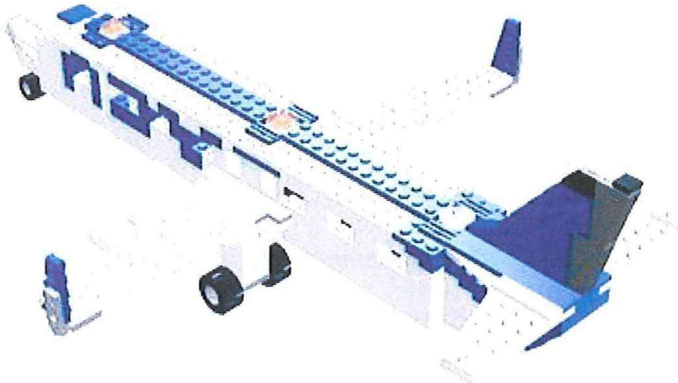
Project Samuraï Pilot

Exploring the Use of Heart Rate Variability Biofeedback as a Stress Management Technique for Commercial Pilots

"To boldly go where no one has gone before" – Captain Picard



D.P. Gorter, TU Delft – 4079159

C.D. Jaeger, University of Leiden – S1437038

August 11th, 2014

TECHNISCHE UNIVERSITEIT DELFT
FACULTEIT LUCHT- EN RUIMTEVAARTTECHNIEK
CONTROL & OPERATIONS

MASTER OF SCIENCE THESIS

Exploring the Use of Heart Rate Variability Biofeedback as a Stress Management Technique for Commercial Pilots

Daniël Pieter Gorter

Geboren te Assendelft, 22 Februari 1983

Ter verkrijging van de graad van ingenieur

in de lucht- en ruimtevaarttechniek

aan de Technische Universiteit Delft

op gezag van de leden van de afstudeercommissie

in het openbaar te verdedigen op 19 augustus 2014 om 10u00

THIS PAGE IS INTENTIONALLY LEFT BLANK

Project Samuraï Pilot

Exploring the Use of Heart Rate Variability Biofeedback as a Stress Management Technique for Commercial Pilots

Daan Gorter

BSc Aviation Studies, 2009

Amsterdam University of Technology

Aerospace Engineering Masters

TU Delft – 4079159

Supervisor: Prof. Dr. Ir. John Stoop

Courtney Jaeger

BSc Psychology, 2010

Exeter University

Applied Cognitive Psychology Masters

Universiteit Leiden – S1437038

Supervisor: Prof. Dr. Guido Band

Afstudeercommissie

Prof. dr. ir.	J. A. A. M. Stoop	TU Delft
Prof. dr.	G. P. H. Band	Universiteit Leiden
Ing.	R. J. den Hertog	NG Aircraft
Msc.	J. F. W. Mohrmann	NLR
Dr. ir.	C. Borst	TU Delft

NLR

EPST European Pilot Selection & Training

HeartMath®

UNIVERSITY OF AMSTERDAM

PREFACE

Aircraft accidents make it increasingly important that new avenues into human factors research are investigated. When a new method such as the Samuraï technique – as has been investigated in this thesis - can be shown to improve the behaviour in critical situations, that could really help. It is also clear that something needs to be done because the accidents over the last 10-15 years have shown that pilots rely too much on flight deck automation. By how much the Samuraï technique may improve cognitive behaviour in such situations is as yet unknown.

There are several arguments to underpin the above, these are:

1. Over the last 20 years, changes to pilots training have been made, but these were of an evolutionary nature, fundamental changes have not been implemented.

2. Training requirements are developed between the authorities and the airlines with virtually no influence from the manufacturers, except for maybe the latest aircraft such as 787 and A350.

3. The number of 737 & A320 aircraft is ever increasing, see http://www.nlr-atsi.nl/fast/aoc/aoc_280.html - to become 2/3 of the worlds commercial jet fleet by 2025. Given these numbers, it is unlikely that Boeing or Airbus will change their flight decks, even if design changes would make sense. Rather, they will continue to refine procedures and training as they have been doing now quite successfully over the years, see http://www.nlr-atsi.nl/fast/aoc/aoc_013.html

4. It follows that pilots are there to stay, and hence human factor impact on operations. To label any accident "pilot error" would really be a major mistake, because this essentially means that the aerospace community has simply not found what went wrong between the human and the machine.

It is recommended to perform follow on testing with dark cockpits like A320, F100, etc. because this is the way to the future.

Rudi den Hertog

Netherlands Aircraft Company

F-120NG Chief Engineer

ACKNOWLEDGEMENTS

This research would not have been possible without the help of many institutes and individuals. We want to thank the people who supported the start of the project, by listening in and thinking with us. They have helped define and design the project, without them it would not have succeeded. Gratitude goes out to prof. dr. Jan van Strien of the Erasmus University who helped with the choice of apparatus and dr. Andries van der Leij who helped structure the experimental design.

Going back to the start, HeartMath's research is what inspired the project setup. Dr. Rollin McCraty, Ph. D., who had conducted most of this research, provided support and encouragement for us. The Institute of HeartMath is an international non-profit research and education organisation, whose main aims are to help reduce stress and self-regulate emotions. HeartMath uphold the intelligence of the heart, how it is connected to the brain and the importance positive emotions are for our cognitive performance, physical wellbeing, creativity and interpersonal qualities. In setting up and conducting the research, Jackie Waterman from HeartMath provided calm, reassuring advice about the apparatus used during times of immense stress. During the intervention, Robert Erdbrink and Renzia from HeartMath Benelux were key players by providing the biofeedback training and the emWaves. In comprehending the jungle of statistics, Mike Atkinson helped us with processing all the heart data.

After the set up, we needed pilots to participate, and simulators to use. We were extremely lucky to meet with Eric Duijkers, MD of EPST. He helped us motivate the young pilots to sign up to the research and provided a state of the art simulator for us to use for testing. His lovely office assistant Angela van Ee helped us with the organisation of the simulator planning, to schedule in all participants. Martijn Niekerk of Multi Pilot Simulators helped us with the recording of the simulator data and managed to tweak some failures to our requirements. Fons de Leeuw has been the messiah of data analysis, helping with programming the code for the protocol.

We could not have done this project without the incredible help of the student instructors, Jacir Soares de Brito, Rick van Os, Wouter Enders, Jeroen van Leeuwen, with special thanks to Casper van Wijngaarden who was always there when we needed him! Alexander van Beek, Niarchilion Lopez, Cas Kuiper and Fabian Bleeksma were present from the beginning, for rigorous test flying of the scenarios and as confederates to the study as pilot monitors. Other pilot monitors who were undeniably an essential part of the successful execution of our large scale project were Jolmer van Brakel, Jeroen Arets and Sigourney Segers. Thank you! Not to forget, we want to thank all of the participants who invested six weeks of their time to attend and practice the training interventions, and sat through 2x three hours of testing at EPST by us.

The link between the research and the industry was provided by Rudi den Hertog of Next Generation Aircraft. For the cognitive test iPad application, we received support from Andries' partner Ilja Sligte, from the University of Amsterdam. Our professors John Stoop and Guido Band helped us through difficult times with moral support and in depth academic guidance. We feel honoured to have had them by our sides. Without the help of professor Rinze Benedictus, we would not have been able to complete this project, he helped us tremendously with the acquisition of equipment.

From the Dutch National Aerospace Laboratory (NLR) scenario support by Freek Mohrmann and Edzard Bolard was much apreciated. Henk van Dijk guided Courtney through her time at NLR. Friends and family played an important role, not just in moral support, but also during the long days of testing. Mrs. Gorter especially, who fed us well, keeping us energised and motivated, thank you. Lastly, we would like to thank the volunteers who proof read the thesis.

THIS PAGE IS INTENTIONALLY LEFT BLANK

ABSTRACT

To counter the detrimental effects a high workload emergency situation can have on a commercial pilot's flying performance, psychophysiology and overall cognitive profile, biofeedback training of heart rate variability (HRV) was tested as a stress management technique. This technique teaches self-regulation of one's heart rhythms, which can lead to a calmer, more rational reaction to a stressful, high workload scenario, by reducing sympathetic nervous system activation. It was hypothesised that after the biofeedback training, participants would demonstrate improved HRV, flying performance, and an improved overall cognitive profile, compared to the control group. The training was administered to a group of Dutch pilots with ATPLs over a period of six weeks. Measures were taken before and after the training. Results indicated that in its current form, the biofeedback training did not improve the experimental group's cognitive, flying or HRV measures over that of the control group. Limitations to this study include the little amount participants in the biofeedback group practiced the technique taught to them and the lack of stress the scenarios flown induced. This explorative research serves as a baseline design for future research in the field of psychophysiology in aviation.

Keywords: biofeedback, commercial aviation, heart rate variability (HRV), self-regulation, cognitive improvement

TABLE OF CONTENTS

ABBREVIATIONS

AFT: Autogenic Feedback Training

ANOVA: Analysis of Variance

ANS: Autonomic Nervous System

ATC: Air Traffic Control

ATPL: Air Transport Pilot license

BEA: Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile

Deg: Degrees

ECG: Electrocardiogram

EPST: European Pilot Selection and Training

FAST: Future Aviation Safety Team

GSR: Galvanic Skin Response

HF: High Frequency

HPA axis: Hypothalamic-pituitary-adrenal axis

HR: Heart Rate

HRV: Heart Rate Variability

Hz: Hertz

IBI: Inter-beat-interval

ILS: Instrument Landing System

LF: Low Frequency

LF/HF: Low Frequency to High Frequency ratio

Ln LF/HF: Natural Logarithm Low Frequency to High Frequency Ratio

MPS: Multi Pilot Simulators

NASA: National Aeronautics and Space Administration

NM: Nautical Mile

PM:	Pilot Monitor
PSNS:	Parasympathetic Nervous System
PTSD:	Post-traumatic Stress Disorder
QRH:	Quick Reference Handbook
RMS:	Root Mean Squared
RSA:	Respiratory Sinus Arrhythmia
RT:	Reaction Time
SI:	Student Instructor
SID:	Standard Instrument Departure
SKR:	Skill-based, Rule-based and Knowledge-based action
SNS:	Sympathetic Nervous System
SPSS:	Statistical Package for the Social Sciences
WM:	Working Memory

THIS PAGE IS INTENTIONALLY LEFT BLANK

1 INTRODUCTION

The research objective of this explorative project is to assess the effectiveness of biofeedback as a stress management technique for pilots in time critical, high workload situations. The methodology could provide the airline industry with a tool to create *Samuraï like* pilots (i.e. pilots with mental skills of martial arts masters). The focus in aviation safety should not be exclusive to preventing accidents from happening again. Instead of looking at a linear relation between problem and solution, a more holistic approach is required, where anticipation and adaptation are basic notions. The industry should not restrict itself to rational formal decision-making and optimisation of utility functions and trade-offs, but fortify the entire scope of actor performance. New domains of interest are dealing with emotions, perceptions, awareness and unconscious responses to time critical situations.

It has been argued that more than 95 percent of human behaviour arises subconsciously, whereas less than 5 percent is conscious (Tiggelaar, 2010). Cognitive performance enhancement through mental attention training programs could be an effective approach to increase control over the subconscious processes.

"Consciousness is not to be identified with any particular perceptual-cognitive functions such as discriminative response to stimulation, perception, memory, or the higher mental processes (...) All of these functions can take place outside of phenomenal awareness. Rather, consciousness is an experiential quality that may accompany any of these functions." – John Kihlstrom (1987).

1.1 Aviation Background

1.1.1 Research Fields

Up until today, extensive research has been performed in the three fields related to cockpit dynamics; the field of cognitive engineering (Hollnagel, 2003), aircraft design and engineering (Borst, 2009) and the field of computer programming and algorithms. All aim for the optimum human machine interaction, but not many researchers are bridging the gap between the three. With the engineering perspective at its core, immersing into the realm of aviation psychology through cognitive psychology and psychophysiology, this research aims at bringing these worlds together. As such this project is adding a new integrative perspective to existing and established scientific domains and disciplinary interests in man-machine interface design.

Aircraft operations include complex interactions between man, machine and the environment. This research is limited to the human operator, specifically the mental abilities and qualities. The 'classical' design school initially developed the flight dynamics and aircraft performance, followed by programs to train pilots the skills and competences required for safe operation of the aircraft. A new design school with a human centred approach, where research into the interaction between man and machine cannot stop with training and grading of pilots, cockpit design will play an essential role. The future of aviation safety depends on the collaboration of academics, researchers and pioneers in the fields of accident analysis, cognitive psychology and aircraft engineering.

1.1.1 Future Outlook

With the exponential advancement of electronics and big data, it seems inevitable that the increase in automation will continue, affecting man-machine interfaces throughout the aviation industry. The consequences for the human operator can only be predicted to a certain extent, since the complexity of ultra-safe systems has evolved beyond full scale human comprehension. Currently, some definitions of the most agreed upon flight crew automation issues are: "Unanticipated situations requiring to manually override automation are difficult to understand and manage, create a surprise or startle effect, and can induce peaks of workload and of stress" and "Basic manual and cognitive flying skills tend to decline because of lack of practice and feel for the aircraft can deteriorate." (European Aviation Safety Agency, 2013, p. 8). These statements suggest that the increase in automation can have an adverse effect on the level of control for the human operator.

In the near future new Air Traffic Control (ATC) systems, such as NextGen in the USA and SESAR in the E.U., may introduce new and unforeseen hazards (Jacobson, 2010). With respect to evolving aircraft and airspace design, hazards due to insertion of the human operator in these increasingly complex environments cannot be predicted. Air traffic is expected to double over the next twenty years, whereas the economic operating conditions are tighter than ever before. Leaning (Forrester, 1995) airlines into low cost carriers, is an ongoing trend. The result being, that pilots are forced into the human and operating envelope boundaries, with possibly detrimental consequences. The operating envelope defines the maximum capabilities within a system or human, in which they can safely operate (Amalberti, 1999). Less sleep allowance, less fuel on board, maximum regulatory working hours and reduced training (legal minima are much less than standards used by large carriers) are common characteristics of 'leaning' an airline operation. These factors will likely be pushing the limits of human performance in an already overly complex environment.

1.1.2 The Need for Flexibility

Since we have crossed the barrier of ultra-safe systems in the commercial aviation industry (Amalberti, 2001), we need to investigate new concepts and their subsequent measures to improve performance. The predicted growth in air traffic leads to more complex situations, where actors operate closer together (Amalberti, 2001). According to a Eurocontrol Safety Whitepaper (2013): "we must accept that systems today are increasingly intractable". Within an intractable system, the operating principles are partly known and these principles change before they have been defined or even identified. This speed of changes makes it cumbersome to design a direct cause-effect related safety measure, test it and incorporate it, since the measure may become obsolete even before it is introduced. Tractable systems can be modified during their development, whereas the environment of intractable systems such as the aviation industry, changes more rapidly than the design and certification process can sustain. A new approach, where flexibility is key, should be adopted in order to improve or even maintain our current safety standards.

Ferris, Sarter and Wickens (2010) suggest that the technical aircraft design process can be improved significantly, but as Jacobson (2010) rightfully mentions, improving safety performance through optimising the technical design is an evident but long term solution. We need to look at short term implementations of performance enhancement, whether it be machine or man. Since the aviation industry, one of the heaviest regulated in the world, does not allow for swift adaptations in the technical design process. We should start looking for alternative methods to increase safety levels. Apart from looking at safety exclusively in the technical design, we should improve overall system resilience (Hollnagel, 2011). This means creating a system which has the capabilities of operating outside of its regular envelope to a safe restoration of its flight path without stalling or crashing. Within this system, not only the machine, but also man plays a vital role. If not, the most vital...

1.1.3 Loss of Control

For now and maybe forever, pilots are required to bridge the unexpected with the expected (Dekker & Woods, 2002). Loss of control accounts for the largest part of fatal aircraft incidents. Up until today an estimate of 70% of incidents is claimed to be attributed to human error (Boeing, 2013), either at the sharp end (pilots, mechanics, ATC) or at the blunt end (design, legislation). When incorporating all factors of human error, this rate may even be higher, close to 100%. Environmental factors and bird strikes are the only exceptions of aviation accidents.

Large research facilities such as the National Aeronautics and Space Administration (NASA) describe the causes of loss of control. Within NASA, Jacobson (2010) quantified that between 1999 and 2008, 38 out of 56 of the loss of control incidents were pilot- or human induced, whereas 9 out of 56 were environmentally induced and the remaining 9 were system induced. The general understanding in the industry is, that these human errors cannot be prevented. The notion of 'error' is a basic human performance requisite, varying across individuals, based on skills, competences, experience and operating conditions. The common view is that man, and not machine, is the problem. In particular, their dynamic interaction in a variety of operating conditions is the critical issue. Rasmussen and Vicente (1989) therefore, have expanded the rational decision making modelling by incorporating the operating environment in the decision making process, thus identifying the 'ecological' environment in this process. According to reports by Airbus SAS (2005 – 2007), eliminating human error is not the solution. Aircraft should simply be better designed, more automated and optimised in order to reduce the chances of human interference.

1.1.4 Human Error Research

A case study of five major incidents over the past year identified incorrect pilot response, due to a possible startle effect and/or automation confusion. Air France Flight AF447, Qantas Flight QF32, Turkish Airlines Flight 1951, Colgan Air Flight 3407 and US Airways Flight 1549 can be distinguished by disaster and survival, but more importantly by (incorrect) pilot responses to unforeseen time critical situations. The possible causes for the crash of AF447 can be traced back to systems engineering, lack of high altitude stall training and inappropriate pilot input (Stoop, 2013a). Especially the latter, a possible cause for which many aviation professionals cannot find a logical reasoning, is of concern for this research. The question remains on why the co-pilot held the control stick back continuously.

Not just during the last decades, but since the start of aviation, human error has been the main contributor of aircraft accidents (Hobbs, 2004). Due to the work of Kahneman (2011) and Slovic (1999), intra-individual aspects of operator performance in decision making, have gained interest in industry in advanced modelling and understanding of the man-machine interface and subsequent engineering design solutions. In the approach of man-machine-interfacing there is emphasis on improving procedures, training, the design and upset recovery but no focus on psychophysiological control through anticipatory mental training for pilots. Errors are inevitably made and approximately 90% of errors are recovered without serious consequences (Amalberti, 2013). Error recovery is the best predictor of performance since it indicates a high degree of flexibility of the human operator during detection and recovery. Thus, if error detection and recovery can be improved, the performance of the human operator can be improved which will increase system resilience to external disturbances (such as triggers for loss of control).

Cognitive psychologists (Rasmussen, 1983, Reason, 1990, Dekker, 2004, Hollnagel, 2011 & Kahneman, 2011) have examined the thought process of pilots at a rational level.

However, as Jensen (1997) mentions, little emphasis can be found on perception and cognition. Pilots' reactions to emergency situations from a psychophysiological response perspective (startle and panic) has received little scientific attention. More importantly, little research has been done in the field of pilot training, whilst the aviation industry is one of the most controlled, regulated and standardised industries in the world. According to Dekker & Johansson (2000), basic civil aviation flight training "has not changed in any fundamental sense since the fifties" (p. 83).

In its final report of 2012 on the AF447 crash, the French Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile (BEA) concluded that the combination of ergonomics of warning design, training conditions, and recurrence training process did not generate expected behaviour and showed limits of current safety models (Stoop, 2013b). One of the conclusions was that pilots should be trained for high altitude manual flying, a skill which is not incorporated in the standard training procedures for commercial pilots. The conclusion that behaviour was not as expected and that current safety models show limits, may indicate that our conventional way of thinking is no longer sufficient for the increasingly sociotechnical complex aviation environment.

Considering AF447, the BEA recommends more training in extreme circumstance, e.g. training high altitude stall. This could be a useful addition to pilot's toolbox, but how does this prepare him for the next, not yet occurred, incomprehensible event? It is a reactive response to accidents, one that has brought the industry to its current levels, but may not be enough for the increasingly complex and dense environment. Techniques that can help pilots maintain full use of their executive functioning, may very well be an effective approach. Techniques that allow pilots to reduce the emotional charge, which can be experienced during novel life threatening situations, in order to maintain full control over their rational thinking.

According to Bor and Hubbard (2006), the role of the commercial pilot has evolved from flight performance specialist to social dynamics crisis manager. Stressors have changed over the past decades, requiring a new selection process. Deficiencies in human behaviour has been labelled by human factor researchers as loss of situational awareness, and complacency, which are abstractions that cannot be observed or measured. Bor and Hubbard (2006) mention that in order to maintain aviation safety levels at triple the amount of current traffic, as expected by 2020, only a significant reduction of human error is capable of achieving this.

1.1.5 Skills, Rules and Knowledge

Rasmussen (1983) identified three categories of pilot behaviour and made a model; skill-based, rule-based and knowledge-based action (SRK), respectively defining basic flying skills, procedural operations and determining through sound judgment whether given rules apply in a specific situation. Skill and rule-based tasks have been replaced by automation, but knowledge tasks such as bridging the gap between reality and unexpected non-normal (emergency) situations. Although of great value, automation architecture has progressed past the point of understanding by the operator (i.e. the pilot), sometimes creating a 'surprise' response. This surprise can have catastrophic consequences on the rational thinking process as described.

From Rasmussen's (1983) model, an extra layer seems to be missing; the physiological response and its connection with cognitive impairment. Additional layers beyond the skills, rules and knowledge level in this model should be defined, enabling the transition from compliance with the skill, rule and knowledge based actions, towards dynamic adaptation and recovery from such a 'surprise' reaction mode. As stated by Rasmussen and Vicente (1989) at the introduction of ecological interface design: ecological design is a concept in which the focus should be on the control of the effect of errors rather than on the elimination of errors per se. Introducing ecological interface design principles does not seem to suffice to fully cover this transition.

The current project expands the control levels to an additional level: the reflex level in order to cope with 'surprise' reaction modes. As such it replaces the notion of managing 'error' by the notion of adaptation and recovery from undesirable mental modes, in particular the reflex mode in cognitive decision making and operator response actions. Such an approach represents an intra-personal perspective on dealing with operator performance. In addition, another additional layer beyond the SRK level is taken into consideration by Mohrmann (2013), at the level of how crew decision making coordination and cooperation can be supported in case of failure of automated systems. Such a modelling expands the Rasmussen SRK model regarding rules and knowledge towards automation attitude and teamwork, under safety critical conditions and the trust operators can have in automation. Such an expansion expands the SRK model to social/team situational awareness issues and represents an interpersonal perspective on multiple crew performance (Mohrmann, 2013). For the purpose of this study, the model description has been added for theoretical contextual completeness. The main focus of this study was the identification and control of stress, as endured by commercial airline pilots.

According to the Future Aviation Safety Team (FAST): "Poor understanding of which operation the flight deck automation is commanding the aircraft to perform, has the potential to increase the stress and fatigue levels of the flight crew. This can have an adverse effect on the decision making process." (FAST, 2006, p. 8). Required pilot performance increases in high workload situations, especially during emergency situations and even more within highly automated aircraft. Through this the human cognitive operating envelope (Amalberti,

1999) is exceeded frequently, usually without, but sometimes with serious lethal consequences.

1.2 Psychophysiology and Stress

1.2.1 Pilots and Stress

The human operators of aircraft are a vital component of safe flying. Automation has proved beneficial for pilots, reducing both physical and mental workload, improving flight path precision through navigational aids as well as the presence of collision and weather warning systems. Automation has served the aviation industry too, with less crew members being required for the operation of aircraft. However, for the pilots monitoring these automated systems, there remains a susceptibility to stress and human error.

The job of a commercial airline pilot is regarded as one of the most stressful (Bourne & Yaroush, 2003). From a day-to-day perspective, the responsibility of passenger safety plus the irregular hours and layover times between long-haul flights can increase levels of stress. On the rare occasion, emergency situations can arise. Such unanticipated situations are dynamic, time dependent and complex (Bourne & Yaroush, 2003). Consequently, some pilots can react with a stress response. The body's initial response to a perceived threat such as this is mediated by the autonomic nervous system (ANS). One branch of the ANS, the sympathetic nervous system (SNS), triggers a physiological response, the recognised fight or flight mode, where the stress hormones epinephrine and norepinephrine are secreted into the blood stream (Taelman, Vandeput, Spaepen & Van Huffel, 2009) via the adrenal glands. The rush of these hormones increase heart rate (HR), blood pressure and breathing, and provides a burst of energy through increased blood flow to the muscles, to react physically to the short-term stressor. If the threat continues to be perceived, the hypothalamic-pituitary-adrenal

(HPA) axis comes into play, subsequently releasing a series of hormonal signals to maintain activation of the SNS. The stress hormone cortisol is released at the end of the HPA axis.

In the meantime, activity in the parasympathetic nervous system (PSNS) branch is reduced. When the stressor is no longer present, this system allows for "rest and digest," which conserves energy, and a negative feedback system halts cortisol production (Taelman et al., 2009). Both the sympathetic and parasympathetic systems are constantly interacting, aiming for a sympathovagal balance through homeostasis (Taelman et al.2009). This balance is between the SNS and the vagus nerve, the latter controlling parasympathetic output.

1.2.2 The Effects of the Stress Response

The effects of stress are not just physiological. Emotional reactions such as fear, anxiety and frustration, as well as impaired cognitive functioning can occur (McCraty, Atkinson, Tomasino & Bradley, 2009). Whilst flying an aircraft, reduced functions such as narrowed attention, longer reaction times to peripheral cures, and decreased vigilance (McCraty et al., 2009) can hamper the efficient operational processes pilots need to follow, in order to overcome and successfully manage the critical situation at hand.

The presence of the stress hormone cortisol can also affect functioning on cognitive tasks associated with prefrontal cortex, such as working memory, sustained attention, behavioural inhibition and general mental flexibility (Lupien, Gillin & Hauger, 1999). This is because cortisol is a glucocorticoid, so is able to pass through the blood-brain barrier and bind to glucocorticoid receptors in frontal lobes of the brain (Lupien, Maheu, Tu, Fiocco & Schramek, 2007), where executive functioning is located, used considerably when flying an aircraft.

Decision making, a vital cognitive skill in aviation, has been shown to remain impaired for up to 30 seconds after a startle event (Thackray, 1988), a duration which can have significant effects on the successful, timely management of an emergency.

It is clear that a stress response to a perceived threat (e.g. an emergency situation, or an aircraft failure which could lead to an emergency) in the cockpit, can have detrimental effects on an individual's performance. One way to illustrate the effect of stress on performance, is the inverted-U shape function, known as the Yerkes-Dodson law, illustrating an empirical relationship between arousal and performance (Yerkes & Dodson, 1908). Performance increases with physiological or psychological arousal induced by stress, but only up to a point. At the top of the curve is where optimal performance occurs (Figure 1). Stress and arousal levels can in fact have positive effects in high-pressure situations, and to a certain degree is healthy and beneficial. When levels become too high, however, performance decreases.

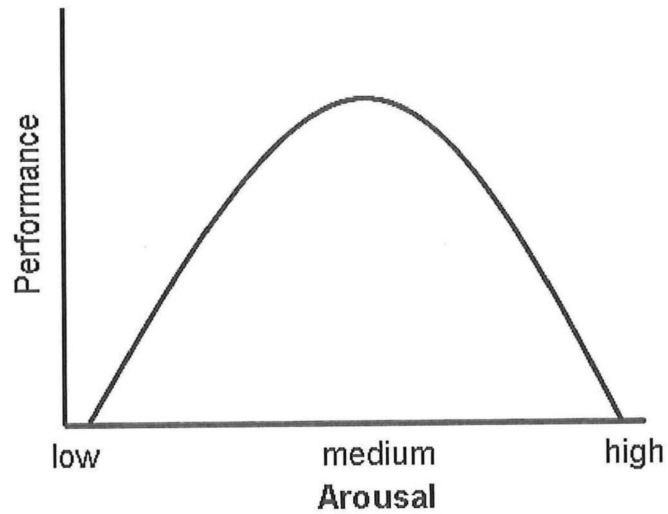


Figure 1: The performance curve, as a function of arousal (stress) levels. ("Yerkes Dodson Law," 2011)

Pilots, required to operate under severe or chronic stress, could be at a greater risk of human error and compromised performance. In some cases, this normal response to stress is undesirable (White & Tursky, 1982). Without fully functioning cognitive skills, pilots may lose control of aircraft as a direct result of reactive stress experienced during emergency flying situations (Cowings, Kellar, Folen, Toscano & Burge, 2001).

In light of this it is evident, that a method for effectively managing and reducing stress symptoms would be of profound benefit for improving pilot's health, mental performance, and for maintaining the safety of crew and passengers. Much like athletes need to train for mental endurance, pilots must train for this (Carr & Montemerlo, 1984), as being mentally prepared is crucial in aviation. While pilots receive ample training in simulators for emergency situations and failures, they do not receive training in which they can effectively self-regulate their physiological responses, in said situations.

1.2.3 Biofeedback

Biofeedback is a method used to provide guidance and reinforcement for successful management of the physiological reaction to stress (Lemaire, Wallace, Lewin, de Grood & Schaefer, 2011). It is defined as: "a means providing immediate information regarding physiological processes about which the individual would normally be unaware" (p. 365 Andreassi, 2000). This information is usually conveyed in real time via a device.

Biofeedback is based on the principles of operant conditioning (trial and error process), in which a response learned is followed by a reward or punishment (contingent reinforcement). The method can be used to gain control over the ANS through contingent reinforcement, in the same way athletic skills are acquired (Cowings & Toscano, 1993). Biofeedback has previously been applied in treating physical and psychological disorders. For instance, research found patients with hypertension had reduced systolic blood pressure

(Alabdulgader, 2012), as well as reductions in stress symptoms and depression, after participating in training which included biofeedback (McCraty, Atkinson & Tomasino, 2003). More recently, the technique has been adopted for use in healthy individuals dealing with, or 'self-regulating' normal changes in their physiological state.

In the United States, NASA conducted a series of research studies on a similar physiological training method, called Autogenic-Feedback training (AFT). The method was tested in the context of space motion sickness, (Cowings & Toscano, 1993) airsickness in military pilots (Cowings, Toscano, Casey & Hufnagel, 2005) and for cockpit performance (Cowings et al., 2001). Participants were trained to voluntarily control and reduce (i.e. self-regulate) several of their own physiological responses in light of environmental stressors (Cowings & Toscano, 1993), with the goal of maintaining operator efficiency and performance. Findings indicated a reduction in autonomous nervous system response levels, and a significant improvement in performance, as rated by instructors, compared to a control group who did not receive the AFT exercise (Cowings et al., 2001). Pilots also demonstrated a particular improvement in knowledge of aircraft procedures and technical proficiency. Although these studies did not assess *commercial* pilots, they indicate that physiological self-regulation can be successful in reducing a psychophysiological response to a stressor, through gaining control over the ANS. In overcoming this influence, critical cognitive functions such as attention and communication, in a high workload scenario, can remain fully functional.

1.2.4 Heart Rate Variability

Biofeedback training can condition individuals to control a number of physiological functions, including HR, galvanic skin response (GSR), blood pressure or muscle tension. This research will focus on heart rate variability (HRV) biofeedback. Over the last few decades, HRV has been used as a non-invasive measure of mental stress (Taelman et al.,

2009) and emotion (Lemaire et al., 2011). It is a form of assessment of the complex interaction between the heart and multiple body systems, in particular the brain (Lloyd, Brett & Wesnes, 2010). HRV is the variation in the cardiac beat-to-beat interval (IBI) between normal heartbeats, also known as the R-R interval. Measured via an electrocardiogram (ECG), which measures the heart's electric signals, the time distance between two consecutive R-peaks is taken. This variance in time reflects the dynamics and status of the ANS, with HRV being regulated by the sympathovagal balance (Taelman et al., 2009). HRV has a spectrum of low frequency (LF) and high frequency (HF), centred around 0.1 Hz and 0.25 Hz respectively (Malliani, Lombardi & Pagani, 1994). HF is linked to increased parasympathetic tone and is a marker of vagal activity (inhibition of sympathetic outflow), whereas low frequency (LF) is a marker of sympathetic activity (Malliani et al, 1994). A response to stress is indicated by a sympathovagal balance shift towards sympathetic predominance (Malliani et al, 1994). Thereupon, information about one's HRV can indicate function and balance of the ANS.

By reflecting the interplay between the sympathovagal systems, HRV can indicate an individual's ability to adapt to stress and environmental demands (Beauchaine, 2001), and to relax, respectively.

As HRV reflects variation in cardiac output, naturally occurring variation in HR during a breathing cycle is termed as respiratory sinus arrhythmia (RSA). It refers to the increase in HR with inspiration and a decrease in HR with expiration (Giardino, Chan & Borson, 2004). During inhalation, vagal activity (associated with parasympathetic activity) is suppressed, which leads to an increase in HR, and during exhalation HR decreases as vagal activity continues. Since vagal tone, from the vagus nerve cannot be measured, RSA and HRV can yield estimates of parasympathetic activity, as HRV is highly correlated with vagal nerve activity (Kuo, Lai, Huang & Yang, 2005). Both HRV and RSA can serve as input for biofeedback on the cardiovascular system, and are both indices of autonomic arousal. The terms are used interchangeably throughout literature (Giggins, Persson & Caulfield, 2013). Hansen, Johnsen and Thayer (2003) demonstrated a relationship between resting HRV and cognitive performance. Those with a higher HRV, performed better on a working memory test, compared to those with a low HRV. The former group also demonstrated superior executive functioning, which involves aspects such as planning, working memory, selective and sustained attention (Robbins, 1996). Thayer, Hansen, Saus-Rose & Johnsen, (2009) also found HRV to be linked to neural structures relating to executive function tasks such as working memory and inhibitory control, all functions required for successfully piloting of aircraft.

HRV of pilots has been monitored before, to deduce workload experienced by them (Wilson, 2002). HR was at maximum levels during different flight stages, such as visual flight take-off and go around and during the landing phase which coincided with subjective workload ratings. HRV showed significant decreases around visual flight landing and take-off. These physiological measurements were taken during real flight, which is a financial and time consuming endeavour. It would be more favourable to experiment in a simulated environment to reduce these operating costs. Magnusson and Berggren (2002) performed a study with five fighter pilots, who flew three simulated scenarios first, followed by three identical scenarios with identical aircraft and identical tactics in real flight. Heart rate as well as HRV analysis yielded "a high degree of correspondence" across the simulated and real flights. Dahlstrom and Nahlinder (2009) examined differences in physiological response to real and simulated flight amongst eight Air Transport Pilot License (ATPL) students and stated that, "simulated flight seems to replicate the workload demands from aircraft flight".

Biofeedback of the heart's activity has been used in various professions in which stress is a major factor of their job, and could impair professional performance.

Police officers and hospital physicians benefitted from this stress management technique, demonstrating successful reductions in physiological and psychological responses to both acute and chronic stress (McCraty & Atkinson, 2012), and a decline in stress (Lemaire et al., 2011).

These professions experience in the moment acute stress during life threatening situations, as well as long-term stress from the emotionally charged incidents experienced. Chronic stress can result in the homeostatic feedback system no longer being able to recalibrate, leading to persistent activity of the sympathetic nervous system (McCraty & Atkinson, 2012). Following an acute stressful situation, the body needs to recover from the extreme physiological arousal (McCraty & Atkinson, 2012). During initial measures before the biofeedback training, it was discovered that it took the police officers' heart rates more than 60 minutes to fall below their baseline, after a stressful scenario had ended. This extended period of time for physiological recalibration means, that a stress response is experienced not just during the scenario, but subsequently for some time too. After training they were able to recalibrate and recover a lot faster.

This research illustrates the potential for the use of biofeedback training to manage stress, in occupations of similar risk and responsibility to commercial pilots.

1.2.5 Coherence

HRV reflects heart-brain interactions (McCraty et al., 2009). Psychophysiological interactions amongst the body's systems can bring about coherence of the heart, where this is increased order, efficiency and harmony in the functioning of the body's systems (McCraty & Atkinson, 2012). Whilst in a coherent state, the individual has increased synchronisation between their higher-level brain systems and reciprocal activity occurring in the ANS branches, SNS and PSNS. During coherence, there is a shift in autonomic balance toward increased parasympathetic activity. The coherent pattern is sine wave-like, at a frequency of around 0.1 Hz (McCraty et al., 2009), thus falling in the HF range linked to parasympathetic tone. An erratic, discordant pattern of activity denotes a system that is incoherent (McCraty & Atkinson, 2012). The pattern of the heart's rhythm is primarily reflective of stress and emotional states (McCraty & Atkinson, 2012), so stressful emotions such as anxiety and fear are observed in an incoherent heart pattern. Coherence can occur at higher or lower heart rates, with changes in rhythm acting independently from heart rate.

Benefits of practicing HRV coherence, include resilience, cognitive flexibility enhanced problem solving (McCraty & Atkinson, 2012) and improved cognitive performance (Ginsberg, Berry, & Powell, 2009, McCraty et al., 2009).

Taken together, this research will endeavour to use HRV biofeedback as a stress management tool for pilots, to learn to self-regulate their physiological system. By shifting into a more coherent state before, during and/or after challenging situations individuals can prevent unnecessary stress reactions, that can impede optimal performance, by depleting resources and recover quickly from these acute challenging situations.

A device developed for biofeedback training is the emWave-*2* (HeartMath, LLC, Boulder Creek, California). The emWave-*2* objectively measures HRV and coherence, with the purpose of enabling users to learn how to change their heart rhythm pattern towards coherence through real time observation. After practice, this state of coherence becomes familiar to the brain and nervous system, and is a skill in which individuals can use where necessary (eventually without the device).

1.2.6 Experimental Structure

This research focussed on the fight, flight or freeze stress response (Werrbach, 2014). We were limited to researching a single emergency situation in a flight simulator, and were unable to investigate long term stress experienced by pilots in general. However, the biofeedback training could improve both kinds of stress, as it is considered a transferrable skill. Furthermore, we were limiting this research to an intrapersonal level, although the biofeedback training could possibly improve interpersonal processes in- and outside the cockpit. The experiment was designed to test pilot responses in a high workload scenario, that was cognitively demanding.

There were two groups, two difficulty levels and two time measurements (Table 1). The groups served as an experimental group and a control group which both received training, respectively a biofeedback training and an ineffective 'sham' intervention. Both groups flew two difficulty levels, a low and a high workload scenario. The baseline measurement was called phase 1, the training period phase 2 and the final measurement phase 3. We chose the workload configuration in order to measure a relative degrading between low and high (to see individual differences when workload increased), chose two phases to measure a pre and post biofeedback training effect and chose two groups to control for effects over time.

Table 1

The experimental design

	Phase 1	**Phase 2**	**Phase 3**
Experimental group	Low/High workload scenario A	Biofeedback training	Low/High workload scenario B
Control group	Low/High workload scenario A	Sham intervention	Low/High workload scenario B

1.2.7 Hypotheses

Based on the literature reviewed, we were testing whether HRV improves after biofeedback training with the emWave-*2*, and whether HRV changes in turn are associated with cognitive skills and flight performance.

Hypothesis 1: The biofeedback group will demonstrate higher HRV during the post training, high workload scenario, compared to the control group.

Hypothesis 2: Pilots who have practiced biofeedback will demonstrate a faster return to their HR baseline, during a recovery period after the scenarios, compared to the control condition.

Hypothesis 3: Pilots who have practiced biofeedback will demonstrate improved flight performance during the high workload scenario, compared to the control group.

Hypothesis 4: Pilots who have practiced biofeedback, (and are able to put themselves in a coherent state) will demonstrate improved cognitive skills, compared to the control group

2 METHOD

2.1 Participants

Fifty-six licensed pilots were recruited. All pilots completed their training at the European Pilot Selection and Training (EPST) centre in Utrecht, Netherlands. Recruitment was conducted through a presentation given about the study by the researchers at EPST, including the potential benefits of taking part. The Managing director of EPST was also present and advocated participation in the research. At the end of the presentation, the pilots signed up voluntarily.

From the 56 pilots who volunteered, seven of these pilots were used as pilot monitors and scenario testers. Four pilots were qualified Student Instructors (SI), who acted as air traffic control and rated each participant's performance in the simulator sessions. These 11 individuals were confederates who did not subsequently participate in the main experiment.

Forty-five licensed pilots (38 males, 7 females) ranging in age from 20 to 38 years old were used as participants for the experimental and control groups. In the simulator, all participants had the role of pilot flying and as captain (making all the decisions and having the last say). They had an average of 150-200 real flying hours and they all had experience in simulators, with a range of 30 to 100 hours, depending on when they completed their training. All participating pilots were able to log the hours spent in the simulator for the research, as simulator flying hours. Due to dropouts between the testing phases, the final number of participants who completed the entire research was 40.

2.2 Design

This study was a mixed between and within-subjects, single-blind design. The between-subject variable was the training intervention, which had two levels: The experimental group, who received training from HeartMath, and the control group who received a 'sham' intervention. This latter intervention was administered, in order to match the amount of contact time between researchers and participants. It was designed so that it would not improve cognition, HRV control or performance over the testing period, and to not have a detrimental effect on their performance either. This consisted of participants using a smart phone app to track their sleep, as well as receiving basic information about stress management and relaxation techniques (progressive muscle relaxation and acupressure). The within-subjects variable had two levels, comparing pre and post intervention results of both the experimental and control group. The time periods of the research was split up into 3 phases. Phase 1 was the pre-intervention tests, phase 2 was the intervention, and phase 3 was the post-intervention tests.

The dependent measures include: heart data, cognitive battery test scores, simulator data, subjective student instructor ratings, and participant self-ratings of the experienced workload.

2.2.1 Scenarios

Scenario – General. A script with instructions was provided to the participants before entering the simulator. To incorporate a representative situation, a normal dual pilot cockpit occupation was chosen. Participants were assigned role of Pilot Flying (PF) and were told to make all the decision regarding flight operations and safety. The confederate Pilot Monitor (PM) was told to act in a purely supportive role and enforce the prescribed script, to make sure participants follow the scenario according to plan. The student instructors who took the role of Air Traffic Control (ATC) and elicited the failures were also given a detailed script to follow. The elements required for psychophysiological response to a high cognitive workload such as acute stressors and anticipatory items, as well as elements of the cognitive test battery, have been used to design the requirements of the scenarios.

The potential emergency situations have been chosen from a list of pre-programmed failures stored in the simulator. Participants were notified that departure and arrival precision, as well as altitude and speed restrictions, were monitored and used for flight performance rating. Scenarios were developed to be as realistic and immersive as possible. A low and high workload scenario were developed for each phase, where the high workload included time critical unforeseen events. Both scenarios were set at clear visibility outside, containing scattered clouds with base at 2000 feet and ceiling at 5000 feet to increase realism (Dahlstrom & Nahlinder, 2009). The difficulty level of the scenarios was set just above the pilot's capabilities, to eliminate any floor or ceiling effects. This was achieved by rigorous test flying of the scenarios by confederate pilot volunteers who stemmed from the same population (i.e. EPST novice pilots). For maps of the low and high workload situations, see appendix A. The following section of this report is abridged, for the complete version see appendix B. The scripts which were provided to SI and PM can be found in appendix C.

Scenario – Low Workload. The low workload situation consisted of a repositioned flight, 12 NM (Nautical Mile) out from Düsseldorf Airport's runway 23L.The Instrument Landing System (ILS) was operational, in the aircraft and on the airport, which was used by pilots to initiate and complete the landing according to standard operating procedures (see appendix D).

Scenario – Phase 1, High Workload. The high workload of phase 1 consisted of a continuous flight departing from and landing on, Düsseldorf Airport runway 23L. Participants were instructed to follow the COLA 2T (see appendix D) Standard Instrument Departure (SID) route after take-off and maintain altitude and speed as instructed. The route was divided into three sections in order to administer acute stressors. The sections were take-off to end of SID, end of SID to start of the approach and the approach. During take-off, a false fire alarm sounded, acting as the first stressor. During the first section, the hydraulics of system B failed, designed as the second stressor. The third stressor was the failure of the electronic trim of the horizontal stabilizer. Both the hydraulics and the trim failure resulted in deferred checklist items which had to be taken into account before the landing, designed to induce maximum workload during the final approach. Finally, just before landing another fire alarm sounded, designed as an acute stressor during the landing.

Scenario – Phase 3, High Workload. The high workload scenario for phase 3 was designed to be unique, yet containing similar stressors to evoke a similar type of cognitive load as the high workload scenario for phase 1. The route was modified but contained the same SID since no other similar SID could be used due to the complexity of the initial SID chosen. Instead of the system B failure, system A failure was chosen. The other acute stressors were an intermittent false take-off configuration warning instead of the fire alarm, the stabiliser trim inoperative and a false fire alarm warning on landing.

In order to determine flying performance over a range of situations, a series of performance variables were chosen. The standard (i.e. pre-programmed) recording parameters formed the foundation for this analysis.

2.2.2 Objective Performance Variables

The main analysis determined the deviation of the approach flight path for both high and low workload. As an attempt to determine nervousness of the participants, the control column inputs were analysed for the approach section. Secondary were analyses of the high workload situation, of how well participants were able to maintain a set altitude and speed, follow the SID and how much control input was applied over the complete route.

In order to determine individual flight performance, the output data of the Boeing 737 NG simulator at EPST has been used as a guideline for performance variables. All data was exported from the simulator as comma separated value files and processed with Python™ version 3.4. The data was plotted and verified visually through these plots (appendix E) for anomalies. The autopilot was programmed to fly the scenarios and this data were used to compare with the flight path flown by participants, in order to determine their flight performance.

Readily available from a modern simulator are the aircraft states such as longitude, latitude, altitude, pitch, yaw, roll and airspeed. The simulator at EPST also recorded the aircraft control inputs such as pitch, roll, yaw, brakes, speed brakes and flaps setting. Furthermore, the automatic flight director system records the speed bug, altitude bug and heading bug settings. For the low and the high workload, the deviation of the final approach segment of the scenarios has been analysed with an identical protocol, in order to compare the high and low workload performance of the participants. For both workload settings, the control inputs have been analysed. The amount of lateral and vertical input applied by the pilots on the control column for each section of the two workload settings was recorded and processed. For the high workload, the deviation from the instructed altitude of 6000 feet and speed of 220 knots has been analysed. Threshold values of respectively 5500 feet and 200 knots have been used to define start and end point for these measurements. Finally, the deviation from the SID was analysed. This has been done through lateral analysis only since the largest part of this part of the route was flown at a constant altitude and because altitude deviation was analysed separately. The deviation from speed, altitude and the SID, which were only recorded during the high workload scenario, were used to compare pre and post effects of the training during the highest cognitively loaded section of this research.

The approach deviation was determined by using the aircraft state and transforming it into an angular deviation from the ILS glide path. This angular deviation was corrected with the root mean squared (RMS) method. A 6NM segment was used, starting at 7NM and ending at 1NM from touchdown point. Segments before and after these markers were not clean enough for analysis. RMS method was also applied to the amount of degrees the yoke was moved and consequently plotted for verification. The other variables such as altitude, speed and SID deviation have been calculated in a similar fashion by importing the raw data into Python™ and converting them to useable output variables. The output of the code was summarised in Excel files which was later used as an input for the IBM SPSS Statistics software. In total, 14 variables have been processed and analysed. An overview of the protocol can be found in appendix F.

2.3 Materials

2.3.1 Simulator

A fixed base Boeing 737 NG aircraft simulator (Figure 2), built by Multi Pilot Simulators (MPS) B.V. in Utrecht, was used to depict the flight scenarios. It is a full replica of the original aircraft, certified for the type rating of pilots. It is the latest model built by MPS and the latest model in use by EPST, containing the most modern features on failure simulation available on the market. It is constructed as a full flight deck replica, has a fully operational flight management system, autopilot, flight director and auto-throttle capabilities. All auxiliary aircraft systems including hydraulics, pneumatics and electrics are functional. The simulator contains a realistic aerodynamic flight model, designed independently by MPS. Visual representation of the flight scenarios is through high definition collimated mirror 200 by 40 field-of-view visual system with Level D image generator. It contains a worldwide navigation database and has Level D fidelity control loading on all flight controls. Summarised, it is the latest state of the art fixed based simulator available. For failure management, the Quick Reference Handbook (QRH) of the 3rd of September 2007 developed by Transavia Airlines for its Boeing 737 NG fleet, was provided to the pilots.

Figure 2: Boeing 737NG simulator, by Multi Pilot Simulations B.V.

2.3.2 Cognitive Battery Test

A cognitive battery test was used to establish the participants' cognitive skills, both before and after the training. The cognitive battery test used was developed at Universiteit van Amsterdam (UvA). The test was designed to be engaging and practical, being administered as an application on an iPad mini. See appendix G for screen shots of each game. Having tested the test battery with athletes, the developers were interested in testing a different population. The app consisted of five games, each testing five different aspects of the individual's cognitive skill set. The test duration was approximately 50 minutes. The cognitive skills tested were:

1. *Working Memory (WM)*: WM enables active maintenance of information held in a readily accessible state (Fukuda, Vogel, Mayr & Awh, 2010). WM is important too, as it correlates highly with fluid intelligence (Fukuda et al., 2010), used when solving problems in novel situations, e.g. aircraft failure.

   The game consisted of eight subtests that are described below. In each subtest, the number of items to remember increased when participants made a correct response. When participants made two sequential mistakes on a subtest, the subtest was dropped from the game. After all subtests were dropped, a double staircase commenced, in which all subtests were repeated again with an easy (-3 items) and a difficult (+3 items) mode. When people were correct on the easy trials, set size increased. When people were incorrect on the difficult trials, set size decreased. This continued per subtest until the easy and difficult mode converged to the same set size.

Subtest 1-3 measured WM by presenting items on screen simultaneously. People had to remember all the items and then tap the locations where all items were presented.

Subtest 4-5 measured WM when items were presented on screen sequentially. People had to remember all the items and then press the location where all items were presented.

Subtest 6-8 measured WM by presenting items simultaneously and in the next screen, one item always changed compared to the initial screen. People had to press the item that changed between screens.

In subtest 1, 4 and 6 only relevant, red items were presented.

In subtest 2 and 7 irrelevant, orange items that had to be ignored were presented simultaneously with the relevant, red items.

In subtest 3, 5, and 8 irrelevant, orange items were presented during the blank retention interval.

The game's *outcome measures* were:

-*Working memory: Capacity.* Average set size over all subtests (set size increases when players are correct on easy trials, and set size decreases when players are incorrect in difficult trials).

- *Working memory: Distractibility.* Average set size over all subtests with distraction (irrelevant items in subtest 2, 3, 5, 7, 8).

2. *Anticipation*: Initially developed with athletes as the end user in mind, this game measures the skill of anticipating the trajectory of an object. During the game, the size of the object, and speed of its movement is manipulated. Different cues can be learned

to anticipate different trajectories too, as well as measuring how well players anticipate the path of an object. The game sets out to establish how well the players can prioritise one action over the other, and continue to respond in a situation of information overflow.

American football players run from a dug-out (1 = left, 6 = right). If the football player gets hit by a participant player when just entering the field, participants got a lot of points (600), but when the football player is hit close to the finish line participants get just a few points (100). After reaching 6000 points, a challenge mode starts where the football players enter the field with increasing speed until three players were missed.

Outcome measures:

- *Anticipation: Performance.* Points scored per second, indicating how good the player is at anticipating.

- *Anticipation: Stress.* Time (seconds/100) it takes for three American football players to pass the finish line, in the challenge mode of the game.

3. *Cognitive Control*: A task-switching paradigm, assessing the adaptation of a planned movement to suit a change in the environment (i.e. a coloured cue). Prepared actions are inhibited and an alternative must be selection (Neubert, Mars, Buch, Olivier & Rushworth, 2010). These actions are based on pre-defined rules, which need to be kept in mind, so executive control becomes utilised. The task requires participants to respond with their left or right hand, in response to visual stimuli presented on the iPad. The manipulation has two trials: stay trials, which measures speed relative to response repetition (same colour cue, exerting prepared response), and switch trials, with the different cue means players must inhibit their prepared response and reprogrammed their action plans, which produces a switch cost (Neubert et al., 2010).

Outcome measures:

- *Control: Motor.* Measuring if efficiency (accuracy/RT: .8 (proportion correct)/.4(RT in s) = 2) changes when a player responds to the left and suddenly they have to switch, because the stimulus switches to the other side. Efficiency score (0-1): 1 = no change in efficiency compared to no switch trials. Lower values = more costs.
- *Control: Cognition.* Assessing whether efficiency changes when a player responds to the left and still has to respond to the left but the stimulus in the middle changes (needs to update the task set). Efficiency score (0-1): 1 = no change in efficiency compared to no switch trials. Lower values = more costs.
- *Control: Reaction time (RT).* Measures how fast a person responds to a stimulus (left/right) that is similar to the stimulus in the middle. RT in ms (subtract 130 ms for the real RT). Only repeat trials (2-5 repeats; for example stimulus matching the middle stimulus appears to the left a couple of times in a row) .

4. *Attention & Concentration*: Using the Attentional Network Task (ATN) (Fan, McCandliss, Sommer, Raz, & Posner, 2002) the efficiency of visual attention and reaction time is measured in this game. Players have to swipe upwards or downwards whenever the middle arrow is pointing upwards or downwards. The surrounding arrows can be congruent (upwards when the middle arrow points upwards) or incongruent (downwards when the middle arrow points upwards). All values are differences in efficiency: E = proportion correct (0-1)/RT (in seconds)

Outcome measures:

- *Attention: Alerting.* The inter-trial time fluctuates. In some trials, there is no hint when the arrows are going to appear (no cue trial). In other trials, an asterisk appears in the middle that indicates that exactly 400ms later the arrows are going to appear (central cue trial). If people are good at maintaining concentration, the difference in efficiency between the no-cue trial and the centre cue trial will be 0. Higher values means there are less efficient on the no cue trial.

- *Attention: Orienting.* In some trials, an asterisk appears in the middle indicating, that exactly 400ms later, the arrows are going to appear (central cue trial), but people do not know whether the arrows are going to appear to the left or the right. In other trials, the asterisk appears to the left or the right (spatial cue trial), indicating where the arrows are appearing. If people are fast at moving spatial attention, the difference in efficiency between the central-cue trials and the spatial-cue trials will be close to 0. Higher values means people are slower to move spatial attention.

- *Attention: Cueing.* On some trials, the asterisk appeared on the left, but the arrows were presented on the right. Therefore, spatial attention needs to be disengaged. Efficiency differences between no-cue trial and these wrong spatial cue trials were computed. If people are bad at disengaging attention, the value is higher than 0.

- *Attention: Cognitive Control.* Difference in efficiency between conflict trials (middle arrow different than flanking arrows) and no-conflict trials.

- *Attention: RT.* Median reaction time of all the trials together.

5. *Integration*: Also initially developed with athletes in mind, this game measures how well a player can use sound to predict where a visual object is, and initiate an action. This measure of audio-visual integration was included in the cognitive test battery administered to pilots in this research, yet it was not given much focus due to the lack of literature supporting it and the specific population it was targeted for (athletes, and not pilots).

   A ball is played from left to middle. The ball disappears under a cloud and is then played to another location. A sound indicates the direction in which the ball is played.

If players use the sound, they will be faster to click on the ball as compared to when they just rely on visual information.

Outcome measures

- *Integration: Performance.* RT difference between locations where object (ball) is to be selected.

- *Integration: Learning.* RT difference between before a certain set of sounds have been learnt, and after the sound mapping has changed.

2.3.3 HeartMath emWave-2

For the intervention technique, emWave-2 (Institute of HeartMath) devices (Figure 3) were provided by HeartMath Benelux to all participants in the experimental condition. The emWave-2 is a portable heart rhythm feedback device. It has audible cues (which can be changed in volume or turned off), and visual cues (flashing lights and graphics), to help the user guide their breathing, and receive real-time feedback on how they are doing. For instance, at the top there is a red, blue or green light that illuminates depending on participant's level of coherence (from poor to good).

These devices were taken home by the pilots in the intervention condition, and they were instructed to use them daily. Along with the devices, a CD-ROM was provided for participants to install a programme onto their computers. This programme helped provide feedback on their progress. During the three sessions of professional training presentations from a HeartMath trainer, there were discussions about how stress interfered with performance. Furthermore, theoretical knowledge behind heart rate variability , the ANS, and the role emotions play was provided. There were practical exercises where participants could practice the taught techniques and be provided with instruction and guidance on the use of the devices.

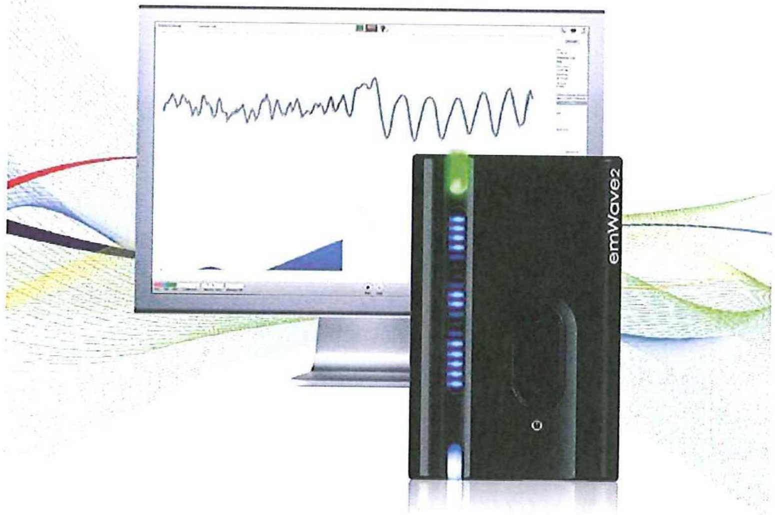


Figure 3: HeartMath emWave-2

2.3.4 Firstbeat Bodyguard-2

A small wireless heart monitor (Figure 4) with two electrodes, called the Bodyguard-2 (Firstbeat Technologies Ltd., Jyväskylä, Finland) was used to take the physiological measurements of the pilots, before, during and after the simulator sessions in phase 1 and 3. The device is effective at recording HRV data during physical activity, as well as rest and recovery (Finni, Haakana, Pesola & Pullinen, 2014). Thus its readings during the high workload scenario would not be impaired by any physical movements, such as carrying out manual trim. It was attached to participant's chests in two locations: one just below the right collarbone, and the other on the bottom of the ribs, on the left hand side. The data were subsequently uploaded onto a computer directly via USB port.

Figure 4: Firstbeat Bodyguard-2

2.3.5 NASA TLX

The computer version of the NASA Task Load Index (TLX) was administered to participants immediately following the flight simulator sessions. The NASA TLX is a self-report rating scale that provides an overall subjective workload rating. Participants are asked about six dimensions: mental demands, physical demands, temporal demands, performance, effort and frustration. There are two parts to the task. 1. Rating: Participants indicated on a visual scale ranging from low to high (and from good to poor for the performance dimension) 2. Weights: assigning a weight to each of the six dimensions, in comparison to one another ("which contributed more to the workload you experienced? Mental demand or temporal demand?")

2.3.6 Subjective Feedback Form

After completing phase 3, participants were asked to fill out a subjective feedback form (appendix H). This form was set up to track the amount of time they had been practicing the techniques as taught by HeartMath Benelux. The emWave-2 is capable of tracking users' progress but connection and storage problems with the devices required an alternative approach to track the training progress.

Secondly it was used to generate feedback for the HeartMath training, to check for future improvement. Finally, the form contained questions related to changes in the participants' personal lives.

2.4 Procedure

22 pilots were in the intervention condition, and 18 participants were controls. In total, the pilots were tested in 3 phases, over 9 weeks. The study was blind, in that participants were not aware of which intervention technique researchers were investigating the effectiveness of.

2.4.1 Phase 1: Pre Training, Baseline Measures

The research was conducted at the European Pilot Selection and Training centre (EPST), in Utrecht, The Netherlands. After arrival, participants were provided with an informed consent form, containing information about the study and the procedures they would be asked to follow. They were told the importance of not engaging in discussion with any fellow participant about their experiences in the study (in particular, about the scenarios they flew). This was important to reiterate, as the participants knew each other from their pilot training. After signing the informed consent form, participants completed a visual analogue scale form, for researchers to get an indication about their current body state (e.g. fatigue, hunger, thirst).

The Bodyguard device was then attached. Participants were asked to sit still, without talking or moving for 5 minutes. This not only provided researchers with a baseline heart rate measure for the individual, but it also provided an opportunity for participants in the experimental group to use this time in the third phase after the training, to put themselves in a coherent state using the emWave-2.

Participants then completed the cognitive battery test on the iPad, followed by a 5-minute vigilance task on the iPad. Before going into the simulator, participants were provided with a minimum of 10 minutes to prepare for their simulator session. Copies of the SID and ILS plates were provided, as well as some brief instructions on how the simulator session was set up (i.e. they were the pilot flying, and captain, and they were not to discuss the scenarios with any other participants).

Participants flew two scenarios (low and high workload) in the simulator, after a few minutes of familiarising themselves with the simulator. This was counterbalanced, with half the participants completing the low workload scenario first, taking around 10 minutes, and the other half flying the high workload scenario first (35 minutes).

After the hour in the simulator, participants were required to sit and relax for 30 minutes. Researchers wanted to keep the Bodyguard on for these 30 minutes following the simulator session, to establish the speed at which their heart rates returned to baseline. Lastly, they completed the NASA TLX.

2.4.2 Phase 2: Training Intervention

Over the course of 6 weeks, participants in the experimental group attended 3 training sessions. The standard training, as used by HeartMath to train CEOs and CFOs of large corporations, was conducted by professional instructors in biofeedback training using the emWave-2 device, from HeartMath (the founding company of the emWave-2 device). Participants practiced biofeedback using the device during sessions, and an accompanying programme for achieving a coherent HRV at home was set, to practice on a daily basis, for 5-

10 minutes, 3 times a day. Participants in the experimental group were also encouraged to use simple techniques to apply in their daily lives, to help reduce stress.

Those in the control group tracked their sleep up to 5 times a week using an application on their smart phones, Sleepbot. They were also provided with online materials to read through, and exercises to practice at home. These materials included a basic stress management document, and the exercises were progressive muscle relaxation and acupressure, both from the guidance of a YouTube video.

Both groups sent weekly updates to researchers, of their emWave-2 data and sleep tracking data.

2.4.3 Phase 3: Post Training

This phase replicated the procedure in phase 1. They flew two different scenarios to phase 1 (low and high workload, order reversed from phase 1 order) in the same simulator. Researchers debriefed the pilots after participation, and answered any questions they had.

THIS PAGE IS INTENTIONALLY LEFT BLANK

3 RESULTS

All statistical tests were conducted using IBM's SPSS statistical analysis software, with $\alpha = .05$, that is, there was a 95% confidence level. As this was an exploratory piece of research, values between .05 and .10 were also reported as a trend. The default analysis was a series of mixed between and within 2 × 2 ANOVAs (analysis of variance) with intervention group (biofeedback and control) as the between subjects factor and time (pre and post training) as the within subjects factor. Exceptions are indicated.

3.1 Heart Rate Variability (HRV)

It was hypothesised that the subjects in the biofeedback group would have higher HRV, during the high workload simulator session, post training, compared to the control group. Heart rate (HR), inter-beat interval (IBI), total power (TP), low frequency (LF), high frequency (HF), and LF/HF ratio were computed from averaged 30-second measurements. LF, HF, TP and LF/HF ratio were naturally logarithmically (ln) transformed before analysis to correct for skewness of distribution. Of all of these HRV parameters, the ln LF/HF figures were chosen to be used exclusively for HRV analysis as an index for sympathovagal balance. These figures were corrected for negatives before analysis. Note that for HRV, more variability was hypothesised (the higher, the better), however with the ln LF/HF measure, lower figures indicate better HRV, indicating less sympathetic (stress response) activation. Measures were taken at different flight phases of the high workload scenario, which were take-off, end of departure (SID), seven and one nautical miles (NM) out during the approach, and touchdown.

There was a significant interaction during take-off, between time and group, (F (1,38) = 4.99, p = .031, η_p^2 = .116). The interaction however, was in the opposite direction to the hypothesis. The biofeedback group had increased levels of stress in the post-test, indicating sympathetic system activation, whilst the control group's sympathetic activity decreased in the post-test.

There was a significant main effect in the high workload 7 NM approach for time (pre/post) (F (1,38) = 7.58, p = .009, η_p^2 = .166). Means dropped for both groups (see Table 2) in the post measure, meaning less sympathetic activation in the post-test during the 7NM approach in the high workload scenarios for all participants.

In Table 2, means reduced across all measures for the biofeedback group, apart from during take-off. This indicates less sympathetic nervous system activation in post-tests, albeit figures are not significant (see interactions in Table 3).

Table 2

Means and standard deviations of ln LF/HF figures for each flight phase in the high workload scenarios.

	Pre, M (SD)		Post, M (SD)	
Flight Phase	**Biofeedback**	**Control**	**Biofeedback**	**Control**
Take-off	1.95 (0.87)	1.74 (0.72)	2.35 (0.83)	1.39 (0.47)
End of SID	1.77 (0.63)	1.87 (0.72)	1.61 (0.68)	1.75 (0.83)
Approach, 7 NMs out	1.98 (0.73)	2.08 (0.66)	1.53 (0.63)	1.79 (0.59)
Approach, 1 NM out	1.82 (0.71)	2.03 (0.81)	1.77 (0.74)	1.82 (0.76)
Touchdown	1.82 (0.75)	1.75 (0.72)	1.71 (0.57)	1.82 (0.54)

Table 3

F, P and η_p^2 time x group interaction values from the mixed ANOVAs of ln LF/HF during different flight phases in the high workload scenarios.

Flight phase	*F (1,38)*	*p*	η_p^2
Take off	4.99	.031	.116
End of SID	0.03	.856	.001
Approach, 7 NMs out	0.34	.562	.009
Approach, 1 NM out	0.30	.585	.008
Touchdown	0.55	.463	.014

3.1.1 Return to Baseline Heart Rate

It was hypothesised that during the post-simulator recovery period, the biofeedback group would return to their baseline heart rates (HR) faster than the control group would. The first 15 minutes of the 'return to baseline' recovery time period, directly after the simulator sessions, was analysed with Microsoft Excel in order to determine a rate of decline. For each participant, an equation was extracted from the data. This equation described the rate of change from the initial heart rate. The outcome y, is defined by a factor times x (rate of change), added to the initial rate; *y = factor* times *x* + *initial HR*. These linear equations added per group and divided over the group total, resulting in a series of linear equations (Table 4) to define means between the control and experimental groups. For both phases 1 and 3, four means were determined, resulting in eight values. For phase 1, there are two equations for the control group and two for the experimental group.

During phase 1 the control group had a mean of y = -924 x + 91 and the experimental group y = -1252 x + 93 indicating that the initial heart rate of the control group was lower and the rate of decline slower during phase 1, compared to the experimental group.
During phase 3 the control group thus had a means of y = -1315 x + 93 and the experimental group y = -879 x + 85 indicating that the initial heart rate of the control was higher and the rate of decline faster during phase 3, compared to the experimental group. The rate of decline of heart rate of the control group increased in phase 3 measurements compared to phase 1, whereas their initial heart rate increased. The rate of decline for the experimental group in phase 3 decreased as well as their initial heart rate, compared to phase 1.

Table 4

Decline rate and starting hear rates across phases and intervention group (CTRL: control group, BIO: biofeedback group)

Linear equations Phase 1			**Linear equations Phase3**		
	Decline rate	Starting HR		Decline rate	Starting HR
CTRL L/H	-801	88	CTRL L/H	-1197	94
CTRL H/L	-1083	94	CTRL H/L	-1434	91
BIO L/H	-1339	92	BIO L/H	-934	87
BIO H/L	-1164	93	BIO H/L	-824	83
CTRL mean	-942	91	CTRL mean	-1315	93
BIO mean	-1252	93	BIO mean	-879	85

Note: The first figure in the linear equation defines the rate of decline where the second figure indicates the initial heart rate at the start of the measurement.

3.2 Flight Performance

The pilot participants who practiced biofeedback were hypothesised to demonstrate improved flight performance during the high workload scenario, compared to the control group. Flight performance measures included deviation from altitude hold, speed hold and the standard instrument departure (SID). Vertical and lateral aircraft control inputs into the control column were also measured (pitch and roll) during key phases during the scenario (take off – end of SID, end of SID – approach, and 7 - 1 NM approach).

Due to missing data from one day of testing, seven participants' simulator data from the pre-test was not logged. These participant's post-test simulator data was taken out of the analysis, leaving N = 33 (biofeedback = 19, control = 14).

From the 2 × 2 ANOVAs, there were no significant interactions between group and time for any of the flight performance measures during the high workload scenarios.

Significant main effects for time (pre-post) were found for pitch input during the flight phase 7 - 1 NM approach ($F\,(1,31) = 5.232, p = .029,\ \eta_p^2 = .114$) and for overall pitch input across all flight phases ($F\,(1,31) = 11.79, p = .002,\ \eta_p^2 = .276$).

Significant main effects for time were also found for the control input roll, during the flight phases take-off - end of SID ($F\,(1,31) = 13.11, p = .001,\ \eta_p^2 = .297$), end of the SID - approach ($F\,(1,31) = 7.26, p = .011,\ \eta_p^2 = .190$) and total roll input across all phases ($F\,(1,31) = 32.09, p < .0001\ ,\ \eta_p^2 = .509$). These findings indicate that performance improved over time between the pre to post tests, with less control inputs required for control of the aircraft. Performance between groups was similar across most measures (see Table 5 of means and standard deviations).

A 3 way ANOVA was carried out on the variables time (pre/post), group (biofeedback/control) and workload (high/low). All measures were taken during the approach to land, assessing the approach accuracy, and the pitch and roll input during the approach. A significant interaction occurred between time and workload for the approach accuracy ($F\,(1,31) = 14.76, p = .001,\ \eta_p^2 = .321$) and the amount of roll input during the approach ($F\,(1,31) = 63.19, p < .0001,\ \eta_p^2 = .671$). This interaction indicates that the accuracy in which participants flew the approach during the high workload improved over time, and there was less roll input during the high workload approach in the post test, compared to the pre-test and low workload scenarios.

There were main effects in time for the approach accuracy ($F\,(1,31) = 17.29, p < .0001,\ \eta_p^2 = .358$), pitch during approach ($F\,(1,31) = 6.42, p = .017,\ \eta_p^2 = .172$, and roll during approach ($F\,(1,31) = 133.06, p < .0001,\ \eta_p^2 = .811$). All of these measures improved in the post test, indicating that fewer inputs were required to control the aircraft simulator in the post-test compared to the pre-test.

Workload also had main effects in approach accuracy ($F\,(1,31) = 57.95, p < .0001,$

η_p^2 =.651) and roll during approach (F (1,31) = 49.71, p < .0001 , η_p^2 =.616). These main effects for workload, indicate a significant improvement in approach accuracy, with less roll input administered, between pre and post-tests.

Table 5

Means (M) and standard deviations (SD) of flight performance during high workload flight phases, pre and post intervention.

	Pre, M (SD)		Post, M (SD)	
Flight Performance Measure	**Biofeedback**	**Control**	**Biofeedback**	**Control**
Altitude hold [feet]	99.84 (28.26)	82.86 (15.91)	91.40 (25.62)	95.02 (34.43)
Speed hold [knots]	7.82 (2.19)	7.05 (1.50)	7.66 (2.06)	6.72 (1.69)
SID deviation [NM]	0.66 (0.25)	0.66 (0.18)	0.59 (0.25)	0.59 (0.39)
Pitch: take-off – end of SID [deg]	0.22 (0.06)	0.22 (0.04)	0.21 (0.05)	0.22 (0.03)
Pitch: End of SID – Approach [deg]	0.13 (0.02)	0.15 (0.04)	0.13 (0.02)	0.13 (0.02)
Pitch: Approach (7 - 1 NM) [deg]	0.18 (0.08)	0.17 (0.05)	0.15 (0.05)	0.13 (0.04)
Pitch: total [deg]	0.17 (0.03)	0.18 (0.04)	1.16 (0.02)	0.16 (0.02)
Roll: take-off – End of SID [deg]	2.21 (0.31)	2.34 (0.37)	2.01 (0.37)	2.02 (0.49)
Roll: End of SID – Approach [deg]	1.60 (0.20)	1.65 (0.33)	1.45 (0.21)	1.53 (0.31)
Roll: Approach (7 - 1 NM) [deg]	1.94 (0.84)	2.04 (0.79)	1.69 (0.49)	1.72 (0.55)
Roll: total [deg]	1.80 (0.27)	1.87 (0.28)	1.59 (0.19)	1.63 (0.27)

3.3 Cognitive Test Battery Scores

This research examined whether biofeedback training was effective in increasing the cognitive profile of participants in the biofeedback group, after the duration of the intervention period, compared to the control participants. Means and standard deviations for the outcome measures in the cognitive test battery are reported in Table 6. Improvements in reaction time scores were observed across the measures, and differences in the anticipation and control measures appear to be of interest.

Table 6

Means (M) and standard deviations (SD) for the cognitive test battery outcome measures. Units are either in seconds (s), milliseconds (ms), seconds divided by 100 (s/100), efficiency (e), points scored (p), and set size (ss).

**The lower the score, the better*

Cognitive Test Battery Outcome Measure	Pre, M (SD)		Post, M (SD)	
	Biofeedback	Control	Biofeedback	Control
Working memory: Capacity (ss)	7.38 (0.77)	7.10 (0.85)	7.48 (0.85)	7.25 (0.92)
Working memory: Distractibility (ss)*	0.92 (1.13)	0.84 (0.83)	1.01 (1.06)	0.89 (0.81)
Anticipation: Performance (p)	0.96 (0.18)	1.02 (0.22)	1.24 (0.23)	1.30 (0.22)
Anticipation: Stress (ms)	0.052 (0.010)	0.055 (0.008)	0.057 (0.010)	0.060 (0.009)
Control: Motor (e)	0.64 (0.18)	0.76 (0.12)	0.55 (0.22)	0.64 (0.13)
Control: Cognition (e)	0.78 (0.09)	0.76 (0.62)	0.83 (0.10)	0.81 (0.08)
Control: Reaction time (ms)*	0.41 (0.04)	0.42 (0.03)	0.38 (0.04)	0.39 (0.03)
Attention: Alerting (s)*	0.01 (0.13)	0.02 (0.11)	0.06 (0.13)	0.06 (0.13)
Attention: Orienting (s)*	0.17 (0.14)	0.24 (0.10)	0.21 (0.16)	0.18 (0.16)
Attention: Cueing (s)*	-0.48 (0.22)	0.04 (0.18)	0.01 (0.19)	-0.15 (0.22)
Attention: Cognitive control (s)*	0.40 (0.17)	0.40 (0.17)	0.41 (0.14)	0.45 (0.25)
Attention: Reaction time (s)*	0.57 (0.04)	0.57 (0.03)	0.55 (0.05)	0.55 (0.04)
Integration: Performance (s)*	0.004 (0.04)	0.01 (0.04)	0.003 (0.04)	0.01 (0.04)
Integration: Learning (s)*	0.12 (0.03)	0.13 (0.04)	0.13 (0.03)	0.14 (0.03)

A series of 2 × 2 ANOVAs were run individually on each specific measure within the five task sections of the cognitive games. Initial tests for outliers were run using box plots. Only one outlying datum was observed. Mixed 2 × 2 ANOVAs were run on the measure which held the outlier, with the outlying case included, and not included. F values of the two outputs were compared, and it was concluded that the F values substantially differed, deeming it necessary to not include this outlying case in the subsequent final analyses. This participant's data were removed across all cognitive test battery scores (leaving N = 39, biofeedback 22, control 17).

From the ANOVAs run on the remaining data, there were significant interactions found between the pre-post tests and intervention group for the attention game (orienting: $F(1,37) = 4.27, p = .046,\ \eta_p^2 = .103$, and cueing $F(1,37) = 6.46, p = .015,\ \eta_p^2 = .149$.) In the orienting measure, the control group's scores improved in moving their spatial attention, while the biofeedback group's scores worsened. In the cueing measure, the direction was the same. When the spatial cue was false, the control group was better at disengaging their attention. The direction of these significant interactions does not support the hypothesis that the biofeedback group would outperform the control group during the second, post measure. The ANOVA results for the remaining outcome measures of the cognitive test battery showed no significant interactions, (Table 7). This suggests that overall, the biofeedback training in its current form did not lead to improvements in the cognitive profile of a subject, over a six week time period, compared to the control intervention.

Significant main effects on time (pre and post training) were found for performance in the anticipation: performance ($F(1,37) = 75.37$, $p < .0001$, $\eta_p^2 = .671$), and anticipation: stress measures ($F(1,37) = 8.94$, $p = .005$, $\eta_p^2 = .195$). Both groups improved in anticipating a target (performance), and particularly fast targets (stress). The same pre/post main effects was found for control: cognition ($F(1,37) = 8.10$, $p = .007$, $\eta_p^2 = .180$) and control: reaction time ($F(1,37) = 24.88$, $p < .0001$, $\eta_p^2 = .402$), meaning they made more efficient responses to a stimulus change (cognition) and faster responses to a similar stimulus (RT), and attention: reaction time $F(1,37) = 17.08$, $p < .0001, \eta_p^2 = .316$.), where both groups became faster in their responses.

No significant interactions or main effects were found for both the working memory and integration tasks, indicating that scores of both the control and biofeedback group were similar during pre and post-tests.

Table 7

F, P and η_p^2 time x group interaction values from the mixed ANOVAs on cognitive test battery measures

Cognitive test battery outcome measure	*F(1,37)*	*p*	η_p^2
Working memory: Capacity	0.08	.777	.002
Working memory: Distractor	0.01	.920	< .001
Anticipation: Performance	0.003	.954	< .001
Anticipation: Stress	0.00	.996	< .001
Control: Motor	0.20	.659	.005
Control: Cognition	.072	.790	.002
Control: Reaction time	0.07	.791	.002
Attention: Alerting	0.04	.847	.001
Attention: Orienting	4.27	.046	.103
Attention: Cueing	6.46	.015	.149
Attention: Cognitive control	0.43	.516	.011
Attention: Reaction time	0.03	.860	.001
Integration: Performance	0.02	.888	.001
Integration: Learning	0.30	.589	.008

3.4 NASA TLX

To get an indication of the amount of workload participants subjectively experienced during the simulator session, namely the high workload scenarios, analyses of their self-reported ratings using the NASA TLX were carried out. Participants were asked to indicate the degree to which six factors contributed to the workload they experienced in the simulator session. These factors were: mental demand, physical demand, temporal demand, performance, effort and frustration.

Data were missing for two participants in the post measure, so their data from the pre measure were removed, leaving N equal to 38 (biofeedback = 21, control = 17). Initial tests were run on the raw scores to check for any outliers (box plots). No data were removed as the box plots did not indicate any values which exceeded three interquartile ranges above the third quartile or below the first quartile.

The numbers in Table 8 illustrates a reduction in workload. The reduction in the overall mean ratings, demonstrates that the 6 factors stated above played a lesser role, contributed less, to the workload experienced by the pilot participants. Note that the reduction in overall workload was more for the biofeedback group, than for the control group. This was in line with the anticipation that the biofeedback group would report to have experienced less workload in the post-test, compared to the pre-test and compared to the control group.

Table 8

Means and SD for ratings of all 6 factors of the NASA TLX overall

Overall ratings of workload	Pre	Post
Biofeedback group	334.14 (54.32)	305.38 (81.69)
Control group	342.59 (45.73)	320.94 (69.62)
Overall	337.92 (50.17)	312.34 (75.93)

For further analysis, 2 × 2 ANOVAs were administered on the ratings of all six factors. Data from the weights part of the task were not used, as the additional data was not deemed necessary. According to Hart, (2006), the usability of raw TLX ratings has been demonstrated to be as equally useful as a weighted TLX workload score. Previous research has also chosen to solely analyse the ratings data (Li, Chiu, Kuo & Wu, 2013). We expected the biofeedback group to report to have experienced less workload in the post-test, compared to the pre-test and compared to the control group.

Sub scores of the NASA-TLX showed group x time interactions for effort (F (1,36) = 4.12, $p = .05$, $\eta_p^2 = .103$), suggesting a lower workload for the experimental group, and interactions for frustration (F (1,36) = 7.06, $p = .01$, $\eta_p^2 = .164$), indicating a higher workload for the experimental group.

For mental demand, there was a significant main effect for time (pre/post), (F (1,36) = 10.04, $p = .003$, $\eta_p^2 = .218$), suggesting both groups experienced less mental demand in the pre-test.

Temporal demand had a significant difference between groups ($F\,(1,36) = 5.23, p = .028,\ \eta_p^2 = .127$). Figure 5 illustrates how the mean ratings for temporal demand remained the same for the control group, across the pre and post-tests (pre: M = 62.18, post: M = 62.12), and the means dropped for the biofeedback group (pre: M = 61.00, post: M = 48.05). A significant interaction was not found, only a trend ($F\,(1,36) = 3.41, p = .073,\ \eta_p^2 = .087$).

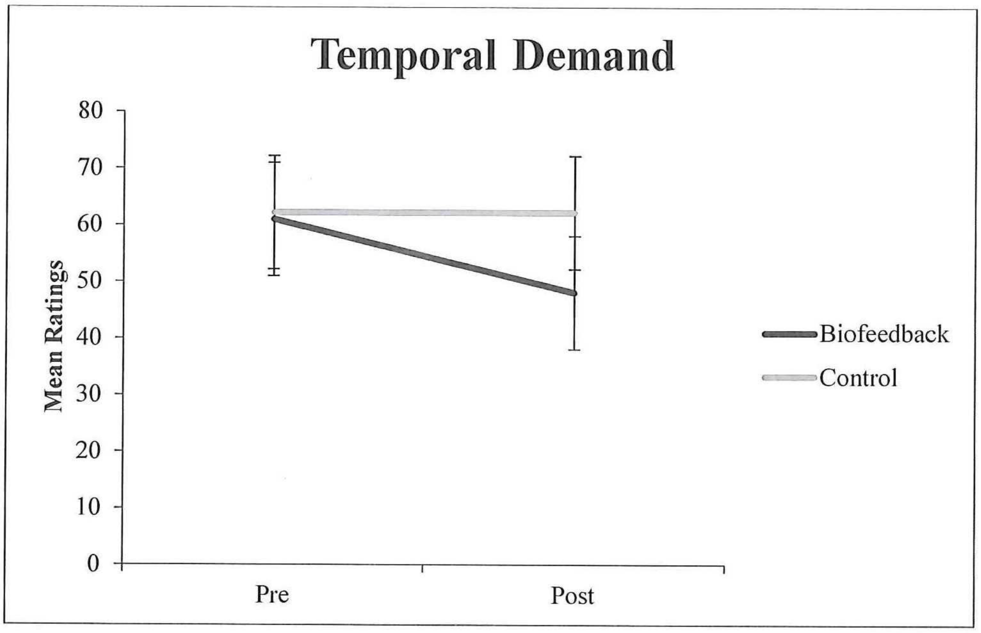


Figure 5: Temporal demand ratings of the NASA TLX for both biofeedback and control group, taken after simulator sessions in pre and post-tests.

No significant interactions or main effects were found for physical demand and performance, indicating that no physical work during the scenario, nor the pressure to perform contributed to the workload experienced by participants.

3.5 Subjective Feedback Form

At the end of their participation, the biofeedback group completed a feedback form, which contained questions regarding the training, the expectations they had and the effects that were noticed throughout their daily life. An overview of all the feedback can be found in appendix H. Many responses towards the content of the training included an increased awareness of their breathing during stressful situations and clearer thinking. One participant stated: "I thought something as simple as breathing didn't have so much influence on the way you relax and think."

Some of the replies towards the effects on the daily lives of participants suggested that the techniques were being applied to situations, where previously participants would feel stressed, such as meeting new customers. One participant commented: "I have a lower and more stable heart rate, and am able to stay calm. If I let myself go emotionally I was able to calm down and reset quicker." Another participant commented on their speed of 'reset:' "Less stressed and hurried/rushed feelings and moments. If it occurred, I could easily feel less stress especially the last week."

4 DISCUSSION

4.1 Main Findings

The research described here explored the potential benefits of biofeedback as a stress management technique for commercial pilots. The hypotheses included increased HRV and flying performance during a high workload stressful scenario, as well an increased overall cognitive profile.

After the intervention training, the biofeedback group's HRV did not differ from the control group's HRV throughout any of the four flight phases in the high workload scenarios. In fact after the training, the biofeedback group experienced more stress than the control group during take-off.

Perhaps due to the similarity in HRV levels, the flying performance of the biofeedback group was also found to not exceed that of the control group. Both groups' performance improved over time, applying less control column inputs (pitch and roll) for control of the aircraft, with improved accuracy in the post-intervention test. Less control column inputs could have been a consequence of participants' familiarity with the ILS (flew during pre and post-tests), and the simulator controls. For the majority, this experiment was the first time they had experience in the simulator used for the experiment, which had been recently acquired by EPST. Many participants commented on how different the controls felt to the other simulators they had experienced.

For the majority of the outcome measures in the cognitive test battery, no differences between the groups were evident. Of the two measures in which differences between groups were observed, it did not support the hypothesis that the biofeedback group would outperform the control group, as it was the control group who outperformed the biofeedback group on these measures (the attention game measures: orienting & cueing). So overall in its current form, the biofeedback training did not lead to improvements in the cognitive profile of a subject, over a 6 week time period, compared to the control group. In this situation a correction of the alpha for multiple tests would have been appropriate, yet this was not applied as the current study was intended to explore the possibilities. That is, no definite inferences are drawn yet.

The biofeedback group reported positive changes in their personal lives, indicating that the biofeedback training had noticeable effects. One participant noted: *"I now know again how it is to be free of stress."* Others reported that they were also less affected by stress, had improved sleep cycles, were better able to reset more quickly after stressful events and experienced improvements in their relationship with others. These responses could suggest that the benefits of the biofeedback at a perceived level were greater than those that could be measured through the dependent variables. Such benefits of the biofeedback training could not be quantified however, as it was out of the scope of this study.

Participants in the biofeedback group indicated via the NASA TLX ratings that they felt more frustrated during the simulator session in phase 3, compared to the control group and to phase 1. This could perhaps be in affiliation with the above finding that the biofeedback group had lower HRV (so more sympathetic system stress response) at the take-off phase, leading to a likelihood of having pressure to perform or a desire to perform well. A possible cause could have been due to the biofeedback group guessing they were in the experimental group. On the other hand, there was a difference in HR at the start of the recovery phase, which suggests that the stressful event had a smaller effect on the biofeedback group. It was difficult to physically isolate the experimental and control group from one another, given that they were from the same flight school (there were established companionships and continued practice in the flight school simulators together). Thus, there was inevitable discussion amongst one another about their experience in the research, despite signing an informed consent form, which instructed participants that discussion of the study's content was prohibited.

An encouraging finding was that the biofeedback group reported that less effort was required in the post-test simulator session, compared the control group, and to the pre-test session, although it is difficult to conclude that this is due to the biofeedback intervention. The reduction in effort may have been due to the familiarity with the route and/or the experimental setting and researchers, who had been present throughout the intervention training sessions with the biofeedback group.
Another finding in the NASA TLX that may also have been due to the familiarity of the route flown in phase 3, was the biofeedback group reporting less temporal demand, compared to the controls.

Participants held the knowledge from phase 1 that they should be able to control the aircraft and land it in time safely in the particular route above Dusseldorf. However this does not explain the difference between the biofeedback and control group.

Although there were some improvements across time throughout all the measures (HRV, flight performance and cognitive profile), this did not differentiate between the groups, meaning the biofeedback group did not improve more so than the control group. Thus, without isolating the biofeedback group from the control group, it cannot be concluded that the improvement in performance was due to the HeartMath training. This lack of effectiveness of the intervention could be down to numerous elements, as will be explored below.

It should be noted that in this situation a correction of the alpha for multiple tests would have been appropriate, but that this was not applied, as the current study was intended to explore the possibilities. That is, no definite inferences have been drawn yet.

4.2 Use of Inexperienced, Newly-Licensed Pilots

The pilot population sample was limited to pilots who had acquired their licenses very recently, or within the last few years, with the majority of the participants in their 20s. These participants had not yet experienced working as a full time pilot over a long period of time (10+ years). More experienced pilots perhaps would have recognised the importance of stress management, due to an increased likelihood of experiencing stress during their careers. Thus, the benefits and the need for the training might have been understood by experienced pilots. Whited, Larkin & Whited (2014) recently discovered that a population sample of university aged students yielded no significant improvements in HRV, in a similar design (2x2, group x time), whilst investigating the efficacy of coherence training using an emWave device. This research supports the possibility that age and life experience may be a limiting factor to the success of the biofeedback training.

In McCraty and Atkinson's (2012) research, the police officers were older with a mean age of 39, and had years of experience on the job (almost half of them with 16-30 years of experience). With more years under the belt 'on the beat,' there was an increased likelihood that the officers had experienced stress, and perhaps were stressed already before the intervention began. The younger, newly qualified pilots in the current research did not have as much, if any experience in the field, and were not working at the time, feeling jet lagged like an operational pilot may be. Future research could attempt to test established, operational pilots. This was not attempted in the current research, due to the difficulty of recruiting more than 50 volunteers with an ATPL over a nine-week period.

4.3 Limitations of Methodology

4.3.1 HeartMath Training

A general training protocol as designed by HeartMath was used for familiarising participants with the emWave and with stress reduction techniques. Some elements of the training led to the pilots not feeling the need for the training to make them a better pilot. As an attempt to reduce the amount of effort required by participants, the trainers indicated that they could practice during daily activities such as driving their car or brushing their teeth. This was later found to be applied by participants. For maximum results reducing stress levels with the emWave, however, participants should practice in a quiet and controlled setting.

Participants in the biofeedback group were instructed to train three times a day, for a minimum of five minutes per session. Similar to learning to drive a vehicle, this training implied that more is better. Previous studies using the emWave have shown significant improvements with 28-day interventions, where participants practiced for 15 minutes per day (Ginsberg et al, 2009, Lemaire et al, 2011).

It was challenging to motivate the participants to train regularly, perhaps due to them having a lack of intrinsic motivation. From the feedback form that participants filled out upon completion of their participation, a summary of the total amount of reported time that participants trained was gained. The total required time of training based on the above set of practice guidelines, over the six weeks for 22 participants was 13,860 minutes, whereas the actual amount participants trained was 9,191 minutes. The required amount of training per participant was 630 minutes over the course of six weeks, whereas the actual average was 418 minutes per participant. Two out of the 22 participants trained above the required amount of which one trained 1,031 minutes. Removing this participant brings down the average to 388 minutes per participant compared to the 630 minutes required. Since the initial analysis did not yield any significant results, flight performance data of the top (amount of time spent practicing) 33 percent of the experimental group was compared with the control group data, without generating significant findings (and therefore not reported in the results section). Specifically the top 33 percent, who appeared to believe in the training the most, based subjectively on their feedback forms, seemed to experience performance anxiety during phase 3 of the experiment, as previously discussed in their frustration levels from pressure to perform. This could be related to the observer effect, i.e. Hawthorne effect (Cook, 1962), where participants showed a change in performance, not due to the changing conditions, but due to the awareness of external observation.

Biofeedback participants' perceived effects of the biofeedback training (appendix H) indicate that they benefitted from the training in day to day life, by being better able to regulate their stress levels than before the training. Although the scenarios were designed to test participants at their highest cognitive capacities during the operation of an aircraft, it is possible that the scenarios were not difficult enough, thus not subjecting the participants to the highest possible stress levels. This may have contributed to the lack of significance in the final results. Most likely the cause was not only the scenario, but also as suggested before, the participants getting used to being tested in an experimental environment, resulting in an increase of performance over time. It is also possible that young airline pilots have pre-existing higher levels of HRV and thus have a healthy physiological response to stress. A population with a lower initial HRV might demonstrate a greater improvement from the biofeedback training.

4.3.2 Resources

Budget was made available through TU Delft for the acquisition of equipment and for the heart data analysis. Combined with the cost free facilities provided by EPST, the free training and equipment by HeartMath Benelux, this study became feasible. Nevertheless, budget for incentive or travel expenses for participating students may have increased their motivation to train more.

The original project plan contained agreed upon cooperation with third parties to develop specialised elements of this study such as the flight scenarios and simulator data protocols, however they were unavailable. Without the help of these experts, researchers had to develop the scenarios and the data analysis protocol themselves, which was not of the highest quality, had the experts helped with the development. The initial design incorporated a trial run of the experiment, which included the generation of a full data readout. Researchers planned to process these readouts with Python™, validate the model and create an integrated objective simulator output model. The researchers received documentation about the simulator architecture after starting phase 1, making this impossible. The lack of documentation created additional time pressure on the scenario design process before phase 1, thus influencing its quality.

The size of the project caused the researchers to make decisions based on personal experience and estimations. The project was fully established by MSc students, their mentors and novice pilots. No experienced aviation professionals were available to guide the process. A future project should incorporate the knowledge of professional pilots and experienced simulator technicians to develop the scenario and the data protocol, in order to guarantee the highest level of data fidelity.

4.3.3 Scenario Development

From the NASA TLX data gathered, no significant interactions or main effects were found for physical demand or performance, indicating that neither physical demand nor the pressure to perform during the scenario contributed to the workload experienced by participants. This confirms that flying an aircraft is not a physically demanding job and the physiological response is mostly due to mental demand (Wilson, 2002).

The initial plan was for four identical sections to be flown by participants. One low and one high workload scenario, for both phases. This would have allowed for a matchup of the physiological response per stressor, possibly resulting in a more accurate comparison between the high and low workload settings, with the scenarios consisting of two identical routes, one with and one without stressors. The scenario without stressors could have functioned as a baseline, where the physiological response is compared with the induced physiological response of the scenario where stressors were included. The low workload scenario would then would then be used as a covariate, to determine the baseline of subjects' flying ability, thereby removing most contributions of individual differences. The post biofeedback measurement (phase 3) could have then determined the relative improvement between a high and low workload setting. Within the current study this approach was not possible, due to a lack of time and professional support during the preparation stage.
The main concern of the flight academy was to prevent negative training effects in its pilots, thus demanding that the scenarios were not surpassing regular training protocol. This limited the scenario development to the failures as stored in the simulator. Ideally, in order to generate a surprise or startle effect to induce acute stress, participants should be subjected to unforeseen emergencies they have never before been experienced to.

Researchers held initial intentions of having two distinguishable yet similar scenarios, called scenario A and scenario B, ready at the start of the experiment. This was going to balance the order, alternating A and B over both phases. Not only could this design have controlled for any learning effects, it could also limit the information exchanged between participants about the details of the scenarios, thus maintaining a higher level of novelty.

4.3.4 Confederate Inconsistencies During Simulator Sessions

Pilot monitors (PM) and student instructors (SI) were selected out of the pilot volunteers. These individuals appeared to have large variety in personality traits such as openness, conscientiousness, extraversion, agreeableness and neuroticism (John, Donahue, & Kentle, 1990) resulting in observable variations of stress levels, when participants were paired to different SIs or PMs. A future study should aim to control for this variation. The atmosphere in the simulator varied from day to day, ranging from informal to highly informal, depending on the instructors and/or pilot monitors. This also seemed to affect stress levels. This could be prevented by having a single confederate as SI and a single as PM, for all measurements. The scripts for SIs as well as PMs were not always followed rigorously due to long working hours in the simulator, perhaps causing individual differences in the flight performance of participants.

4.3.5 Technical Complications with Equipment

Data from the 5th of April was not recorded for unknown reasons, despite normal protocol being followed, thus resulting in a loss of data of seven participants during phase 1. Another technical complication was the plotted flight paths of participants, which contained sharp angles instead of smooth lines. This is possibly due to round off errors of the longitude and latitude, as programmed in the simulator architecture. These stepwise lines may influence the final grading of the flight path precision.

Participants experienced problems with the biofeedback equipment too, such as malfunctioning sensors and main units, connectivity issues with PC or Mac, and unrecorded data. This potentially influenced the motivation levels and use of the emWave. It was planned to track coherence levels, scores and thus motivate participants who were not training as well but without the fully functioning equipment this was not possible. Secondly in order to reduce the pressure on participants, the trainers allowed participants to train without the emWave which made tracking of their progress difficult.

4.4 Future Research Suggestions

Repetition of the project with dedicated professionals with experience on the job could significantly improve the output reliability of the dependent variables. The biofeedback training should be developed specifically for aviation professionals, which directly relates to their work in the cockpit. Previous studies which trained police officers (McCraty and Atkinson, 2012) and veterans with PTSD (Ginsberg et al, 2009) did exactly that. Specifically designed interventions for the target group could improve intrinsic motivation. The input for the scenario development as well as the flight performance motivation could add to the potential success of a future study. The scenarios should be set up to match the physiological response with the various flight segments. Wilson (2002), was able to synchronise each flight phase, from take-off to landing, with a physiological response. In order to test the physiological response of pilots to acute stressors (i.e. emergency situations), the same method should be applied in future scenario design for the effectivity of biofeedback training. Future research should incorporate an integrated objective simulator rating protocol, where objective performance can directly be determined from the simulator output, thus eliminating complicated calculations through external software such as Python™. Validating the data analysis protocol, the evoked physiological responses and training the confederate pilots is essential for a high fidelity simulation.

It could suffice to reduce the experiment to a single phase measurement, i.e. post biofeedback training, phase 3 only. This would prevent the presence of learning effects and would limit the possible exchange of information between the control and the biofeedback group.

The initial design contained a method to assess relative degrading of performance due to increasing workload, per participant. This within-subject analysis could provide more in depth knowledge, as to which participants are more prone to stress and could thus benefit more from the biofeedback training. Moreover, it increases the statistical power.

Hansen, Johnsen and Thayer (2003) demonstrated a relationship between resting HRV and cognitive performance. Those with a higher HRV, performed better on a working memory test, compared to those with a low HRV. The former group also demonstrated superior executive functioning, which involves aspects such as planning, working memory, selective and sustained attention (Robbins, 1996). Future research could look at resting heart rate, to determine if there is a difference in the effectiveness of biofeedback training relative to subjects' resting HRV.

Other forms of biofeedback, which have proven to increase executive and cognitive functions, should also be investigated with regards to pilots' resilience. Neurofeedback can increase IQ scores (Thompson & Thompson, 2012), autogenic feedback training can improve pilot performance in emergency situations (Kellar et al, 1993) and direct transcranial stimulation, where low voltage electric currents are run through the brain, can improve memory functions (Fregni et al, 2005). Diet, supplements, sleep training programs and the like are all important factors of pilot's stress resilience. More research into interventions which improve cognition and stress response, may significantly improve pilot resilience in high workload situations.

4.5 Support of Hypotheses

In its current form, the biofeedback training did not lead to the hypothesised effects. Neither the physiological response, nor the cognitive profile, nor the flight performance of participants in the biofeedback group improved. Without a difference in HRV between the groups (as observed), it was inevitable that there would be no differences in flight performance or cognitive profile. Previous studies using different populations for this training have been successful, where police officers showed significant improvements in their physiological response to a simulated stress evoking scenario, after using the emWave over a 28 day period (McCraty & Atkinson, 2012). Another study with the emWave, this time with physicians, showed a significant stress reduction, also over a 28 day trial period (Lemaire et al, 2011).

4.6 Possible Causes

The current study was unable to show relationships between HRV training and a reduced stress response. This is most likely due to the participants' background, who were young, self-assured pilots, who had not yet experienced any occupational stress. The training required the introduction of a daily habit, which was asked of participants, rather than being chosen by participants to take on.

The job of a commercial airline pilot is considered as one the most stressful ones in the world (Bourne & Yaroush, 2003). Pilots are responsible for many souls on board, including their own, as soon as setting foot on the aircraft. Moving at high velocities in a metal cocoon, high up in the sky or low to the ground, and knowing that the smallest mistake could have lethal consequences, is undoubtedly a fertile ground for distress. The safe continuation of the operation is not only dependant on the pilot's skills, but even more so on third parties who are safeguarding the operation (ATC for guidance, co-pilot for judgement, mechanics for maintenance of systems). These factors could result not only in acute stress but also chronic stress accumulating in the body. The young pilots in this study have not yet been in contact with these life lessons, which could be a cause for their scepticism and minimal training efforts. Experienced pilots with more awareness of stress on the job may have more intrinsic motivation to practice.

Another reason why the training did not support the hypotheses, could be accounted to the informal settings of the experiment, since the tests were executed at the flight academy, where participants were in a familiar environment and working with their friends and colleagues (student instructors and pilot monitors). It was observed that the participants were not genuinely stressed and there was no pressure to perform, such it being an examination or having the managing director present. The mindset of participants was not as if they were in a real life situation, thus possibly affecting how immersed they were resulting in lower measurable stress levels.

One could also reason, considering some significant improvements of both groups over time, that the phase 3 scenario did not evoke as much stress as was intended, especially compared to the first scenario. The second scenario was most likely not novel enough, thus not causing identical amounts of acute stress as observed during the first phase of the experiment. The overall relaxed atmosphere during the second measurement, due to familiarity of the procedure by the researchers and the confederate pilots, could also have resulted in weaker physiological responses. However, when visually analysing the heart data graphs, responses to the most cognitive demanding sections of the flight are evident (appendix I). This indicates that the scenarios were evoking higher stress levels during these sections, compared to less demanding parts of the flight. It would also not explain why there were no differences between the control and the biofeedback group.

4.7 Hypotheses for Future Research

The findings above can be combined into new hypotheses which could be tested using a similar approach, identical tools and equipment, as used in this study.

Hypothesis 2.1: Pilots with a lower resting HRV will show a stronger increase in flight performance, compared to those with a higher resting HRV, after HRV biofeedback training. This could be assessed through a within-subject design, to determine relative degrading of flight performance due to increasing workload, per individual.

Hypothesis 2.2: Inexperienced pilots will demonstrate higher flight performance improvements compared to experienced pilots, as a result of HRV biofeedback training. The differences between newly qualified and experienced pilots could be explored further. Yao et al. (2008) demonstrated that inexperienced pilots encounter a higher mental workload compared to experienced pilots.

Hypothesis 2.3: HRV biofeedback intervention given during APTL training will increase the learning curve of pilots.

Biofeedback training may reduce the amount of experienced stress during pilot training, thus accelerating the learning process by reducing performance anxiety (Wells, Outhred, Heathers, Quintana & Kemp, 2012). Inexperienced pilots have higher heart rates compared to expert pilots, during demanding flight sections, thus experience higher mental load (Yao et al, 2008). By defining which individual is more prone to performance anxiety and subsequently subjecting to stress reduction training, flight training could be optimised.

Hypothesis 2.4: Mindfulness training will increase flight performance of airline pilots.

The subjective feedback reported by participants, indicates that in general they became more aware of their stress levels, body and general wellbeing. Self-awareness and self-regulation are qualities linked to mindfulness. Mindfulness is described as "bringing a quality of attention to daily activities and moments" (Kabat-Zinn & Hanh, 2009). A mindfulness study within the Norwegian Air force F-16 squadron suggests that 'being in the moment' improves flying capacities of high performance individuals (Meland & Fonne, 2012). Since biofeedback training with the emWave is a form of mindfulness, it may improve the same qualities as reported by the fighter pilots. Mindfulness studies suggest, that training may accelerate the four step learning process from unconscious incompetent, to unconscious competent (Darwin & Melling, 2011).

Hypothesis 2.5: HRV biofeedback training will improve commercial airline pilots' ability to cope with stress accumulated through irregular working hours.

Long distance commercial pilots are continuously subjected to varying time zones, irregular working hours and sleep disturbances. The results of Kim et al. (2002) suggest that "shift-workers suffer more from physical and psychological distresses, sleep problems and stress than non-shift workers." Sleep disturbances lead to stress (Åkerstedt, Fredlund, Gillberg & Jansson, 2002). Biofeedback training can improve sleep quality (Kim et al, 2002), thus reduce stress levels, thus may improve flight performance. Negative consequences of shift work for airline pilots may be reduced through biofeedback training.

Hypothesis 2.6: Pilots who have received HRV biofeedback training will have improved interpersonal skills.

These skills are essential for cooperating in a cockpit to make sound judgement of situations and be acceptant to alternative points of view. Emotional intelligence (Goleman, 2006) is highly important from a crew resource management perspective. Emotional intelligence does not only increase due to stress reduction, it also improves due to increased self-awareness, self-regulation and empathic abilities, which may be achieved through HRV biofeedback (Bradley et al., 2010).

Hypothesis 2.7: Other forms of biofeedback training will improve pilots' abilities to cope with acute and chronic stress, thus improve overall system resilience to disturbances.

Apart from HRV biofeedback training, cognitive functioning of pilots can be improved through neurofeedback training (Hanslmayr, Sauseng, Doppelmayr, Schabus & Klimesch, 2005, Hammond, 2007) and direct transcranial stimulation (Fregni et al, 2005), all leading to a more psychological and cognitive resilient pilot. In depth literature studies are required to support the further definition of this hypothesis.

4.8 Aviation Industry Implications

Since the initial design by the Wright brothers, technical improvements have been inspired from lessons learned through accident analysis. Instead, we should look at the operator, the human in the cockpit, and design the aircraft around the current knowledge of psychophysiology and neurobiology, applying a human centred design. Cockpit design should incorporate the diversity in personalities but also the diversity in stress response of pilots to emergency situations. Aircraft designers should approach the technical design process from a completely new angle. Instead of a top down philosophy where the brains develop aircraft architecture and cockpit ergonomics, it should incorporate a bottom up approach from the perspective of the operator, such as applied in lean production process optimisations (Forrester, 1995). This could very well lead to a new blueprint for aircraft design, which is far 'out of the box' of what we currently use in the industry. Knowledge based engineering (Studer, Benjamins & Fensel, 1998), which has proven to be successful up until today, may not suffice to maintain and improve current safety levels. Instead, a collaborative based engineering should be included in the design process.

An interesting finding of this project was the resistance experienced during the design phase. Some experts within the fields of aircraft design, cognitive psychology and flight training were sceptical about the theoretical framework, but changed their perspective towards the end of the project. It was difficult to introduce a new perspective, which did not relate to the current analytic approach of pilot training. Academics of subsequent research fields approach the challenges in the aviation industry from an individual perspective, not always integrating a holistic view. In order to continue and surpass the current levels of safety which are needed to support the predicted growth in air transport, it will be essential to create a cohesive cooperation between research fields in an open source format. A format where researchers function as a team, where the qualities of the individual are amplified by the qualities of their team members.

4.9 Conclusions

This explorative study could function as an inspiration for future projects, dedicated to measuring and managing the stress response of airline pilots in emergency situations. The explorative nature of this research and the steps taken, may lay the foundations for a new level of human factors research within the aviation industry, unlike anything attempted before in the field of commercial aviation pilot training. New parameters were measured, by using existing equipment and methodologies, which have been proven successful in the past. This joint project between major universities, the national aerospace laboratory, one of the best flight academies of the Netherlands and Next Generation Aircraft, set an example for future collaborations.

A group of young aviation pilots has been exposed to the possibilities of stress reduction through self-regulation, tools they otherwise may never have learned. Eyes have been opened, lives have been changed. The subjective feedback of participants indicates that the training has a broad applicability. Not just for improved flight performance but also as a transferable skill for their current occupations. Less stress at work, at school, less tired, more in control of their emotional state, better sleep, increased awareness of their physiological state and better relationships are only some of the reported benefits. These responses could suggest that the benefits of the biofeedback at a perceived level were greater than those that could be measured through the dependent variables.

REFERENCES

Airbus (2005-2007). Flight operations briefing notes, human performance

Åkerstedt, T., Fredlund, P., Gillberg, M., & Jansson, B. (2002). Work load and work hours in relation to disturbed sleep and fatigue in a large representative sample. *Journal of psychosomatic research, 53*(1), 585-588.

Alabdulgader, A. A. (2012). Coherence: A novel non-pharmacological modality of lowering blood pressure in hypertensive patients, *Global Advances in Health and Medicine, 1* (2), 54-62

Amalberti, R. (1999). Automation in aviation: A human factors perspective. *Handbook of aviation human factors*, 173-192.

Amalberti, R. (2001). The paradoxes of almost totally safe transportation systems. *Safety science*, *37* (2), 109-126.

Amalberti, R. (2013). *Navigating Safety: Necessary Compromises and Trade-offs--Theory and Practice*. Heidelberg: Springer.

Andreassi, J. (2000). *Psychophysiology*, 4th Ed., Lawrence Erlbaum Associated, Inc., New Jersey

Beauchaine, T. (2001) Vagal tone, development, and Grey's motivational theory: toward an integrated model of autonomic nervous system functioning in psychopathology, *Developmental Psychopathology, 13* (2), 183-214

BOEING, Statistical Summary of Commercial Jet Airplane Accidents 2012, August 2013. Retrieved on the 7th of August from: http://www.boeing.com/news/techissues/pdf/statsum.pdf

Bor, R., & Hubbard, T. (Eds.). (2006). *Aviation mental health: Psychological implications for air transportation*. Ashgate Publishing, Ltd..

Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile (2012). Final report on the accident on 1st June 2009 to the airbus a330-203 registered F-GZCP operated by Air France flight AF447 Rio de Janeiro – Paris

Bouso, J. C., González, D., Fondevila, S., Cutchet, M., Fernández, X., Barbosa, P. C. R., ... & Riba, J. (2012). Personality, psychopathology, life attitudes and neuropsychological performance among ritual users of ayahuasca: a longitudinal study. *PloS one*, *7*(8), e42421.

Borst, C., Ecological approach to pilot terrain awareness. Delft repository: Borst, 2009

Bradley, R. T., McCraty, R., Atkinson, M., Tomasino, D., Daugherty, A., & Arguelles, L. (2010). Emotion self-regulation, psychophysiological coherence, and test anxiety: results from an experiment using electrophysiological measures. *Applied psychophysiology and biofeedback*, *35*(4), 261-283.

Cowings, P. S., & Toscano, W. B. (1993) *Autogenic-Feedback Training (AFT) as a preventative method for space motion sickness: Background and experimental design*, (NASA Technical Memorandum Number 108780). Moffett Field, CA: National Aeronautics and Space Administration, Ames Research Center.

Cowings, P. S., Kellar, M. A., Folen, R. A., Toscano, W. B., & Burge, J. D. (2001). Autogenic feedback training exercise and pilot performance: Enhanced functioning under search-and-rescue flying conditions, *The International Journal of Aviation Psychology, 11* (3), 303-315. DOI: 10.1207/S15327108IJAP1103_04

Cowings, P. S., Toscano, W. B., Casey, C., & Hufnagel, J. (2005) Autogrenic feedback training exercise: A treatment for airsickness in military pilots, *The International Journal of Aviation Psychology, 15* (4), 395-412.

Cook, D. L. (1962). The Hawthorne effect in educational research. *Phi Delta Kappan*, 116-122.

Dahlgren, A., Kecklund, G., Theorell, T., & Åkerstedt, T. (2009). Day-to-day variation in saliva cortisol—relation with sleep, stress and self-rated health. *Biological Psychology, 82*(2), 149-155.

Dahlstrom, N., & Nahlinder, S. (2009). Mental workload in aircraft and simulator during basic civil aviation training. *The International journal of aviation psychology, 19* (4), 309-325.

Dekker, S. (2004). *Ten questions about human error: A new view of human factors and system safety.* CRC Press.

Dekker, S. W. A., & Johansson, B. (2000). JAR-FCL and pilot knowledge needs. *Report from the Swedish Centre for Human Factors in Aviation (now renamed Swedish Network for Human Factors). Linköping Institute of Technology, IKP/IAV, SE-581, 83.*

Dekker, S. W., & Woods, D. D. (2002). Maba-maba or abracadabra? Progress on human–automation co-ordination. *Cognition, Technology & Work,4* (4), 240-244.

European Aviation Safety Agency (2013). Develop an Automation Policy. Retrieved on August 6, 2014 from: http://www.easa.europa.eu/system/files/dfu/sms-docs-EASp-SYS5.6---Automation-Policy---28-May-2013.pdf

Eurocontrol (2013). Safety whitepaper - safety I and safety II, towards a Resilience Engineering perspective. Retrieved on August 6, 2014 from http://www.eurocontrol.int/sites/default/files/content/documents/nm/safety/safety_whitepaper_sept_2013-web.pdf

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience, 14* (3), 340-347

FAST (2006). Summary of the FAST Analysis of the Top Priority Area of Change "Increasing Reliance on Flight Deck Automation"

Ferris, T., Sarter, N., & Wickens, C. D. (2010). Cockpit Automation. 479-503

Finni, T., Haakana, P., Pesola, A. J., & Pullinen, T. (2014) Exercise for fitness does not decrease the muscular inactivity time during normal daily life, *Scandinavian Journal of Medicine and Science in Sports, 24,* 211-219, doi: 10.1111/j.1600-0838.2012.01456.x.

Forrester, R. (1995). Implications of lean manufacturing for human resource strategy. *Work Study, 44*(3), 20-24.

Fregni, F., Boggio, P. S., Nitsche, M., Bermpohl, F., Antal, A., Feredoes, E., & Pascual-Leone, A. (2005). Anodal transcranial direct current stimulation of prefrontal cortex enhances working memory. *Experimental Brain Research,166*(1), 23-30.

Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity, not quality: The relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin & Review, 17* (5), 673-679. doi:10.3758/17.5.673

Giardino, N. D., Chan, L., & Borson, S. (2004). Combined heart rate variability and pulse oximetry biofeedback for chronic obstructive pulmonary disease: preliminary findings. *Applied Psychophysiological Biofeedback, 29* (2), 121–133

Giggins, O. M., Persson, U. M., & Caulfield, B. (2013). Biofeedback in rehabilitation, *NeuroEngineering and Rehabilitation, 10* (60), doi:10.1186/1743-0003-10-60

Ginsberg, J. P., Berry, M. E., & Powell, D. A. (2009). Cardiac coherence and posttraumatic stress disorder in combat veterans. *Alternative therapies in health and medicine, 16* (4), 52-60.

Goleman, D. (2006). *Emotional intelligence*. Random House LLC.

Hammond, D. C. (2007). What is neurofeedback?. *Journal of Neurotherapy,10*(4), 25-36.

Hanslmayr, S., Sauseng, P., Doppelmayr, M., Schabus, M., & Klimesch, W. (2005). Increasing individual upper alpha power by neurofeedback improves cognitive performance in human subjects. *Applied psychophysiology and biofeedback, 30*(1), 1-10.

Hart, S. G. (2006). Nasa-task load index (nasa-tlx); 20 years later. Technical report, NASA-Ames Research Center.

Hobbs, A. (2004). Human factors: the last frontier of aviation safety?. *The International Journal of Aviation Psychology, 14* (4), 331-345.

Hollnagel, E. (Ed.). (2003). *Handbook of cognitive task design.* CRC Press.

Hollnagel, E. (2011). Epilogue: RAG: The resilience analysis grid. *Resilience engineering in practice: a guidebook,* 275-96.

Jacobson, S. R. (2010). Aircraft Loss of Control Causal Factors and Mitigation Challenges. *American Institute of Aeronautics and Astronautics, 8007,* 2-5.

Jensen, R. S. (1997). The boundaries of aviation psychology, human factors, aeronautical decision making, situation awareness, and crew resource management. *The international journal of aviation psychology, 7*(4), 259-267.

John, O. P., Donahue, E. M., & Kentle, R. (1990). 'The "Big Five. Factor Taxonomy: Dimensions of Personality in the Natural Language and in Questionnaires." In *Handbook of Personality: Theory and Research,* ed. Lawrence A. Pervin and Oliver P. John, 66-100.

Kabat-Zinn, J., & Hanh, T. N. (2009). *Full catastrophe living: Using the wisdom of your body and mind to face stress, pain, and illness.* Random House LLC.

Kahneman, D. (2011). *Thinking Fast and Slow.* New York: Farrar, Strauss and Giroux.

Kihlstrom, J. F. (1987). The cognitive unconscious. *Science, 237* (4821), 1445-1452.

Kim, Y. G., Yoon, D. Y., Kim, J. I., Chae, C. H., Hong, Y. S., Yang, C. G., Kim, J. M., Jung, K. Y. & Kim, J. Y. (2002). Effects of health on shift-work: general and psychological health, sleep, stress, quality of life. *Korean Journal of Occupational and Environmental Medicine, 14*(3), 247-256.

Kuo, T. B., Lai, C. J., Huang, Y. T., & Yang, C. C. (2005). Regression analysis between heart rate variability and baroreflex-related vagus nerve activity in rats, *Journal of Cardiovascular Electrophysiology, 16* (8), 864–869

Lemaire, J. B., Wallace, J. E., Lewin, A. M., de Grood, J., & Schaefer, J. P. (2011). The effect of a biofeedback-based stress management tool on physician stress: a randomized controlled clinical trial. *Open Medicine, 5* (4), 154-163

Li, W-C, Chiu, F-C., Kuo, Y-.S, & Wu, K-J (2013). The investigation of visual attention and workload by experts and novices in the cockpit, *Engineering Psychology and Cognitive Ergonomics, 8020*, 167-176

Lloyd, Brett & Wesnes, (2010). Coherence training in children with ADHD

Lupien, S.J, Gillin, C.J., & Hauger, R.L (1999). Working memory is more sensitive than declarative memory to the acute effects of corticosteroids: A dose-response study in humans. *Behavioral Neuroscience, 113,* 420-430.

Lupien, S. J., Maheu, F., Tu, M., Fiocco, A., & Schramek, T. E. (2007). The effects of stress and stress hornomes on human cognition: Implications for the field of brain and cognition, *Brain and Cognition, 65*, 209-237. doi:10.1016/j.bandc.2007.02.007

Magnusson, S., & Berggren, P. (2002, September). Dynamic Assessment of Pilot Mental Status. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 46, No. 24, pp. 1997-2001). SAGE Publications.

Malliani, A., Lombardi, F., & Pagani, M. (1994). Power spectrum analysis of heart rate variability: a tool to explore neural regulatory mechanisms, *British Heart Journal, 71* (1), 1-2

Meland, A., & Fonne, V. (2012). Effects of a 12 month mindfulness based mental training intervention on F-16 fighter pilots. *Aviation, Space, and Environmental Medicine, 83*(3).

McCraty, R., & Atkinson, M. (2012). Resilience training program reduces physiological and psychological stress in police officers, *Global Advances in Health and Medicine, 1* (5), 42-64

McCraty, R., Atkinson, M., & Tomasino, D. (2003). Impact of a workplace stress reduction program on blood pressure and emotional health in hypertensive employees, *The Journal of Alternative and Complimentary Medicine, 9* (3), 355-369

McCraty, R, Atkinson, M., Tomasino, D., & Bradley, R. T. (2009). The coherent heart. Heart-brain interactions, psychophysiological coherence, and the emergence of system-wide order, *Integral Review 5* (2), 10-115

Mohrmann, J. F. W. (2013). Investigating flight crew recovery capabilities from system failures in highly automated fourth generation aircraft. *Delft repository*

Neubert, F. -X., Mars, R. B., Buch, E. R., Olivier, E., & Rushworth, M. F. S. (2010). Cortical and subcortical interactions during action reprogramming and their related white matter pathways. *Proceedings of the National Academy of Sciences, 107* (30), 13240-13245. doi:10.1073/pnas.1000674107

Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *Systems, Man and Cybernetics, IEEE Transactions on, 3*, 257-266.

Rasmussen, J., & Vicente, K. J. (1989). Coping with human errors through system design: implications for ecological interface design. *International Journal of Man-Machine Studies, 31* (5), 517-534.

Reason, J. (1990). *Human error*. Cambridge university press.

Robbins, T. W. (1996). Dissociating executive functions of the prefrontal cortex, *Philosophical Transactions of the Royal Society B: Biological Sciences, 351,* 1463–1471

Slovic, P. (1999). Trust, emotion, sex, politics, and science: Surveying the risk-assessment battlefield. *Risk analysis, 19*(4), 689-701.

Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge engineering: principles and methods. *Data & knowledge engineering, 25*(1), 161-197.

Stoop, J. A. (2013a). To certify, to investigate, or to engineer, that is the question. *Proceedings of the 5th annual Resilience Engineering Association conference in Soesterberg, NL.*. Retrieved August 6, 2014 from: http://www.rea-symposium.org/download/symposium2013/Stoop%20(REA% 20 2013).%20To%20certify,%20to%20investigate%20or%20to%20engineer,%20that%20is%20the%20question.pdf

Stoop, J. A. (2013b). Towards a failsafe flight envelope protection; The recovery shield. In *Advances in Risk and Reliability Technology Symposium* (p. 160).

Taelman, J., Vandeput, S., Spaepen, A., & Van Huffel, S. (2009). Influence of mental stress on heart rate and heart rate variability, *4th European Conference of the International Federation for Medical and Biological Engineering, IFMBE Procedings, 22*, 1366-1369

Thackray, R. I. (1988). *FAA Performance Recovery Following Startle; A Laboratory Approach to the Study of Behavioural Response to Sudden Aircraft Emergencies*, Oklahoma City, OK: Civil Aeromedical Institute, Federal Aviation Administration.

Thompson, L., & Thompson, M. (1998). Neurofeedback combined with training in metacognitive strategies: effectiveness in students with ADD. *Applied psychophysiology and biofeedback, 23*(4), 243-263.

Tiggelaar, B. (2010). *Dromen, Durven Doen*. Spectrum.

Werrbach, M. (2014). Fight, Flight or Freeze: The Stress Response. *Psych Central*. Retrieved on August 4, 2014, from http://psychcentral.com/blog/archives/2014/07/31/fight-flight-or-freeze-the-stress-response/

Wells, R., Outhred, T., Heathers, J. A., Quintana, D. S., & Kemp, A. H. (2012). Matter over mind: a randomised-controlled trial of single-session biofeedback training on performance anxiety and heart rate variability in musicians. *PloS one*, *7*(10), e46597.

Whited, A., Lerkin, K. T., & Whited, M. (2014). Effectiveness of emWave biofeedback in improving hear rate variability reactivity to and recovery from stress, *Applied Psychophysiological Biofeedback, 39*, 75-88. doi: 10.1007/s10484-014-9243-z.

Wilson, G. F. (2002). An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *The International Journal of Aviation Psychology, 12* (1), 3-18. doi: 10.1207/S15327108IJAP1201_2

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit formation, *Journal of Comparative Neurology and Psychology, 18* (5), 459-482

Yerkes Dodson Law. (2011), Retrieved August 4, 2014 from Wiki of Science: http://wikiofscience.wikidot.com/quasiscience:yerkes-dodson-law

APPENDIX A

Scenario Overview – Phase 1 and 3

Both phases of the experiment consisted of a low and a high workload situation. The low workload scenario was a repositioned Instrument Landing System (ILS) assisted approach and landing (Figure A1). The high workload scenario for phase 1 consisted of a departure from runway 23L of Düsseldorf Airport, following the COLA 2T departure (red line), after which pilots were redirected northbound back to runway 23L via radar vectors (Figure A2). The COLA2T was used again for phase 3, and participants were redirected southbound back to Düsseldorf, for the ILS LOC only approach on runway 05L (Figure A3).

Figure A1: Low workload phase 1 and 3, ILS RW23L Düsseldorf Airport

Figure A2: High workload phase 1, COLA 2T SID – RW23L

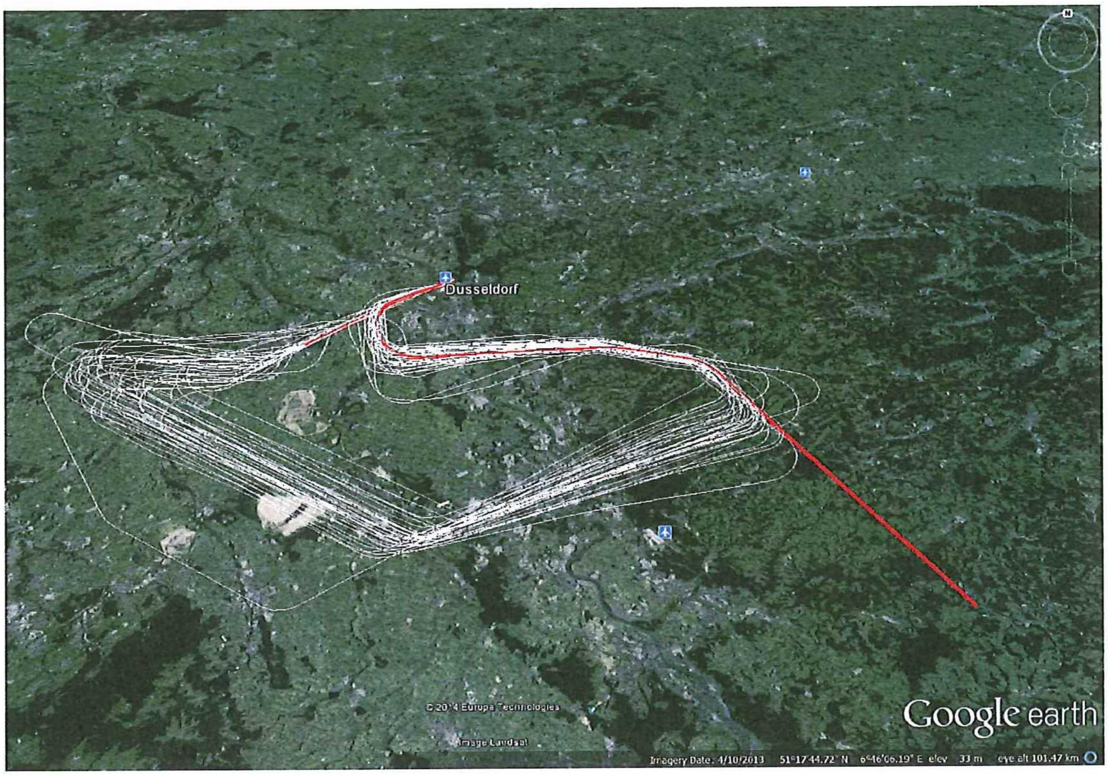


Figure A3: High workload phase 3, COLA 2T SID – RW05L

APPENDIX B

Detailed Scenario Description

Phase 1 and Phase 3 – Low Workload

Low workload consisted of a 12 Nautical Mile (NM) repositioning of the aircraft, lined up straight before runway 23L of Düsseldorf Airport, in which pilots were asked to fly an Instrument Landing System (ILS) approach. This is an approach procedure with lateral and vertical guidance as well as a glide path (i.e. 3 degrees sloped line, projected into the sky when viewed from the runway) on the Primary Flight Display (PFD). Calm weather conditions such as clear skies, no wind, no icing and normal atmospheric pressure, were applied. After unfreezing the simulator, the PF briefed the PM on weather and approach procedure. Pilots followed checklists; extended flaps and landing gear and continued to fly. PM guided through visual landing guidance systems next to the runway, while PF was to look at instruments. After following normal procedures, the aircraft was landed safely. Low workload scenario for phase 1 was identical to that of phase 3. The 12 NM approach is a standard procedure trained thoroughly by each participant over the course of their training, thus not needing variation over the two phases since it was used as a baseline measure.

Phase 1, High Workload

The high workload situation consisted of a complete continuous flight; take-off, cruise and landing under normal EPST operating procedures. ATC provided weather conditions and stated the altitude restriction of 6000 feet and speed restriction of 220 knots. Clear visibility was set for realism and the wind conditions were set at 20 knots gusting 25, from a direction perpendicular to the runway (full crosswind). Aircraft was fully loaded at maximum take-off weight and positioned at the start of runway 23L at Düsseldorf Airport, ready for take-off. Pilots discussed the Standard Instrument Departure (SID), which had to be flown after departure. The most complex route available in the simulator at EPST was chosen, the COLA 2T departure from Düsseldorf Airport. This route comprises of tuning into a variety of navigational aids during the first 15 minutes of flight, thus increasing cognitive demand. Four minutes of preparation time before departure were given, for pilots to discuss the SID and set the aircraft configuration ready for take-off. When clearance was provided by ATC, the PF followed normal operating procedures for take-off. At lift-off speed an intermittent fire alarm with a 5 second interval sounded, designed as first acute stressor. The alarm turned off by itself, the PM was instructed to reassure the PF that no fault was present. After the first corner on a southward heading, hydraulic system B stopped working. Pilots continued to fly the SID whilst proceeding into failure management. Flying the complex SID and dealing with failure management was designed to maximise cognitive load for the PF. The hydraulics system B failure has a longer than average checklist and requires anticipatory capacities to deal with deviated landing items. Hydraulics system B failure resulted in a delayed flap extension, which had to be anticipated by the PF in the final approach. It also affects critical flight control systems such as flight spoilers, autopilot, yaw damper, nose steering and brakes. In the midst of failure management, a confederate cabin attendant called the cockpit with the question for assistance. This was done to interrupt the thought process of the PF and create confusion. After completion of the hydraulics system B failure, the PF called a PanPan distress signal. ATC replied with a radar vector approach back to Düsseldorf. During this phase, the electronic stabiliser trim was failed by the SI (this equipment is used to reduce control forces in various segments of the flight). This was the second acute stressor and increased the difficulty to control the aircraft. The checklist for the stabiliser trim inoperative is one of the longest and contains delayed actions for the landing configuration of the aircraft, required in the approach phase thus increasing time pressure during the final phase of flight. This checklist was also interrupted by a cabin call in order to create confusion. The PM was instructed to assist the PF in calling out the QRH and to notify the PF of manual trim options. During the last part of the route until the approach, nothing significant occurred. It gave all participants enough time to arrive at the final approach course, with an identical aircraft configuration and completed failure management. The loss of hydraulics system B and loss of electronic stabiliser trim generated an unusual high workload on the last 10 NM of the route. Flap extension had to be anticipated and the inoperative stabiliser trim increased control difficulty, which combined were designed to maximise stress levels on the most difficult part of flying the aircraft. Clearance was provided for a non-precision approach on runway 23L on Düsseldorf airport. This approach comprises of tuning into a beacon located on the airfield, which can provide lateral guidance only. Vertical and thus glide path guidance was provided by the PM by calling required heights at specific distances from the airfield. This type of approach is the most cognitive demanding of all approaches, yet not unfamiliar to the participants due to flight school training. It requires controlling the aircraft on a precise path, monitoring instrumentation and communicating with the PF at the same time. The PF had to correct continuously while the PM set up the aircraft for landing. Checklists were completed (with deviated items for hydraulic B and stabiliser trim failure) during this process as well. To increase control difficulty, full crosswind was (perpendicular to flight path) was set at 20kts. Fuel levels were cut down before the approach, to ensure that the PF was committed to land the aircraft and could not opt for a missed approach. The PM would confirm the fuel status of the aircraft on the approach, increasing pressure to land the aircraft. At a height of 250 feet above the runway, another intermittent fire alarm sounded to create a third acute stressor. Normal braking was inoperative due to hydraulics system B failure so the pilot was forced to brake manually upon touchdown.

Phase 3, High Workload

The high workload scenario for phase 3 was designed to be unique, yet containing similar stressors to evoke a similar type of cognitive load as the high workload scenario for phase 1. The route was modified but contained the same SID. No other programmable route complied with the requirements of the initial scenario. Instead of flying an anti-clockwise circuit, after finishing the SID pilots were instructed to turn southward and fly clockwise, returning to Düsseldorf to land on runway 5L. The acute stressors were an intermittent false take-off warning at V1 (decision speed for aborting the take-off), the stabiliser trim inoperative and a false fire alarm warning on landing. Instead of the system B failure, only system A failure could be selected to evoke similar anticipatory qualities of participants. This consisted of an identical amount of systems being inoperative such as spoilers, autopilot and brakes but instead of the manual flap extension, a manual gear extension was required. Checklists were interrupted by the confederate SI acting as ATC, similar to phase 1. An extra questionnaire was added to the process with additional performance variables since the instructor rating form did not supply sufficient variance in ratings per student to be useful for this study. The final approach was on a different runway, using a different navigational aid but with similar characteristics of that of phase 1. Weather and wind conditions were configured identical, i.e. challenging crosswind of 25 knots on take-off and landing.

APPENDIX C

Scripts for Student Instructor and Pilot Monitoring – Phase 1

NOTE: Before each takeoff. Clear failures, route and path
NOTE: let student do a simple takeoff without clearance and without PM, to get a feel for simulator! No more than 2 mins
NOTE: USE QRH OF SIM 3. SEPTEMBER 24, 2007 (TAV)

NOTE: ACTIVATE DEBRIEF MANAGER

Scenario A Low workload (landing checklist only)
CALLSIGN: SPEEDBIRD012 FOB: 3.5 ZFW CG: 22%
- ATIS: 140/05, +10K, 5/3, SCT020, QNH989 (confusion risk:998), LIGHT ICING
- AP OFF, A/THR OFF, FD OFF
- Position 12NM final EDDL RW23L
- Briefing from PF (ILS23L, EDDL) (while flying

- At 9NM final; ATC: Clear to land RW23L

NOTE: DEACTIVATE DEBRIEF MANAGER

NOTE: ACTIVATE DEBRIEF MANAGER

Scenario B High workload (should be stored in instructor screen)
Place A/C at EDDL RW23L FOB: 3T **A/C WEIGHT GROSS @ MTOW**

DEP ATIS: 140/20G25, +10K, 4/1, SCT020, 998, LIGHT ICING, WS reported, RW state: WET
ARR ATIS: 160/20G25, +10K, 4/1, SCT020, 998, LIGHT ICING, WS reported, RW state: DRY

- *DEP CLR*: SPEEDBIRD012, You're cleared for a local training flight via the COLA 2T SID, initial climb clearance FL060, speed restriction for complete route SPD220, squawk 4663, you have 4 minutes before departure. (they must adhere to the 4 minutes, or you will run short on time for the sessions)

(pilots set up aircraft, discuss SID)

Before take-off check list

Advance trust levers

- **FAILURE** →At V1 (@140 kts), FIRE ENG1 (intermittent). TURN OFF AFTER TWO AUDIBLE SOUNDS finished

-if PF says return, PM ensures there is no failure!

- **FAILURE** → On HDG174 (aircraft level!!) , LOSS OF SYSTEM B→ Failure management
- PF should say short term plan

- *CAB CALL*: interrupt checklist after **"FEEL DIFF PRESS light illuminated"**.
Do you guys want some coffee yet? (PM SHOULD ANSWER → Script)

Let pilots finish checklist LOSS OF SYSTEM B
PF will set up short term plan. PanPan. Pilots will ask to return to Dusseldorf
(if pilots ask for return to EDDL *before* on course 148 from BAM to COLA; ATC: SPEEDBIRD012 continue on SID due to heavy traffic until further notice) FL060
(IF request alternate airport, state thunderstorms present. RETURN TO EDDL)

- *ATC* give guidance where required. Give clearance for runway EDDL VOR RW23L
-Give weather
ATIS: 160/20G25, +10K, 4/1, SCT020, 998, LIGHT ICING, WS reported, RW state: DRY
(do not change in SI screen yet)

-*ATC* give radar vector 360 (no earlier than @WYP)
Set new ATIS in computer

Make pilots have enough time to finish QRH and setting up aircraft for landing. (PM should enforce quick set up of a/c for landing!!)

Pilots swap controls:
(PM verkeerd trimmen)
FAILURE → fail electronic stabilizer trim
Pilots swap controls back:

PM reads QRH of electronic stabilizer trim

- ATC CALL AFTER "**Note; the handles should be folded inside the stabilizer trim wheel when not in use**" SPEEDBIRD012
State number of souls on board!!

Once PF is stating aircraft condition
ATC: turn left heading 270

(PF is explaining system status to PM)

Turn left heading 140 (aim for <u>start</u> of dotted line, extended of runway)

(DESCENT REQUEST TO 3000ft, in order to compare descent profile with defective stab trim)
(deferred items stated and processed)

ATC: Right turn heading 200, intercept VOR
-<u>TURN DOWN FUEL</u> to 1500kg

- When overhead FAP, ATC CALL:SPEEDBIRD012, you're number 1 to land, standby clearance

- At RA 1000' ATC CALL: SPEEDBIRD012, wind 160/15G25, RW23R, cleared to land

-At RA 500' FIRE ENG1 (intermittent). TURN OFF AFTER ONE AUDIBLE SOUNDS

TOUCHDOWN

NOTE: DEACTIVATE DEBRIEF MANAGER

STUDENT INSTRUCTOR!!!! BEFORE REPOSITION MAKE SURE STAB TRIM HANDLE IS STOWED

Fill in SI rating form. Please be honest and critical. Data is and will be anonymous
NOTE: Tell participant to remain confidentiality throughout the experiment!! DO NOT talk to anyone about the scenarios flown. You guys like to give a heads up to others, please do not do this as it may influence the results of the experiment

Scripts for Student Instructor and Pilot Monitoring – Phase 3

NOTE: Before each takeoff. Clear failures, route and path
NOTE: USE QRH OF SIM 3. SEPTEMBER 24, 2007 (TAV)

NOTE: ACTIVATE DEBRIEF MANAGER

Scenario A Low workload (landing checklist only)
CALLSIGN: SPEEDBIRD012 FOB: 3.5 ZFW CG: 22%
- ATIS: 140/05, +10K, 5/3, SCT020, QNH989 (confusion risk:998), LIGHT ICING
- AP OFF, A/THR OFF, FD OFF
- Position 12NM final EDDL RW23L
- Briefing from PF (ILS23L, EDDL) (while flying)

- At 9NM final; ATC: Clear to land RW23L

NOTE: DEACTIVATE DEBRIEF MANAGER

NOTE: ACTIVATE DEBRIEF MANAGER

Scenario B High workload (should be stored in instructor screen)
Place A/C at EDDL RW23L FOB: 3T **A/C WEIGHT GROSS @ MTOW**

DEP ATIS: 140/20G25, +10K, 4/1, SCT020, 998, LIGHT ICING, WS reported, RW state: WET
Comment: Can we turn on Rain?
ARR ATIS: 150/20G25, +10K, 4/1, SCT020, 998, LIGHT ICING, WS reported, RW state:

DRY

- *DEP CLR: SPEEDBIRD012, You're cleared for a local training flight via the COLA 2T SID, initial climb clearance FL060, speed restriction for complete route SPD220, squawk 4663, you have 4 minutes before departure.*
NOTE: They must adhere to the 4 minutes, or you will run short on time for the sessions

(pilots set up aircraft, discuss SID)

Before take-off check list

Advance trust levers

- **FAILURE** →At V1 (@136 kts), FALSE T/O warning (intermittent)

-if PF says return, PM ensures there is no failure!

- **FAILURE** → On HDG174 (aircraft level!!) , LOSS OF SYSTEM A→ Failure management
- PF should say short term plan

- *CAB CALL*: interrupt checklist after **"FEEL DIFF PRESS light illuminated"**.
Do you guys want some coffee yet? (PM SHOULD ANSWER → Script)

Let pilots finish checklist LOSS OF SYSTEM A
PF will set up short term plan. PanPan. Pilots will ask to return to Dusseldorf
NOTE: If pilots ask for return to EDDL before on course towards NOR; *ATC: SPEEDBIRD012 continue on SID due to heavy traffic until further notice, FL060 (IF request alternate airport, state thunderstorms present. RETURN TO EDDL)*

- *ATC* give guidance where required. YouGive clearance for runway EDDL LOC only RW05L
-Give weather and change in SI screen
ATIS: 150/20G25, +10K, 4/1, SCT020, 998, LIGHT ICING, WS reported, RW state: DRY

-ATC*:* direct NOR (no earlier than @WYP)
Set new ATIS in computer

-If pilots ask for radar vectors to return to EDDL after reaching NOR, reply: *Roger*

-PM selects NOR

Pilots swap controls:
(PM verkeerd trimmen)
FAILURE → fail electronic stabilizer trim
Pilots swap controls back:

-ATC: @ DME ±4NM from NOR, *turn right heading 300.*
NOTE: If pilots ask for descent to 3000ft. Do not allow yet!

NOTE: Only let pilots make turn after controls are swapped back!

PM reads QRH of electronic stabilizer trim

- CABIN CALL AFTER "**Note; the handles should be folded inside the stabilizer trim wheel when not in use**" SPEEDBIRD012
The aircraft is moving a lot in the back, what is going on?

-Fail both GS receivers (on DME 9.0 NOR, Northwest of BCN)

(PF is explaining system status to PM)

ATC; @ DME ±24 from DUS: *Turn right heading 090 for final approach intercept*
(aim for start of dotted line, extended of runway!!)
NOTE: After this turn DESCENT REQUEST TO 3000ft is approved

(deferred items stated and processed)

-TURN DOWN FUEL to 800kg when on final approach heading

- When overhead FAP, ATC CALL:*SPEEDBIRD012, you're number 1 to land, standby clearance*

- At RA 1000' ATC CALL: *SPEEDBIRD012, wind 150/15G25, RW05L, cleared to land*

-At RA 350' FIRE ENG1 (intermittent). TURN OFF AFTER ONE AUDIBLE SOUNDS

TOUCHDOWN

NOTE: DEACTIVATE DEBRIEF MANAGER

STUDENT INSTRUCTOR!!!! BEFORE REPOSITION MAKE SURE STAB TRIM HANDLE IS STOWED

Fill in SI rating form. Please be honest and critical. Data is and will be anonymous
NOTE: Tell participant to remain confidentiality throughout the experiment!! DO NOT talk to anyone about the scenarios flown. You guys like to give a heads up to others, please do not do this as it may influence the results of the experiment

APPENDIX D

Plates

Approach and departure plates are used to guide pilots to and away from the airport, via preset routes and navigational guidance beacons. The plate for the ILS approach RW23L (Figure D1), used for the low workload scenario, contains frequencies and procedures to use the ILS instrumentation for lateral and vertical guidance during the approach and landing. The plate with the COLA 2T SID (Figure D2), contains frequencies and distances of the 5 beacons which are required for accurate guidance.

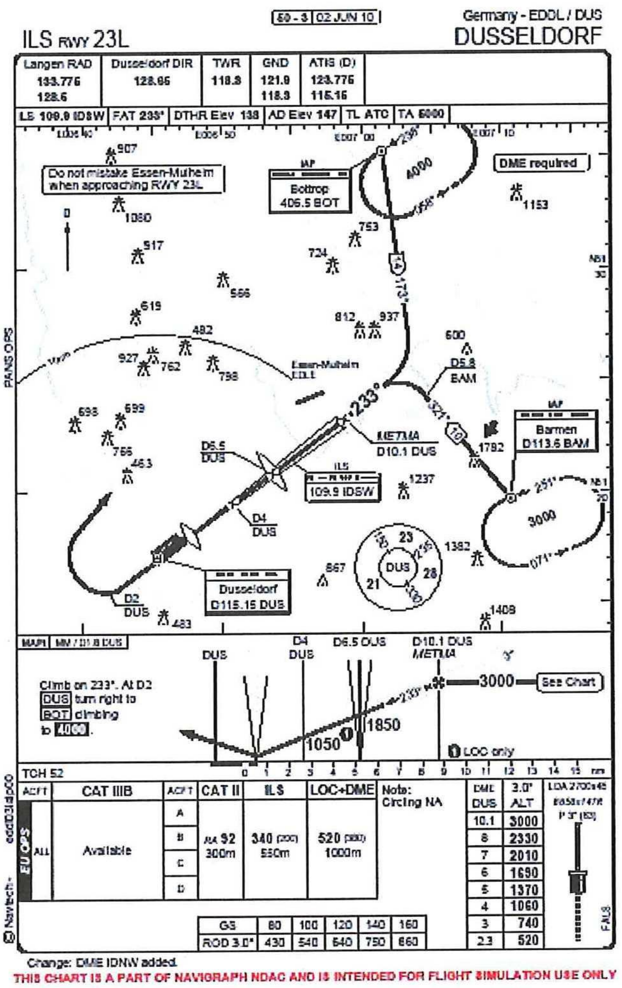


Figure D1: Plate for ILS 23L

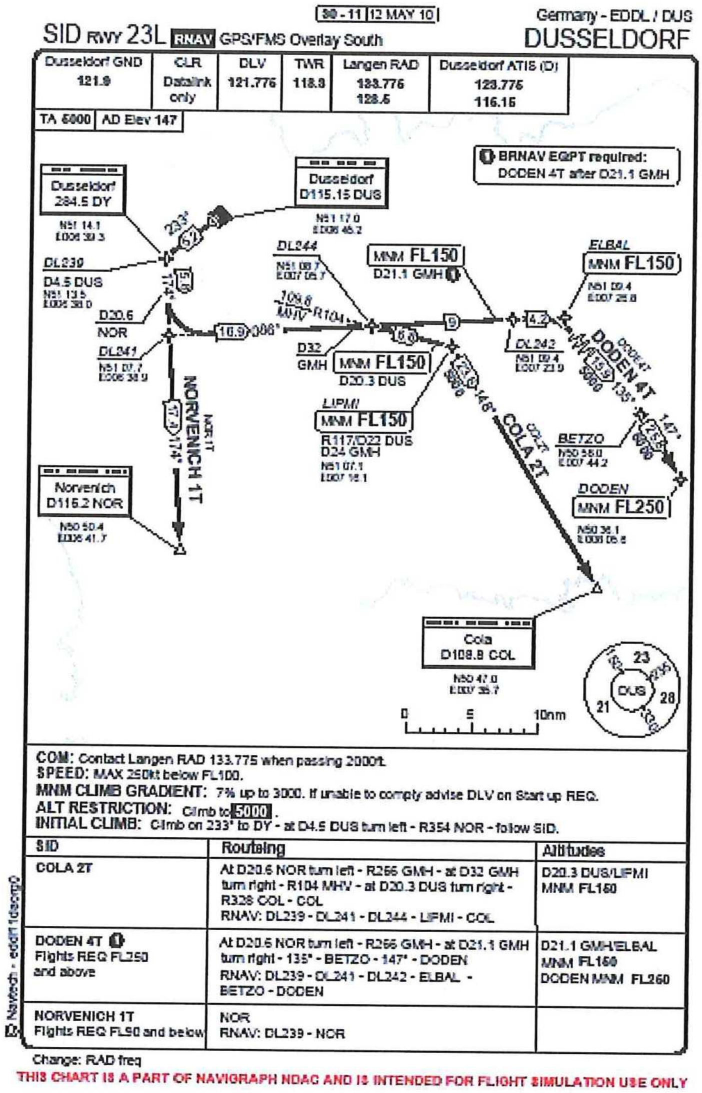


Figure D2: Plate for COLA 2T Departure

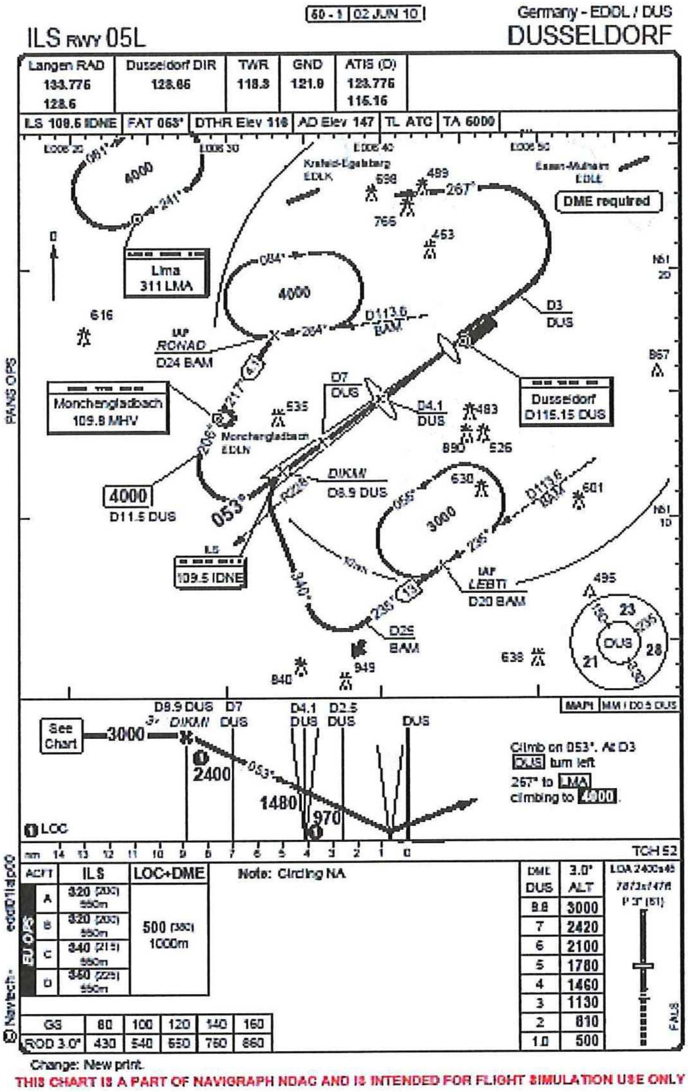


Figure D3: ILS RW 05L

APPENDIX E

Flight Performance Plots

The output generated by Python™ was visually verified by plotting graphs of the raw data. This was done for the ILS approach deviation (Figure E1), the SID deviations (Figures E2 – E3), altitude and speed (Figures E4 – E5) and for the control inputs (Figures E6-E7).

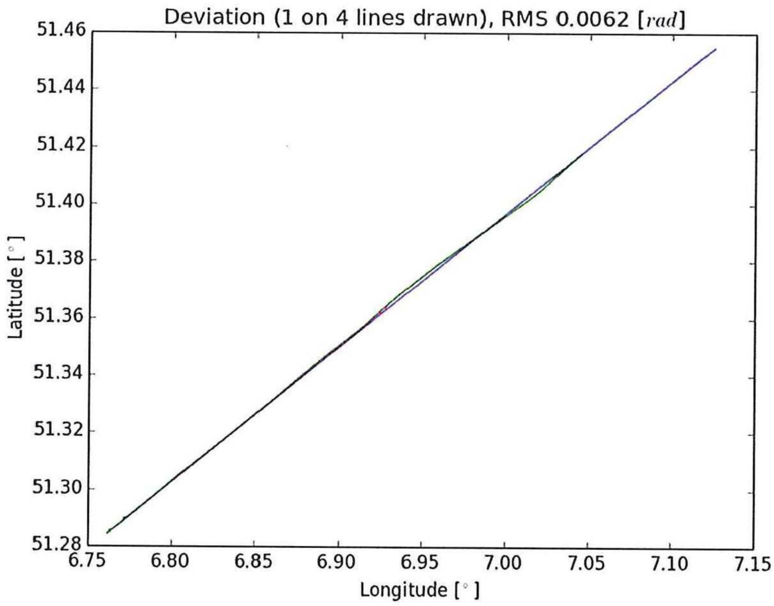
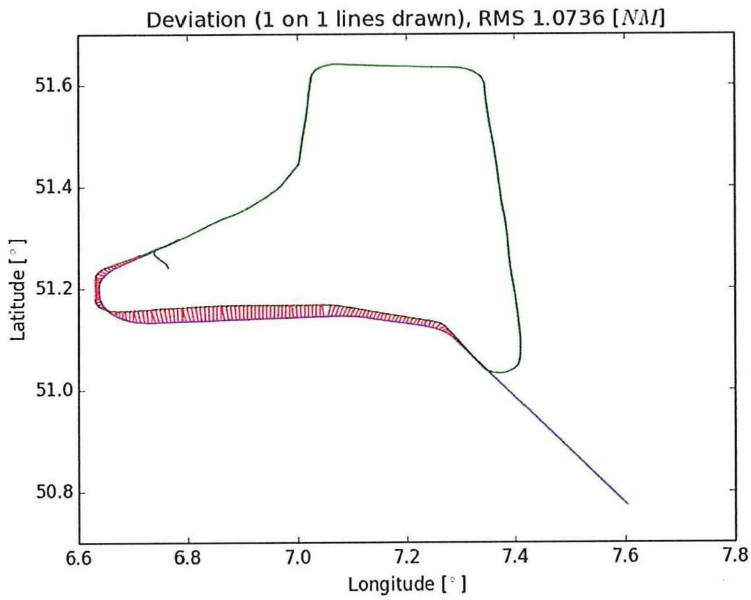


Figure E1: ILS approach deviation



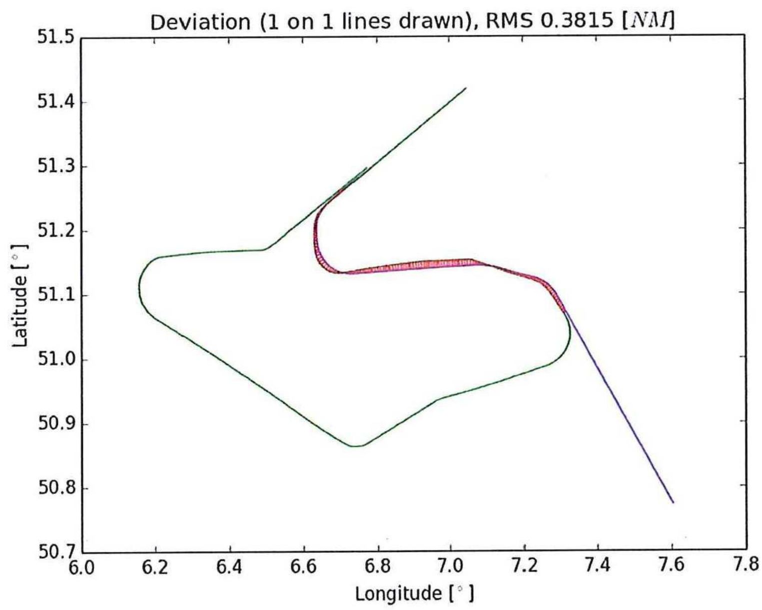
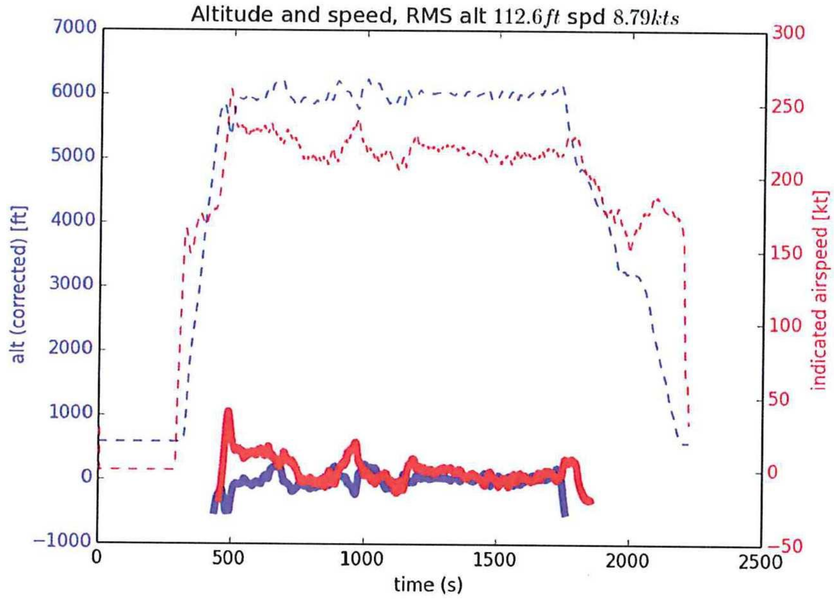


Figure E2 and E3: SID deviations for phase 1 and phase 3



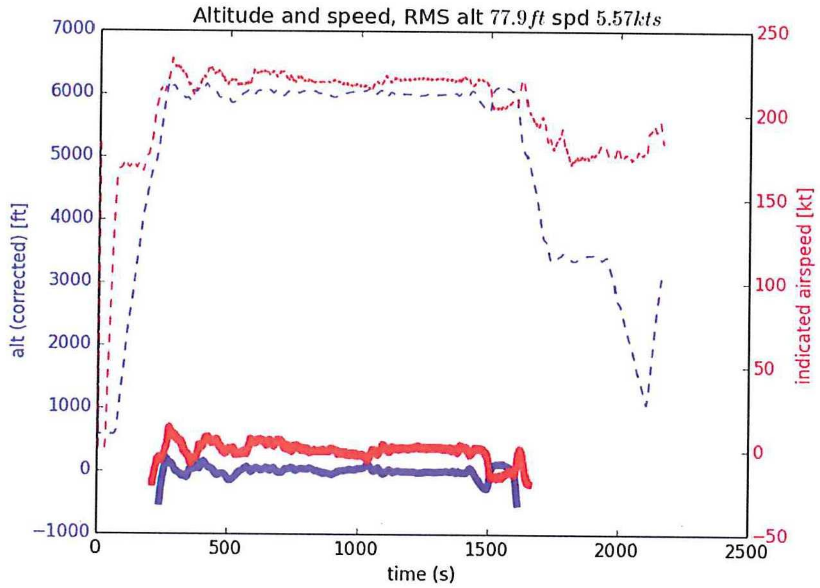
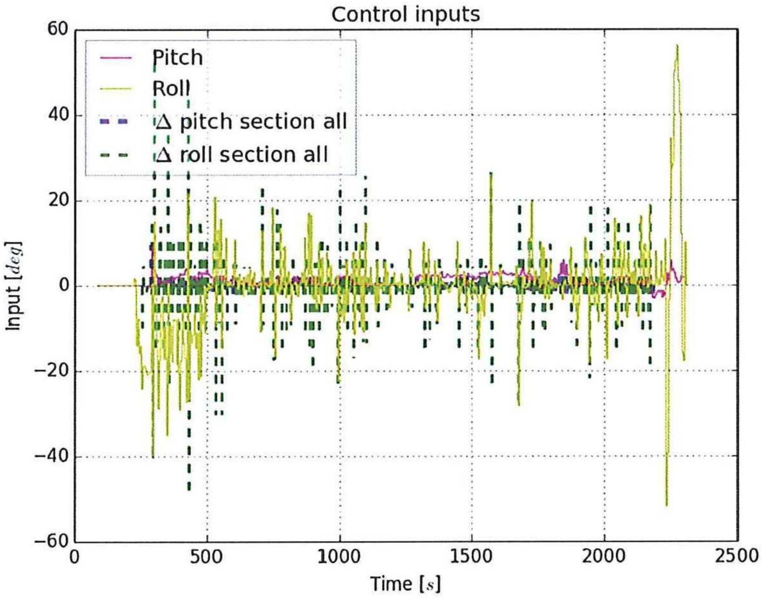


Figure E4 and E5: Plots of altitude and speed



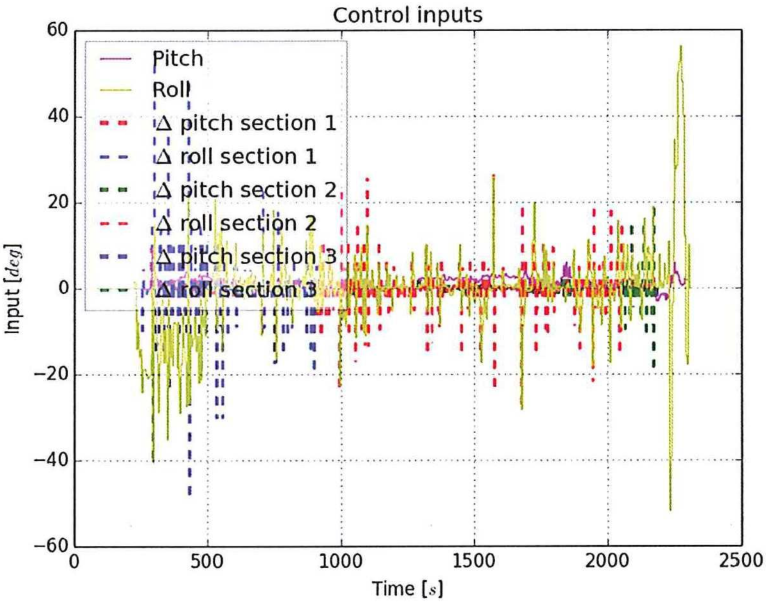


Figure E6 and E7: Plots of the control inputs

APPENDIX F

Performance Variables – Available Data

The performance variables development was limited to readily available information in the simulator architecture. The following states were available to choose from. These led to the simulator data protocol which can be found on the following page.

- Aircraft State
  - Latitude
  - Longitude
  - Altitude
  - Pitch
  - Yaw
  - Roll
  - Airspeed
- Aircraft Control Inputs
  - Pitch
  - Roll
  - Pedal (yaw)
  - Brake L
  - Brake R
  - Speed brake
  - Flaps
- AFDS Data
  - Speed bug
  - Altitude bug
  - Heading bug
- Multi-mode receiver
  - Glideslope deviation degrees
  - Localiser deviation degrees
- Radio Altimeter
  - Height in feet
- Aircraft surfaces
  - Elevator, aileron, rudder, TE and FE flaps and slats and
- Deviation from SID

Performance Variables – Data Analysis Protocol

Low workload

- ILS approach deviation
  - ✓ Deviation from glide path
- Average control input to test nervousness
  - ✓ Delta pitch for 7NM – 1NM section
  - ✓ Delta roll for 7NM – 1NM section

High workload

- Departure SID
  - ✓ Deviation from SID
- Approach non precision (for high workload situations)
  - o Deviation from glideslope, normal line to autopilot ILS, 7NM to 1NM
- Maintain altitude. Filtered data first, started measuring when altitude was 500ft below 6000ft so @ 5500ft. Stopped measuring when alt hit 5500ft
- Maintain speed. Plot speed for all flights. Start at >200kts IAS, stop at <200kts IAS
- Control inputs
  - ✓ Delta pitch and roll for all three sections: take-off to end of the SID, end of the SID to approach and the approach (7NM -1NM out from runway)
  - ✓ Delta pitch and roll for total flight

APPENDIX G

Cognitive Test Battery Screenshots

A series of games, developed by Andries van der Leij and Ilja Sligte of the University of Amsterdam, was used to test pilot's working memory (Figure G1), anticipatory abilities (Figure G2), their control capacities (Figure G3), how well they could keep attention and concentration (Figure G4) and how well they were able to integrate audio cues into a visual exercise (Figure G5).

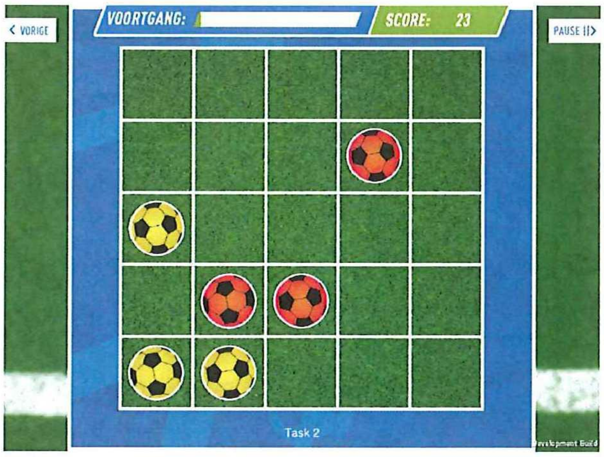


Figure G1: Working memory

Figure G2: Anticipation

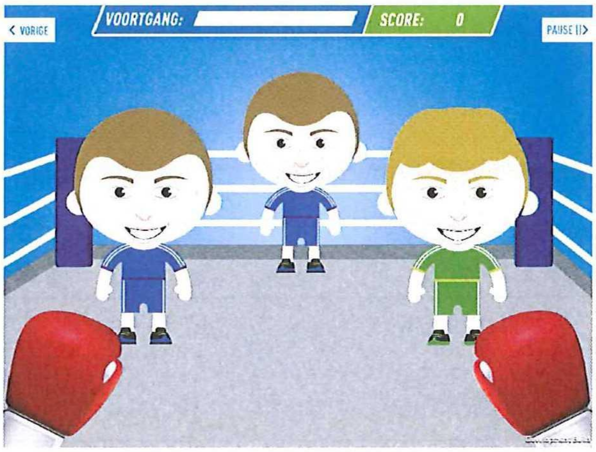


Figure G3: Control

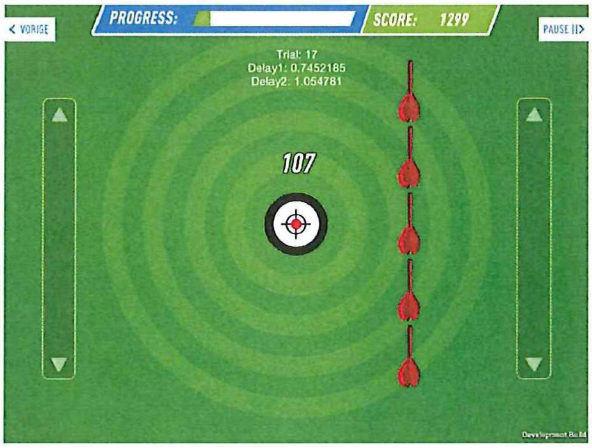


Figure G4: Attention and concentration

Figure G5: Integration

APPENDIX H

Subjective Feedback Form – blank

How many days on average have you practiced over the last 6 weeks?

How many times per day on average?

How long each time on average?

What did you think of the HM training? Was it what you expected?

Noticed any changes in your daily life (if you've minimally trained the prescribed amount)?

Any additional feedback or comments you would like to share?

Subjective Feedback Form - Answers

What did you think of the HM training? Was it what you expected?

- It was useful. Helps me relax more and deal with situations
- Very interesting. More interesting than expected. Very curious about outcome of the research. The training was a trigger for me to read up on this subject more and also do other trainings/exercises.
- It was something new for me, the training was very well organised and professional
- It seems pointless, as there are no direct obvious results that the training is actually working and improving performance. Therefore, it is what I expected.
- It was nice to participate. The emWave however did not always work proper. A lot of times it couldn't find my heartbeat.
- I thought something as simple as breathing didn't have so much influence on the way you relax and think.
- I did not really know what to expect at the beginning, but found it a useful technique to be more relaxing during demanding moments.
- I enjoyed the class but found it hard to discipline on training during the day/week
- The biggest challenge was to practice 3 times a day but when I did I noticed improvement every 2nd or 3rd time I tried that day. It is interesting to see how you can train your coherence and improve mentally. Not what I expected and sometimes a bit vague but it was alright.
- It was a good training. Clear and good advice.
- Yes I am more aware of my breathing and mental state during stressful situations.
- I think it is a good technique to reset your emotions and energy. In the end of the last training I understood better what the objective was, so I expected something else in the beginning.
- Yes it was what I had expected. It was easy to do and it was nice to learn a bit more about heart rhythm and how it is all connected.
- I think it can be helpful for this profession but when it is a real emergency, you don't have time to get in a coherent state.
- It was not what I expected but overall a good/interesting training.
- I didn't expect much, but I noticed that I am thinking clearer and I am more aware during stressful situations.
- No, I didn't expect anything, because normally I don't believe in this kind of training but I think the results will be good.
- I think it was really helpful/useful. Not sure what I expected exactly, but I think it worked out for me after all the practice.
- Took too much time. 1 Session in my opinion is enough. Had no expectations because I never heard of it.
- It was interesting to find out what is happening when you're stressed out and how to control that to a certain extent. It was not what I expected. I thought it would comprise of exercises like arithmetic and perhaps some 'games' like we played on the iPad.

Noticed any changes in your daily life?

- I can deal better with situations that would normally irritate me.
- At my current job it is easier to work together with that particular colleague which I was not working together with well before.
  Private life; something in which I find difficult to do, or don't want to do, do a neutral and find out that it won't be difficult or that it will be easily done.
- Not very much, apart from taking some time now and then every day to 'step back' has been a good experience.
- Negative.
- I think it is a good training to do before you have an interview. Because it makes you calm and less stressed.
- Definitely. I use it all the time when I know I'm going to meet new customers so I'll relax more and don't stress.
- Not in my standard day to day life but I used it to prepare a presentation and noticed afterwards that I was more relaxed during the presentation.
- No but I found the lessons very helpful.
- I've noticed it not in daily life but when playing some numbers game on my iPad I did have a clearer view right after I practiced with the emWave.
- I now know again how it is to be free of stress. I now use this as a baseline to refer to when I am stressed.
- Yes, more awareness of state of mind. More relaxed and in control during stressful situations.
- Yes, I have a lower and more stable heart rate, able to stay calm better and if I let myself go emotionally was able to calm down and reset quicker.
- More aware when I'm not coherent and at that moment I'll try to relax a bit to be sure that I will react/perform to a better standard.
- Less stressed at work/school.
- Less annoying customers at work. I think it is because people react more relaxed when you focus on yourself, on your own behaviour.
- Less tired. More aware of heart rhythm.
- Less stressed and hurried/rushed feelings and moments. And if it occurred, I could easily feel less stress especially the last week.
- Nope
- Sometimes it helps to take a second, think and act then. It also helps before going to bed, to fall faster asleep.

APPENDIX I

Physiological Response – Phase 1

Physiological response to the first scenario, without a reset moment. It can be observed that after event 5, the low workload touchdown, this participant does not return to its initially lower heart rate (Figure H1).

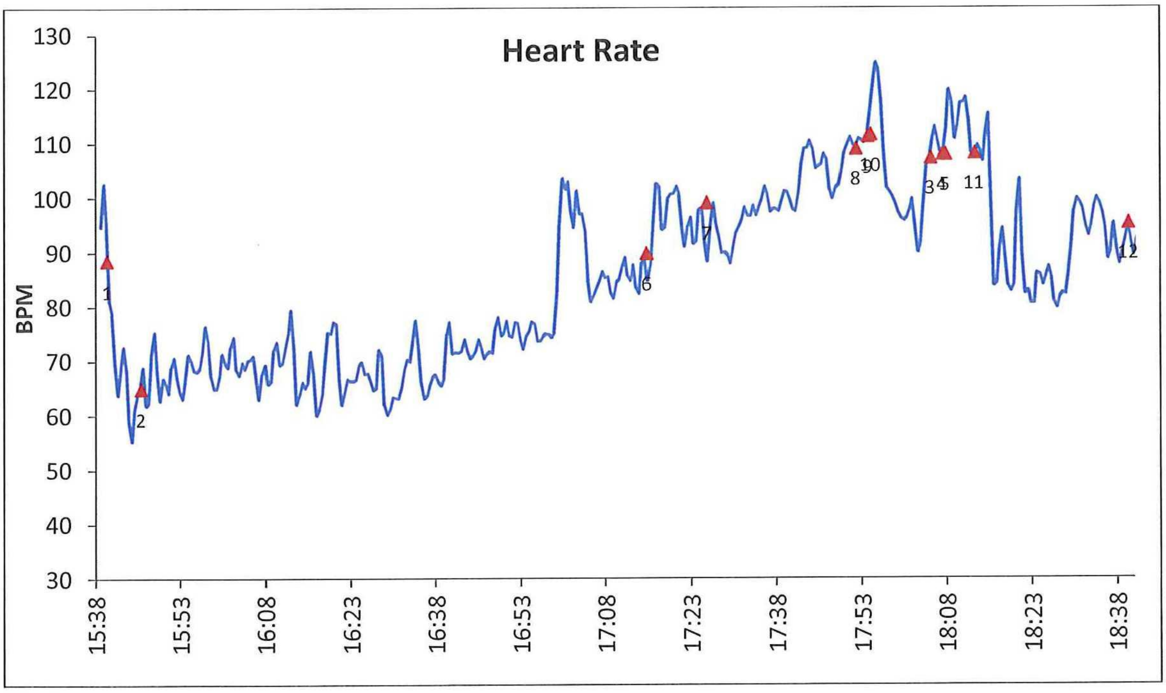


Figure H1: Heart rate of one participant during phase 1

Table H1

Event key legend

Event key						
1	Baseline start	6	HWL takeoff	11	Recovery start
2	Baseline end	7	HWL end of SID	12	Recovery end
3	LWL App 7 NM	8	HWL App 7 NM		
4	LWL App 1 NM	9	HWL App 1 NM		
5	Touchdown	10	Touchdown		

Physiological Response – Phase 3

Physiological response to the second scenario (Figure H2). It is evident that after the recovery period starts, this participant was able to return to baseline almost instantly, with the help of the emWave.

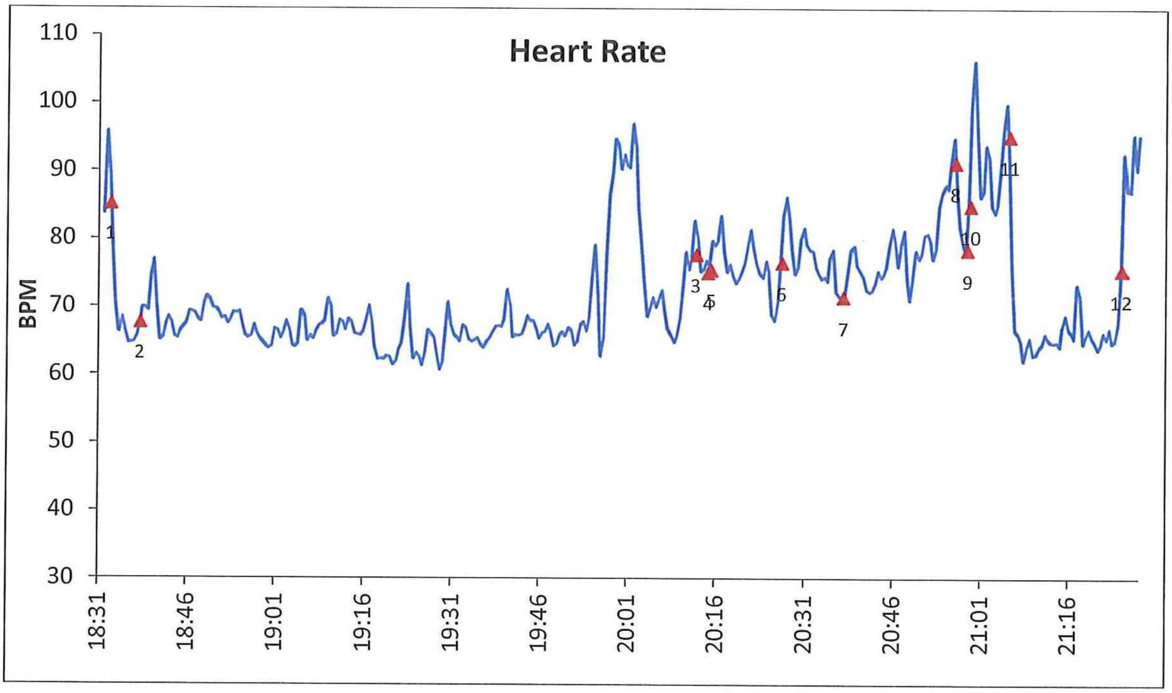


Figure H2: Heart rate of one participant during phase 3

Table H2

Event key legend

Event key					
1	Baseline start	6	HWL takeoff	11	Recovery start
2	Baseline end	7	HWL end of SID	12	Recovery end
3	LWL App 7 NM	8	HWL App 7 NM		
4	LWL App 1 NM	9	HWL App 1 NM		
5	Touchdown	10	Touchdown		