Perceived attitude (left turn)

(b) Indicated attitude (right turn)

Fig. 3 Situation in which a horizon control reversal can be made. The perceived aircraft attitude (a) is seen from behind and is opposite to the real orientation shown (b).

The objective of the study presented in this paper is to evaluate pilots' responses and see whether the control strategy is guided by expectations. These expectations are manipulated by simulating motion-induced disorientation based on the 'leans' cues as they occur in real flight.

III. Method

A. Participants

A total of eighteen professional airline pilots participated in the experiment, which were all familiar with flying medium- to large-sized aircraft using a control column. The experiment was approved by the research ethics review board of the university and informed consent was obtained from each participant.

To evaluate the influence of experience on making a RRE, two groups were compared. The first group was labeled 'experienced pilots', as all had more than 10,000 flight hours, from which most in airplanes using a control column. The other group were the 'novice pilots', who all had 5,000 or less flight hours. More information about the pilots and their experience can be found in Table 2 and 3, and in Figure 4. A more detailed description is given in Appendix B.

Table 2 Participant information for the two groups: gender and rank.

	Experienced (No.)	Novice (No.)
Gender, female	0	1
Gender, male	10	7
Rank, captain	9	0
Rank, first officer	1	4
Rank, second officer	0	4

Table 3 Participant information for the two groups: flight hours, years employed and age.

	Experienced Mean (SD)	Novice Mean (SD)
Flight hours (hrs)	16,989 (3,213)	2,300 (1,930)
Years employed (yrs)	30.4 (5.6)	4.0 (3.7)
Age (yrs)	55.3 (6.4)	28.8 (7.1)

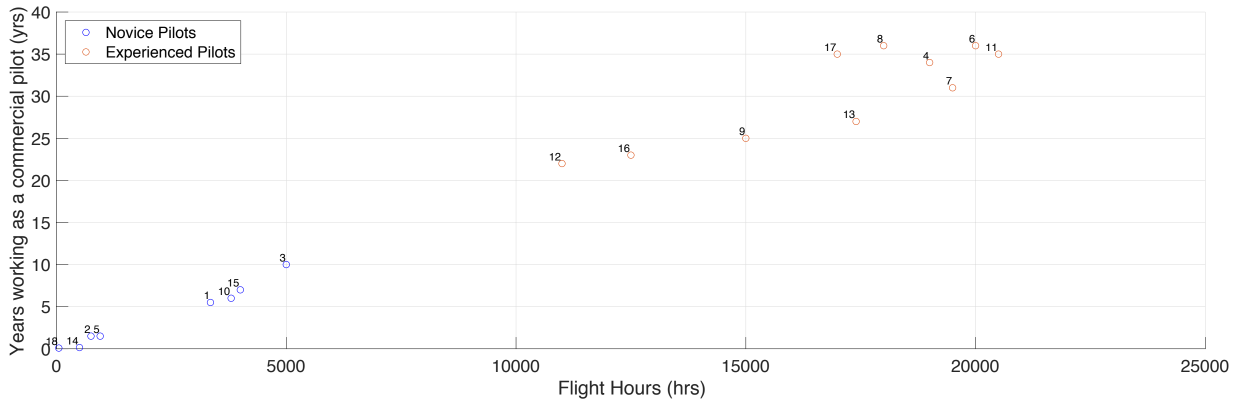


Fig. 4 Flight experience of the participants in years and flight hours.

B. Apparatus

The experiment was conducted in the Simona Research Simulator (SRS) at the Faculty of Aerospace Engineering of TU Delft. The SRS, shown in Figure 5, is a six-degrees-of-freedom full-motion simulator with a hydraulic hexapod motion system which can realize accelerations below human vestibular perception [30]. The pilot was seated in the left hand seat of the cockpit in front of a collimated 180 degrees horizontal by 40 degrees vertical field of view (FOV) screen. The images were rendered by FlightGear software and projected with the use of high-resolution computer generated images using three DLP projectors.

The aircraft flight dynamics model that was used in the simulation resembles that of an A320 since these flight dynamics correspond most with the airplane in which the participants have experience [31]. Participants were able to control only roll whenever an input was desired using a control-loaded hydraulic column. Throttle was controlled by the autopilot during the entire flight, which resulted in a speed around 230 knots and altitude of 10,000 feet. The only instrument provided was a simplified digital Primary Flight Display (PFD), showing the attitude indication, speed and altitude as can be seen in Figure 6.

Fig. 5 Simona Research Simulator.

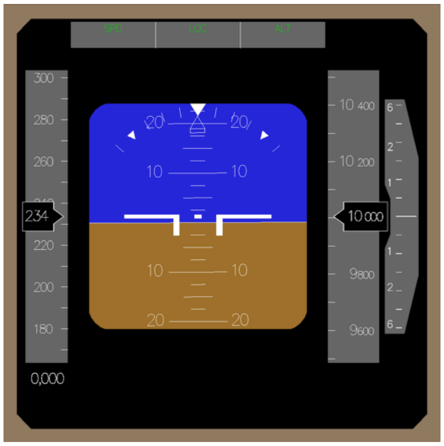


Fig. 6 The simplified PFD.

C. Tasks

During most of the experiment, the autopilot was controlling the aircraft. The primary task of the pilot was to monitor the aircraft attitude and make sure it remained in level flight. At times, the outside visuals and AI would disappear, and consequently the pilot could only use motion cues as a reference. Shortly after the autopilot off alarm sounded, the AI was shown again, from which the pilots had to turn the aircraft back level manually, if needed. To elicit an intuitive and fast response, they were told that the objective of this study was to look into the immediate reaction of pilots after a period of inattention or a period with high workload and therefore that their reaction time after the AI was shown again would be measured.

Also, during the entire experiment, a secondary task was performed. This was a Multi Attribute Test Battery (or MATB) task, which is a computer-based task designed to evaluate operator performance and workload. MATB provides a benchmark set of tasks comparable to activities that aircraft crew-members perform during a flight and requires the simultaneous performance of monitoring and dynamic resource management [32]. A full description of the task is given in Appendix C.

D. Experiment Runs

The experiment runs all consisted of two phases. ‘Phase 1’, the ‘motion setup’, entailed the motion cues given to the pilots, manipulating the expectation about the aircraft attitude. ‘Phase 2’, which immediately follow Phase 1, was the ‘input phase’, in which the autopilot was switched off and the pilot had to respond to the AI shown to turn the aircraft back level. Both phases are explained in more detail below.

Phase 1: The Motion Setup

Figures 7 and 8 illustrate the two motion scenarios used for the experiment. The left two columns show the simulator with outside visuals and the AI, respectively. The black aircraft symbol, in the third column, represents the aircraft model running in the simulation. The red aircraft symbol, in the fourth column, shows the intended expectation of the pilot due to the motion given. The indication on the AI followed the airplane model as they are linked throughout the entire simulation. The grey airplane symbol shows that here the attitude of the aircraft model could be adjusted without the pilot getting notice of the changed angle. This is further explained in the following paragraphs.

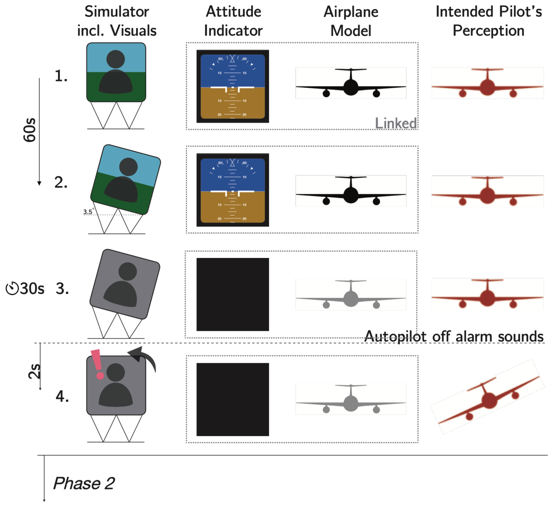


Fig. 7 **The 'motion' scenario used to manipulate the expectation.**

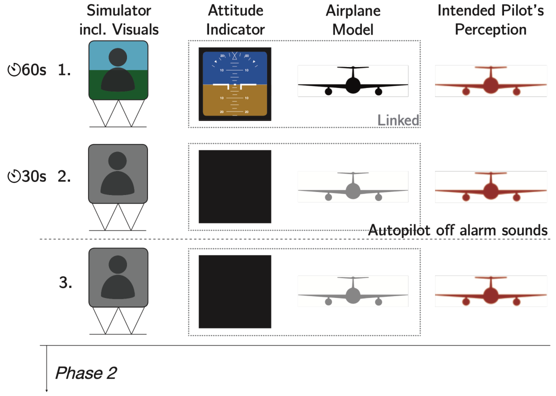


Fig. 8 **The 'no-motion' scenario used as a baseline condition.**

The 'Motion' Scenario

The 'motion' scenario was developed based on the 'leans' cues and had to elicit a motion-induced expectation regarding the bank angle. This scenario had to provide a fast cue to the participants, leading to the perception of being in a turn. Also, after this cue, the simulator was preferably upright, since the pilot had to give a response while still being in the 'frame' of perceiving a banked angle. Consequently, the challenge in creating this scenario originated from the fact that ground-based hexapod flight simulators cannot create the sensation of being in a coordinated turn with a tilted simulator cab, since this tilt can easily be noticed by the pilot. Therefore, for the scenario to have the desired effect, a combination had to be found between a masked tilt angle for pre-positioning and afterwards a fast cue, which would be sufficiently noticeable for making the participants perceive a banked angle. The scenario consisted of four steps, which are indicated on the vertical axis in Figure 7 and are elaborated upon hereafter.

1) The autopilot controlled the aircraft and the outside visuals showed a slightly cloudy day, but with a visible horizon. The pilot had to monitor the aircraft attitude while performing a secondary task on a touchscreen provided on the right hand side. This distraction could lead to a phenomenon called 'inattentional blindness', which is defined as "*the very strong tendency to fail to notice prominent environmental changes if they are not directly noticeable at the time that the change occurs*" (p.5) [33].

2) With the autopilot still flying and the pilot continuing the secondary task, the simulator cab was tilted to 3.5° (0.06 rad) in 60 seconds (either to the left or right). This led to a roll acceleration of $1.696{\cdot}10^{-5}$ rad/s^2 and angular velocity of 0.001 rad/s, both well under the human perception thresholds of 0.0349 rad/s^2 and 0.002 rad/s, respectively [30, 34].

   Meanwhile, the visuals and AI remained indicating level flight. Figure 9 illustrates the concept of 'visual dominance' which was used here. It shows that cues from the visual environment are used to determine what will be perceived despite the presence of (potentially strong) proprioceptive or vestibular cues [35]. Also, in our experiment, pilots were already leaning slightly to the right hand side in order to reach for the secondary task. Hence, pilots could interpret the information given by the proprioceptive system regarding the tilt of the simulator as being a result of the slightly odd position in which the secondary task had to be performed.

   To hide the sounds of the hydraulic pumps when setting up this bank angle, noise cancelling headphones were used through which a constant aircraft engine noise was played. Also, a continuous light turbulence was added only on the vertical z-axis of the simulator throughout the whole experiment. This helped in further masking the sounds and might also have had a small effect in masking the simulator tilt as the cab was already constantly moving.

3) The outside visual turned black, simulating loss of outside visibility. A few seconds later, the PFD was switched off to simulate operator's inattention, forcing him or her to rely only on the vestibular and proprioceptive system to sense the aircraft's attitude.

   These conditions were held for 30 seconds based on the results of a previous study. Here, it was observed that prolonged tilt attributed to adaptation of the proprioceptive system [36]. After maintaining a tilted position for more or less 30 seconds to one side, subjects experienced an overshoot to the other side when the chair was turned upright again. Figure 10 illustrates this phenomenon, where the grey simulation position is the initial one which was held for 30 seconds and the simulator is eventually turned quickly to the upright (black) position.

   In this 30 seconds, the aircraft model roll rate and acceleration were disconnected from the motion system. Consequently, since the pilot was also watching two black screens, the aircraft model could be turned to any desired bank angle without the pilot noticing it.

4) The autopilot disconnect alarm sounded, directing the pilot's attention to the motion of the simulator. One second later, the simulator was turned back to upright position in two seconds (corresponding with a roll rate of 1.75 deg/s or 0.03 rad/s). As the pilot believed the aircraft was still flying wings level before, this high roll rate would give the sensation of an initiated turn to the other side. One second later, Phase 2 started.

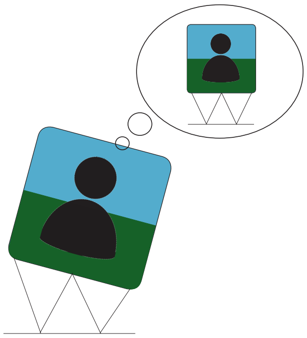


Fig. 9 The concept of visual dominance.

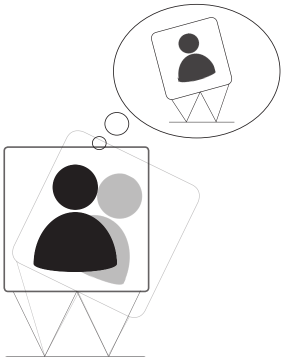


Fig. 10 The concept of adaptation.

The 'No-Motion' Scenario

The 'no-motion' scenario, shown in Figure 8, was used as a condition in which the pilots' expectation would *not* be influenced by the motion. The three steps, as shown on the vertical axis in the figure, are shortly elaborated on below.

1) The first step was the same as in the 'motion' scenario. Only here, these conditions were held for 60 seconds. Where in the 'motion' scenario that time was used to slowly tilt the simulator cab, now the simulator stayed in the upright position.

2) This step is similar to the third step from the 'motion' scenario, in that the outside visuals disappeared first, after which also the PFD turned black. This step lasted 30 seconds so that the two scenarios took the same amount of time.

3) After the autopilot off alarm sounded, also no motion cues were given, resulting in no motion-induced expectation for the pilot. One second later, Phase 2 started.

Phase 2: The Input Phase

Figure 11 shows the attitude indications given to the participants in Phase 2, resulting in five different conditions (which could be given both to the left and right). The left-most columns refer to the motion setup from Phase 1 with the intended pilot's perception about the bank angle shown. In the middle columns, the AI to which the participants had to respond in Phase 2 are shown. All turns showed a bank angle of 30° from which the pilot had to recover. This was chosen as such since a bank angle below 30° was not considered a severe situation. The right-most column in the figure shows the real-life in-flight situation that is simulated by the subsequent scenario.

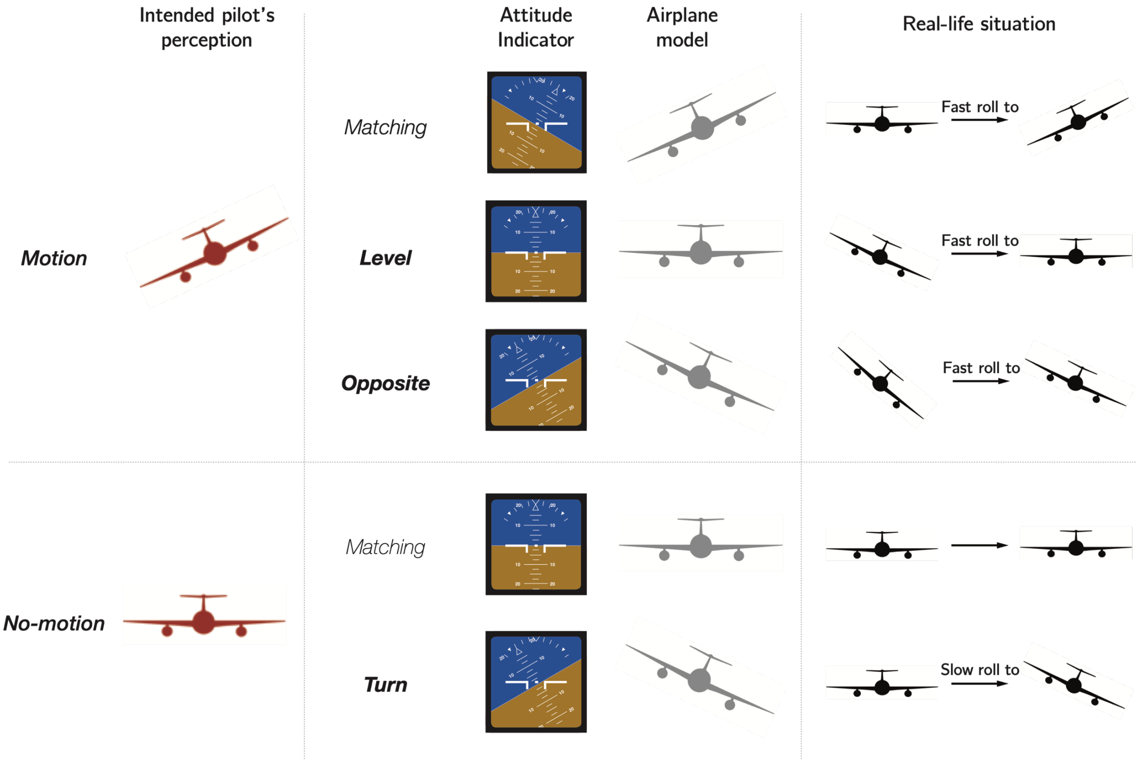


Fig. 11 The five conditions (test conditions shown in bold) with a schematic representation of the real-life situation it is simulating. The aircraft is seen from behind in all figures.

In the two matching conditions, *'motion-matching'* and *'no-motion-matching'*, the AI shown at the point where the participants needed to give an input matched the expectations created in the subject by the motion in Phase 1. Those conditions were used as 'filler runs' to remove any learning effect or anticipation for mismatching cues and for the participants to gain trust in the motion of the simulator.

The other three conditions were the test conditions used in the experiment. First, the *'no-motion-turn'* condition corresponded to a situation in which the aircraft had rolled with a subthreshold motion to the angle of 30°. Since no rolling motion cues were used, the participant's expectation about the attitude was consequently not affected. Therefore, this condition could be used as a baseline condition for evaluating the effects of a manipulated expectation. A RRE made would then point to a misinterpretation of the attitude indicator without the influence of expectation. The setup of this scenario was similar to previous studies evaluating the ambiguity of the AI in that no motion cues and a moving-horizon type AI were used [9–11, 13].

The *'motion-level'* condition mimicked a flight in which the aircraft had rolled to a certain bank angle with a subthreshold motion and then quickly rolled back to level flight. Here, the AI was not meant to trigger a misinterpretation of the instrument, meaning that any input given by the pilot could be an indication that he or she was being guided only by the motion cues, instead of the information on the visual display.

Similar to the ‘motion-level’ condition, the *‘motion-opposite’* condition also showed an incongruent AI indication when comparing it to the earlier given motion cues, except that for this condition the simulated situation was that of an airplane having slowly rolled to an excessive angle (> 30°) from which it then rolled quickly back to 30°. The AI then indicated that the airplane had not turned level yet and that an input was desired in the same direction as the preceding motion cue. A RRE being made in this condition could refer to an expectation-induced misinterpretation of the attitude indicator.

E. Experiment Design

The conditions were given in two sets of eighteen runs, given back to back, with a short break between the two sets. Every set consisted of six test runs (each test condition twice) and twelve filler runs (twice the ‘no-motion-matching’ condition and ten times the ‘motion-matching’ condition). All conditions including motion or a turn were given an equal number of times to the left and right. Each set started with six filler runs after which the filler and test conditions were balanced in the same order for every participant. A Latin Square Design was used in the experimental design to minimize the systematic correlation between any of the conditions.

The secondary task had two event sequences. All participants received both versions divided over the two sets of runs. The order of those two sequences changed for each participant to eliminate potential effects of one secondary task sequence being more demanding than the other. Also, a free practice flight of approximately three minutes to get familiar with the simulator motion and two practice runs in which the participant could get familiar with the primary and secondary task were given prior to the experiment. The experiment design is illustrated in Figure 12.

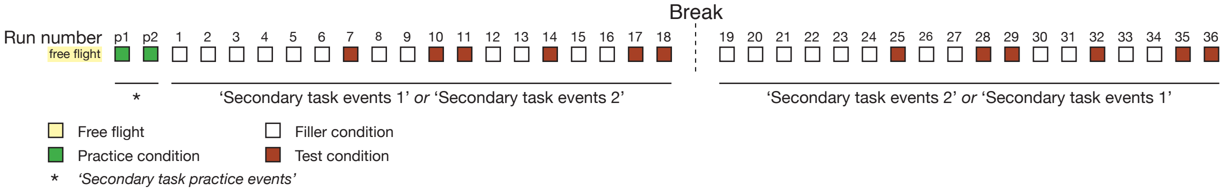


Fig. 12 Illustrating the experiment design used for each participant.

F. Dependent Measures

The outcomes were compared between the ‘no-motion-turn’ (as baseline), ‘motion-level’, and ‘motion-opposite’ conditions. Even though the matching conditions could not be used in the analysis, the results are still shown as they gave a more complete view of the results. The following variables were taken as dependent measures for the experiment.

Error Rate - The direction of the first input done after the AI was shown indicated if a RRE (or undesired input) was made. This could be due to wrong expectation in combination or not with a horizon control reversal.

Error Severity - When an error was made, the maximum bank angle deviating from level flight was measured. This gave an overview of how severe the errors were, which could be used in further analysing the effect of expectation on the control strategy.

Reaction Time - From the moment after the autopilot switched off and the AI was shown again, the reaction time to the first input (|column deflection rate| > 0) done was measured. A comparison of reaction times between the test conditions could point to a pilot who needed to think twice, or ‘re-frame’, when referring to the attitude indication, even when no error was made.

Subjective Workload - The subjective workload assessment was done with the NASA Task Load Index (TLX) deriving an overall workload score based on a weighted average of the ratings of mental demand, physical demand, temporal demand, performance, effort and frustration [37].

G. Data Analysis

The aircraft state, pilot inputs and simulator motion data from the SRS were processed using custom MathWorks MATLAB scripts from which the error rate, severity and reaction time could be computed. Various analyses were performed using these data, which are stated below.

Error Threshold - Not all pilot inputs done to the other side than required could be called a roll reversal error because also a small input could be made without any intention to change the aircraft attitude accordingly. Hence, looking at the error rate and error severity, a threshold was determined after which an erroneous input could be called a RRE.

Learning Effect - The expectation of the participants was influenced in such a way that the attention was drawn away from the slow tilt and amplified when the fast motion cue occurred in Phase 1 of the experimental runs. A comparison was made between the first experimental runs and the ones following after based on the number of errors made and the severity of those errors to see whether a learning effect was present.

Distraction - Analyzing the amount of inputs done in the secondary task together with the workload rating gave an indication on how distracted pilots were by performing the secondary task.

Statistical Analysis - Statistical analysis was performed using IBM SPSS software. In the case of non-parametric data, as for the error rates, the Generalized Estimating Equations were used allowing for analysis of repeated measurements. Reaction times were processed using a factorial ANOVA after normality and equality of variances was proven, using the Kolmogorov-Smirnov and Levene's test, respectively.

Differences in error rate between the repetitions of the test conditions was tested with Cochran's Q test, which helped identifying a significant learning effect. The Wilcoxon signed-rank test with a Bonferroni correction was then performed for pairwise comparison between the repetitions. To check for significance in error severity between the groups, a non-parametric Kruskal-Wallis test was used. A learning effect regarding error severity for the 'motion-opposite' condition was tested using the non-parametric Friedman's ANOVA. Lastly, a one-way ANOVA was performed to identify differences in the secondary task rating.

H. Hypotheses

Looking at the error rate, it was expected that some errors were made in the baseline 'no-motion-turn' condition, similar to previous studies evaluating the ambiguity of the moving-horizon type display [9–11, 13]. For the 'motion-level' condition it was hypothesized that professional pilots would not just act on the motion-induced expectation without scanning the instrument first. Their training and experience with the AI would result in a quick conclusion that no input was needed, resulting in a minimum or no errors made. The 'motion-opposite' condition facilitated the expectation-induced misinterpretation of the AI and therefore more errors were expected to be made in this condition when comparing it to the baseline. To summarize, it was hypothesized that an expectation about the bank angle would not directly influence the control strategy of pilots, but it would make them more vulnerable to misinterpretations of the AI, leading to more RREs being made.

The inputs made by the participant automatically gave feedback on their chosen control strategy, as they would know it was correct or incorrect. Consequently, after the first test run (and perhaps first error made) a pilot might have made sense of what was happening during the experiment. Therefore, for both the error rate and error severity, a learning effect was expected to be found in each of the conditions given. This means that the most and most severe errors would be made in these first repetitions of these conditions.

Both the experienced and novice group of pilots received different training, from which the training of novice pilots was more focused on Upset Prevention and Recovery Training [38]. This could result in this group making less and less severe errors than the experienced pilots, as their intuitive reaction in upset situations is well-trained. On the other hand, pilots with more flight hours have more hands-on flying skills and are more experienced in using the AI, which could result in them being less vulnerable to misinterpretations. Consequently, no statement was made regarding the differences in error rate or severity between the experienced and novice group of pilots.

Regarding the reaction time, no significant differences between runs or conditions were expected to be found since it would be dependent on many factors (e.g., inattention, distrust, overconfidence, general control strategy). Between the groups it was hypothesized that experienced pilots would respond slightly slower than novice pilots since the group of experienced pilots also consisted of older participants and reflexes are known to slow with age [39].

Lastly, between the two groups also no significant difference in workload rating based on the NASA TLX workload assessment was expected, nor between pilots who made and pilots who did not make errors. However, it was expected that the workload rating would indicate that the secondary task was sufficiently distracting.

IV. Results

A. Example Run

Figure 13 shows the time history for two subsequent runs from one of the participants. Both scenarios provided this pilot with the same motion cues, as can be seen from the dashed red line indicating the simulator's specific force (f_y). The black dot indicates the moment the 'autopilot off alarm' sounded and the black dotted vertical line shows the moment at which the PFD was shown again. The first condition was a 'motion-matching' condition. From the thick blue line (ϕ) it can be seen that the aircraft was rolled level in one smooth motion. The second condition was a 'motion-opposite' condition and therefore showed a PFD that was not matching the direction of the motion. Here, the participant mistakenly responded by increasing the bank angle up to 41° before turning the wings back level, which is a typical example of a pilot making a RRE in our experiment.

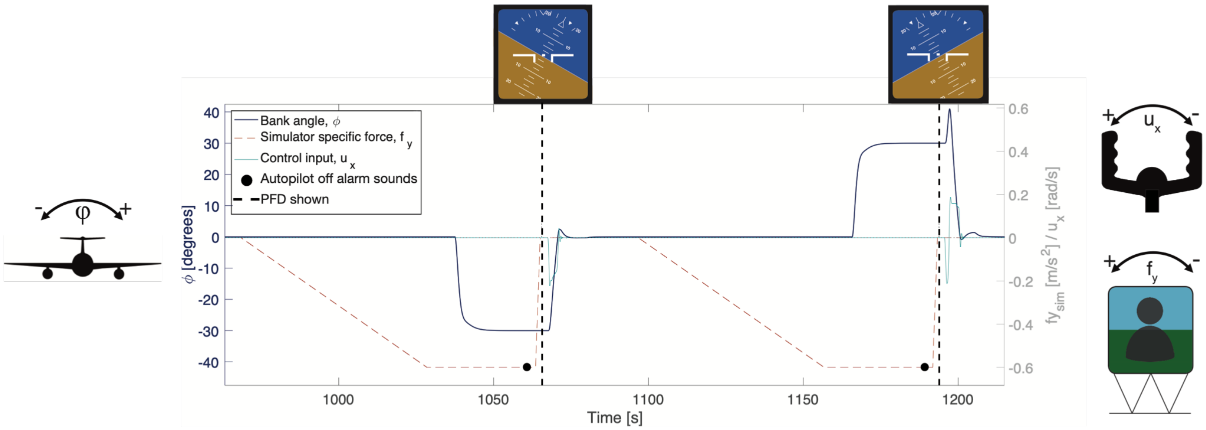


Fig. 13 Example of a 'motion-matching' run followed by a 'motion-opposite' run where a roll reversal error was made. All direction indications are seen from behind.

B. Error Rate

An input was considered erroneous when the bank angle error exceeded 1.5° *and* the column deflection exceeded 2.5°. Appendix D elaborates further on how this combined threshold was determined.

Table 4 lists an overview of the roll reversal errors made, which are illustrated in Figure 14. It can be observed that the error rate in the 'motion-opposite' condition was higher compared to the baseline 'no-motion-turn' condition. In the 'motion-level' condition, no errors were made at all. Also, from Figure 15, showing the error rates for the two groups, it can be noted that the group of experienced pilots made slightly more errors in almost all conditions. The only error in the 'no-motion-matching' condition was made by a novice pilot. An overview of all conditions given to the participants with an indication of the errors made can be found in Figure 16. From this figure it can be observed that six out of eighteen of the participants (33.3%) made a RRE in the first 'motion-opposite' condition, which was considered beforehand as the best representation for the leans illusion as it is experienced in real flight since it involved the most surprising effect. In total, the errors were made by twelve participants, from which eight experienced and four novice pilots.

Table 4 Amount of errors made, shown in number and percentage.

		motion-matching	no-motion-matching	no-motion-turn	motion-level	motion-opposite
Total runs	(No.)	360	71	70	71	72
Errors	(No.)	8	1	4	0	12
Errors	[%]	2.2	1.4	5.7	0.0	16.7

The Generalized Estimating Equations procedure, allowing for statistical analysis of repeated measurements, showed no significant effect on the error rates between the two groups and also no significant interaction effect between the conditions and groups. There was a significant effect of condition on error, $\chi^2(1,18) = 4.535$, $p = 0.033$. The number of errors made in the 'motion-opposite' condition were significantly higher compared to both the 'motion-level' scenario, $Z = -2.972$, $p = 0.003$, and the 'no-motion-turn' scenario, $Z = -2.309$, $p = 0.021$. No significant effect was found between the two latter scenarios, $Z = -1.633$, $p = 0.102$.

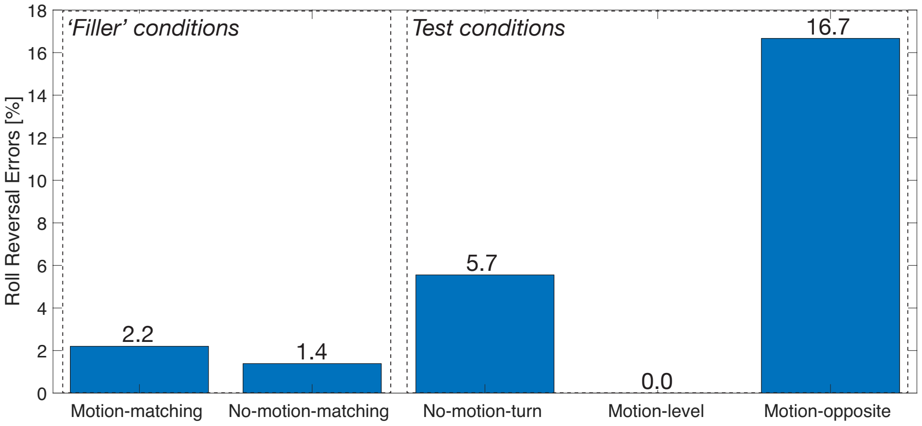


Fig. 14 The mean error rates for all five conditions for all pilots.

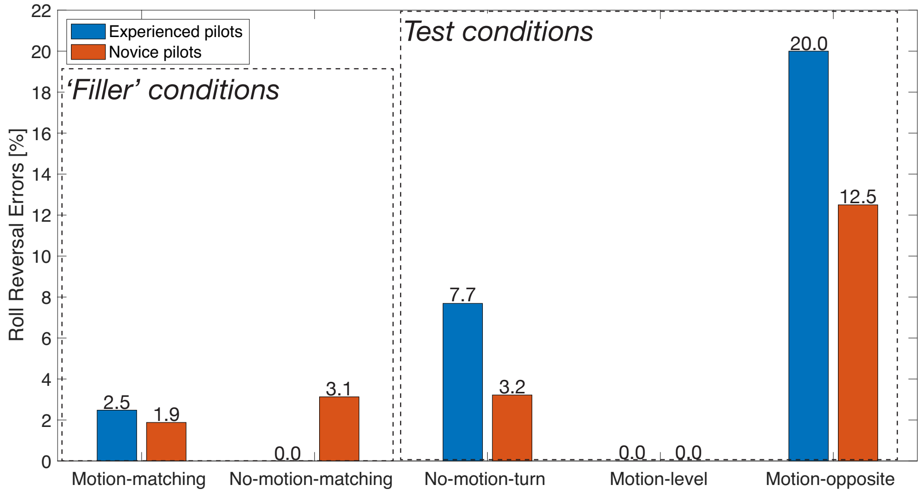


Fig. 15 The mean error rates for all five conditions per group.

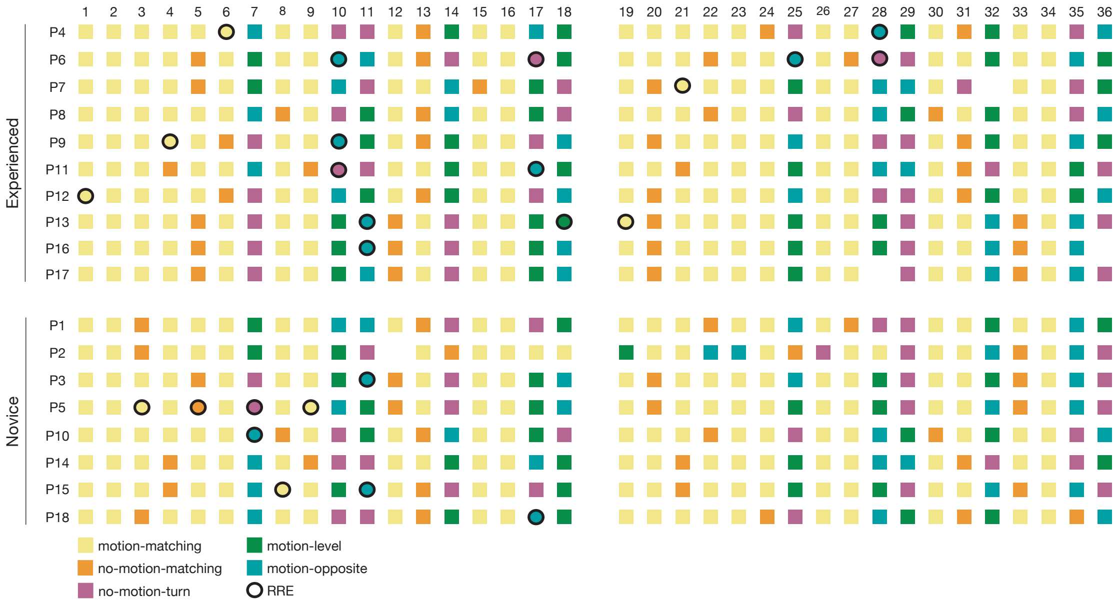


Fig. 16 Order of conditions given to the participants, with all RREs indicated. The four white coloured runs were not executed, and therefore no data is available for those runs.

C. Error Severity

The box plots in Figure 17 show the maximum error made in degrees of bank angle for all scenarios divided over the two groups. The numbers between the parentheses on top of the box plots show the number of RREs. The crosses to the left of the box plots show the individual data points. From the ‘motion-matching’ and ‘motion-opposite’ plots it can be found that experienced pilots deviated more from the intended flight path than novice pilots. Only for the ‘no-motion-turn’ scenarios this observation is not consistent, but it should be noted that there was only one data point available for the novice pilots in this scenario.

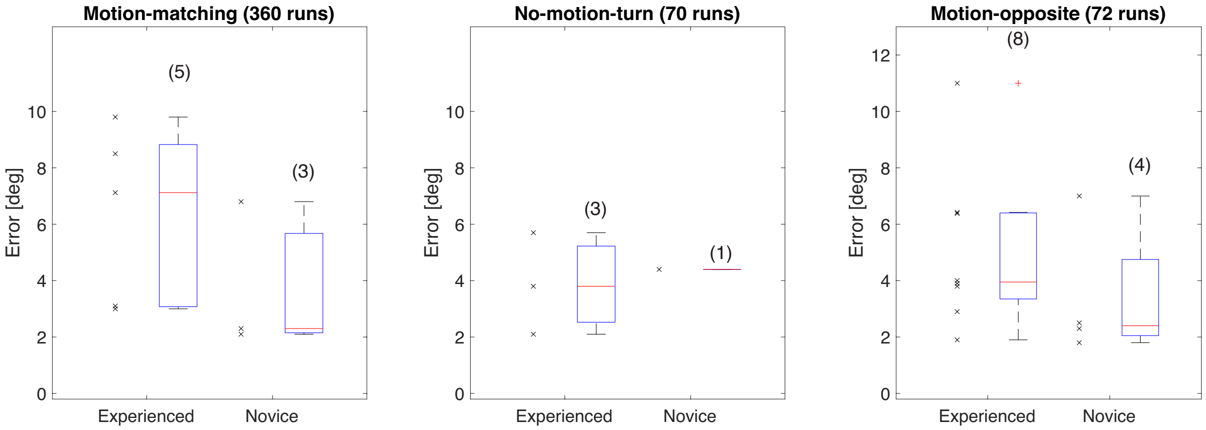


Fig. 17 Effects of experience on error severity for three of the conditions.

Using the non-parametric Kruskal-Wallis test, it was found that there was no significant difference in error severity between the groups for the ‘motion-opposite’ condition, $\chi^2(1) = 0.7$, $p = 0.403$. This test was not performed for the ‘no-motion-turn’ condition since there were not enough data points to draw a conclusion. The ‘motion-matching’ condition is only shown as a reference, but was not a test condition, and is therefore not further analysed.

D. Reaction Time

When looking at the average reaction times for each condition, as shown in Figure 18, it can be observed that the differences between conditions within both groups are small. For the conditions involving a banked angle of 30° from which the pilots had to recover, the reaction times did differ between the groups. The mean reaction time and standard deviation for each condition are shown for the experienced and novice group of pilots in Table 5. When looking at the average reaction time over all runs, it can be observed that the experienced pilots responded approximately half a second later than the novice pilots.

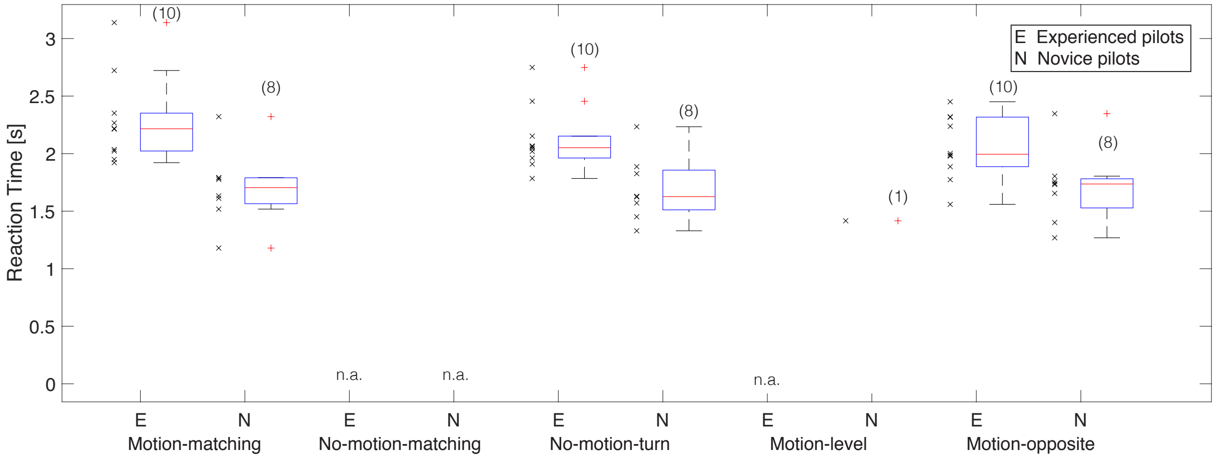


Fig. 18 Average reaction time per participant for each condition.

Table 5 Mean reaction times for the two groups.

		motion-matching	**no-motion-matching**	**no-motion-turn**	**motion-level**	**motion-opposite**	**all runs**
	N	Reaction Time - Mean (SD)					
Experienced pilots	10	2.28 (0.32)	n.a.	2.12 (0.28)	n.a.	2.05 (0.28)	2.29 (0.37)
Novice pilots	8	1.70 (0.32)	n.a.	1.69 (0.28)	0.17 (-)	1.71 (0.32)	1.75 (0.30)

Using the Kolmogorov-Smirnov and Levene's test, it was found that the total reaction times for both the novice group, $D(8) = 0.21$, $p = 0.2$, and for the experienced group, $D(10) = 0.25$, $p = 0.08$, were normally distributed and that the variances of the two groups were equal, $F\ (1,16) = 2.32$, $p = 0.15$. Consequently, a factorial ANOVA analysis was performed, showing that the difference between the reaction times of the two groups was significant, $F(1,32) = 31,68$, $p < .05$. There was no significant difference in reaction time between the 'no-motion-turn' and 'motion-opposite' scenario, $F(1,32) = 0.002$, $p = 0.962$.

E. Learning Effect

A large number of errors made was either in the first (6 out of 18 participants) or second (4 out of 18 participants) occurrence of the 'no-motion-turn' or 'motion-opposite' conditions, as shown in Figure 19. Moreover, even though during the break one was instructed not to talk or discuss about the experiment, eight out of the eighteen pilots were thinking out loud, expressing suspicion about the simulator motion in some of the runs. During the de-briefing after the experiment twelve pilots indicated that their control strategy changed after the break, such that they would 'rely more on the instrument' while trying to eliminate the perception of the simulator motion.

Figure 20 shows the error severity for the 'motion-opposite' condition, split over the first three repetitions of the condition (no errors were made in the last repetition). From this plot, it can be observed that the errors made in the first 'motion-opposite' run also led to larger deviations compared to those in later repetitions of this condition.

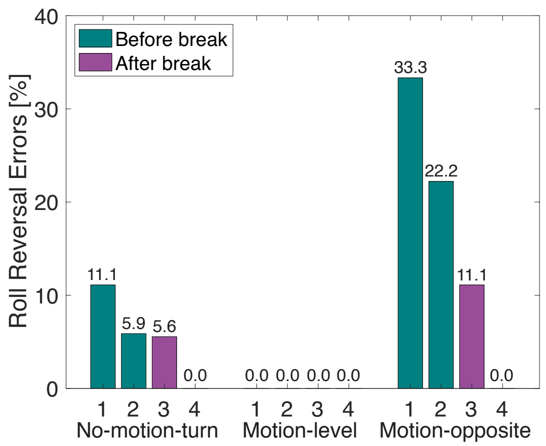


Fig. 19 The mean error rates for the three test conditions, divided over the four repetitions. Error percentages are calculated over the number of participants.

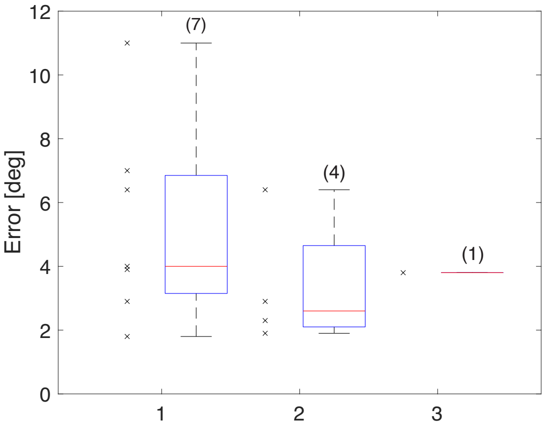


Fig. 20 Error severity for the 'motion-opposite' condition shown for the first three repetitions.

Using a Cochran's Q test, it was found that the error rate did not significantly change between the four repetitions of the 'no-motion-turn' condition, $\chi^2(3) = 2.4$, $p = 0.875$. For the 'motion-opposite' condition, there was a significant change, $\chi^2(3) = 9.41$, $p = 0.024$. A Wilcoxon signed-rank test with a Bonferroni correction was used to make a pairwise comparison between the repetitions. It showed that there were no significant changes between the first two repetitions, $T = 13.5$, $p = 0.48$, nor between the second and third repetitions, $T = 3.0$, $p = 0.18$, nor between the third and final, T = 0.0, $p = 0.32$. Only when comparing the number of errors made before the break (first two repetitions) and after the break (last two repetitions), a significant difference was found, $Z = -2.31$, $p = 0.021$. Using a Friedman's ANOVA, no significant differences were found between the first three repetitions when focusing on the error severity, $\chi^2(3) = 5.10$, $p = 0.078$. The fourth repetition was left out of this analysis, since here no error was made.

F. Distraction

Figure 21, showing the TLX workload rating, indicates that the experienced pilots rated their workload higher than the group with novice pilots. Also, the pilots who made at least one error rated their workload to be slightly higher than those who did not. Also, pilots seemed sufficiently distracted by the secondary task, since the average workload rating for both groups lays around the middle.

A one-way ANOVA showed that the differences between the experienced and novice pilots, or between the pilots who made an error and those who did not, were not significant, with $F = 3.076$, $p = 0.099$, and $F = 2.244$, $p = 0.154$, respectively. An overview of the amount of inputs done by all participants, together with the full workload rating, can be found in Appendix E.

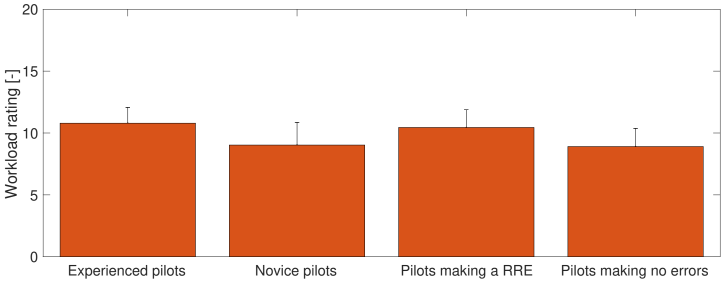


Fig. 21 Weighted subjective workload for different groups.

V. Discussion

The experiment design led the participants to have an intuitive reaction when responding to the AI. As Bainbridge(1983)[3] wrote, "*the operator who has to do something quickly can only do so on the basis of minimum information, he will not be able to make decisions based on a wide knowledge of the plant*" (p.776). It was already shown in both a fixed-base simulator and in-flight experiment that, in those situations, an expectation about the bank angle influences how the information on the AI is interpreted for non-pilots [6, 14]. The difference between the 16.7% of RREs made in the 'motion-opposite' scenario to 5.7% in the 'no-motion-turn' scenario in the current study, shows that also professional pilots tend to misinterpret the AI due to expectations created by the motion.

The 5.7% error rate from the baseline condition corresponds well with the 4.5-8.7% error rate in an experiment done using a motion-based simulator evaluating the ambiguity of the display by Müller (2018)[11], where also the expectations about the bank angle were tried to be minimized. Errors made in this condition suggest that the pilot was confusing the horizon symbol with the aircraft symbol, meaning that the occurrence of these errors reflect the ambiguity of the indicator. The observation that no errors were made in the 'motion-level' condition suggests that the errors made during the 'motion-opposite' scenarios were also due to a misinterpretation of the attitude indicator. The increase with a factor 2.9 between the 'motion-opposite' and baseline condition shows a significant influence of expectation in making a horizon control reversal (i.e., reading the horizon as being the aircraft). The results from this study therefore show that expectation shapes the way information is being interpreted when using an intuitive reaction. Consequently, the hypothesis regarding error rate can be accepted, that is, whereas an expectation about the bank angle does not directly influence the control strategy of pilots, it does make them more vulnerable to misinterpretations of the AI.

As expected, most errors were found in the first repetition of the 'motion-opposite' condition. Up to 33.3% of the pilots made a RRE the first time the condition was presented, which were on average also more severe compared to later repetitions. This shows that pilots quickly learned to change their control strategy after the first test condition. Consequently, it was found that pilots still could be better prepared for unexpected and surprising situations. Results from a study by Landman et al.(2018) [40] show that training could be offered in a more unpredictable and variable way, as this improves transfer of training to unexpected situations.

The average reaction time showed to be significantly higher for the experienced pilots compared to the novice pilots. It should be noted that the pilots were told that the experiment looked into how fast they could respond after a moment of inattentiveness or when experiencing high workload. Therefore, this could either point to better reflexes of novice (and also younger) pilots, or to the 'older' generation being less 'trigger-happy'. It can also mean that experienced pilots are a bit calmer in their control strategy as they have experienced spatial illusions or airplane upsets more often (in-flight and in simulator training), and therefore are more confident when using the AI in these situations. This lays in line with a conclusion drawn from a study of Hasbrook et al.(1973)[9] saying that "*experienced pilots subjectively were more 'at ease' with the moving horizon instrument*" (p.2). From an operational perspective the difference might originate from the different types of training the pilots received. The younger generation had a more extensive training in how to handle recovery from airplane upsets and might therefore be more confident in using an intuitive response. The experienced pilots were told to have a moment of rest before responding, although this was more in general than specifically for recovering from upsets.

The man-in-the-loop spatial disorientation scenario designed for a hexapod simulator was effective in that it successfully shaped the expectation of pilots regarding the bank angle. Even though it was based on the true 'leans' cues, it does not exactly simulate this illusion as it will happen in real flight. However, in the de-briefing of the experiment, one pilot did indicate that he had recently experienced the 'leans' and that this sensation was very similar to what he experienced in the simulator. Another pilot indicated that he did not actively notice the motion cues and that his expectation was not necessarily changed by it. It was, however, quite noticeable that this same pilot did make two RREs, exactly in the two 'motion-opposite' conditions given before the break. This could indicate that the scenario did put him in a certain frame regarding the bank angle.

This study was designed to examine the influence of expectation on a pilot's control strategy. In the experiment design also some limitations were involved, which should be considered when interpreting the results or which could be a good framework for future research. First of all, the pilots had to recover from the same bank angle for many repetitions. In the first 'motion-opposite' run, surprise still may have been a present factor which could have affected the response. Later, these effects decreased as a possible ‘mismatch’ was expected. Also, pilots had to recover from static bank angles only. As mentioned by Landman et al.(2020)[6], “*situations in which a pilot attempts to correct a roll motion away from level might be more hazardous*”(p.13). As this also adds a time restriction, it might induce acute stress, which has been found to particularly impair slow appraisal and reframing as these processes are relatively more analytical [18]. Furthermore, for this study only roll inputs could be made in order to control the measurements in such a way that they could be used for the evaluation of the effect of expectation. Care should be taken when using the results as a general performance measure for pilots in recovering from bank angles since for this experiment the normal recovery protocols could not be executed. Lastly, it was found that all first repetitions of the ‘motion-opposite’ scenario in this study provided the pilots with a motion cue in the same direction as the ‘motion-matching’ run given just before, but now with a different AI shown in the second phase. The run given before a test condition might have affected the frame of the pilot, which could have attributed to the extensive amount of RREs made in the first repetition of the ‘motion-opposite’ condition.

VI. Conclusion

The spatial disorientation scenario developed for a hexapod simulator was successful, in that it led a large proportion of both novice as experienced pilots to have a wrong sensation about the bank angle, resulting in errors being made. Also, some of the participants indicated that the sensation was similar to experiencing spatial disorientation in-flight. Consequently, the scenario could be used in training and further research into the effects of expectation or spatial disorientation in general.

Furthermore, the first time the incongruent motion cues were given was the most representative for real in-flight spatial disorientation, as here the ‘surprise’ effect was still the most salient. In later repetitions, a clear learning effect was found. Results of the experiment suggest that having a wrong expectation about the aircraft attitude makes pilots more vulnerable to misinterpretations of the AI. The possibility of developing a (wrong) expectation is a common human factor that can not be completely eliminated. Therefore, automation and technologies used by pilots should be developed in such a way that they guide pilots in their decision-making strategy, minimizing the chances of misinterpretations.

For future research it would be recommended to incorporate the full recovery scenarios into an experiment looking into expectation, as those recovery strategies might be more intuitive for pilots.

References

[1] Strauch, B., “Ironies of Automation: Still Unresolved after All These Years,” *IEEE Transactions on Human-Machine Systems*, Vol. 48, No. 5, 2018, pp. 419–433. https://doi.org/10.1109/THMS.2017.2732506.

[2] Billings, C., *Aviation Automation: The Search for a Human-Centered Approach*, 1^{st} ed., CRC Press, 1996.

[3] Bainbridge, L., “Ironies of Automation,” *Automatica*, Vol. 19, No. 6, 1983, pp. 775–779. https://doi.org/10.1016/0005-1098(83)90046-8.

[4] Ranaudo, R. J., “Startled Increasing automation has introduced new situational awareness challenges for pilots,” Tech. rep., 2019. URL https://flightsafety.org/asw-article/startled/.

[5] Technical Commission of Inquiry - Republic of Cameroon, “Technical investigation into the accident of the B737-800 crash registration 5Y-KYA,” http://www.aviation-accidents.net/report-download.php?id=200, 2007. Online; accessed 8 January 2020.

[6] Landman, A., Groen, E. L., van Paassen M. M., Bronkhorst, A. W., and Mulder, M., “Incongruent expectation induces roll reversal errors when using the artificial horizon,” *Perception*, Vol. 0, No. 0, 2020, pp. 1–14. https://doi.org/10.1177/0301006619901053.

[7] Johnson, S. L., and Roscoe, S. N., “What Moves, the Airplane or the World?” *Human Factors*, Vol. 14, No. 2, 1972. https://doi.org/10.1177/001872087201400201.

[8] Roscoe, S. N., and Williges, R. C., “Motion relationships in aircraft attitude and guidance displays: A flight experiment,” *Human Factors*, Vol. 17, No. 4, 1975, pp. 374–387. https://doi.org/10.1177/001872087501700409.

[9] Hasbrook, A. H., and Rasmussen, P. G., “In-flight performance of civilian pilots using moving-aircraft and moving-horizon attitude indicators,” Tech. rep., 1973. URL https://apps.dtic.mil/dtic/tr/fulltext/u2/773450.pdf.

[10] Beringer, D. B., Williges, R. C., and Roscoe, S. N., “The Transition of Experienced Pilots to a Frequency-Separated Aircraft Attitude Display,” *Human Factors: The Journal of Human Factors and Ergonomics Society*, Vol. 17, No. 4, 1975, pp. 401–414. https://doi.org/10.1177/001872087501700411.

[11] Müller, S., Sadovitch, V., and Manzey, D., “Attitude Indicator Design in Primary Flight Display: Revisiting an Old Issue With Current Technology,” *The International Journal of Aerospace Psychology*, Vol. 28, No. 1-2, 2018, pp. 46–61. https://doi.org/10.1080/24721840.2018.1486714.

[12] Ince, F., Williges, R. C., and Roscoe, S. N., “Aircraft simulator motion and the order of merit of flight attitude and steering guidance displays,” *Human Factors*, Vol. 17, No. 4, 1975, pp. 388–400. https://doi.org/10.1177/001872087501700410.

[13] Singer, G., and Dekker, S. W. A., “the Effect of the Roll Index (Sky Pointer) on Roll Reversal Errors,” *Human Factors and Aerospace Safety*, Vol. 2, No. 1, 2002.

[14] Landman, A., Davies, S., Groen, E. L., van Paassen M. M., Lawson, N. J., Bronkhorst, A. W., and Mulder, M., “In-flight Spatial Disorientation Induces Roll Reversal Errors when Using the Attitude Indicator,” *Applied Ergonomics*, Vol. 81, No. 7, 2019, pp. 1–14. https://doi.org/10.1016/j.apergo.2019.102905.

[15] Lampton, A. K., Klyde, D. H., and Schulze, P. C., “Evaluation of a Missed Approach/Go-Around Spatial Disorientation Demonstration Scenario for Commercial Pilot Training,” *Paper presented at the AIAA Guidance, Navigation and Control conference, Grapevine, Texas*, January 2017. https://doi.org/10.2514/6.2017-1080.

[16] Lewkowicz, R., Bałaj, B., and Francuz, P., “Susceptibility to Flight Simulator-Induced Spatial Disorientation in Pilots and Non-Pilots,” *The International Journal of Aerospace Psychology*, Vol. 0, No. 0, 2020, pp. 1–13. https://doi.org/10.1080/24721840.2019.1696680.

[17] Klyde, D. H., Lampton, A. K., and Schulze, P. C., “Development of Spatial Disorientation Demonstration Scenarios for Commercial Pilot Training,” *American Institute of Aeronautics and Astronautics*, Vol. 6, No. 1, 2016, pp. 1–19. https://doi.org/10.2514/6.2016-1180.

[18] Landman, A., Groen, E. L., van Paassen, M. M., Bronkhorst, A. W., and Mulder, M., “Dealing With Unexpected Events on the Flight Deck: A Conceptual Model of Startle and Surprise,” *Human Factors*, Vol. 59, No. 8, 2017, pp. 1161–1172. https://doi.org/10.1177/0018720817723428.

[19] Neisser, U., *Cognition and reality: Principles and implications of cognitive psychology*, New York, NY: W.H. Freeman, 1976.

[20] Kahneman, D., “A perspective on judgment and choice: Mapping bounded rationality,” *American Psychologist*, Vol. 58, No. 9, 2003, pp. 697–720. https://doi.org/10.1037/0003-066X.58.9.697.

[21] Martin, W. L., Murray, P. S., and 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*, September 2012, pp. 387–394.

[22] Klein, G., Phillips, J. K., Rall, E. L., and Peluso, D. A., “A Data-Frame Theory of Sensemaking,” *Lawrence Erlbaum Associates*, Vol. 0, No. 0, 2007, pp. 113–155.

[23] Beach, L., and Connolly, T., *The Psychology of Decision Making: People in Organizations*, SAGE Books, 2005.

[24] Rankin, A., Woltjer, R., and Field, J., “Sensemaking following surprise in the cockpit—a re-framing problem,” *Cognition, Technology and Work*, Vol. 18, No. 4, 2016, pp. 623–642. https://doi.org/10.1007/s10111-016-0390-2.

[25] Delk, J. L., and Fillenbaum, S., “Differences in perceived color as a function of characteristic color,” *The American Journal of Psychology*, Vol. 2, No. 78, 1965, pp. 290–293. https://doi.org/10.2307/1420503.

[26] Draper, D., “Bayesian statistics,” *Computational Complexity: Theory, Techniques, and Applications*, Vol. 9, 2012, pp. 254–274. https://doi.org/10.1007/978-1-4614-1800-9_17.

[27] Navathe, P. D., and Singh, B., “An operational definition for spatial disorientation,” *Aviation, space, and environmental medicine*, Vol. 65, No. 12, 1994, pp. 1153–1155.

[28] Nooij, S. A., and Groen, E. L., “Rolling into spatial disorientation: Simulator demonstration of the post-roll (Gillingham) illusion,” *Aviation Space and Environmental Medicine*, Vol. 82, No. 5, 2011, pp. 505–512. https://doi.org/10.3357/ASEM.2946.2011.

[29] Ledegang, W., Groen, E., and Wentink, M., "Motion Fidelity of Uncoordinated Simulator Roll for Simulation of Coordinated Aircraft Roll," *Paper presented at the AIAA Modeling and Simulation Technologies Conference, Minneapolis, Minnesota*, August 2012, pp. 1–12. https://doi.org/10.2514/6.2012-4494.

[30] Heerspink, H., Berkouwer, W., Stroosma, O., van Paassen, R., Mulder, M., and Mulder, B., "Evaluation of Vestibular Thresholds for Motion Detection in the SIMONA Research Simulator," *Paper presented at the AIAA Modeling and Simulation Technologies Conference, San Francisco, California*, August 2005. https://doi.org/10.2514/6.2005-6502.

[31] Lombaerts, T. J. J., Looye, G., Seefried, A., Neves, M., and Bellmann, T., "Development and Concept Demonstration of a Physics Based Adaptive Flight Envelope Protection Algorithm," *International Federation of Automatic Control*, Vol. 49, No. 5, 2016, pp. 248 – 253. https://doi.org/10.1016/j.ifacol.2016.07.121.

[32] 'NASA', "The Multi Attribute Test Battery," https://matb.larc.nasa.gov/test-2/291-2/, 2014. Online; accessed 8 January 2020.

[33] Wickens, C., "Attention to safety and the psychology of surprise," *Presented at the 11th International Symposium on Aviation Psychology, Columbus, Ohio*, March 2001.

[34] Gundry, A., "Thresholds to roll motion in a flight simulator," *Journal of Aircraft*, Vol. 14, No. 7, 1977, pp. 624–631. https://doi.org/10.2514/3.58832.

[35] Gillingham, K. K., and Previc, F. H., "Spatial Orientation in Flight," Tech. rep., Brooks Air Force Base, TX, November 1993.

[36] Ledegang, W. D., and Groen, E. L., "Spatial Disorientation Influences on Pilots' Visual Scanning and Flight Performance," *Aerospace Medicine and Human Performance*, Vol. 89, No. 10, 2018, pp. 873–882. https://doi.org/10.3357/amhp.5109.2018.

[37] 'Human Performance Research Group', "NASA Task Load Index (TLX)," Tech. rep., 1980. URL https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20000021488.pdf.

[38] ICAO, "Manual on Aeroplane Upset Prevention and Recovery Training," 2014, pp. 1–94.

[39] Tun, P. A., and Lachman, M. E., "Age Differences in Reaction time and Attention in a National Telephone Sample of Adults: Education, Sex, and Task Complexity Matter," *Developmental psychology*, Vol. 44, No. 5, 2008, pp. 1421–1429.

[40] Landman, A., van Oorschot, P., van Paassen, M. M., Groen, E. L., Bronkhorst, A. W., and Mulder, M., "Training Pilots for Unexpected Events: A Simulator Study on the Advantage of Unpredictable and Variable Scenarios," *Human Factors*, Vol. 60, No. 6, 2018, pp. 793–805. https://doi.org/10.1177/0018720818779928.

[41] Landman, A., Groen, E. L., van Paassen, M. M., Bronkhorst, A. W., and Mulder, M., "Dealing With Unexpected Events on the Flight Deck: A Conceptual Model of Startle and Surprise," *Human Factors*, 2017. https://doi.org/10.1177/0018720817723428.

II

Book of Appendices

A

Conceptual Model of Startle and Surprise

The paper from Landman (2017)[18], 'Dealing With Unexpected Events on the Flight Deck: A Conceptual Model of Startle and Surprise', aimed to conceptualize startle and surprise responses to understand the potential incapacitating effects during unexpected events.

For the model, startle was defined as "*the initial short-term, involuntary physiological and cognitive reaction to an unexpected event that commence the normal human stress response*". The startle response will mostly arise due to sudden exposure to an intense stimulation, such as a flash of light or a sudden loud sound that generates a physiological reflex with an emotional response. Notice that also the first reaction to the realization of spatial disorientation can evoke a startle response. Surprise is defined as a cognitive-emotional response to something unexpected, which results from a mismatch between one's mental expectations and perceptions of one's environment.

The current study looked into the pilot's response to an information mismatch due to wrong expectations in spatial disorientation flight. The intensity of the stimuli was not amplified to elicit a startling response, nor did the task involve a (persisting) threat potentially leading to an increased stress level. Some startling effects or stress responses might have been present, but they were not shown in the reduced version of the model, as they were not taken into account in the analysis of the actions following the mismatch.

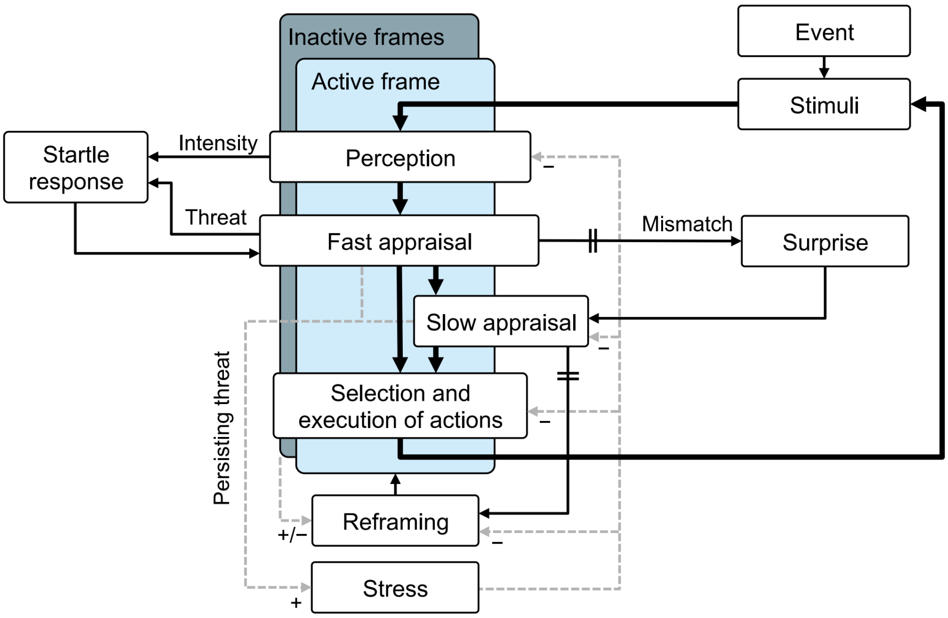


Figure A.1: Full conceptual model of startle and surprise as directly copied from Landman (2017). Solid lines indicate sequenced events. Dashed lines indicate potential influences, with plus signs indicating an increasing effect and minus signs indicating an impairing effect. Double lines indicate thresholds [18].

B

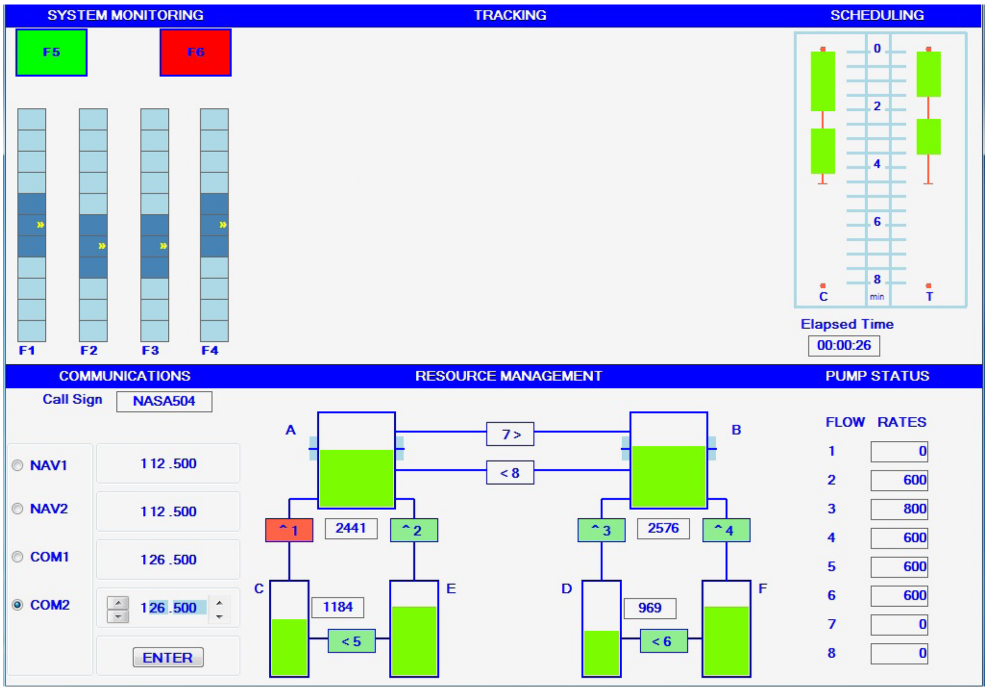
Participant Information

Participant number	**Sleep in last 24h**	**Experience**		**Position**	**Extra experience**
	[hrs]	[yrs]	[hrs]		
Experienced pilots					
12	16	22	11000	Captain	Instructor
16	11	23	12500	First officer	Glider, instructor, Tiger Moth
9	10	25	15000	Captain	Aerobatics, instructor, Flight testing, tow, glider
17	15	35	16990	Captain	Glider, aerobatics, instructor
13	14	27	17400	Captain	Aerobatics, military, instructor, taildragger
8	15	36	18000	Retired	Instructor
4	14	34	19000	Captain	Glider, instructor
7	14	31	19500	Retired	Glider
6	16	36	20000	Captain	Instructor
11	15	35	20500	Retired	Test pilot after heavy maintenance
Novice pilots					
18	16	0.08	50	Second officer	Upset Prevention and Recovery Training (5hrs)
14	15	0.14	500	Second officer	-
2	15	1.5	750	Second officer	Glider, glider-aerobatics, upset recovery single engine
5	12	1.5	950	Second officer	Glider
3	15	10	5000	First officer	Glider
1	14	5.5	3350	First officer	Simulator instructor
10	18	6	3800	First officer	-
15	12	7	4000	First officer	Upset prevention and aerobatics

C

Revised Multi Attribute Test Battery

Most of the following information is directly copied from https://matb.larc.nasa.gov/, NASA [32].



Elapsed Time 00:00:26

Call Sign NASA504

○ NAV1 112.500

○ NAV2 112.500

○ COM1 126.500

● COM2 126.500

FLOW	RATES
1	0
2	600
3	800
4	600
5	600
6	600
7	0
8	0

Figure C.1: The Multi Attribute Task Battery, with the tracking task disabled [32].

The system monitoring task

In the normal state the green light is on, the red light is off and the four scales move randomly within the central range. In the non-normal state, one or more of the following occur: the green light is off, the red light is on and a scale or scales move randomly either above or below the central range. The subject uses either mouse clicks or function key selections to return systems to the normal state. If the correct response is not made within a timely manner the system returns to the normal state.

Scheduling task

The scheduling window allows the subject to look ahead for up to eight minutes, at expected activity of the communication and tracking task. The scheduling window can display the beginning, end, and duration of these two tasks. The timelines are identified by ‘c’ for the communications and ‘t’ for the tracking task. The bars indicate times when the subject can anticipate having to perform on of these tasks. The thin lines indicate times during which the tasks do not usually require input from the subject.

Communications task

The ATC assigns a new frequency for one of the four installed radios. When the instruction is directed to ‘NASA504’, the subject response by selecting the appropriate radio and adjusting the frequency. When the instruction is directed to another call sign, the subject ignores it. The event will time out if a subject response is not made in a timely manner. This task was slightly adjusted for this study, as instructions could be given over the headset. Therefore, the participants received a list with radios and frequencies that had to be changed at a certain time.

Resource Management task

In the normal state the subject uses any or all of the eight pumps to maintain the fluid level in tanks A and B within the target range, which is indicated by the light blue borders on the sides of the tanks. In the non-normal state, one or more of the pumps fail and is unusable by the subject. Failed tanks may be fixed by an event and returned to the normal state during a run.

D

Error Threshold Determination

Not all inputs done to the other side than required can be called a roll reversal error because also a small input could be made without any intention to change the aircraft attitude accordingly. This can be related to the movement with which the pilot reaches the column, or it can be originated from a pilot's control strategy where the column is simply moved slightly to (double) check the working of the instrument, without being guided by expectations. To determine the threshold for which an erroneous input could be counted as a RRE, a closer look was taken into the inputs made.

The inputs done on the control column translated directly to the aircraft roll rate. To take the control aggressiveness into account, the maximum stick deflection is obtained by assuming a linear relationship between the maximum input and time. Figure D.1 shows the maximum bank angle error made after the AI was shown again, against the absolute value of this maximum stick deflection. It can be observed that in the lower error region, there are two groups of errors divided by the lines of 1.5° with a stick deflection not above 2.5°. Therefore, this has been chosen as the threshold above which an input can be called a RRE.

Table D.1 shows all runs in which the bank angle was increased to over 30° for all conditions involving a bank angle and over 0° for the conditions in which a level AI was shown. The runs shown between the two dashed lines are under the determined threshold. It can be seen that most of those inputs are made to the left (positive). Looking at the experiment design, where the pilot was slightly leaning to the right in order to reach the secondary task, it is more probable for the column to also be moved to the left when grabbing it as this is in the same direction the pilot moves to sit up straight again before giving an input.

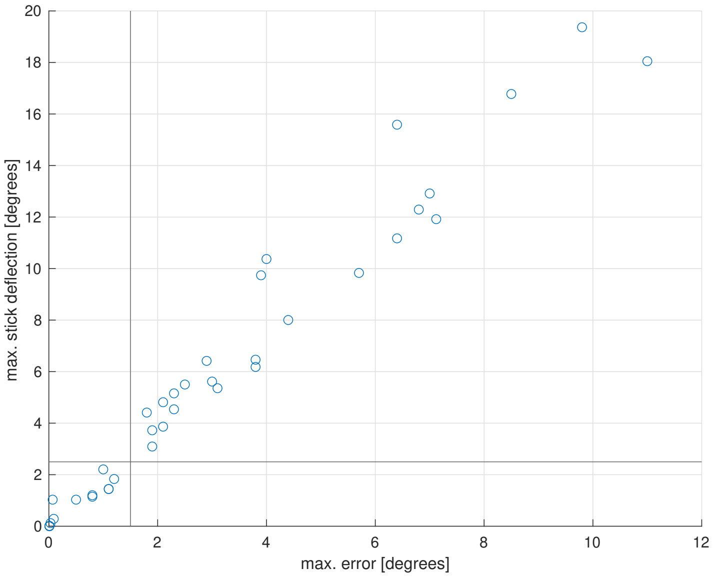


Figure D.1: The maximum error made against the maximum stick deflection.

Participant	Run Number	Scenarios	max. Bank [deg]	max. Error [deg]	Error Duration [s]	max. Stick Deflection Rate [rad/s]	max. Stick Deflection [deg]
6	10	MO	41	11	1.5	-0.21	-18.05
7	21	MM	38.5	8.5	1.2	-0.244	-16.78
13	11	MO	36.4	6.4	1.7	-0.16	-15.58
4	11	MO	37	7	1.4	-0.161	-12.91
5	3	MM	36.8	6.8	1.3	-0.165	-12.29
12	1	MM	37.12	7.12	1.3	-0.16	-11.92
6	17	NMT	35.7	5.7	1.3	-0.132	-9.83
6	25	MO	33.8	3.8	1.2	-0.094	-6.46
16	11	MO	32.9	2.9	0.7	-0.16	-6.42
11	10	NMT	33.8	3.8	1.3	-0.083	-6.18
6	28	NMT	32.1	2.1	1.4	-0.06	-4.81
5	10	MO	31.2	1.2	0.8	-0.04	-1.83
5	21	MM	30.07	0.07	1	-0.018	-1.03
2	22	MO	30.03	0.03	0.15	-0.014	-0.12
5	14	NMT	30.01	0.01	0	0	0.00
10	28	MO	-30.01	0.01	0.1	0.0005	0.00
2	32	MO	-30.01	0.01	0.1	0.0027	0.02
5	16	MM	-30.09	0.09	0.5	0.01	0.29
15	9	MM	-30.5	0.5	0.6	0.03	1.03
5	13	MM	-30.8	0.8	1	0.02	1.15
18	11	NMT	-30.8	0.8	0.4	0.06	1.38
5	18	MO	-31.2	1.1	1.2	0.021	1.44
5	19	MM	-31.1	1.1	1.2	0.021	1.44
4	2	MM	-31	1	0.7	0.055	2.21
5	5	NMM	-1.9	1.9	1.8	0.03	3.09
11	17	MO	-31.9	1.9	1.3	0.05	3.72
5	9	MM	-32.1	2.1	0.95	0.071	3.86
10	7	MO	-31.8	1.8	0.7	0.11	4.41
15	11	MO	-32.3	2.3	0.9	0.088	4.54
15	8	MM	-32.3	2.3	0.9	0.1	5.16
13	19	MM	-33.1	3.1	1.1	0.085	5.36
18	17	MO	32.5	2.5	0.8	0.12	5.50
4	6	MM	-33	3	0.7	0.14	5.61
5	7	NMT	-34.4	4.4	1.1	0.127	8.00
9	10	MO	-33.9	3.9	1	0.17	9.74
4	28	MO	-34	4	1	0.181	10.37
13	18	MO	-36.4	6.4	1.5	0.13	11.17
9	4	MM	-39.8	9.8	1.3	0.26	19.37

Table D.1: Information about the errors made (MM = ‘motion-matching’, NMM = ‘no-motion-matching’, NMT = ‘no-motion-turn’, ML = ‘motion-level’, MO = ‘motion-opposite’).

E

Secondary Task Performance

This appendix gives an overview of the influence of the secondary task on the perceived workload. The main goal of the task was to distract the pilots from the simulator motion and to slightly increase the workload. The overall workload was measured using the NASA TLX workload scale. The average scores for pilots, non-pilots, pilots who made an error and pilot who did not make an error are given in Table E.1. This table also indicates the weighting factor which is used in the calculation of the total workload. The scores are visualized in Figure E.1, from which is can be observed that (on average) experienced pilots rated their workload higher than novice pilots, and pilots who made errors also experienced a slightly higher workload than the other pilots.

Table E.2 on the next page shows all inputs made in the secondary task, divided over the two sets given and the three different tasks. Also the weighted workload ratings are shown. Besides the observations as indicated above, no relationship between the secondary task performances could be identified.

	Weighting factor	**Experienced pilots**	**Novice pilots**	**Pilots who an RRE**	**Pilot who did not make an error**
Mental demand	5	13.1	11.5	12.7	11.3
Physical demand	3	8.6	5.6	8.1	5.9
Temporal demand	1	9.7	8.4	9.4	8.3
Performance	2	8.4	7.3	8.5	6.7
Effort	3	12.2	11.5	12.1	10.8
Frustration	1	7.7	3.6	6.2	5.3
Total		10.8	9.0	10.4	8.9

Table E.1: (Weighted) subjective workload ratings.

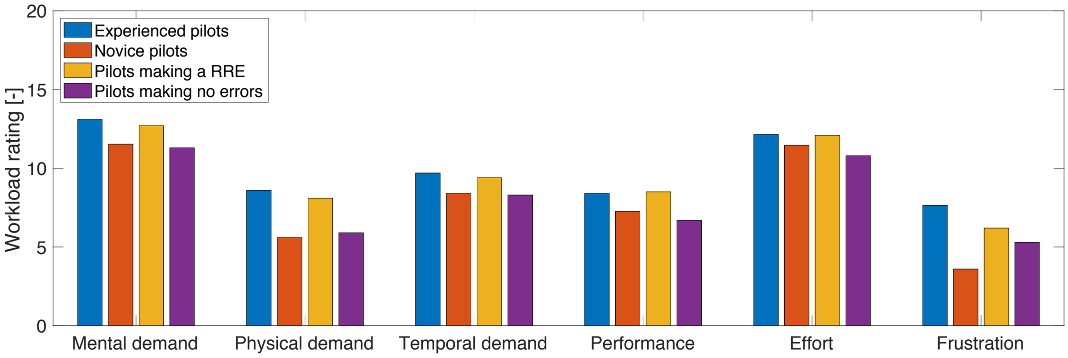


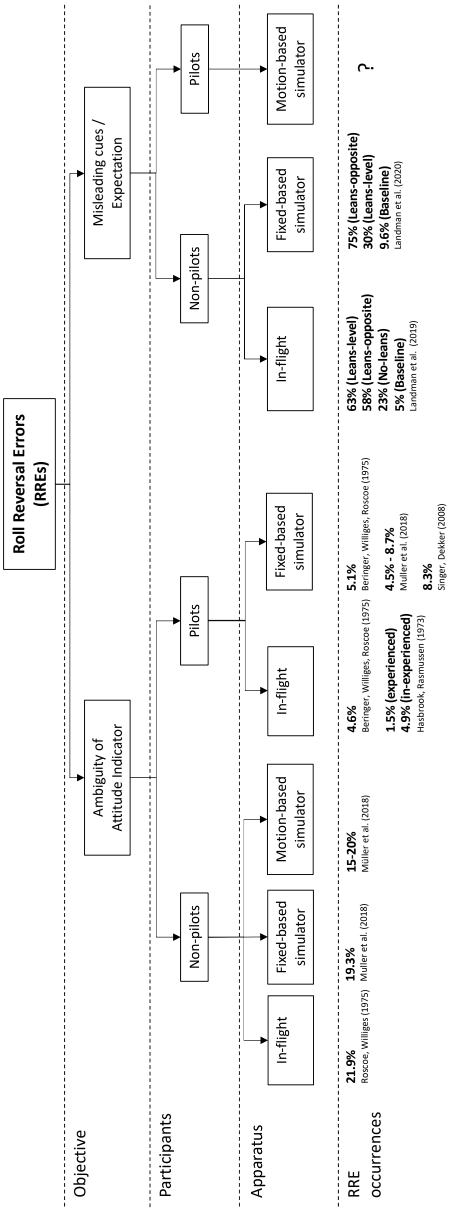
Figure E.1: Visualisation of the workload ratings for different groups.

	Set 1							Set 2						
	COM		RMAN	SYSM				COM		RMAN	SYSM			
	Correct inputs	Incorrect inputs	Inputs	Correct inputs	Incorrect inputs	Total	Workload	Correct inputs	Incorrect inputs	Inputs	Correct inputs	Incorrect inputs	Total	Workload
P4	25	23	51	30	2	131	13.93	47	5	58	44	0	154	13.67
P6	35	15	74	19	1	144	12.60	49	7	114	24	6	200	11.67
P7	36	16	77	21	4	154	12.33	44	8	77	29	0	158	9.47
P8	18	26	59	35	4	142	11.40	28	26	62	38	1	155	7.13
P9	29	17	101	33	1	181	8.80	48	4	102	36	1	191	4.87
P11	26	26	46	34	5	137	12.20	48	10	79	38	4	179	11.33
P12	47	9	72	10	1	139	10.20	34	22	123	49	2	230	10.13
P13	29	23	82	34	1	169	12.40	59	15	150	65	0	289	10.40
P16	n.a.	n.a.	78	47	1	126	11.53	n.a.	n.a.	93	46	0	139	7.60
P17	53	2	74	32	3	164	11.80	55	11	84	50	2	202	12.40
P1	30	12	141	37	1	221	6.83	51	5	304	49	2	411	6.57
P2	29	14	158	54	3	258	n.a.	58	0	165	55	0	278	9.67
P3	29	21	92	32	2	176	12.73	46	4	71	48	2	171	8.40
P5	30	24	105	40	0	199	12.27	49	5	92	44	3	193	10.27
P10	40	12	82	39	4	177	7.00	50	0	70	44	0	164	5.60
P14	45	5	85	39	3	177	6.73	50	14	108	52	2	226	6.60
P15	41	13	100	44	5	203	12.47	44	14	120	61	0	239	11.47
P18	52	4	79	34	1	170	8.67	52	18	103	63	1	237	10.13

Table E.2: Amount of inputs made by each participant in the three parts of the secondary task. The ’bold’ coloured participants are those who made an error. The table is split up in two parts; set 1 and set 2.

F

Overview of Research on RREs



G

Upset Prevention and Recovery Training Evaluation - Experiment Briefing

This document gives information regarding the experiment into the evaluation of upset prevention and recovery training. The experiment will be conducted in the Simona Research Simulator operated by the Control and Simulation division of the faculty of Aerospace Engineering of Delft University of Technology. The purpose of this document is to provide general information to the participant regarding the experiment, the task that will be performed and the general setup of the experiment.

Goal of the experiment

The goal of the experiment is to evaluate pilot handling performance in line with research done into Upset Prevention and Recovery Training.

Experiment task

The research looks into the reaction time of pilots after a moment of being out of the loop and in situations with a possibly higher workload. Most of the flight will be with autopilot engaged. The participant will be performing communication tasks and has to monitor system performance. Whenever the autopilot is disengaged, the aircraft has to be turned level as soon as possible.

Experiment procedures

During the experiment all tasks will be explained in more detail. Also, there will be a short practice run to get familiar with the Simona Research Simulator. In between the test runs there is time for a short break. The whole experiment, including briefing and breaks, will last approximately three hours.

Apparatus

The experiment will be conducted in the Simona Research Simulator (SRS), where motion will be enabled. The inside and outside view of the simulator can be found in Figure G.1 and G.2. You will be seated in the left seat of the cockpit and use a control column. Also, you will be able to control only roll whenever the autopilot switches off during the experiment and their input is desired. The throttle will be controlled by the autopilot during the entire flight.

Participant rights

Participation in the experiment is voluntary. This means that the participant can decide to stop with the experiment at any time.

Also, the data that will be collected during the experiment will remain confidential and anonymous and will be treated such that only the experimenter can link the results to a particular participant. This means that the results can not be traced back to you. By participating in the experiment you do agree that this anonymous data may be published. Lastly, we ask you to not discuss any details of the experiment with anyone until the complete experiment has finished. This prevents other participants to be biased.

To make sure you have understood all of the above, you are asked to sign an informed consent form at the end of the experiment briefing.

Figure G.1: The Simona Research Simulator.

Figure G.2: The flight deck.

H

Informed Consent Form

I hereby confirm that:

☐ I volunteer to participate in the experiment conducted by the researcher (**Annemarie van den Hoed**) under supervision of **Olaf Stroosma** from the Faculty of Aerospace Engineering of Delft University of Technology. I understand that my participation in this experiment is voluntary and that I can withdraw from the participation at any time, for any reason.

☐ I have read the experiment briefing (dated 05-09-2019) and I understand the experiment instructions. Furthermore, I have had all my questions answered to my satisfaction.

☐ I understand that in the experiment my primary task is to make sure the airplane performs a level flight and that I also have a secondary task to perform.

☐ I confirm that the researcher has provided me with a detailed safety and operational instructions for the Simona Research Simulator.

☐ I understand that the researcher will not identify me by name in any report or publication that will result from this experiment and that my confidentiality as a participant in this study remains secure.

☐ I understand that this research study has been reviewed and approved by the TU Delft Human Research Ethics Committee (HREC). I am aware that I can report any problems regarding my participation in the experiment to the researchers using the contact information below or, if necessary, the TU Delft HREC (hrec@tudelft.nl).

☐ I have been given a copy of this consent form.

☐ I accept that any audio recordings made during the experiment can be used when presenting the results

______________________	______________________
My name	Date
______________________	______________________
My signature	Signature of Researcher

I

Post-experiment Questionnaire Responses

All pilots were given a questionnaire after the experiment, before explaining what the experiment was about, containing questions that gave us a better understanding of the control strategy used. The answers given can be found below. In case the answer of a participant is not shown, that means that the question was not answered or the answer was simply ‘no’.

1. What do you think the experiment was about?
P1. Reaction time in a UPRT situation, when busy with another task.
P2. Reaction time to recover from an upset attitude.
P3. Startle and surprise reaction time.
P4. Recovery UPRT time.
P5. Reaction time during upset recovery including confusion and task division.
P6. Distraction by a task, during a stress situation.
P7. Influence of stress on cockpit tasks e.g. unusual attitudes.
P8. Reaction time in recovery vs. performing tasks.
P9. Division of attention, division of tasks, prioritising, efficient scanning in accordance with level of automation.
P10. Reaction time and task management.
P11. Concentration, task devision, prioritising
P12. Spatial desorientation.
P13. Reaction time to sudden unexpected situations.
P14. Response tim in UPRT, combined with tasks to distract you and influence the reaction time.
P15. Distraction during/before upset recovery and recovery reaction time and accuracy.
P16. Reaction time, compared to workload.
P17. It has to do something about scanning/reaction time.
P18. Testing of your reaction time during different levels of workload.

2. Did you notice something unusual with the motion during the experiment? If yes, what?
P1. At times the AI indication did not seem to correspond with the feeling from the simulator.
P2. Opposite motion compared to actual attitude.
P3. Sometimes counter-intuïtive.
P5. Seemed that motion was not always in line with cues on PFD.
P6. Something got a feeling of being in a bank/roll, but this didn’t show on the PFD.
P8. Sometimes during level flight I had a feeling the simulator was slightly banking.
P9. Yes, motion sometimes not consistent with onward attitude display (like SD) and slight roll tendency.
P11. Yes, felt motion, but no sign on the PFD. Feeling of motion with a level indication.
P12. Yes, counter-intuitive.
P14. Yes, sometimes the motion seemed to contradict the PFD.
P15. In some cases the motion was in opposite direction of the upset.
P18. Yes, sometimes the motion was not corresponding with the artificial horizon.

3. Was it sometimes more difficult to respond than at other times?
P3. Yes, half way I got confused one time.
P4. Yes, when there was a combination of multiple tasks.
P5. Yes, confusion between motion and indication on PFD.
P6. Yes, sometimes a ‘vertigo’ feeling. Sometimes busy with the tablet task.
P7. Yes, when multiple tasks appeared in short time under working stress, punching values in tablet for frequencies was more difficult.
P9. No (but I did notice a learning curve).
P12. Yes, the very first practice run.
P13. Yes, when tasks were separated with little time.
P14. Yes, recovering the upset sometimes led to missing second task.
P15. Yes, multiple actions in a short period of time together with flying task.
P17. Yes, but only for the task on the tablet.
P18. Yes, if the motion makes it feel like a left turn and the horizon shows a right turn it takes longer to respond.

4. Did you experience a situation in which your reaction time was lower? If yes, when?
P2. When distracted by the time of sidetask.
P3. When I continued the task on the tablet too long, before autopilot disengaged.
P5. When I was confused during the first session.
P6. When being busy on the IPAD.
P7. During a multitude of tasks, also when scanning circle was down due to attention to other tasks.
P9. Forgot inputs on tablet on occasions, due to prioritising.
P11. Sometimes, when busy with other tasks but steering had priority.
P13. When accomplishing a task during the upset.
P14. When a frequency change was coming up and attention needed to be divided. However, priority was given to flying task.
P16. Yes, when the secondary task needed more attention.
P18. If the horizon comes back whilst you are changing frequency it can take a moment longer to react.

5. Have you ever received spatial disorientation training? If yes, in what form?
P2. Static simulator training.
P5. Yes, during flight training (Single engine, Upset Prevention and Recovery Training.
P9. Yes, aerobatic and UPRT. Fast jet, optical illusions in combination with sensory illusions.
P10. Theoretically, not especially in a simulator or aircraft.
P11. Yes, simulator and aerobatic training.
P14. UPRT in aerobatics aircraft.
P18. Yes, I have had extensive UPRT training in simulators and in real aircraft with disorientation/vertigo training.

6. Have you ever experienced noticeable spatial illusions during flight? If yes, what?
P1. Rarely.
P2. Yes, during flight training.
P5. Once, vertigo.
P6. Yes, cloud leans, vertigo, leans.
P7. Once, leans.
P9. Yes, leans.
P10. Yes
P11. Yes, vertigo due to cloud layers.
P13. Yes, cloud leans.
P15. Yes, similar to what was experienced in the experiment.
P18. Yes, many. Mostly in simulator and on time in actual flight.

7. What would be your control strategy during such an illusion?
P1. Always focus on your instruments and specifically PFD.
P2. Refer to AI indication.
P3. Rely on the instrument.
P5. Engaging the autopilot to reset and gain right information again.
P6. Fly on instruments!
P7. Applying more input from the PFD.
P9. Priority on aircraft stabilisation by reference to (working) instruments, lockout false instrumentation and ‘seat-of-the-pants’ feeling, until control is recovered and ‘see-feel-steer’ are back.
P10. Observe PFD.
P11. Change controls to other pilot.
P13. Focus on instruments.
P15. Upset recovery technique using instruments.
P18. Trust your instrument, never what you feel.

J

Plots Reaction Times

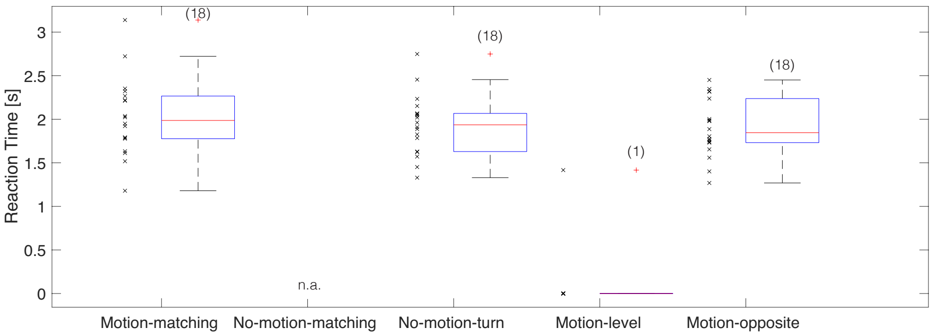


Figure J.1: Average reaction times from all participants divided over the conditions.

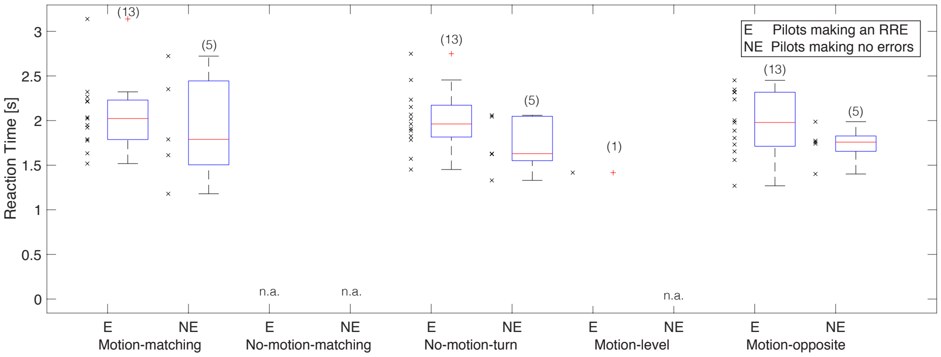


Figure J.2: Average reaction times divided over pilots who did and did not make errors.

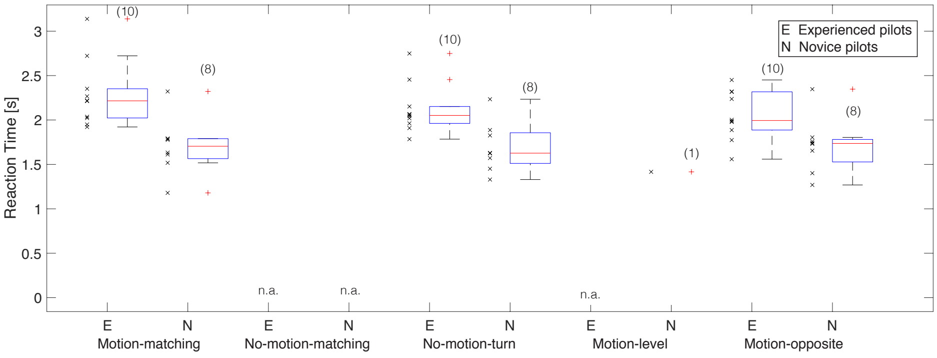


Figure J.3: Average reaction times divided over experienced and novice pilots.

K

Plots Learning Effect

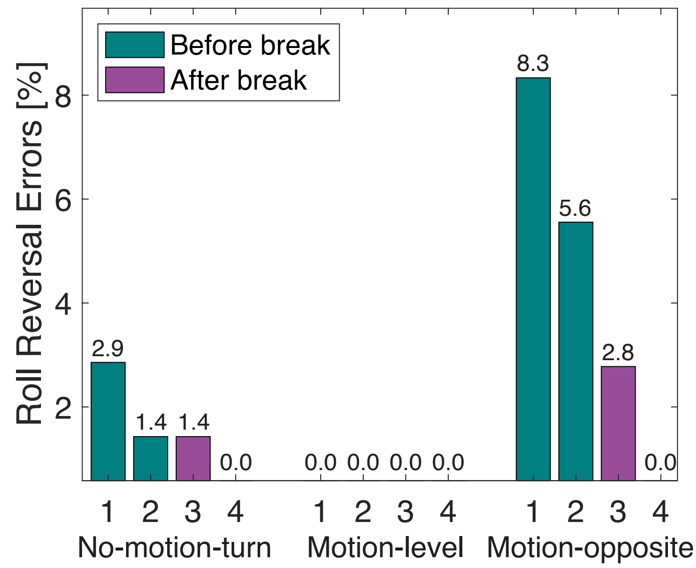


Figure K.1: The mean error rates for the three test conditions, divided over the four repetitions. Error percentages are calculated over the total number of runs for each condition in the experiment.

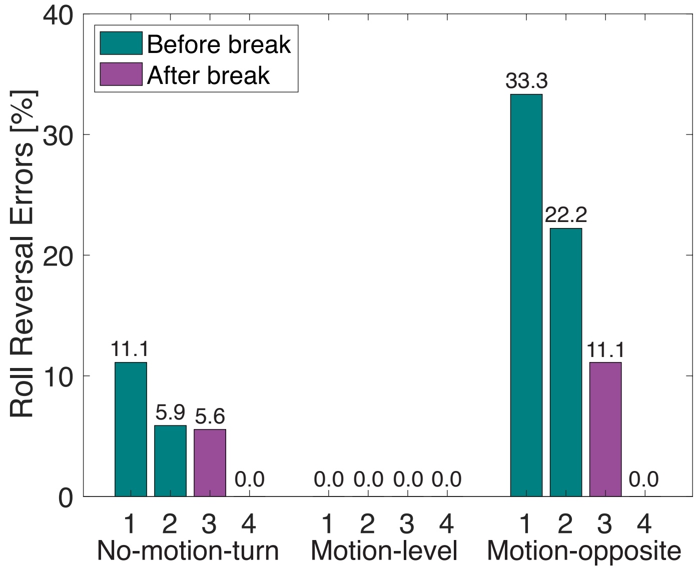


Figure K.2: The mean error rates for the three test conditions, divided over the four repetitions. Error percentages are calculated over the number of participants.

L

Results Preliminary Experiment

The ‘motion’ scenario from phase 1 of the experimental runs had to put the pilots in a certain frame regarding the aircraft attitude by changing their perception of the visual and motion cues given. To evaluate the effect of the designed man-in-the-loop scenario, a small preliminary experiment had been performed. This appendix shows the experiment design and results.

Participants and Instruction

Nine students from the Control and Simulation section of the Aerospace Engineering faculty of TU Delft participated in the experiment. This was the maximum possible amount due to time limitations. The simulator cab would be put in a certain position, after which their task was to put the simulator upright in one fluent motion shortly after the autopilot off alarm sounded. It should be noted that for all scenarios an input was desired with no AI and no outside visual, meaning that only vestibular and proprioceptive information could be used. A roll input done on the controls translated into a roll rate of the simulator. Or in other words, the bigger the input, the faster the simulator would roll. Making no input, thus thinking the simulator was already upright, was also a possibility. The participants were also working on secondary tasks consisting of flight planning exercises. These tasks were designed to distract the participants.

Scenarios

Using the following four scenarios more information about the working of the motion cues used to elicit the expectation of the bank angle could be obtained. All scenarios are explained briefly below, together which their influence to the experiment.

Slow-Motion

Figure L.1 shows a schematic representation of this scenario. The simulator slowly tilted to a 5.2° angle after which the outside view and AI disappeared. The simulator remained in that position for 30 seconds after which the alarm sounded and the participant had to decide to make an input in order to put the simulator upright again. With this condition the effect of the concepts of adaptation and visual dominance while being distracted was tested. The desired effect was for the participants to indicate no input was needed, meaning that in their perception the simulator had not moved.

No-motion

In this condition nothing happens apart from the outside view and the PFD disappearing as can be seen in Figure L.2. Here also no visual information is given for 30 seconds before the alarm sounded, making the run very similar to the ‘slow-motion’. This scenario was chosen as such to act as a baseline to the ‘slow-motion’ scenario, since it is the true case in which the simulator had not moved.

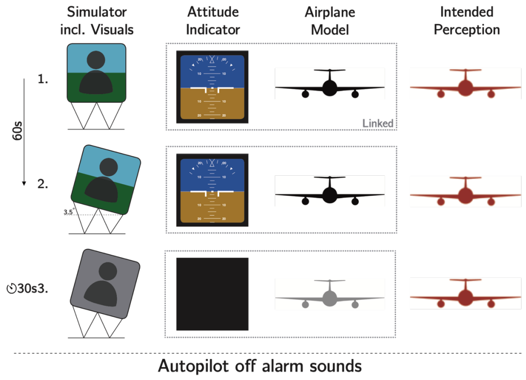


Figure L.1: The ‘slow-motion’ scenario

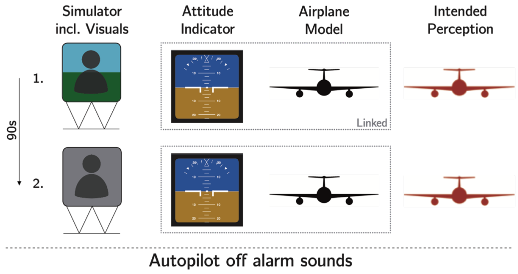


Figure L.2: The ‘no-motion’ scenario

Full-motion

The ‘full-motion’ scenario, as shown in Figure L.3, tested the effect of all motion cues given in the ‘motion’ scenario from phase 1 in the main experiment. The scenario is similar to the ‘slow-motion’ scenario, only now after 30 seconds of adaptation the simulator is put upright again with a fast and noticeable motion being above the human perception threshold. If the slow tilting for pre-positioning went unnoticed, which can be concluded from the results from the ‘slow-motion’ scenario, the motion should give the participant the feeling that the simulator is tilting to the other side than were it came from. In this case, an input back towards the original side it was tilted to confirmed the working of the motion cues.

Fast-motion

Finally, the ‘fast-motion’ scenario is depicted in Figure L.4. Here, first the outside visuals and AI disappear. This again remains like this for 30 seconds, after which the simulator is tilted to a 5.2° angle with the same fast motion as used in the ‘full-motion’ scenario.

This scenario acts as a baseline to the ‘full-motion’ scenario. In case the same inputs are given, this validates the working principle of the ‘full-motion’ cues.

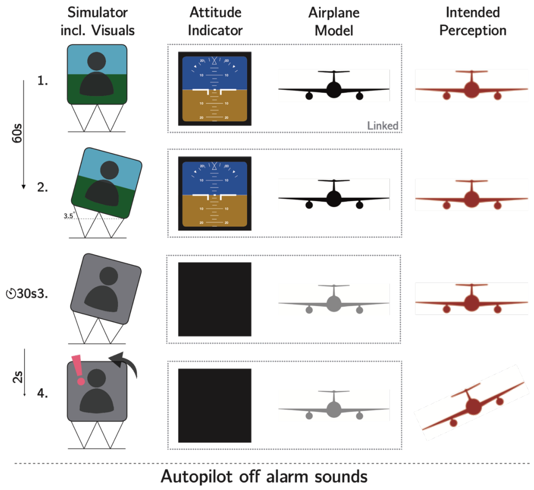


Figure L.3: The ’full confuser’ scenario.

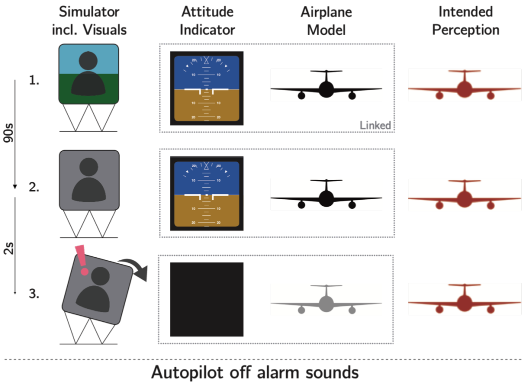


Figure L.4: The ’fast motion’ scenario.

Scenario Sequence and Procedures

After having received the briefing explaining the tasks to perform together with the safety briefing, the participants were seated in the left hand seat of the Simona Research Simulator. They received two sets of eight runs, leading to a total of 16 runs. In those 16 runs, each scenario was given four times, equally divided over the two sets. The scenarios including motion were done twice to the left and twice to the right. The order of the scenarios was randomized under the participants by using a Latin Square Matrix adjusted to the amount of participants that took part in the experiment. Each participant received turbulence on the vertical z-axis in one half and no turbulence in the other half of the test runs.

Measurements

For this experiment only the control input and the final simulator position were be measured. Together, this indicated the perception the participants had about the position of the simulator. Based on the answers given in a questionnaire which was filled out afterwards, the effect of the turbulence and their control strategy was determined. The participants were also given the opportunity to give comments and feedback on the scenarios or their performance and to share their experiences. With this information it could for instance be concluded whether the tilt angle was noticed during the runs or not.

Results

In the end, the preliminary experiment was performed in an iterative manner, where small adjustments to the experiment design were made between participants and evaluated later. To explain the results and adjustments made, the results are split into groups of participants. Each subsection given below shows an iteration of the experiment design and the corresponding results. All final positions deviating from the true upright position below 10% would be considered a correct input. This number is based on a rough estimation based on the average under- or overshoot in all runs.

Participant 1 to 4

The first four participants made no mistakes. On all conditions given, the correct initial input to put the simulator upright was given. Also most of the final positions fell in the range of 10% error.

Even though the ‘full-motion’ scenario did not seem to have the desired effect, some interesting conclusions could be drawn from the results of those first four participants. First of all, in 37.5% of the ‘slow-motion’ scenarios, the participants only put the simulator approximately half way back. Where in all the ‘no-motion’ and ‘fast-motion’ conditions the simulator was correctly put up straight. Some of the participants also indicated that you could feel the gravity factor at some point during the slow rotation. Or that they suspected that the simulator had turned slowly, and that the fast cue in the ‘full-motion’ condition gave a confirmation of this fact. This indicated that part of the tilt is masked but also that the tilt angle chosen might be slightly too high.

Most of this 37.5% appeared in the runs with turbulence enabled. Not all participants noticed the turbulence, but the ones that did indicated that this was in their eyes better to reach the goal of the ‘full-motion’ scenario. Moreover, one participant said, “*the turbulence was not very noticeable, it could have been more*”.

Based on those results it had been decided to decrease the tilt angle. The next participants received varying angles of 3.0°, 3.5° and 4.5° to evaluate the best angle for both the slow as the fast motion. Also, turbulence was provided in all conditions, only now with a gain of 1.0 (the original setting) in one of the sets and a gain of 1.25 in the other.

Participant 5 to 7

With the changes made, the following participants were given the same four scenarios. The correct initial responses were given in all the ‘fast-motion’ conditions, and for the ‘no-motion’ scenario a few small inputs were sometimes given but these, being only small, could be neglected.

However, in the ‘half-confuser’ scenario now 78% of the runs was responded to wrongly. Either a very small correction was made or no input was given at all. In the ‘full-motion’ scenario there was an error rate of 72% using the new settings. Both indicating great results concerning the effect of these motion cues to generate the ‘leans’ cues in a hexapod simulator.

Turbulence combined with a slightly smaller tilt angle prevented the participants from clearly distinguishing tilting. Also, the fast motion cues were still rated as ‘clearly noticeable’ by all three participants in all conditions, despite the smaller angle. Furthermore, the participants seemed more distracted by the secondary task than before and they showed confidence in the PFD. The latter was concluded from the call-outs they made when suspecting a tilt. The PFD would be consulted after which statements were made like, "*still flying straight*".

Figure L.5 and L.6 show the results from one of the participants. This participant received non of the 'no-motion' scenarios, since it was deemed more important to gather more information about the working of the motion scenario in these runs. The peak in the same direction of the motion cues coming directly after the autopilot off alarm (black dot) indicated an erroneous input, pointing to the correct working of the motion scenario. From the figures it can clearly be seen that this input was given for all of the 'full-motion' scenarios given, without a clear learning effect. This shows that the concept works, as long as the participant is not specifically focusing on the motion but just responds intuitively as was the case for this participant.

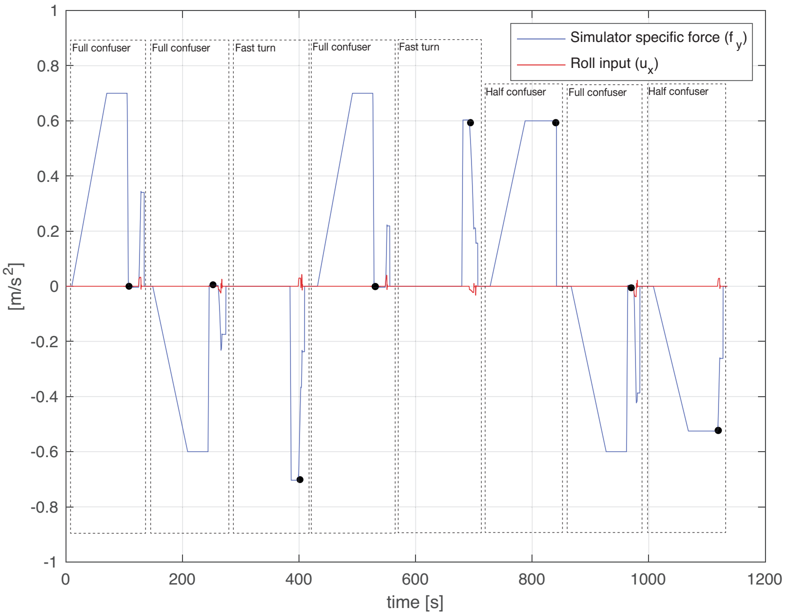


Figure L.5: Specific force of the simulator (f_y) and the participant's input - Participant 6, run 1. The black dot indicates when the alarm sounded. Turbulence gain = 1.00.

Participant 8 and 9

The final two participants showed different results, where one made a few small errors and the other none. The most interesting to learn from this is, '*why did they suddenly give the correct inputs?*', since nothing had changed regarding the experiment design comparing to the previous runs.

First of all, this was because both participants suspected something would happen with the motion. They gave clear call-outs with every little bit of motion they felt, also the slow motion cues. They also made the decision to discard the information given by the outside visual and PFD, which made them even more focused on the motion.

The results for one of the participants is shown in Figure L.7. In the first 'full-motion' scenario, it can be seen that a clear erroneous input was given. Only then the participant mentioned that "*the motion can not be trusted*", after which a clear learning effect could be observed.

Conclusions

Initially, the goal of the preliminary experiment was to evaluate the motion cues given in the 'motion' scenario from phase 1 of the experiment runs in the main experiment, but this experiment resulted in a more iterative process towards the best angles and turbulence setting to be used.

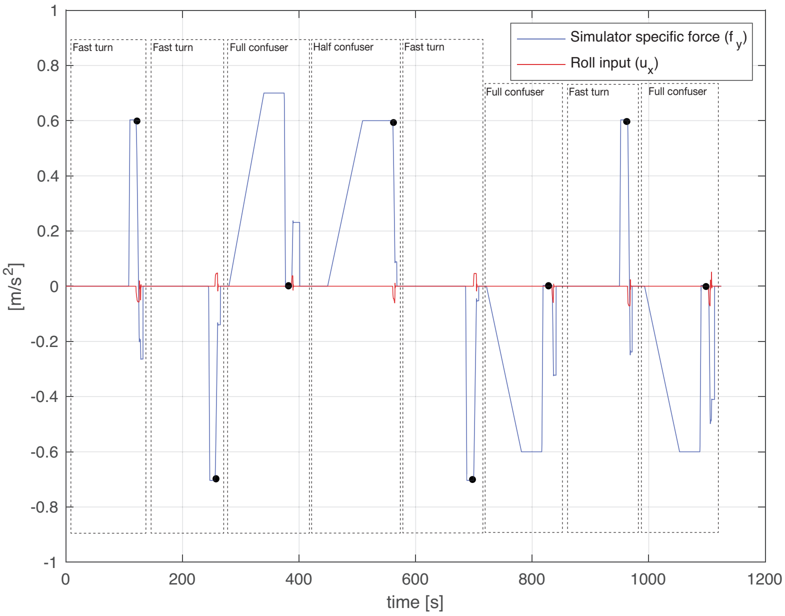


Figure L.6: Specific force of the simulator (f_y) and the participant's input - Participant 6, run 2. The black dot indicates when the alarm sounded. Turbulence gain = 1.25.

The 'full-motion' scenario indeed only had the desired effect when also the small tilt angle was not noticed. It can therefore be concluded that it is important to avoid that participants focus too much on the motion cues. This can first of all be realized by giving a good briefing, as the pilots will only know what is being told about the experiment. Consequently, their way of thinking about the experiment can be manipulated in a smart way as to make them not paying attention to the slow tilting of the simulator. Secondly, there has to be adequate distraction through the secondary task to deviate the attention from the motion. Also, having this task at one specific side of the participant is desirable. Any felt tilt could be related to already sitting slightly leaning towards this side.

From the results it had been decided to use an angle of 3.5°. This angle was mostly left unnoticed during the slow motion and the fast motion cues remained giving a sufficient rolling sensation to lead to the illusion of leaning to the other side when this motion stopped. Adding turbulence only showed a little effect on the masking of the tilt and was less than expected. It has been decided to use a gain of 1.25 for the turbulence, since no negative effects presented themselves either.

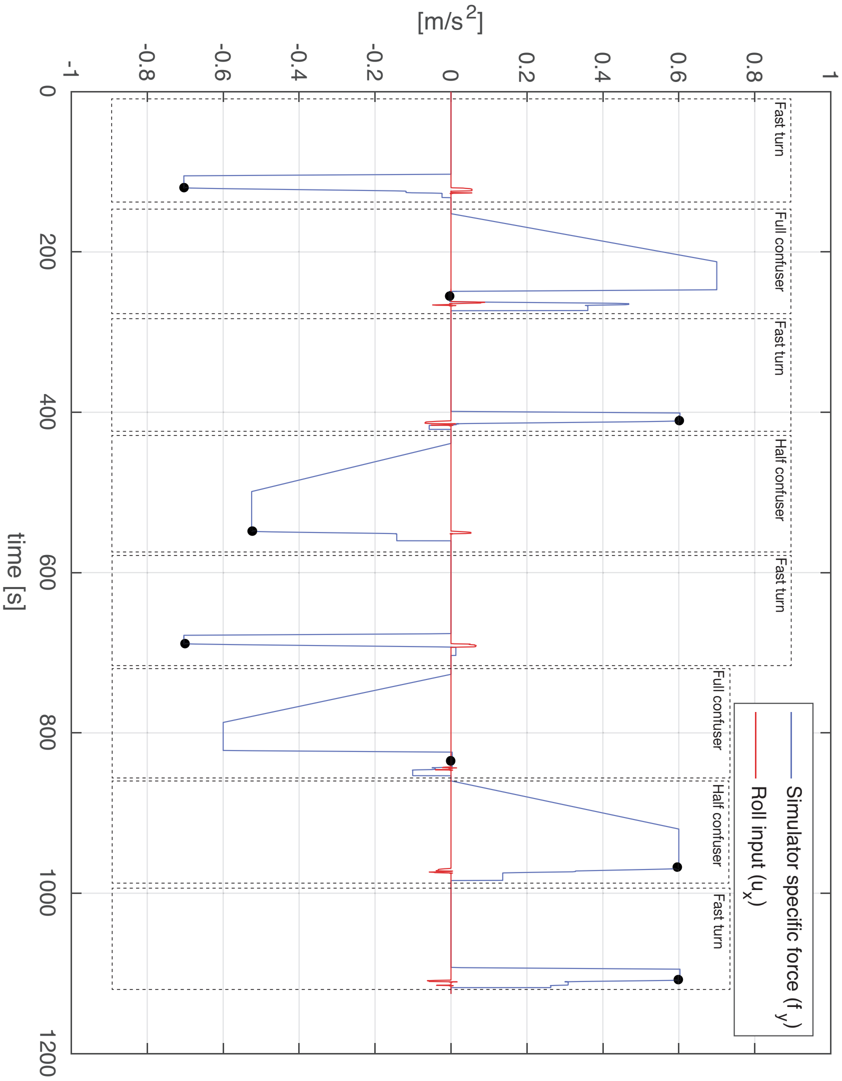


Figure L.7: Specific force of the simulator (f_y) and the participant's input - Participant 9, run 1. The black dot indicates when the alarm sounded. Turbulence gain = 1.00.