DOCTOR OF PHILOSOPHY

Fighter Pilot's performance and mental workload

Mansikka, Heikki Petteri

Award date:
2016

Awarding institution:
Coventry University

Fighter Pilot's Performance and Mental Workload

PhD

By

Heikki Petteri Mansikka

April 2016

Coventry University

Fighter Pilot's Performance and Mental Workload

By

Heikki Petteri Mansikka

April 2016

A thesis is submitted in partial fulfilment of the University's requirements for the Degree of Doctor of Philosophy

PREFACE

While it is not possible to list all those that have supported me during this journey, there are a few that should be specifically recognised.

First of all, it is impossible to express my appreciation to Professor Don ‘Uncle D’ Harris, Professor Kai ‘Käytöskukka’ Virtanen and Doctor Petteri Simola for their invaluable help. For the record, everything that is good in this thesis is because of their advice. Everything that is not so good, is because of me. Thank you for your patience, guidance and most importantly, your friendship. It has been a true pleasure to work with and to get to know you.

I would like to use this opportunity to thank General Kim Jäämeri and Brigadier General Sampo Eskelinen for their advice and support during my whole career. I hope my work will help you in your efforts to further improve the quality of the Finnish Defence Forces. In addition, a special thank you goes to Major (eng.) Matti ‘Muumi’ Jalava for all the inspiration he has given me and for the intellectual wrestles we have had.

This thesis is about fighter pilots and it was ultimately written for the fighter pilots; for them to be safe and for them to excel. For everything I have achieved as a pilot, I owe to those fellow fighter pilots that I have had a great honour to fly with. It is with the greatest possible humbleness that I pass my thank you to the brave men of the Fighter Squadron 21: Spokesman, Kwak, Dr Schneider, Midnight, Amex, Tango, Cash, Shaft, Snack, Jeltsin, Excel, Hoff and other BMFs I did not mention here. If you, dear reader, see any of those above mentioned names as your sweep lead or mission commander, be assured they will make your life easy. And a lot of fun.

Thank you Alina, Akseli, Valtteri and Verneri. You have kept me sane by reminding what really is important in life. I hope I have been able to give you even a fraction of the love and support you deserve. I am incredibly proud of you.

Finally, a special thank you goes to my intelligent and beautiful wife, Eriika. Without you, this – or anything else – would have no meaning. I love you.

Abu Dhabi, 15.4.2016 Heikki ‘Duracell’ Mansikka

ABSTRACT

Human information processing consists of multiple and limited resources; some of them are shared while some are separate and non-interchangeable. High pilot mental workload (PMWL) - and the subsequent decline in performance - results from the imbalance between the mental resources available to perform the task and the amount of resources needed to perform it. When the pilot's proficiency is evaluated, s/he should deliver an acceptable performance while being able to reserve enough mental capacity for the unexpected, additional resource demands. The task demands and cognitive stressors of air combat have potential to degrade pilot performance to an unacceptable level. Therefore, it is important to understand the amount of mental workload the pilots are experiencing and how much spare capacity they have available to cope with the possible additional resource demands. This thesis was aimed at understanding the relationship between PMWL and performance. The approach presented in this thesis was expected to support the development of reliable metrics for predicting the pilot performance under the stress of combat. In terms of practical applications, this thesis contributed to the development of the methodological principles that could help assuring the pilots' ability to cope with the task demands higher than those experienced during training or proficiency checks.

Heart rate (HR) and heart rate variation (HRV) were used as indexes of PMWL. The selection was done for several reasons. HR and HRV measures were accepted by the pilots as they were non-intrusive and they appeared to be objective. In addition, the implementation requirements were by no means excessive. Considering the aims of this thesis, the low diagnosticity of HR/HRV was not an issue. Finally, HR and HRV proved to be sensitive measures of varying task demands – especially when measured together with the pilots' awareness of the mission requirements. Simulated fighter missions were used to

manipulate the pilots' task demand and to measure their performance and HR/HR.

The thesis is constructed around three studies. In the first study, the subjects were required to fly instrument approaches in a high fidelity simulator under various levels of task demand. The task demand was manipulated by increasing the load on the subjects by reducing the range at which they commenced the approach. HR and the time domain components of HRV were used as measures of PMWL. The findings indicated that HR and HRV were sensitive to varying task demands. HR and HRV were able to distinguish the level of PMWL after which the subjects were no longer able to cope with the increasing task demands and their performance fell to a sub-standard level. The major finding of the first study was the HR/HRV's ability to differentiate the sub-standard performance approaches from the high performance approaches.

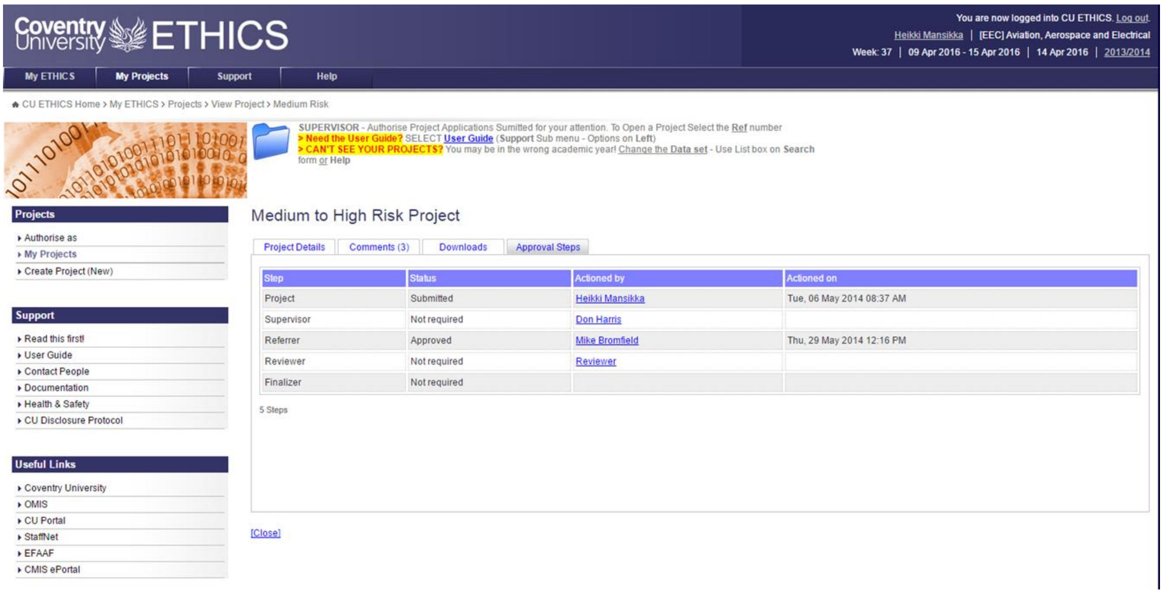
In the second study, fighter pilots' performance and PMWL were both measured during a real instrument flight rules proficiency check in an F/A-18 simulator. PMWL was measured using HR and HRV. Performance was rated using Finnish Air Force's official rating scales. Results indicated that HR and HRV were able to differentiate varying task demands in situations where variations in performance were insignificant. It was concluded that during a proficiency check, PMWL should be measured together with the task performance measurement.

In the third study, fighter pilots' HRV and performance were examined during instrument approaches and air combat. The subjects' performance was rated by a weapons instructor. In addition, the subjects' HRV was measured and used as an indicator of PMWL. During the instrument approaches, low performance was associated with high PMWL as expected. However, during the combat phases of the mission, low performance was associated with low PMWL. When the subject's awareness of the mission requirements was studied, it was found that the combination of low performance and low PMWL was associated with the subjects' low awareness of the mission requirements. The major finding was that unless the subjects' awareness of the mission requirements is examined, the relationship between the mental workload and performance during a complex combat mission may be difficult to explain.

It is concluded that HR and HRV are sensitive measures of PMWL in a simulated fighter aviation environment. HR and HRV proved to be associated with the changes in task demands and pilots' performance during simulated instrument approaches and air combat. However, the results of this thesis suggest that measuring just PMWL and performance is not sufficient – especially if the task of interest is complex and dynamic. To fully understand the pilot performance in such environment, the relationship between awareness of the mission requirements, workload and performance needs to be untangled. While this thesis provides encouraging results to understand this phenomena, further research is still needed before awareness of the situation requirements (or more broadly, situation awareness), performance and mental workload can be measured simultaneously, objectively and in real time.

This research project (Ref: P23260) was submitted for Ethical Approval on 6 May 2014 and approved by Coventry University on 29 May 2014. The Ethical Approval covered all three studies reported in this thesis. Ethical approvals from other institutions or organisations were not required. For details, visit:

https://ethics.coventry.ac.uk/App/Projects/View/medium_high_All.aspx?upd=23260



PUBLICATIONS RESULTING FROM THIS WORK

Journal Articles:

Mansikka, H., Simola, P., Virtanen, K., Harris, D. and Oksama, L. (2016). Fighter Pilots' Heart Rate, Heart Rate Variation and Performance during Instrument Approaches. Ergonomics, 1-9.

Mansikka, H., Virtanen, K., Harris, D. and Simola, P. (2016). Fighter Pilots' Heart Rate, Heart Rate Variation and Performance during an Instrument Flight Rules Proficiency Test. Applied Ergonomics, 56: 213-219.

Mansikka, H., Virtanen, K., Harris, D. and Simola, P. (2016). Fighter Pilots' Mental Workload and Performance: A Comparison of Simulated Instrument Approaches and Air Combat. Submitted for Publication.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

TABLE OF APPENDICES

NOMENCLATURE

ANOVA	Analysis of variance
ANS	Autonomic nervous system
AR	Autoregression
ATC	Air traffic control
bpm	Beats per minute
BVP	Blood volume pulse
BVR	Beyond-visual-range
CNS	Central nervous system
DH	Decision height
EEG	Electroencephalography
ECG	Electrocardiogram
EDA	Electrodermal activity
EFTP	Europe-Finland Tampere-Pirkkala
EMG	Electromyography
EOG	Electrooculography
ERP	Event-related brain potential
F/A	Fighter/attack
FFT	Fast fourier transformation
FinAF	Finnish Air Force
fNIR	functional near-infrared

ft	Feet
G	Gravitational
GS	Glide slope
HF	High frequency
HFnu	Normalised HF component of HRV
Hz	Hertz
HOTAS	Hands on throttle-and-stick
HR	Heart rate
HRV	Heart rate variation
HRVTRI	Heart rate variation triangular index
IAC	Instrument approach chart
IBI	Inter beat interval
IFR	Instrument flight rules
ILS	Instrument landing system
IMC	Instrument meteorological conditions
ISA	Instantaneous self-assessment
km	Kilometer
LF	Low frequency
LF/HF	Ratio between LF and HF components of HRV
LFnu	Normalized LF component of HRV
LLZ	Localizer
m	Meter

M	Mean
MANOVA	Multivariate analysis of variance
Max	Maximum
MCH	Modified Cooper-Harper
M.D.	Doctor of medicine
MEANHR	Mean heart rate
MEANRR	Mean of the time intervals between successive heart beats
Min	Minimum
min	Minute
MRT	Multiple resource theory
ms	Millisecond
m/s	Meter(s) per second
N	Number of subjects in a sample
NM	Nautical mile
NN	Normal-to-normal
NN50	Number of successive NN interval pairs that differ more than 50 ms
NASA	National aeronautics and space administration
OT&E	Operational testing and evaluation
PAR	Precision approach radar
pNN50	NN50 divided by the total number of NN intervals
PMWL	Pilot mental workload

PSD	Power spectral density
QRS	Combination of three graphical deflections seen on a typical electrocardiogram
RMSSD	Root square of the mean squared differences between successive NN intervals
RVR	Runway visual range
RR	Normal-to-normal, measured from R-wave peaks
s	Second
SA	Situation(-al) awareness
SD	Standard deviation
SDNN	Standard deviation of NN intervals
SE	Standard Error
SEM	Single engine manoeuvering
SWAT	Subjective workload assessment technique
TLX	Task load index
TTP	Tactics, techniques and procedures
VLF	Very low frequency
WM	Working memory
VOR	Very high frequency omni directional radio range
WTSAT	Weapon tactics and situation awareness trainer
#1	Leader of the formation
#2	Wingman of the formation lead
#3	Leader of the formation’s second element

#4	Wingman of the formation’s second element

1. INTRODUCTION

> *"I belong to a group of men who fly alone. There is only one seat in the cockpit of a fighter airplane. There is no space allotted for another pilot to tune the radios in the weather or make the calls to air traffic control centers or to help with the emergency procedures or to call off the airspeed down final approach. There is no one else to break the solitude of a long cross-country flight. There is no one else to make decisions. I do everything myself, from engine start to engine shutdown. In a war, I will face alone the missiles and the flak and the small-arms fire over the front lines. If I die, I will die alone."*
>
> Richard Bach (1963)

An air combat mission requires the management and control of, for example; air to air missiles, air to ground weapons, onboard sensors, off-board sensors, radios, multi-function displays, tactical contracts, rules of engagement, commit geometries and electronic warfare suite. On a single-seat fighter aircraft the personnel available to perform these tasks includes, and is limited to; a pilot.

A single-seat fighter aircraft is one of the most challenging man-machine systems ever created. A great deal of effort and resources is invested in recruiting and training to ensure the pilots are capable of operating their aircraft effectively even in the most complex combat environments. Despite the strict selection criteria and exhaustive training, the mission contingencies and the fog of war may expose the pilots to task demands that exceed their mental capacity. Should this occur, both the flight safety and the mission success become at risk of being compromised. As reported by Shappell and Wiegmann (2001), 70 to 80% of the aviation accidents are associated with human error – an outcome of human information processing failure. Billings and Reynard (1984) have come up with a same kind of conclusion; in their study of over 35 000 aviation accidents they found that human error contributes to over half of all aviation mishaps. In general aviation, there are numerous tools and mechanisms in place to manage the pilot mental workload (PMWL) from increasing to an unacceptable level. These include, but are not limited to; aircraft design, safety standards, operating limitations and crew resource management. Unfortunately, a single seat fighter pilot flying a combat mission does not enjoy the luxury of many of those safety mechanisms; there is nothing one can do to improve the cockpit and aircraft design, no matter how terrible it might be; the mission requirements may override the safety standards and limitations; only seldom there is an option to delay or alter the mission and finally, there are no co-pilots, first officers or navigators with whom to manage the workload. While the fighter pilot is expected to deliver an above standard performance, the enemy force is doing its best to complicate the mission and to deny the success of the friendly air assets.

Just a several fighter (technology) generations ago, flying a fighter aircraft required mostly basic piloting skills; effective stick, rudder and throttle control coupled with good gunnery skills. The mental demands of flying resulted mainly from the poor controllability of the aircraft and the ergonomically challenging cockpit designs. Piloting a modern fighter aircraft, on the other hand, has very little to do with the traditional tasks of flying. While the improvements in aircraft design have made the fighter aircraft relatively easy to fly, a modern fighter platform has turned into a hybrid of a bomber, fighter, command and control post, communications hub and an electronic warfare centre. The same technological advancements which give the fighter aircraft their capability edge, are the ones pushing the pilots to the limits, and sometimes beyond, their mental abilities. It is no longer a question of what is technologically possible, but whether the human information processing is capable of keeping up with the technology.

For an experienced fighter pilot, the basic aircraft control is a simple task; the essential information is constantly available either on the head-up display or on the helmet mounted display, the flight control computers' control laws provide an excellent manoeuvering capability and the high-thrust engines are extremely responsive. However, during the fighter operations, even the very basic flying tasks may have to be performed under seriously disruptive or non-permissive conditions. For example, an instrument landing system (ILS) is a radio navigation system which provides the pilot with a horizontal and vertical guidance just before and during landing. For an experienced pilot, flying the ILS approach is a relatively simple task. However, if the ILS is preceded with an intense combat or combat training mission, the pilots may have limited time for the approach preparations, they may have to conduct tactical tasks during the approach or they may have to troubleshoot complex combat related aircraft failures while flying the approach. For example, the pilot may be requested to give tactical mission reports during the approach, or the pilot may have to perform evasive manoeuvres and actions to mitigate aerial or surface threats. These relatively simple tasks, when performed simultaneously, generate potential to degrade the pilot performance if managed improperly.

Even without the external disturbances or disruptions, the fighter missions are unique in nature. As pilots are prepared to perform in extreme situations and with very small safety margins, they are expected to handle and manoeuvre the aircraft with a high control accuracy. For example, an instrument flying rules (IFR) proficiency check – consisting of basic manoeuvres, air combat manoeuvres and instrument approaches – requires very precise aircraft control for the mission to be rated as satisfactory. In addition, as poor performance may have a negative impact on pilot's career progression, the nature of the proficiency check has potential in itself to impact negatively on pilot's performance. On the other hand, the ease of aircraft control also allows aggressive and almost effortless air combat manoeuvering. However, combat and combat training missions have other, and more profound, challenges. In these missions, the pilots are required to perform multiple simultaneous and complex tasks accurately, in a compressed timeline, in a non-permissive environment and often with limited, overwhelming or contradictory information. The relevant information has to be identified, selected, combined with the past and current information and properly interpreted and integrated. Only after this can the information be used to build an understanding of the mission requirements, and the different elements and activities within the environment can be given their meaning. Once the pilot has gained a correct, or false, awareness of the mission requirements, s/he has to build a mental model of the future events and to make decisions that will best support the flight safety and the achievement of the mission objectives (Endsley 2001, 1999). Depending on how the pilot perceives the operating environment and his/her own status as part of it, either increased or decreased PMWL may result (Parasuraman et al. 2008; Endsley 1995).

Flying a basic or tactical mission with a modern fighter aircraft may expose a pilot to high, and sometimes extreme, task demands. While the physical demands of the modern fighter missions should by no means be overlooked, the majority of the tasks during the fighter mission stress especially the pilots' information processing capacity. As reported by Diehl (1991), 50 to 90% of all aviation accidents are associated with human factors issues. That being said, it is worth noting that Diehl's findings do not include minor events or mission failures. In the context of the fighter mission, the aircraft design, operating environment, task demands and performance requirements are typically given and cannot be altered at the pilots' discretion. Therefore, it is a duty of a responsible human factors researcher to safeguard the appropriate level of performance by making sure the 'man' does not become the limiting factor in the man-machine system.

This thesis is aimed at understanding the relationship between the fighter pilots' mental workload and their performance during simulated missions. To achieve this, six research questions were addressed. The research questions are listed below. In addition, Figure 1 illustrates how the three studies of this thesis (Chapters 5, 6 and 7) address the research questions.

1: Are the selected time domain components of heart rate variation (HRV) and heart rate (HR), as measures of PMWL, related to variations in the pilot performance during a simulated flying mission?

2: Can HR and HRV identify the level of task demands that lead to a sub-standard performance?

3: Can HR and HRV differentiate the task demand differences between the different mission segments of the mission?

4: Can HRV identify the differences between the mission segments of the proficiency check even when there are no significant performance differences between these segments?

5: Is HRV associated with the changes in the pilots' performance during the instrument approaches and air combat phases of the simulated mission?

6: Can the subject's awareness of the mission requirements during the air combat phases of the simulated mission be used to further explain the possible association between PMWL and performance?

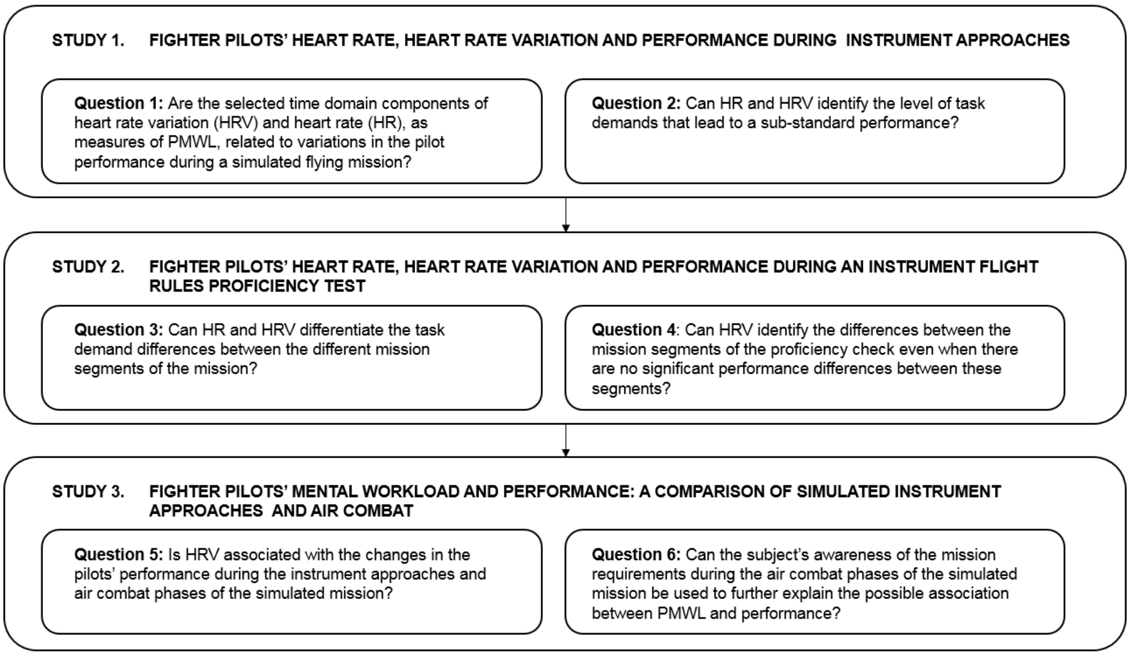


Figure 1. A summary of the research questions and how the studies of this thesis address them.

The practical applications of this thesis include developing the methodological principles for assuring pilots' ability to cope with the task demands higher than those experienced during training or proficiency checks. Ultimately, the research documented in this thesis was undertaken to help minimise the losses of the friendly fighter pilots' lives due to unacceptably high PMWL during training and combat.

2. PILOT’S INFORMATION PROCESSING

The pilot’s information processing system consists of multiple independent and limited capacities, or resources, that are not interchangeable (Wickens 2002a; Wickens 1991; Navon and Gopher 1979). These resources are used to process the information; to perceive and transform data into something meaningful, to select the responses and to respond. The expenditure of most of these resources is not automatic, i.e., it requires voluntary effort (Robert and Hockey 1997; Mulder 1986). During a proficiency test, the pilots are typically willing to use a lot of mental effort to invest their information processing resources in response to the varied task demands. However, despite the pilots’ willingness to invest their mental resources, they do not always deliver the kind of performance expected. When the performance is not limited by insufficient information, the pilots’ performance deficiencies are considered to result from their information processing limitations and their inadequate awareness of the mission requirements. In other words, the task is resource limited; degraded performance is considered to result from the imbalance between the mental resources *available* to perform the task and the amount of resources *needed* to perform it (Wickens 2008, 1991; Norman and Bobrow 1975). However, it is acknowledged that the relevant data may not always be available to the pilot as a result of physical limitations; visual cues may be outside of the pilot’s field of view, the volume of the cockpit auditory system may be turned too low or the cockpit displays may not be configured to present the required data. Even then, if the pilots’ awareness of the mission requirements is adequate, they are able to configure the cockpit in a fashion that supports the acquirement of the necessary data and they are able to focus their attention on the correct data source. Information processing, and expenditure of the mental resources is needed to gain and maintain the necessary awareness of the mission requirements (Wickens 2002b). Therefore, within the context of this thesis, the pilot is considered to have access to all necessary data to complete his/her tasks, i.e., none of the tasks discussed later are considered to be data-limited due to lack or poor quality of the data (Norman and Bobrow 1975).

The cockpit of the fighter aircraft can, at times, be a highly data intensive environment. The data is presented using both visual (e.g., multifunction displays and ambient cues) and auditory (e.g., radios and auditory tones of the aircraft's systems) modalities (Duncan et al. 1997). Unless actively limited by the pilot, the data are presented unfiltered, i.e., all radios receive at the same time, the systems' auditory tones go off indiscriminately and the displays may get cluttered while displaying whatever inputs are being received from the sensors and other aircraft systems. The pilot senses these data mainly through the sensory receptors of the visual and auditory channels. Once sensed, the unfiltered sensory data arrive into the modality specific, very high capacity short term memory store, or sensory register (see Figure 2) (Wickens 1980). As a result of the high capacity and autonomic nature of the sensory register, it does not generate limitations on the pilot's information processing capacity, nor does it require active effort or focus of attention to function at its full capacity.

This item has been removed due to 3rd Party Copyright. The unabridged version of the thesis can be found in the Lancester Library, Coventry University.

Figure 2. Adapted form of Wicken's schematic model of the Human Information Processing incorporating Baddeley's proposed structure of the Working Memory. From: Harris, D. (2011). Human Performance on the Flight Deck. Aldershot: Ashgate (p. 21). Reprinted with permission.

Unlike the operation of the sensory register, the other functions of the information processing are not autonomous. In other words, voluntary effort is required and mental resources are expended when data is drawn from the sensory register and taken through the different stages of the information processing; perceptual encoding, central processing and responding (see Figure 2). During the perceptual encoding stage, the auditory and visual data are given meaning and, based on their qualities, coded either as spatial or verbal (Wickens 1991; Wickens and Liu 1988). Both spatial and verbal coding use dedicated resources. In other words, coding of the visual information does not generate resource limitations for the coding of the verbal information. Depending on the mission phase and the tactical situation, the same piece of coded information may be either irrelevant or of great importance. Higher level mental processing is required before the data can be put into the correct context and its underlying meaning can be ascertained. These activities take place during the central processing stage of the information processing, where activities such as decision making, judgement and reasoning are conducted. Figure 3 illustrates the different stages, modalities and codes within the human information processing. During the central processing stage, data modalities are no longer separated. However, separate resources are dedicated to the processing of spatial and verbal data. During a fighter mission, the central processing stage typically requires a lot of cognitive resources as the pilot has to keep a large amount of inputs in the working memory (WM) while, at the same time, combining the selected pieces of information into a meaningful whole, evaluating the relative importance of the information, reasoning about its meaning and making decisions about the most appropriate responses (Endsley and Bolstad 1994). During the responding stage the responses are selected and executed. Separate processing, and resources, are used for manual and vocal responses. While the execution of simultaneous vocal responses has its physical limitations and the responses may have to be sequenced, many manual responses can, to some extent, be performed concurrently; the hands on throttle-and-stick (HOTAS[1]) system allows the pilot to make aircraft control inputs while simultaneously operating the tactically essential sensors and systems. As a result, during even the most dynamic situations most of the pilot's responses can be effectively time-shared.

---

[1] HOTAS refers to an aircraft cockpit design concept where switches and buttons are placed on the flight control stick and throttle lever(s). HOTAS allows the pilot to operate the most essential tactical systems without removing his/her hands from the flight control stick and throttle lever(s). HOTAS allows quick and accurate control of tactical systems especially in turbulent air and under high G-forces. In addition, the use of HOTAS removes the need to visually locate the switches and buttons inside the cockpit

The different stages of processing, types of codes and modalities utilise both separate and shared information processing resources (Vergauwe et al. 2010; Wickens et al. 1983). The tasks requiring the use of separate resources cause less interference than the tasks requiring shared resources (Pashler 1994). The requirement to use the shared resources limits – as a result of the limited capacity of the resources – the amount of information that can be processed simultaneously (Yeh and Wickens 1988). As different resources are allocated for the processing of different types of information, the pilot typically has no difficulties in conducting such simultaneous tasks as listening to the radio and changing the aircraft's attitude or reasoning about the optimal intercept geometry while looking at the radar display.

In addition to the non-interchangeable resources, WM provides an additional, shared pool of resources which is required during the perception and central processing. WM refers to a fractioned, limited capacity system capable of concurrent manipulation and short-term storage of information (Baddeley 2012). This capacity to simultaneously store and process information is a widely used definition of WM (Baddeley and Hitch 2000; Baddeley et al. 1975). WM is needed to perform such complex tasks as learning, comprehension and reasoning (Baddeley 1992). The limited capacity of WM is probably the most critical bottle-neck within the fighter pilot's information processing system, as the rapidly changing tactical situation requires constant (re-)assessment, calculation, evaluation and decision making. WM includes a central executive which acts primarily as an attention controller for its slave systems – each with their own subsystems and functions (Baddeley 1996a). As illustrated in Figure 2, these subsystems include at least the phonological loop, which is assumed to be responsible for the processing of the speech and digit based information, and a visuo-spatial sketch pad, which is assumed to perform similar activities as the phonological loop but for the visuospatial imaginary (Baddeley 1996b).

This item has been removed due to 3rd Party Copyright. The unabridged version of the thesis can be found in the Lancester Library, Coventry University.

Figure 3. An adapted form of Wickens's (1984: 2002) Multiple Resource Theory (MRT) showing the different stages, modalities and codes within the human information processing. From: Harris, D. (2011). Human Performance on the Flight Deck. Aldershot: Ashgate (p. 22). (Harris 2011). Reprinted with permission.

The phonological loop has at least two components; a phonological store and an articulator control process (Baddeley 1998; Baddeley et al. 1984). The phonological store is used for the short term storage of speech based information. Acoustic signal strings lasting two to three seconds can generally be stored in the phonological store, but unless rehearsed continuously, the memory traces will spontaneously fade away (Baddeley and Hitch 1974). The articulatory control process uses repetition of the auditory signals to maintain the information in the phonological store. The rehearsal process is sub-vocal and happens at about the same speed as it takes to utter a memorised word (Cowan et al. 1998; Atkinson and Shiffrin 1968). Once the signal string's rehearsal time exceeds the autonomous storing period of the phonological store, the first signals start to fade before they are being refreshed. The combined capacity of the phonological store and phonological loop activity comprise an individual's overall WM capacity known as a memory span (Baddeley 1998). A congested radio frequency can often stress the phonological loop's capacity to its limits; a pilot may have to memorise a received radio command, such as a wordy targeting directive, while the following vocal responses may make the auditory rehearsal impossible. In a tactical radio transmission, the most critical information may be in the beginning, middle or in the end of the transmission. Unfortunately, the pilots have a tendency to memorise the first few words and the last few words of the radio transmission while forgetting the words in the middle. The early words of the radio transmission are stored in a long term memory and can be recalled if rehearsed. The recall of the early words is known as a primacy effect (Hendrick and Costantini 1970). The most recent words (typically up to seven chunks of words or digits) of the radio transmission are stored in WM (Shiffrin and Nosofsky 1994; Miller 1956). The recall of the late words is known as a recency effect (Vallar and Papagno 1986).The words in the middle of the radio transmission cannot be held either in a short term or in the long term memory. This phenomenon is known as an asymptote. (Atkinson and Shiffrin 1968; Glanzer and Cunitz 1966; Murdock Jr 1962)

The visuo-spatial sketch pad is used to maintain the temporary representations of visuo-spatial information (Logie and Marchetti 1991). Where the phonological loop is related to language processing, the visuo-spatial sketch pad is assumed to perform similar functions in acquiring visuospatial symbols and semantics (e.g., digital symbols on a screen or shapes on a radar) (Logie 2014). Results from the experiments where spatial and verbal processing has been disrupted separately, support the theory of separate visual and verbal processing within

WM. Some later studies suggest that the visuo-spatial sketch pad is further fractioned and includes a separate kineasthetic component (Baddeley 2000). Challenges in developing an adequately precise methodology to separate different subcomponents of the visual WM make it somewhat difficult to study. However, existing experimental data suggests that the visual and spatial data are processed separately, although with strong interconnections (Klauer and Zhao 2004; Hecker and Mapperson 1997; Tresch et al. 1993). For example, if a subject is tasked to memorise spatial information, such as a location of an object and information concerning its visual appearance while simultaneously conducting a spatial tracking task, a remarkable interference can be observed as the shared resources are used for the visual encoding. The non-spatial visual signals are memorised by the use of rote verbal rehearsals and the processing of such information does not interfere with the spatial task. The auditory noise disturbs the verbal rehearsal activity and causes performance degradation both for the rote rehearsal and for the articulatory control processes. However, a visual noise disturbs only the visually coded spatial processing, leaving the processes related to the verbal activity intact (Cocchini et al. 2002; Scholl and Xu 2001). The visual suppression effects have shown to be smaller than those of the verbal suppression. Visuo-spatial sketch pad is best suited for storing information related to a single, complex pattern, while it lacks a serial recall performance. The visuo-spatial sketch pad (visual semantics), the phonological loop (language) and the episodic long-term memory interlinking the two, form a system capable of creating knowledge. Figure 2 summarises the different elements of human information processing by combining the elements of WM to the main components and phases of the information processing.

The majority of the information processing functions require active allocation of resources, or attention, before the mental resources can be used. For the information processing to be effective, attention may have to be focused on, divided or sustained between different inputs (Norman 1976). In addition, most information processing functions are not automated and thus require effort, or voluntary investment of the resources (Kahneman 1973). An effective active and voluntary allocation of resources requires awareness of the task requirements; if the task requirements are not known, or they are perceived incorrectly, the pilot has no means of directing his/her attention effectively, nor does s/he know how much effort should be invested on the task.

In summary, a pilot uses past and current information to build a mental picture of the current and future events, which is used to define, or refine, what is required from the pilot. An understanding of the mission requirements defines how effectively the pilot is able to direct his/her attention and justifies how much it is considered feasible to invest effort on the task. Finally, the relationship between the available mental resources and the resource demands of the task state the effectiveness of the information processing and the subsequent pilot performance. Each time mental resources are expended, it is done at a certain cost (Robert and Hockey 1997). The level of effort and resources invested to meet both the objective and subjective performance criteria - mediated by the task demands - constitutes a cognitive stressor known as PMWL (Paas and Van Merriënboer 1993; Camp et al. 2001).

3. MENTAL WORKLOAD AND ITS MEASUREMENT

3.1 Mental Workload

Mental workload characterises the demands imposed by the tasks on the limited mental resources when the desired performance is to be maintained (Vicente et al. 1987; Wickens 2008). As discussed in Chapter 2, the expenditure of mental resources is considered to vary for three reasons. First, the variations in the task demand cause variations in the amount of mental resources required to satisfy the demand. Second, the available mental resources define the portion of the overall mental resources required to achieve a desired level of performance. Third, the level of desired or acceptable performance dictates the amount of the voluntary mental resource investment or effort. In other words, the performance variations between two pilots conducting the same task may result from unequal cognitive resources or different levels of effort. Likewise, if the same pilots generate equal levels of performance, they may need to invest different levels of effort and may have to expend different proportions of their mental resources. When the task demands are kept similar for both pilots and they are equally willing to invest effort on the tasks, the resulting performance differences are caused by the differences in their information processing capacity. When the pilots are exposed to high or extreme task demands, some pilots will deplete their mental resources sooner than the others. Once there is no more mental capacity left to compensate the increasing task demand, the performance will begin to degrade – regardless of the level of effort. When the performance degradation is significant enough, the flight safety and operational effectiveness are at risk of being compromised.

If performance is used as a sole measure to assess the pilot proficiency, it is possible that two pilots will be rated as equally proficient when actually they are not: for a fighter pilot to be able to operate safely and effectively, s/he should be able to deliver an acceptable performance while being able to reserve enough mental capacity for the unexpected, additional resource demands. Therefore, it is important to understand the amount of PMWL the pilots are experiencing and how much spare capacity they have available to cope with the possible additional resource demands. In other words, the pilots' proficiency assessment should be two-dimensional: pilots should achieve a minimum acceptable performance level without exceeding a maximum acceptable PMWL. This kind of proficiency assessment requires a suitable measure of PMWL.

3.2. Mental Workload Measures

A variety of measures are available to assess the mental workload. Most empirical measures can be categorised either as behavioural (or performance based), subjective or physiological. Not all measures are applicable for all purposes. When the different techniques for PMWL measurement are considered for a particular application, five major criteria should be considered (O'Donnell et al. 1986; Wierwille and Eggemeier 1993):

- **Sensitivity**. A sensitive measure should reflect the external stimuli that it is supposed to assess. In the context of workload measurement, such a measure should be capable of detecting the task imposed variations in the operator's level of workload, arousal or resource demand. The mental workload assessment measure should also be selective in nature; it should not react to changes unrelated to the mental workload.

- **Diagnosticity**. Diagnosticity of a workload measure refers to its ability to trace the task demands to the different cognitive resources of the operator. Global workload measures have a low diagnosticity and provide an assessment of the overall workload without distinguishing the exact phase or modality (see Figure 3) of the information processing that is being loaded. The choice between the diagnostic and global measures depends on the objectives of the assessment.

- **Intrusion**. Intrusion describes the degree of disturbance the measuring imposes to the operator. Measures with low intrusiveness are typically desired. When the mental workload is measured in an operational, safety critical environment, intrusion is categorically unacceptable.

- **Implementation requirements.** Implementation requirements refer to the amount of training required to collect reliable data and the complexity of the measurement instrumentation. The measuring environment dictates what kind of implementation requirements are acceptable.

- **Operator acceptance.** Operator acceptance may reduce the utility of the otherwise effective measure. Operator acceptance and the measure's face validity can often be increased by educating the subject population. Face validity, or appearance validity, refers to a subjective assessment of the measure; does it appear to be a valid measure of a given variable.

The different characteristics of the measures make some of them more suitable for the aviation environment than others. When PMWL is measured in an operational environment, e.g., during a pilot proficiency check, the utility of some measures become highly limited.

3.2.1 Behavioural Measures

The behavioural measures can be broadly divided into primary and secondary task measures (Paas et al. 2003). Primary tasks are the duties an operator needs to perform in order to achieve the desired performance on the task of interest. The primary task measures assess directly the operator's behavioural output in a given system. For a primary task measure to be sensitive, it should capture all relevant pilot behaviours. The primary task measures are based on the assumption that the pilot's primary task performance is related to workload. When exposed to a very low perceived workload, a pilot may not be performing optimally as a result of boredom (or ‘underload’) caused by the dullness of the task at hand (Young and Stanton 2002a). Once the perceived workload is increased, the pilot’s arousal is elevated and the performance is enhanced. When the workload is further increased, the pilot can – to a certain degree – compensate it by investing more effort on the task and by adjusting the operating strategy (Williges and Wierwille 1979). Once the workload is increased beyond the subjectively sensed optimum, the pilot’s mental state begins to shift from arousal to stress and the performance begins to degrade despite the amount of effort invested or the strategy selected. The relationship between the operator’s performance and the mental workload is often described, although in a highly simplified fashion, as an inverted U-curve (Muse et al. 2003; Kavanagh 2005; Yerkes and Dodson 1908). The generic description of the inverted U-curve is illustrated in Figure 4.

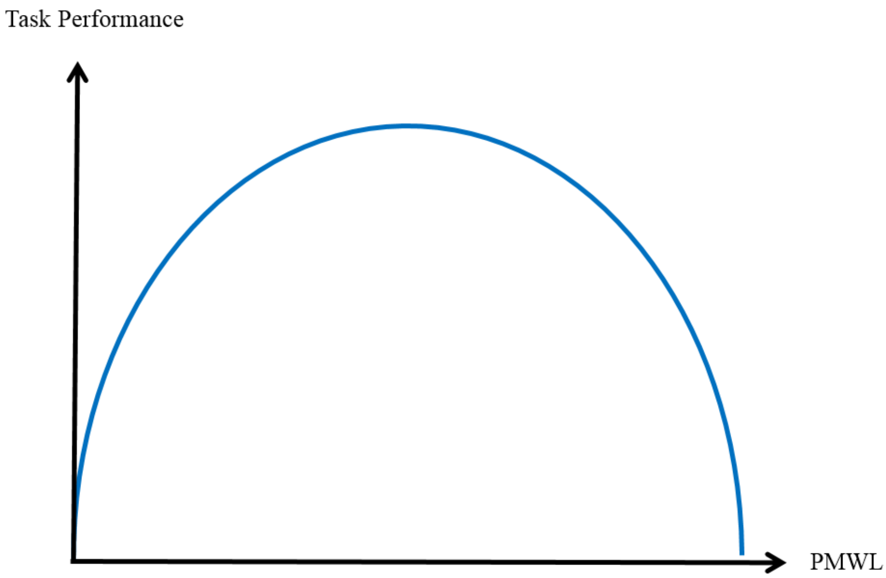


Figure 4: A generic illustration of the inverted U-curve

The primary task measures can be further divided into two sub-categories. Type-1 measures assess the combined operator and system output. For example, if an instrument approach would be the task of interest, the Type-1 measures could assess the pilot's control errors observed as deviations from the optimal localizer (LLZ), glideslope (GS) and airspeed parameters. Type-2 measures provide more direct measures of the operator's workload. Using the same example, the Type-2 measures could assess the actual pilot control inputs or control strategy that lead to the observed deviations (Lysaght et al. 1989). Due to the operator's capability to regulate the invested effort, primary task measures are generally considered sensitive workload measures only at the moderate and high workload levels. Primary task measures are equally insensitive when the workload becomes overwhelming, as at that point the amount of expended effort makes no difference on the system output. In addition, the performance decrement caused by an 'underload' – while being a real problem in some real life tasks – can be somewhat difficult to measure with the primary task measures. Authors like Hart and McPherson (1976) and O'Donnell et al. (1986) argue that the primary task measures have an important value of their own and should not be used as workload measures.

The secondary task measures build on the fundamental assumption that the operator has a limited mental capacity, which can be allocated to different tasks (Kerr 1973). When the secondary tasks are used, an operator performs a primary task while performing a secondary task using the spare mental resources left over from the primary task. A performance decrement in a secondary task is assumed to reflect the depletion of cognitive resources used in the primary task (Ogden et al. 1979; Brown 1978). The secondary task measures, when selected properly, can be highly diagnostic: a secondary task may be selected to stress a specific cognitive resource, thus reflecting the spare capacity of that specific resource (Verwey and Veltman 1996). The secondary tasks typically include either logical or arithmetic reasoning. Classical secondary tasks used in a laboratory environment include, but are not limited to, monitoring (Spyker et al. 1971), shadowing (Anderson and Toivanen 1970), mental arithmetic (Green and Flux 1977), classification (Huddleston and Wilson 1971), memory scanning (Gomer et al. 1976), tapping (Michon 1966) and time estimation (Hart and McPherson 1976; Wierwille et al. 1985). Ogden et al. (1979), together with Williges and Wierwille (1979) provide more comprehensive reviews of the different secondary tasks.

The operator performance in the secondary task is used as a metric of the cognitive workload in the primary task (Casner and Gore 2010). The secondary task measures assume that the operator performance in the primary task can and will be held constant. This can create challenges for the measurement validity if the primary task is complex and dynamic. When the overall mental load is low, the secondary task measures cannot distinguish mental workload variations as the operator has enough capacity to perform satisfactorily in both tasks. On the other hand, the operator performance may vary due to the peak workloads or shift of effort between the tasks. Also, when a secondary task is introduced, the operator needs to modify his/her time sharing strategy. This may lead to performance degradation even if there would be enough mental capacity available to conduct both tasks (Meshkati and Hancock 2011). The secondary task measures assess the estimated average workload rather than the momentary peak workload. The secondary tasks are, by their nature, disruptive and may lack operator acceptance especially if used to measure performance in a real-life task. As a result, the secondary task measures cannot be safely utilised in an operational, high risk environment when the expected mental workload is high (O'Donnell et al. 1986; Casali and Wierwille 1984).

3.2.2 Subjective Measures

The subjective measures utilise the operator's subjectively experienced workload, i.e., how a person feels when doing a task (Johanssen et al. 1979). The non-intrusiveness, ease of use and low-cost implementation of the subjective workload measures are some of the features that motivate their usage. For subjective PMWL measuring, there are many different methods to choose from. These include, but are not limited to, methods such as NASA TLX (Task Load Index) (Hart and Staveland 1988), modified Cooper-Harper Scale (MCH) (Wierwille and Casali 1983), Subjective Workload Assessment Technique (SWAT) (Reid and Nygren 1988) and Bedford Scale (Roscoe and Ellis 1990). The subjective measures have been widely employed to assess PMWL (Casner 2009; Lee and Liu 2003; Vidulich and Tsang 1986; Battiste and Bortolussi 1988; Wierwille et al. 1985; Casali and Wierwille 1984). While being widely used, the subjective workload measures have some critical limitations.

The subjective mental workload ratings are typically collected after the task execution. While this avoids the primary task intrusion, it generates a source of time error (Young and Stanton 2002b). Subjects have to memorise the past events and recall the type and extent of the personal sensations felt during those events that are to be measured. Subjects have a tendency to bias their subjective assessments towards the moments of peak workload and the final phases of the task. In addition, the number of different task features and the phasing of high and low task demand events affect the subjective perception of workload (Wierwille et al. 1985). Techniques such as NASA-TLX and SWAT are multidimensional measuring scales, which use several different dimensions to assess workload. On the other hand, the unidimensional scales, such as MCH and Bedford scale, utilise just one dimension and their overall sensitivity may be questioned (Hill et al. 1992). Regardless of the type of scale used, the subjective techniques have been criticised for their inability to discriminate the minor variations in the workload during the task. Even if the subjects are able to provide an accurate estimate of the average workload during a t ask, the workload peaks are likely to remain undetected. This limitation is critical during PMWL assessment, where even a momentary mental under- or overload condition in a man-machine system may have critical effects on performance and safety.

All subjective measures count on the observer's introspective experiences for the evaluation of the impact of external stimulus on their internal emotions and feelings. Independent observers may disagree on the external stimulus, the semantic meaning of the scale used to reflect the impact and most importantly, the internal sensations and their meaning (Veltman and Gaillard 1998). For example; multidimensional scales may require the operators to make judgments on notions such as spare capacity. The validity of this type of scales can be reduced as the operators may view the meaning of the spare capacity differently. This may result in completely different responses (Casner and Gore 2010). As a consequence, a significant experimental error may result from this lack of intersubjectivity; the subjects may report equal workload when it's unequal, or vice versa. The combined effect of the time error and the subjects' dissimilar preconceptions of both the external and internal meanings generate between-subject variability and make the relative mental workload comparisons between different tasks meaningless (Annett 2002a; Hart and Staveland 1988). Considering these limitations, the essential instrument of any subjective mental workload measure – the human operator – challenges their sensitivity.

When a person gives a mental workload rating on any subjective measure's scale, the result is an expression of the strength of the agreement between the person's internally generated scale and that expressed in the measure's scale (Annett 2002b). The operators' way of perceiving a scale may differ; some operators may naturally view the scale's mid-point as a representation of an ideal workload where there is neither under- nor overload condition, whereas others may use the lowest extreme as a personal ideal. Even though some multidimensional subjective mental workload measuring scales take into account the operator's performance as a separate dimension, this distinction is not necessarily enough to reveal the interactions between the performance and mental workload as a person may interpret poor performance as high workload. In a similar fashion, a person performing well may underestimate the amount of mental workload (Casner and Gore 2010).

The attributes of the subjective mental workload measurement scales are ordinal in nature. While the operator's inner sensations may be represented on a numerical scale, the scale itself is ordinal. All subjective workload measuring scales lack the interval and ratio properties and there are no universal units for the scales. That is, even if the subjective feelings or sensations are given numeric values, the distance between the values is not equal. For example; a 'Temporal Demand' of 8 in a multidimensional scale does not represent twice as great 'Temporal Demand' as a value of 4 (Casner and Gore 2010). The scientific value of this kind of quantification of fundamentally qualitative data is seriously questioned (Baber 2002; Michell 1997).

The unidimensional subjective rating scales ignore the multidimensional nature of the human information processing and do not even attempt to distinguish the task's processing demands on different cognitive modalities or stages (Wickens 2008). The multidimensional workload measures have scales with multiple dimensions, which attempt to capture the different resource demands within the human information processing. Such scales put exceptional requirements on the operator's capability to differentiate between the processing demands of the different modalities or stages. In addition, if a multidimensional rating scale is being used, the subjects have to compare and arrange their past sensations to the different dimensions of a rating scale (Annett 2002b).The accuracy of such measure is therefore reliant of the subjects' ability to memorise their perceived workload in retrospect. Depending on the complexity and duration of the task and the short term memory capacity of the subjects, their ability to recall the past events accurately may be limited. It is argued that the subjective measures are a better test of a WM capacity than that of mental workload (O'Donnell et. al. 1986). In a comparison study of SWAT and NASA-TLX methods Vidulich and Tsang (1986) concluded that the subjective measures indeed are sensitive to the processing loads of WM. However, they also found that the subjective workload measures cannot capture those variations in workload which are not well represented consciously or represent processing demands to response execution.

The basic assumption of the subjective workload measures is that if a person experiences workload, stress or frustration, then s/he has workload, stress or frustration – regardless of the indications of the other measures. This approach might be feasible when the workload measures are used for 'fitting the task on the man' (Grandjean and Kroemer 1997). When such tasks as instrument approaches or air combat engagements are considered, the alteration of the task is not an option. In these occasions the man has to be fitted, or selected, on the task. If the amount of workload is used as an evaluation or selection criteria, it is possible that while a subject reports no workload, stress or frustration, s/he may still experience excessive workload, stress or frustration. In such environments the operator is a highly unreliable measuring instrument, thus making the subjective measures insensitive and unreliable (Gopher and Donchin 1986).

3.2.3 Physiological Measures

Variations in arousal, effort and general activation level cause physiological changes. This has motivated the use of various physiological measures as indexes of mental workload (Ryu and Myung 2005; Ursin and Ursin 1979; Wierwille 1979). The physiological mental workload measures can be classified either as central nervous system (CNS) measures or peripheral nervous system measures. CNS measures include, for example, electroencephalographic (EEG) activity (Berka et al. 2007), event-related brain potentials (ERPs) (Kramer et al. 1987) and electrooculographic (EOG) standing potentials (Ryu and Myung 2005). In comparison, peripheral nervous system measures include such measures as heart rate (HR) (Reimer and Mehler 2011), heart rate variabiliy (HRV) (Mehler et al. 2011), electrodermal activity (EDA) (Setz et al. 2010) and electromyography (EMG) (Roman-Liu et al. 2013).

Different tasks generate different physiological responses. As a result, physiological measures may be highly sensitive on one type of task, but insensitive in other type of tasks. In addition, as the physiological responses seldom take place in isolation, the second- and third order physiological effects and bodily interactions need to be carefully considered; a seemingly sensitive measure may eventually turn out to be a measure of something other than what it is was supposed to be. For example, a pupillary diameter may be affected not only by the variations in the information processing demands (Beatty 1982), but also by the variations in the eye’s fixation distance or ambient lightning. In a similar fashion HRV can be affected by the blood pressure variations, body temperature and arterial pressure (Lutfi and Sukkar 2011; Stauss 2003). In fighter aviation, factors like extreme cockpit temperatures, exposure to direct sunlight and high G-loads can generate physiological responses which can, if not properly controlled, be falsely interpreted as PMWL responses.

The physiological measures can be highly diagnostic or they may reflect more global aspects of the mental workload. Whereas blink rate, HRV and EDA provide a global assessment of the subject's arousal, measures like ERP and magnetoencephalographic are sensitive to variations of the central processing demands (Isreal et al. 1980; Salvia et al. 2015). Selecting the most appropriate physiological measure means balancing between the qualities of the different measures and the objectives of the measurement itself. When an overall understanding of the task demands' impact on human information processing is needed, a global measure would be sufficient. On the other hand, during system design more diagnostic measures may be required to capture the potential overloading of a specific stage or modality.

Most physiological measures generate little or no intrusion and can thus be utilised in operational, or simulated operational environments (Wilson and Russell 2003; Eggemeier 1988; Kramer et al. 1987). In addition, the physiological measures allow continuous, real time monitoring of the operator state. This is typically a desired characteristic when the mental workload is measured in real life environments where the task demand may abruptly vary and the sudden overload conditions are likely to have critical impact on operator's performance. Another advantage of the physiological measures is their objectivity, which increases their utility in scenarios where it is reasonable to expect that the subjects' subjective opinions are not very accurate.

The increased computing power and the reduced size of the instrumentation components have improved the utility of many physiological measures in the real-time mental workload measurement. However, when the physiological measures are applied in an aviation environment, there are other limitations to the applicability than just the size of the instrumentation. For example, the measure needs to have the aviation authority's approval and it has to be accepted by the pilots. When PMWL is measured during flight or proficiency check in a simulator, the number of possible measures is greatly reduced. HR and HRV measuring instrumentation have been approved by many aviation authorities and are widely accepted by the pilot population. In addition, HR and HRV have been successfully used to measure task demand variations both in a flight simulator and in actual flight (Mansikka et al. 2016a, 2016b, 2016c; Veltman and Gaillard 1998; Svensson et al. 1997; Wilson 1993; Roscoe 1993, 1992; Jorna 1993; Aasman et al. 1987). HR and HRV are reactive to sudden changes in task demands and are thus useful indicators of possible peak overloads (Stuiver et al. 2014).

For this thesis to achieve its aims (see Chapter 1, pages 4-6 and Chapter 4), PMWL was measured during simulated flying missions. One mission was specifically designed to test the selected PMWL measures, while other missions were real proficiency checks, or proficiency checks under operational testing and evaluation (OT&E). This kind of test setting placed exceptional demands on the PMWL measures used. While the primary task measures provide a more direct measure of PMWL, they are not very sensitive at low or extremely high levels of PMWL. Within the flying missions used, the task demands varied from one extreme to another. The primary task measures were considered incapable of distinguishing the PMWL variations at those extremes. The secondary task measures, on the other hand, are highly sensitive to PMWL variations. However, their inherent intrusiveness did not support their use during the real proficiency checks. Subjective PMWL measures are not intrusive at all and could had been easily implemented into the flight simulator environment. In addition, subjective PMWL measures have been widely used in the aviation domain and have proven to be highly sensitive to varying levels of task demands. Unfortunately, as discussed in Chapter 3.2.2., the subjective PMWL measures are not considered suitable for the proficiency test scenarios, especially as the PMWL was supposed to be introduced as a possible pass/fail- criterion. Compared to subjective measures, the physiological measures have the advantage of objectiveness. Also, with the physiological measures the physiological responses can be recorded continuously throughout the trial. In addition, many physiological measures have demonstrated sensitivity to varying task demands both in aviation and in other domains. The intrusiveness and the implementation requirements of some otherwise powerful physiological measures made them less attractive for this kind of study. The HOTAS system does not allow the use of any kind of finger sensors, the pilot's helmet greatly limit the use of skull attached instrumentation and most importantly, the measuring instrumentation may not interfere with the simulator instrumentation. Also, if the methodology proposed in this thesis is later applied in a wider scale, the measuring instrumentation's cost, availability and ease of use must be considered. After the different aspects of sensitivity, diagnosticity, intrusion, implementation requirements and operator acceptance were considered, HR and HRV were selected as measures of PMWL. Selection was based on those measures' non-intrusiveness, affordability, accessibility, sensitivity and flight-worthiness. In addition, their global sensitivity was a desired characteristic as this thesis concentrated on determining if a PMWL differs from one task condition to another and how PMWL is associated with the pilot performance.

3.3 HR and HRV

3.3.1 Physiology of HR and HRV

The nervous- and endocrine systems are responsible for most of the body's control mechanisms. The nervous system can activate body responses quickly and precisely, whereas the endocrine system is responsible for much slower and undiscriminating metabolic regulation. Physiologically, the nervous system can be divided into a CNS and peripheral nervous system. The latter can be further divided into the sensory division and motor division. The sensory division transmits nerve action potentials between the central nervous system and sensory receptors, glands and muscles. The motor division is responsible for the voluntary body movement and active motion control activities which are collectively called the motor functions of the nervous system. Sensory receptors detect sensory stimuli and transfer them into nerve signals. The sensory division and the voluntary motor division together form the somatic nervous system (Hall 2015; Saladin 1998; Fox 1996).

Parallel to the somatic motor division operates an autonomic nervous system which controls the smooth muscle tissue and internal organ functions. The autonomic nervous system (ANS) is further divided into sympathetic and parasympathetic nervous systems. The sympathetic stimulation of organs generally causes an excitation effect, whereas the effects of the parasympathetic nervous system are typically inhibitory. The sympathetic stimulation, together with the Frank-Starling mechanism, increases the heart rate, reduces heart rate variability and increases the heart muscle's force of contraction. The opposite effects occur after a parasympathetic stimulation. The stimulation of the parasympathetic nerves can reduce the cardiac output to near zero, whereas the sympathetic stimulation can increase it by almost 100% (Hall 2015; Katz 1977; Horwitz et al. 1972; Levy 1971).

The continuous balancing of the two branches of the ANS causes changes in the contractile strength of the heart muscle and fluctuations in the cardiac output (Hall 2015; Camm et al. 1996). An increase in mental workload causes an activation of the sympathetic nervous system. When there is a mass discharge in the sympathetic nervous system, the nervous system is preparing the body for vigorous physical activity and increased mental arousal. Expressions 'sympathetic stress reaction', 'sympathetic alarm reaction' and 'fight or flight reaction' all reflect the same phenomenon. Although the workload cannot be measured directly, the heart's responses to the neurological modulation provide an indirect method for its measurement; the nerve activity causes electronic impulse transmissions in and around the heart which can be recorded and interpreted with an electrocardiograph (ECG) (Hall 2015; Silverthorn et al. 2009; Opie 2004).

3.3.2 Analysis of HR and HRV

A normal ECG consists of a P- wave, a QRS complex, followed by a T- wave and U- wave - each representing different de- and repolarisation phases within the heart's muscular cells. When the QRS complexes are detected from the ECG, the normal-to-normal (NN) interval (or HR) and differences between the NN intervals (or HRV) can be determined. The NN interval refers to the normal beat-to-beat sinus rhythm (Sampson and McGrath 2015; Houghton and Gray 2014; Ivanov et al. 1999). While it is normally measured from the middle of the QRS complex, the NN intervals and NN interval differences can be measured from the P-wave, T-wave or from any component of the QRS- complex. However, it is a common practice to use the R-wave peak as a reference point in measurements as it is typically the strongest wave and can therefore be easily detected even in noisy conditions. To emphasise the reference point used, the literature typically uses terms RR interval and RR interval difference (ChuDuc et al. 2013; Opmeer 1973). A generic ECG and RR interval are illustrated in Figure 5.

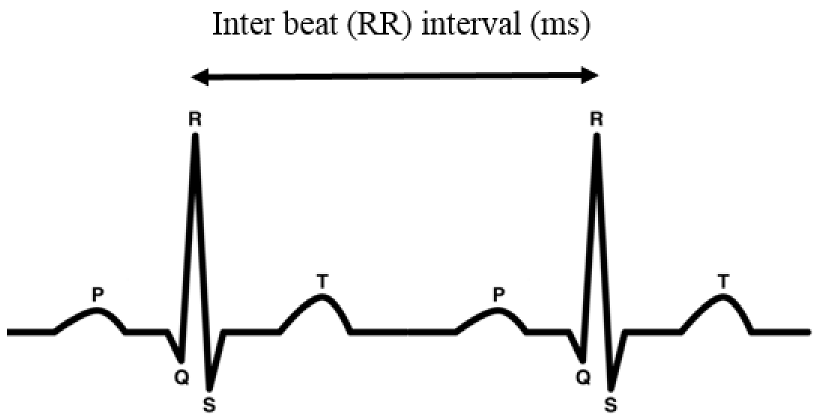


Figure 5. RR interval and the different waves of the ECG.

The artefacts in the RR interval times may cause significant errors to analysis, especially when short term (shorter than five minutes, but not less than two minutes) ECG samples are used. Based on their origin, the artefacts can be classified as technical or physiological. The physiological artefacts include baseline wander, atrial fibrillation, ectopic beats, irregular heart rate and multiplied or masked R- waves caused by arm and chest muscle contractions. The technical artefacts include erroneous detections of the QRS complexes and line interference (50/60 Hz noise) caused by the electrical devices near the actual data acquisition device (ChuDuc et al. 2013). A low sampling frequency (less than 500 Hz) further increases the likelihood of technical artefacts. It is strongly suggested that ECG recordings are manually checked for artefacts (Camm et al. 1996). Interpolation methods can be used to reduce the artefacts and, when possible, the corrupted ECG sectors should be completely excluded from the further analysis. A low sampling rate is another source of R- wave occurrence time errors. Low sampling rates (below 250 Hz) may cause jitter in the estimation of the R- wave fiducial point and disrupt the spectral analysis of HRV.

Both ECG and blood volume pulse (BVP) can be used to detect the RR intervals. The BVP technique requires the use of the finger sensor, which can be somewhat intrusive in an aviation environment as the aircraft's hands on throttle-and-stick system requires the use of both hands' fingers. This limitation, combined with the BVP's inferior time resolution, made ECG the preferred method to detect cardiac ANS responses (Tan et al. 2011; Selvaraj et al. 2008).

HR has been widely used as a measure of PMWL and has been able to differentiate between flight phases of differing task demands (Dahlstrom and Nahlinder 2009; Svensson and Wilson 2002; Svensson et al. 1999; Ylönen et al. 1997; Roscoe 1993, 1992, 1975). HR is typically expressed in beats per minute (bpm) or as time interval between beats (in milliseconds). The increased bpm, or the reduced inter beat interval (IBI) is indicative of higher cognitive workload. The beat to beat rhythm of the heart is seldom constant. The continuous modulation of the different components of the ANS results in variations in beat to beat intervals. When HRV analysis is conducted, the ANS's modulation of the sinus rhythm is examined. Once the QRS complexes (or R- peaks) have been detected, the differences between the successive R- wave occurrence times, or HRV, can be obtained and analysed using the time domain (HRV changes over time), frequency domain (spectrum of oscillatory components) or geometrical methods. While the different methods provide remarkable different outputs, they all illustrate the variations in beat to beat intervals, where a lower HRV is indicative of higher cognitive workload.

Many time domain measures are based on the statistical analysis of the series of successive RR intervals. In its simplest form, the statistical analysis is used to measure the mean value of RR intervals or the standard deviation of RR intervals. Other analysis methods include, for example, measuring the square root of the mean squared differences of successive RR intervals and evaluating the total number of pairs of consecutive normalised IBIs that differ by more than some defined time (50ms is typically used). In addition to the statistical time domain methods, geometrical measures are available. When geometrical measures are used, the NN intervals are converted into a geometric pattern – typically into a histogram. The HRV is then analysed by examining the geometric and/or graphic properties of the pattern.

When the frequency domain measures are used, the power spectral density (PSD) estimate is calculated from the RR interval series. Using the PSD estimate, it is possible to evaluate how the spectral variances distribute as a function of frequency. Both parametric and nonparametric methods can be used to calculate the PSD. The time series of RR- waves are presented as a function of time. As a result, the R-waves are non-equidistantly sampled. For the frequency analysis to be possible, the non-equidistant RR interval time series are first converted to equidistantly sampled series by interpolation methods. Fast Fourier Transformation (FFT) or autoregressive (AR) modeling methods are used to carry out the PSD estimation. The power spectrum is typically divided into several frequency bands. The most commonly used frequency bands include very low frequency (VLF, 0-0.04 Hz), low frequency (LF, 0.04-0.15 Hz) and high frequency (HF, 0.15-0.4 Hz) (Bailón et al. 2007; Niskanen et al. 2004; Camm et al. 1996). The spectral analysis is based on the spectral power differences between these frequency bands. Figure 6 shows an example of the frequency domain spectrum visualisation.

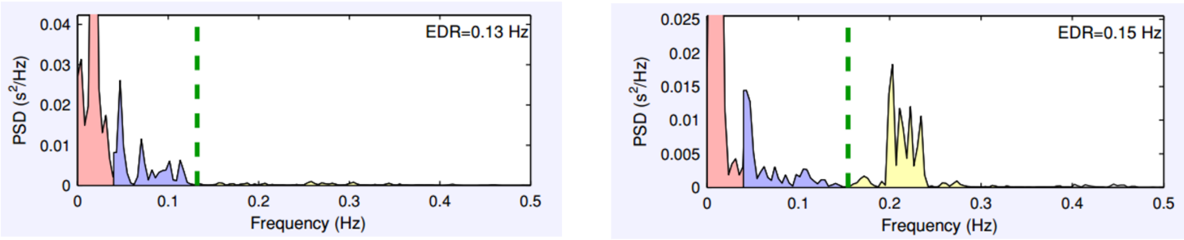


Figure 6. A visualisation of the subject's HRV spectrum. The sample on the left was taken during a trial and has a LF/HF power of 6.924. The sample on the right was taken during a rest and has a LF/HF power of 0.603. An elevated LF/HF value indicates higher PWML.

The question of the relative strengths and weaknesses of different PSD and time domain measures is not clear-cut; a measure that has been reported as being sensitive in one test setting may have been found insensitive in others. When PSD methods are being used in real-world settings the data may easily become confounded by extraneous interference. HR and the time domain HRV measures of HRV, while providing somewhat more rough estimates of the ANS activity, seem to suffer less of this type of interference. The numerous different ways to analyse the spectral power differences and the IBI differences complicates the comparison of older HR/HRV studies; for a long time there was no general agreement about the methods used in the analysis. However, in 1996 The European Society of Cardiology and The North American Society of Pacing and Electrophysiology set up a Task Force to set the standards for the measurement, physiological interpretation, and clinical use of HRV. The HRV analysis and interpretation of this thesis follow the guidelines of the Task Force. For additional information about the Task Force's standards, see Camm et al.1996.

In this thesis the NeXus-10 system was used to detect the cardiac ANS responses (see Figure 7). The NeXus-10 comes with a Biotrace+ analysis software. While the Biotrace+ is suitable for casual use, it has some critical limitations. For example, it does not allow manual artefact detection and correction.

Figure 7. NeXus-10 MkII multichannel biofeedback device

To overcome this limitation, the raw ECG signal was transferred to Kubios HRV software for further analysis (see Figure 8). Kubios HRV enables the user to conduct manual detection and correction of artefacts. Also, it allows the user to define the sample rate used. In addition, Kubios HRV permits the use of various analysis techniques, including time- and frequency-domain analysis of HRV. Both nonparametric and parametric spectrum estimates can be produced for the frequency domain variables.

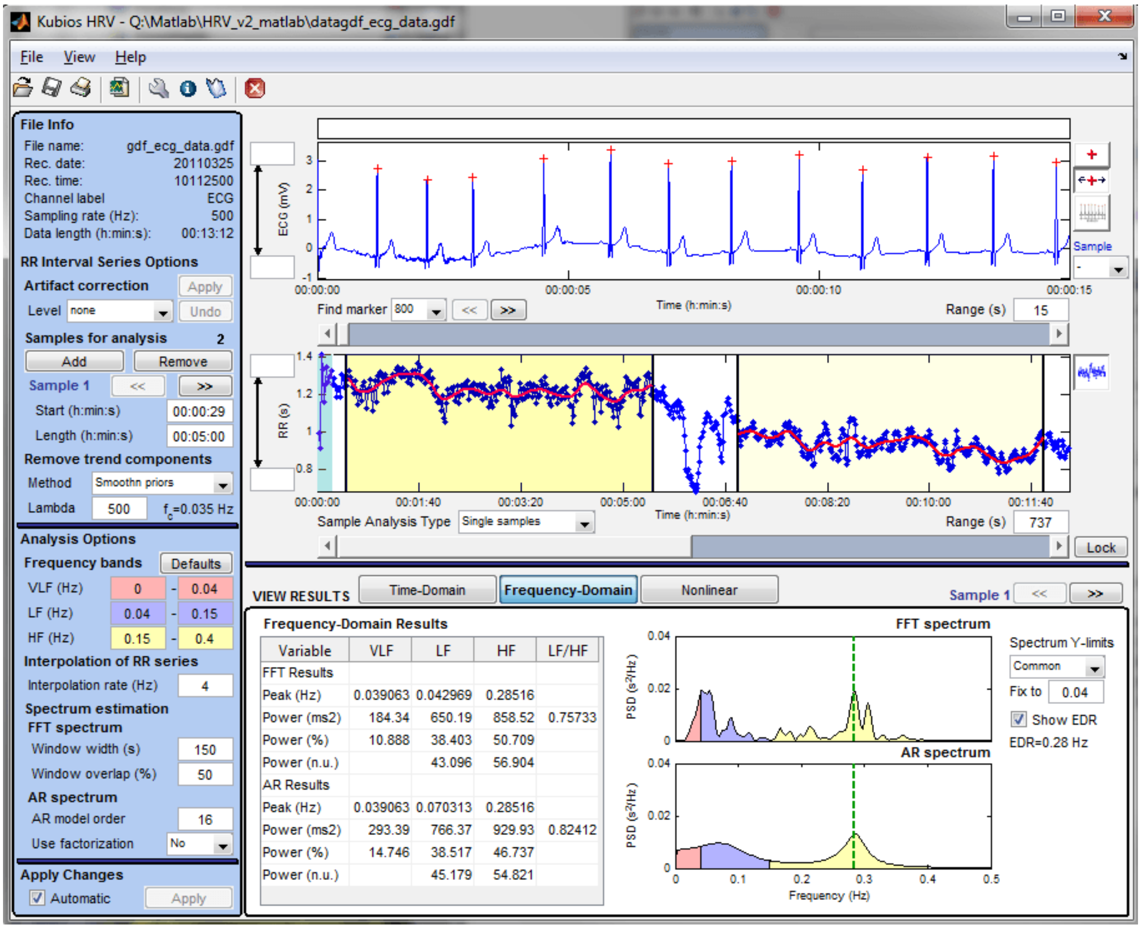


Figure 8. A screenshot from the user interface of the Kubios HRV analysis software.

Figure 9 shows the ECG sample as it is seen in the Kubios HRV software. The upper window in Figure 9 shows the whole ECG sample, whereas the lower window is a zoomed view of the selected time period within that sample. As can be seen from the extreme RR-peaks, the artefacts have not yet been removed. Figure 10 shows another ECG sample where the artefacts have been removed.

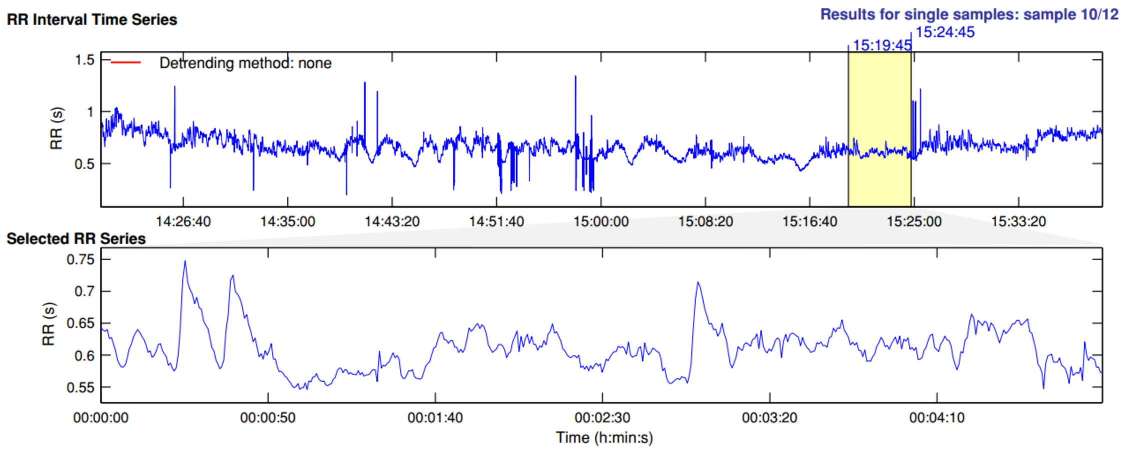


Figure 9. A sample view of raw ECG data as seen in the Kubios HRV analysis software's user interface.

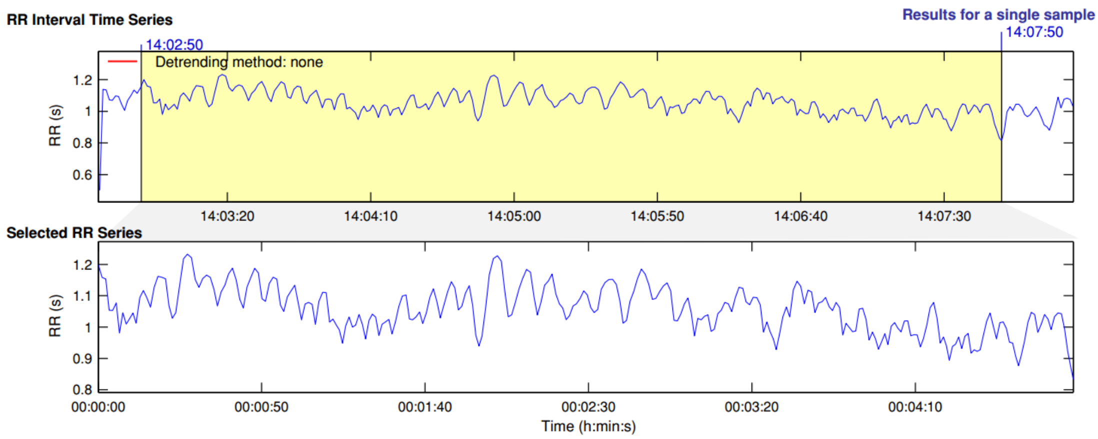


Figure 10. A sample view of the ECG data once the artefacts have been removed.

3.4 PMWL and Performance

The most intensive phases of fighter operations are typically the ones that have high, simultaneous resource demands on cognitive processing. The instrument approaches, especially when flown under difficult weather conditions, with minimum fuel and under the influence of distracting external inputs, expose the pilots to high, and sometimes excessive, workload. In a similar fashion, air combat manoeuvering with dynamically changing aerial situation, high approach speeds and very small margins of safety, has potential to exceed the limits of the human information processing. The studies discussed later in this thesis describe how these cognitive demands can and should be taken into account when the pilot performance is assessed.

There are several reasons to evaluate pilot performance; to determine the pilot's proficiency for his/her current or planned flying duty, to assess and evaluate the training programs and in the case of military forces, to manage the personnel and the allocation of flight hours. Measuring a pilot's task performance without considering PMWL is likely to provide an incomplete understanding of the pilot's actual or predicted combat proficiency. For example, without PMWL measurement two pilots with equal and satisfactory task performance scores during a check ride would always be considered equally competent (see Figure 11).

However, the uncertainty of the combat missions and the unexpected airborne incidents during the peacetime operations can generate additional stressors that negatively affect the pilot performance (Seck et al. 2005; Wickens and Huey 1993; Hancock and Warm 1989). An extreme PMWL during an above-standard check ride performance generates a high potential for sub-standard performance if the task demands are increased. Using the above example, differences in the pilots' cognitive capacity would provide them with an unequal potential to cope with those demands. Figure 12 illustrates how the situation described in Figure

11 could change if PMWL measurement would be combined with the task performance assessment. For a pilot to be able to deliver an above standard performance while exposed to additional stressors, s/he should be able to demonstrate enough spare mental capacity during the check ride – while delivering an above-standard performance. In other words, a satisfactory task performance during a check ride together with an unacceptably high mental workload should be considered sub-standard (see Figure 13).

Figure 11. The pilots' proficiency check scores when evaluated using task performance only. The square on the Y-axis illustrates pilot A and a circle illustrates pilot B. Both pilots have received equal task performance scores and are considered equally competent.

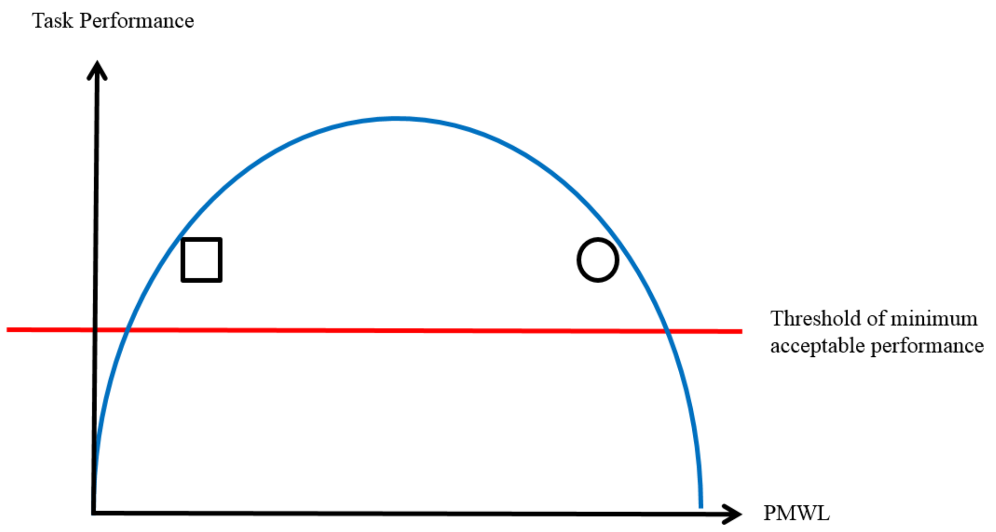


Figure 12. The pilots' proficiency scores when evaluated using both task performance and PMWL. The square illustrates pilot A and the circle illustrates pilot B. Pilots have both received equal, above standard task performance scores, but their PMWL is different. The blue arc illustrates the inverted U.

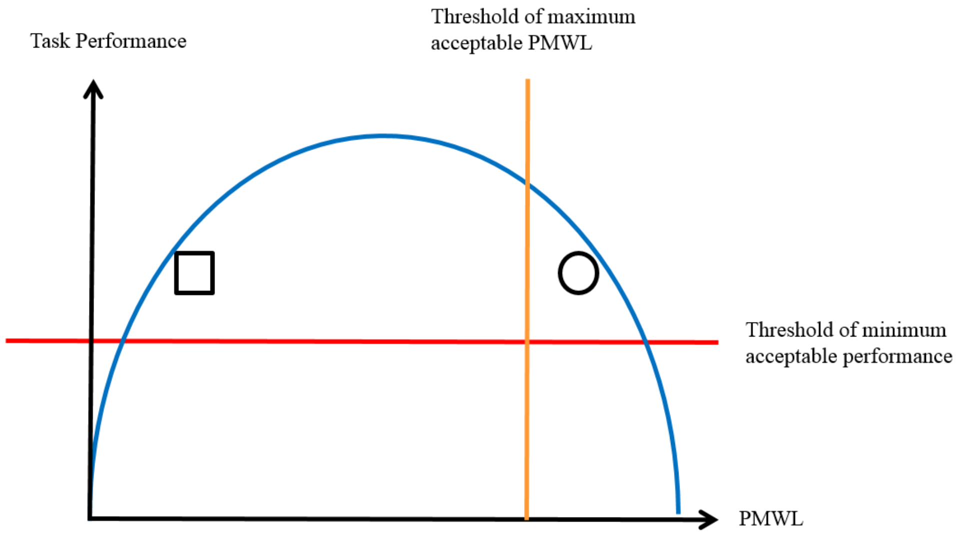


Figure 13. The pilots' proficiency scores when evaluated using both task performance and PMWL. The square illustrates pilot A and the circle illustrates pilot B. Both pilots have received equal, above standard task performance scores, but their PMWL is different. The blue arc illustrates the inverted U. Pilot B fails the proficiency check due to an unacceptably high PMWL.

4. AIMS AND OBJECTIVES

The approach presented in this thesis supports the development of reliable metrics for predicting the pilots' performance under the stress of combat (Prophet 1976). An acceptable workload during an above-standard performance is considered to reflect the pilot's spare mental capacity, which should help the pilot to maintain the above-standard performance even when exposed to additional stressors or task demands. The pilots' individual cognitive capacity differences and the aircraft types' and missions' different task demands make it impossible to define an universal relationship between PMWL and pilot performance. In addition, the lowest acceptable performance levels and the required standards for operational fighter pilots are service and nation dependent – and typically classified. As a result, determining the amount of an acceptable PMWL during a check ride, or the required excess mental capacity, is left for the practitioners to judge.

This thesis is about understanding how PMWL affects the pilot performance. The operational F/A-18 pilots and typical F/A-18 missions were used to study the association between PMWL and pilot performance. The national or Air Force performance standards are not discussed. Before it was feasible to evaluate the utility of HR and HRV in a check ride setting, their sensitivity to varying task demands in a simulated fighter mission had to be assessed. In addition, their ability to differentiate the pilots' sub-standard performance from the above standard performance had to be verified.

The following chapters report three independent, but interrelated studies. The first study (Chapter 5), *“Fighter Pilots’ HR, HRV and Performance during Instrument Approaches”* (Mansikka et al. 2016a), documents how HR and HRV were used as measures of PMWL during simulated instrument approaches and to what extent they were able to differentiate different temporal demands and pilot performances. The second study (Chapter 6), *“Fighter Pilots’ HR, HRV and Performance during an Instrument Flight Rules Proficiency Test”* (Mansikka et al. 2016b), builds on the findings of the first study and reports how HR/HRV were used as measures of PMWL during a real instrument flight rules proficiency test. The aim of the study was to evaluate if HR/HRV can differentiate the task demands between the different mission segments. The third study (Chapter 7), *“Fighter Pilots’ Mental Workload and Performance: A Comparison of Simulated Instrument Approaches and Air Combat”* (Mansikka et al. 2016c), compares pilots’ HRV responses between the simulated instrument approaches and air combat. The aim of the study was to evaluate whether the findings of the previous two studies could be extended to an air combat environment and what other variables, if any, might be needed to explain the association between PMWL and performance.

5. FIGHTER PILOTS' HR, HRV AND PERFORMANCE DURING INSTRUMENT APPROACHES

5.1 Introduction

The cockpit of a modern multirole fighter is one of the most cognitively demanding work environments, exposing the pilot to extreme physical and psychological stress and fatigue (Driskell and Salas 1991). Pilots' failure to cope with the task demands may degrade flight safety and compromise mission success with fatal results (O'hare 2000; Shappell and Wiegmann 1997; O'Hare et al. 1994; Sheridan and Simpson 1979). During aircraft and system development, a great deal of effort is placed on managing PMWL through the design of the human-machine interfaces, i.e., fitting the task to the man (Grandjean and Kroemer 1997). As discussed in Chapter 1, once the platform is released for operational use the workload management becomes an issue of fitting the man to the task, e.g., through selection and training.

To make sure pilots are competent for their flying duty, air forces conduct mandatory proficiency checks for their flight crews (Mavin and Roth 2014). These proficiency checks or 'check rides' are conducted to assess the pilots' performance against standards with the aim of guaranteeing their acceptable operational performance.

A check ride usually consists of mission critical task elements where pilot performance is evaluated by an instructor pilot or examiner. For the pilot to pass a check ride, s/he needs to score a predefined number of points on a specific grading scale. Based on performance the pilot may be given a certificate to operate a specific platform or piece of equipment, a qualification to operate in certain weather conditions, or the pilot may - or may not - be given an appropriate readiness status.

Even though a live aircraft mission is in some instances the recommended method of conducting a check ride, a high fidelity simulator is often the preferred platform. This is as a result of the lower operating costs of the simulator, the easily adjustable environment and system conditions. Also, missions including critical emergency procedures, use of deadly force or operations with a minimal safety margin are almost impossible to conduct realistically or safely in a live flying environment

While a simulator mission can be designed to be extremely mentally demanding, it will inherently lack the stressors of a real flying mission such as the sense of risk and the fear of collision, injury or death. Consequently, a simulator mission is generally less cognitively demanding than a similar mission in a real flying environment (Svensson et al. 1997; Jorna 1993). If PWML is not part of the performance assessment criteria, a check ride conducted in a simulator may provide misleading indications of pilot's performance in real-life situations; a pilot may show acceptable performance in the simulator but executes the same tasks to a sub-standard level in a similar live mission as a result of increased PMWL or stress (see Chapter 3.4) (Young et al. 2014; Lieberman et al. 2005; Berkun 1964).

Several studies have used laboratory environments to evaluate the relationship between an operator's mental workload and performance (Iani et al. 2007; Morris and Leung 2006; Kaber and Endsley 2004; Vitense et al. 2003; Zakay and Shub 1998; Jorna 1993). However, these studies provide only a limited understanding of the cognitive demands of real life systems. On the other hand, studies conducted in operational environments provide inadequate insights concerning the level of PMWL leading to pilot's sub-standard performance (Lahtinen et al. 2007; Magnusson 2002; Veltman 2002; Svensson et al. 1997).

There are many sophisticated physiological measures of individual differences in regulated emotional responses available, such as EDA (Collet et al. 2014), functional near-infrared (fNIR) spectroscopy (Ayaz et al. 2010) and eyelid closure (Mallis and Dinges 2004). These measures are often difficult to implement into a flight simulator environment without unacceptable levels of pilot intrusion and/or disturbances to simulator instrumentation. Other measures such as HR and HRV, although somewhat less sophisticated and novel, have successfully been applied in a flight simulator environment. For a more detailed discussion regarding the physiological measures, see Chapter 3.2.3.

As discussed in Chapter 3.3, HR and HRV represent the activation of the ANS (Stuiver et al. 2014; Hayward et al. 2014; Xhyheri et al. 2012). The time domain methods of HRV analysis involve determining the intervals between successive normal QRS complexes (i.e., NN). From the NN, other HRV components can be derived and used as measures of mental workload, for example; MEANHR (Saperova and Dimitriev 2014; Roman-Liu et al. 2013; Pérusse-Lachance et al. 2012), SDNN (Tran et al. 2010; Terkelsen et al. 2005), RMSSD (Mehler et al. 2011; Li et al. 2009; Orsila et al. 2008), NN50 (Deepak et al. 2014), pNN50 (Taelman et al. 2011), MEANRR (Terkelsen et al. 2005; Sun et al. 2012) and HRVTRI (Cinaz et al. 2013). Table 2 (page 56) provides definitions for the mentioned HRV components. Several studies have been able to demonstrate the changes in pilots' HR and HRV during different flying mission and phases of missions (Veltman and Gaillard 1998; Svensson et al. 1997; Wilson 1993; Roscoe 1993; 1992; Aasman et al. 1987). Furthermore, pilots' primary task performance has been successfully linked to PMWL, HR and HRV (Svensson et al. 1997). However, little is known about the relationship between pilot performance and PMWL during a real training mission or a check ride. As a result, practically no attempts have been made to introduce PMWL as an additional performance criterion for a check ride. It is therefore necessary to study the relationship between PMWL and pilot performance in a real, or realistically simulated operating environment, using representative tasks and associated with existing operational performance standards (Jorna 1992; Rasmussen and Jensen 1974). HRV and HR have been proven to be sensitive measures of PMWL at the higher levels of workload. As the present study was focused on the higher end of PMWL, the possible sensitivity limitations of HRV/HR at the lower levels of workload were not an issue.

The aim of this study was to investigate if HR and the selected time domain components of HRV are related to variations in pilot performance during a simulated flying mission (see research question one, Chapter 1, Figure 1). It was hypothesised that pilot performance was associated with HR and the time domain components of HRV. Ultimately the objective of this study was to evaluate if HR and the HRV measures could identify the level of task demands leading to a sub-standard performance (see research question two, Chapter 1, Figure 1). This finding would suggest that the MEANRR could be a useful measure of PMWL if an actual PMWL redline could be defined in a fighter simulator environment (Young et al. 2014; Brookhuis et al. 2003). To this end, it was necessary to study the dependence between HR, the time domain components of HRV and performance measures. To achieve this, a realistic mission with varying levels of task demands was developed in a high fidelity F/A-18 flight simulator. Operational F/A-18 pilots were recruited as subjects and real air force operational standards were used to assess pilot performance. By utilising such a test design, pilot performance was measured together with HR and the time domain components of HRV in order to describe the inter-dependence between the pilot performance and PMWL.

5.2 Method

5.2.1 Participants

Thirty-five Finnish Air Force (FinAF) F/A-18 pilots participated in the study. The subjects' average flying experience with the F/A-18 was 598 flight hours. Subjects were randomly selected from the fighter squadrons' pilot population. Pilots' backgrounds ranged from wingman to air combat instructor, which resulted in large variation in their flying experience (standard deviation of 445 hours). Written, informed consent was obtained from each subject (see Appendix 1). A structured proforma (see Appendix 2) was used to collect subjects' background data and information concerning their relevant activities for the 12 hours prior to participating. The proforma was prepared with the assistance of an aeromedical professional (M.D.) from the Satakunta Air Command, Finland. All subjects had gone through an extensive aeromedical testing within the last 12 months and were fit to fly at the time of the study. The study was reviewed and approved through the Coventry University's Ethical Review Process (see page iv).

5.2.2 Study Design

As a result of the time consuming, repetitive nature of the test design used, it was not possible to utilise an actual instrument proficiency test. Instead, an ILS approach, one component of the proficiency test, was used for the task demand manipulation. The subjects completed 12 full test procedures each consisting of an ILS approach with different level of task demand. The task demand was manipulated by increasing the temporal demand on pilots by reducing the range at which they commenced the trial. The trial order was randomised between subjects.

A Boeing built weapon tactics and situational awareness trainer (WTSAT) was used for the piloting task (see Figure 14). The WTSAT is used at the FinAF's fighter squadrons for basic and advanced F/A-18 pilot training. The WTSAT is a non-motion, high fidelity flying simulator, with a 135 degree field of view and a fully functional cockpit. The WTSAT replicates the F/A-18 flying characteristics with such a high accuracy that the FinAF F/A-18 pilots can use it to fly their annual instrument check rides.

Figure 14. WTSAT simulator setting

For the study the wind was set to 320 degrees, 10 knots (5.14 m/s) with moderate gusts. Clouds were set to overcast with the cloud top at 30,000 ft (9,144 m) and the cloud base at 200 ft (60 m) from the ground level. Instrument meteorological conditions (IMC) visibility was set to 0 ft (0 m) and the runway visual range (RVR) was set to 700 m (2,296 ft). Light conditions were set to mirror the average light at the Tampere – Pirkkala airport (International Civil Aviation Organization code: EFTP) in Finland on 1st June at 12:00 o´clock local time. Runway was dry and the braking action was good. The arresting cable and the net barrier were not available.

Before commencing the trials a baseline ILS mission was flown. For this mission, the simulator was initialised to 2,000 ft (607 m) above ground level, 9.5 NM (17.6 km) from the touchdown point, minimum approach speed, straight and level flight as well as 0 ft (0 m) azimuth and heading error for the standard ILS approach. The cockpit settings were, however, set incorrectly for the approach and landing; for example, the radios were set to wrong frequencies, the altimeter setting was incorrect and the platform was not configured for landing. By using this kind of mechanisation, the baseline ILS mission closely mirrored the actual trials. For the baseline ILS mission, the objective was to fly a simple, undisturbed ILS approach.

Each trial consisted of an ILS task and additional flying related sub tasks. The ILS task was a standard ILS approach to EFTP runway 24. The pilots were tasked to fly the ILS approach using a platform specific minimum approach speed and a flight profile established for the approach in the official instrument approach chart (IAC). The ILS task started at the GS intercept range and ended at 0.5 NM (0.9 km) from the touchdown point, which was the range at which the standard GS met the ILS decision height (DH). The platform specific DH was the same as the cloud base, thus allowing subjects to land after the successful ILS approach. The sub tasks comprised of carefully selected activities relevant to F/A-18 operations. These included tasks such as setting up the cockpit instruments for the specified approach and landing, flying from the DH to touchdown, communicating with the air traffic controller (ATC) and reacting to in-flight emergencies requiring immediate pilot actions. The sub tasks and the different components of the ILS task used in the study are listed in Table 1 where the ILS task components are marked with a shaded background. To force the subjects to study the IAC, during each trial a different IAC was used (only the ILS flight profile, runway altitude and ILS LLZ frequency were kept identical). The subjects had to copy three altitudes from the IAC to their knee pads and to study the IAC's frequencies as well as to tune six radio presets accordingly. With the exception of the IAC, the pilots were highly familiar with the sub tasks so they had no need to refer to other check lists or supporting documentation to undertake them. Although the sub tasks were standard procedures for any F/A-18 pilot, the subjects practiced each sub task before the trials.

Task	Required Pilot Action
ATC clearance 1	Read back the clearance. Copy the clearance on a knee pad.
ATC clearances 2-3	Read back the clearance and swith to an indicated frequency
ATC clearance 4	Read back the clearance
ATC inquiries 1-3	Check the requested flight parameter and report it to the ATC
ATC directives 1-3	Set the cockpit instrument to a directed value
Engine warnings 1-4	Initiate a related emergency procedure
Icing warning	Select an anti-ice switch to 'ON'
Flight control system warning	Initiate a related emergency procedure
Environmental control system warning	Initiate a related emergency procedure
Fuel level warning	Reset a fuel level warning
Mental task	Calculate the landing speed in km/h based on an indicated fuel state
Display failure	Swith to use an alternate display
IAC radio frequencies 1-6	Tune a radio preset to a frequency indicated in the IAC
IAC parameters 1-3	Copy a value indicated in the IAC to a knee pad
Land	Fly from DH to touchdown point and make a full stop landing
ILS localizer	Maintain an approach course in accordance with the ILS localizer
ILS glidepath	Maintain a glidepath in accordance with the ILS glidepath
Approach speed	Maintain a minimum approach speed

Table 1. The sub tasks and the components of the ILS task. The ILS task components have been shaded.

For the trials the simulator was initialised to the same parameters (with the exception of the starting range) as for the baseline ILS mission. As the task demand was manipulated by varying the starting ranges of the trials, each trial started from a trial specific starting range. The starting ranges were measured as a horizontal distance from the touchdown point. The starting ranges varied from 5.5 NM (10.2 km) to 15.5 NM (28.7 km) at 1 NM (1.9 km) increments. The minimum value of the starting range variable (5.5 NM) equaled the ILS glideslope intercept range, i.e., the starting point of the ILS task. The maximum value of the starting range variable (15.5 NM) was defined during the pretesting as a maximum range it took for the pilots to perform the sub tasks when done undisturbed and at pilots' own pace. Subjects were tasked to fly each approach at a constant minimum approach speed. As a result, the time pressure for the ILS task and the sub tasks varied from 6 minutes and 35 seconds (15.5 NM starting range) to 2 minutes and 20 seconds (5.5 NM starting range). Each trial ended at touchdown or attempted touchdown. A 1 NM decrease in the starting range reduced the available time to conduct the ILS task and the sub tasks by 25.5 seconds.

Triggering times for the sub tasks, except for the landing itself, were randomised between the trials and could potentially occur anywhere between the start of the trial and the landing. Within a trial the sub task triggering times were same for each subject. Tasks including radio transmissions were prepared as audio files and activated based on the elapsed time from the start of the trial. The audio files were played through the pilot's headset. When two or more ATC radio transmissions were to be triggered simultaneously, they were manually separated during the audio file preparation. The manual deconfliction of the sub task triggering times was limited to ATC radio transmissions only. Figure 15 illustrates how the starting range variable, the ILS task and the sub tasks were related. In Figure 15, the 5.5 NM and the 15.5 NM starting ranges are shown to highlight the difference between the trials of highest and the lowest temporal demand.

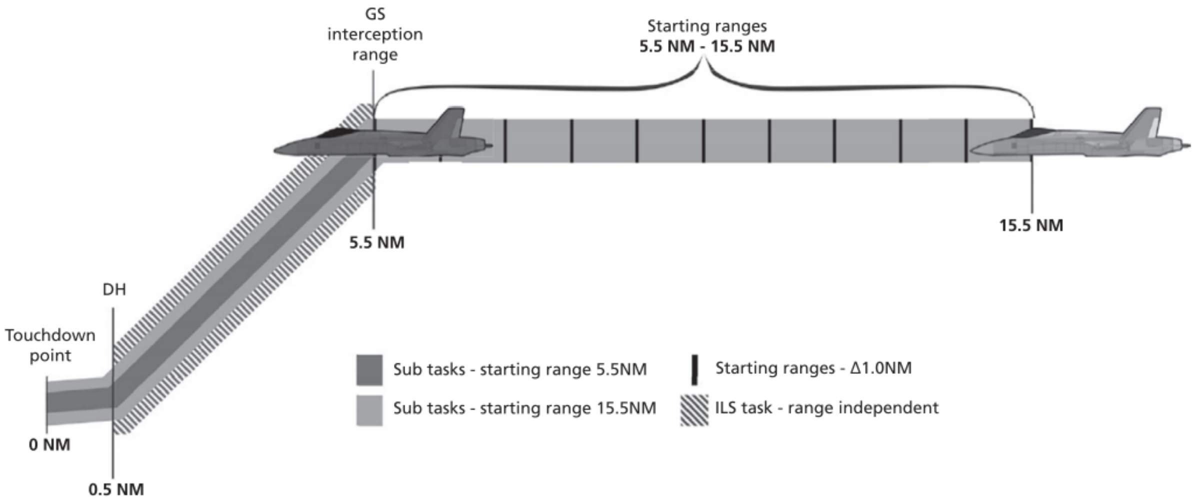


Figure 15. The starting ranges for ILS task and sub-tasks.

Each trial was separated by a rest period lasting approximately three minutes. During the rest period the simulator was re-initialised for the next approach. The flying mission used for the study was treated as a flight curriculum's training sortie and the subjects prepared for the mission accordingly.

A pilot study was conducted in order to evaluate the test design. Five pilots flew the mission using various starting ranges. In addition, each pilot conducted all the sub-tasks. The findings of the pilot study were used to finalise the triggering times of the sub-tasks and to adjust the data collecting instrumentation.

5.2.3 Procedure

The ratio between the time needed for task completion within a trial and the time available to complete them, or the time pressure, was used as an independent variable. To increase the sense of authenticity of the flying mission the subjects were free to select their individual piloting and problem solving strategies.

ECG was recorded with Mind Media NeXus-10 MKII system supported by Biotrace+ software (version V2012C). Three electrodes were placed below the left (negative) and right (ground) clavicle and the left costal cartilage (positive) respectively (see Figure 16). The Biotrace+ samples were exported to Kubios HRV 2.2 software for further analysis and RR interval artefact removal (see also, Chapter 3.3.2). All artefacts were detected and removed manually and noisy data was excluded from the further analysis. A specialist of internal medicine was consulted when necessary. ECG measuring, manipulation and interpretation were done in accordance with the guidance in Task Force of the European Society of Cardiology and The North American Society of Pacing and Electrophysiology (Camm et al. 1996). After the last trial the subjects were asked about the level of intrusion caused by the NeXus-10 MKII system. None of the subjects reported intrusion of any kind.

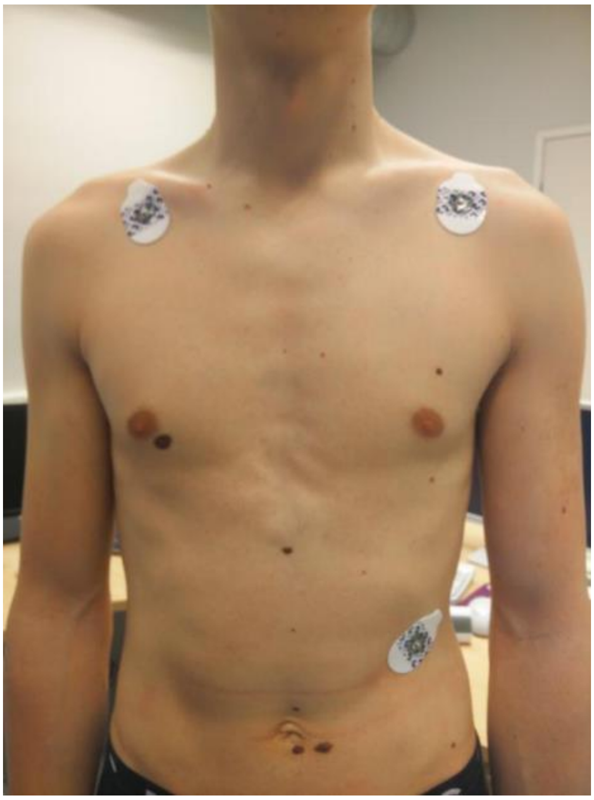


Figure 16. Electrode placement.

A five minute pre-trial rest period was used to record the rest baseline HR/HRV and a three minute sample was taken from it for further analysis. During the rest baseline recording the subjects sat undisturbed in the simulator. As there are great differences in the individual cardiac activity, the subjects' cardiac responses to varying task demands were compared within each subject and not across subjects (Roscoe 1993).

A three minute sample was taken from the end of the baseline ILS and each trial's ECG recording. The values of the HR/HRV components recorded during each trial were compared to the ECG data from the subject's other trials, the baseline rest condition and the baseline ILS mission. HR/HRV components used for measuring PMWL are listed in Table 2.

Measure	Unit	Description	Expected change due to PMWL increase
MEANRR	[ms]	The mean of NN intervals	Decrease
SDNN	[ms]	The standard deviation of NN intervals	Decrease
MEANHR	[1/min]	The mean heart rate	Increase
RMSSD	[ms]	The square root of the mean squared differences between successive NN intervals	Increase
NN50	[count]	The number of successive NN interval pairs that differ more than 50ms	Decrease
pNN50	[%]	The NN50 divided by the total number of NN intervals	Decrease
HRVTRI	[-]	The integral of the NN interval density distribution divided by the maximum of the distribution.	Decrease

Table 2. Summary of HR/HRV components.

The ILS task performance was rated between the ILS GS intercept range and the ILS DH using an official FinAF instrument check ride rating scale. The rating was based on a deviation from the target speed along with the LLZ and GS errors. The values of the rating scale ranged from 5 (best performance) to 0 (worst performance). The ILS scoring was conducted by using the simulator's mission replay. Between the GS interception range and the DH the mission playback was stopped at every 0.5NM (0.9km). While stopped, the deviations from the GS, LLZ and target speed were recorded and scored. The mean of the scores was used as an ILS task performance score (see Appendix 3). The ILS task performance was rated by a qualified F/A-18 examiner pilot. The examiner pilot's ILS performance scoring was based solely on the deviations from the target flight parameters (deviations from target speed, GS and LLZ). More subjectively rated aspects of performance (such as smoothness of aircraft handling) were not scored. The ILS task performance score and the values of the HR/HRV components were used as dependent variables.

5.3 Results

For a pilot to achieve a 1st class instrument rating on a real instrument check ride, s/he needs to achieve at least 60% of the ILS maximum score. In this study the threshold for the sub-standard performance was set to 60% of the absolute maximum ILS score, which mirrored the FinAF standards for the official instrument check ride.

The ILS scores were used to form three different performance categories. A *high performance* category was formed by selecting each pilot's ILS performance score from the baseline ILS mission. For the formulation of the *sub-standard performance* category only the trials with the sub-standard performance were considered. Out of these trials, the trial with the highest ILS performance score was selected for each subject. A *low performance* category was formed by selecting each pilot's trial that had the lowest ILS performance score. The ILS performance scores and the respective values of the HR and HRV components were plotted for each trial.

Data were analysed using IBM™ SPSS™ software (version 22). The 5.5 NM (10.2 km) and the 6.5 NM (12 km) trials were left out from the final analysis as their durations were too short for a reliable HRV analysis. The 7.5 NM (13.9 km) trial was the most frequent candidate for the low performance category. It also had the most missing data, which reduced the number of subjects to 23. To increase the sample size, the 7.5 NM (13.9 km) trial was excluded from the analysis and replaced with the 8.5 NM (15.7 km) trial resulting in a sample size of 28 subjects.

Values of each subjects' HR/HRV components were retrieved for the analysis from four measurement points. The measurement points comprised of the last three minutes of the baseline rest, the baseline ILS mission (i.e., the high performance category), the trial with the highest sub-standard ILS performance score (i.e., the sub-standard performance category) and the trial that had the weakest ILS performance score (i.e., the low performance category).

The HR/HRV components' values were analysed using the repeated measures MANOVA. Post-hoc pairwise comparisons were carried out with the paired t-test. Violation of sphericity and homoscedasticity was handled with the Greenhouse-Geisser correction when necessary.

Table 3 presents the descriptive statistics of the HR/HRV components for each measurement point. There were statistically significant overall HR/HRV differences between performance categories; $F(7,21)=3.9$, $p<0.05$, $\eta^2_p=0.94$. Significant HR/HRV differences between performance categories were found on: MEANRR $F(3,81)=47.1$, $p<0.001$, $\eta^2_p = 0.64$; SDNN $F(3,81)=6.5$, $p<0.01$, $\eta^2_p=0.19$; MEANHR $F(3,81)=31.6$, $p<0.01$, $\eta^2_p=0.54$; NN50 $F(3,81)=18.1$, $p<0.001$, $\eta^2_p=0.40$; pNN50 $F(3,81)=8.4$, $p<0.01$, $\eta^2_p=0.24$; HRVTRI $F(3,81)=17.2$, $p<0.001$, $\eta^2_p=0.38$.

		Baseline rest		Baseline ILS		High perf. category		Satisfactory perf. category		Low perf. category	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
MEANRR	[ms]	840.4	181.8	722.0	164.5	687.0	126.5	656.4	129.0	660.7	126.1
SDNN	[ms]	69.5	21.8	52.8	28.6	49.3	24.3	48.2	21.2	47.3	22.7
MEANHR	[1/min]	75.3	16.7	87.7	19.1	90.6	15.8	94.0	17.9	94.4	16.8
RMSSD	[ms]	44.9	23.4	32.1	29.6	32.0	33.0	26.7	25.1	26.7	27.5
NN50	[count]	75.0	54.4	20.5	31.0	20.8	41.1	22.4	50.3	21.3	48.1
pNN50	[%]	20.5	16.3	9.1	12.5	8.5	14.6	8.4	15.6	8.0	15.3
HRVTRI	[-]	15.6	4.9	10.6	4.8	10.0	3.7	10.2	3.9	10.5	4.3

Table 3. Means and standard deviations of the HR/HRV components at the measurement points (N=28).

The results of the pairwise comparisons are summarised in Table 4. All HR/HRV components showed significant difference between the baseline rest and the high performance category ($p<0.05$). The task demand and the ILS performance changes between the sub-standard performance category and the low performance category were not differentiated by the HR/HRV components used.

		Baseline rest - High performance category			High performance category - Sub-standard performance category			Sub-standard performance category- Low performance category		
		Mean Diff	Std. Error	t-value	Mean Diff	Std. Error	t-value	Mean Diff	Std. Error	t-value
MEANRR	[ms]	103	11.1	9.3 ***	58.4	16.2	3.6 **	3	6.4	0.5
SDNN	[ms]	15.9	4.7	3.4 **	2.4	4.9	0.5	-0.7	3	-0.2
MEANHR	[1/min]	-11.4	1.5	-7.2 ***	-6.8	2.1	-3.2 **	-1	1.2	-0.8
RMSSD	[ms]	9.4	4.2	2.2 *	1.1	4.6	0.2	-2.9	2	-1.4
NN50	[count]	51.5	8.2	6.3 ***	-1.7	5.8	-0.3	-4.6	3.3	-1.4
pNN50	[%]	9.4	2.1	4.4 ***	0.9	2	0.4	-1.1	1	-1.2
HRVTRI	[-]	4.8	0.8	5.8 ***	-0.1	0.8	-0.2	0	0.6	0.1

Table 4. The values of the test statistics and changes in pairwise comparisons between measurement points; ***$p<0.001$; **$p<0.01$; *$p<0.05$ (N=28)

MEANRR ($p<0.01$) and MEANHR ($p<0.01$) were able to differentiate the high performance category from the sub-standard performance category. Figure 17 illustrates the MEANRR values across the measurement points.

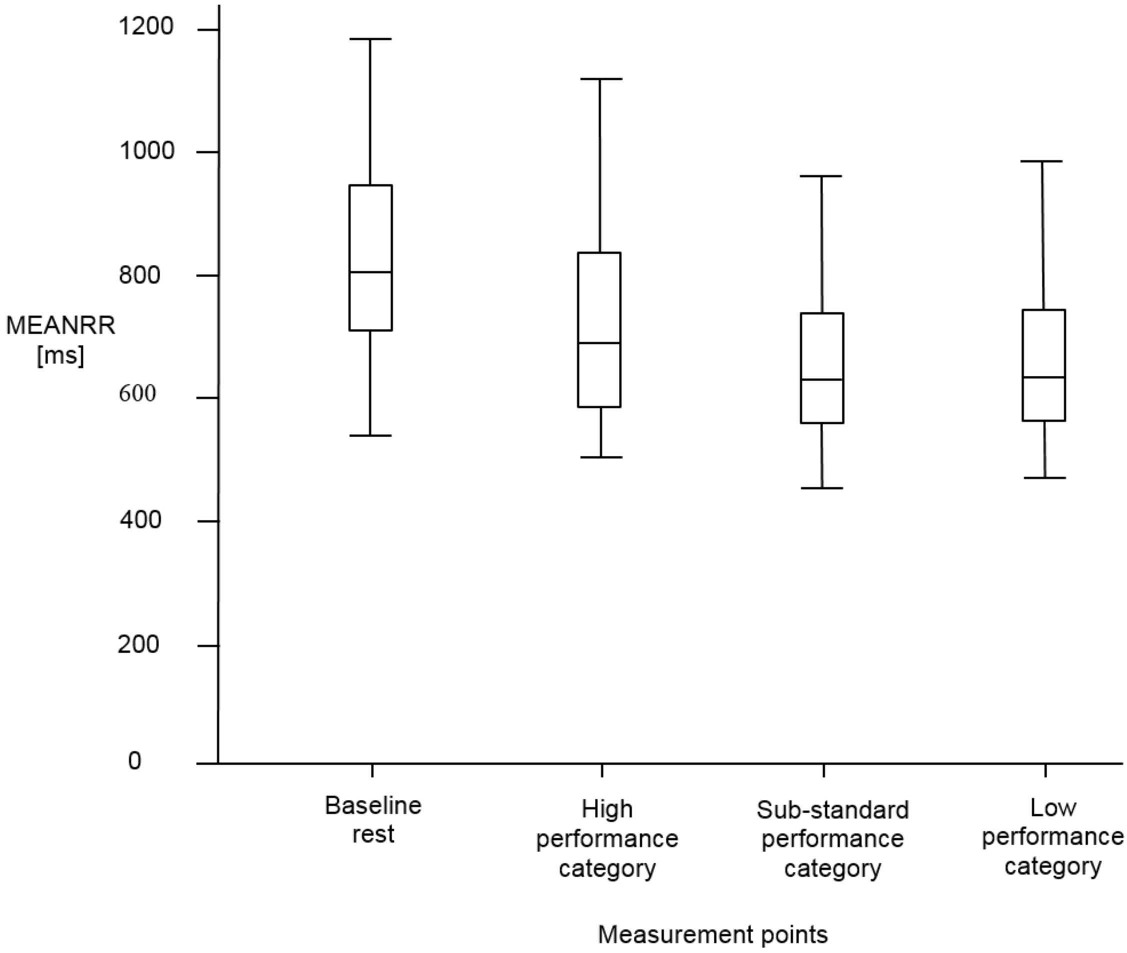


Figure 17. MEANRR for the baseline rest, the baseline ILS mission, the high performance category, the sub-standard performance category and the low performance category.

5.4 Discussion

This study extended the findings of the earlier studies by investigating the associations between the PMWL and performance. The main finding of this study was MEANRR's ability, as a measure of PMWL, to differentiate between the high- and low performance of the pilots. In this context, such a finding has not been previously reported. The results (see Chapter 5.3) clearly indicated that HR and HRV are sensitive to varying ratios between the time available and the time required for completing the tasks in the fighter aviation environment. The MEANRR was able to differentiate the high performance and the sub-standard performance categories, i.e., their MEANRR averages were significantly different across the pilot population (see Figure 17). The sub-standard performance category and the low performance category were not differentiated by the HR/HRV components used (see Table 4). The results of this study provided an encouraging basis for testing HR/HRV components' sensitivity in more realistic and complex fighter missions.

It is possible that the secondary task measures could have been able to identify the task demand manipulation conducted in this study. Unfortunately, as discussed in Chapter 3.2.1, the use of secondary tasks would had destroyed the illusion of 'free play'. However, the secondary task measures can still be useful as they can identify the exact mental resource that becomes a limiting factor. That being said, it seems likely that different PMWL measures should be used during mission or system design, OT&E and execution.

Multiple resource modelling of task interference can be used to evaluate how the planned tasks potentially interfere with the different cognitive processes (Horrey and Wickens 2003). While this technique can reveal the potential resource conflicts, it cannot predict the expected performance, as it does not take into account the operators' individual mental capacity differences. Use of multiple resource modelling of task interference was considered during the preparation phase of this study. However, during the mission testing it was soon realised that there were remarkable task demand variations even within a same trial when the trial was executed by different pilots; while each ILS starting range was related to an exact temporal demand, the pilots' different reactions during the task execution generated variations to the overall, subjective task loading. For the reasons discussed above, the use of task interference modelling to predict task demand or PMWL was considered impractical and was not utilised.

To gain a more profound understanding of the subjective task demands, it might had been valuable to investigate the pilots' attention sharing. As the triggering times of the additional stimuli (i.e., sub-tasks) were scripted, it might had been possible to assess the time it took from the stimuli to be presented to the moment the pilot responded to it. Such an approach could had revealed those stimuli that the pilots did not notice and did not therefore affect their cognitive resource demand or expenditure. However, such a methodology would not had been able to reveal the time from the triggering of the stimuli to the moment it was perceived. That is, the simultaneous or almost simultaneous sub-tasks, together with the continuous primary task, did not necessarily allow perfect time sharing between the tasks. Therefore, the pilots had to prioritise the task responses; some responses were executed immediately while the others were postponed. It is obvious that such delayed responses generated very high demands on WM, especially as the high frequency of auditory inputs and vocal responses complicated the rehearsal process (see Chapter 2). Finally, even if the stimuli had never been associated with a relevant response, it would had been impossible to define whether this was due to the limitations of the central processing (WM) or perception (improperly focused attention). In addition, the possible lack of response may have resulted from a conscious decision; the pilot may have decided not to spend time and resources on a certain response if other, higher priority tasks were present.

When the triggering times of the sub-tasks were randomised, the differences in the order of occurrence affected the overall task demand increase. In other words, the task demand increase was related not only to the reduced starting range, but also to the sequence that the sub-tasks occurred. However, had the sub-task occurrence sequence been kept the same between trials, the learning effect might have had a significant impact on PMWL. This unwanted impact of the learning effect became obvious during the pre-study testing.

When different methods are considered for PMWL measurement, practical restrictions may limit the use of some otherwise useful methods. Despite the valuable efforts of evaluating PMWL during the system and mission design, testing and evaluation, the pilot flying the mission ultimately defines what the resulting PMWL and man-machine system performance is. HR and HRV proved to be highly useful measures for this purpose.

PMWL is typically not measured during the pilots' IFR proficiency test. Therefore, little is known about the PMWL during the proficiency test and the pilots' potential to cope with the higher task demands than those experienced during the proficiency test. The next study introduces how the fighter pilots' performance and PMWL were measured during a real IFR proficiency test in an F/A-18 simulator. A successful PMWL measurement during a proficiency test would motivate the utilisation of HR and HRV in more complex, tactical scenarios.

6. FIGHTER PILOTS' HR, HRV AND PERFORMANCE DURING AN IFR PROFICIENCY TEST

6.1 Introduction

Pilots' IFR performance is an essential contributor to operational effectiveness and safety of flight. European Aviation Safety Agency requires pilots to pass an annual revalidation flight, or a check ride, in order to maintain their IFR currencies (https://easa.europa.eu/regulations). During an IFR check ride, the pilots' performance is assessed against the predefined performance criteria with the intent of verifying their proficiency to operate in IMC. In military aviation, similar IFR (re-) validation check rides are used (Mavin and Roth 2014). Modern, high fidelity simulators allow IFR check rides to be flown in a simulated environment, which reduces risk, allows for more precise data logging and performance feedback and increases aircraft availability (Sarter, Mumaw and Wickens 2007; Weitzman et al. 1979; Valverde 1973).

When task demand is increased during an IFR flight, pilots may compensate for it by investing more effort which in turn increases PMWL. (Shaw et al. 2013). Once the mental capacity and/or willingness to invest more effort is exceeded, at some point pilots' performance begins to degrade (Young et al. 2014; O'Donnell et al. 1986). There is a great risk of compromising flight safety and mission success if these conditions occur during live flying. Measuring PMWL during an IFR check ride can give valuable information about the pilots' ability to maintain the desired performance during events of high task demand. Two pilots with an equal task performance during an IFR check ride may have significantly different cognitive spare capacities, which reflects their potential to cope with subsequent task demand increase (O'Donnell et al. 1986; Yerkes and Dodson 1908). For a more complete discussion regarding the information processing, mental workload and performance, see Chapters 2 and 3.4. PMWL or spare mental capacity is typically not evaluated during an IFR check ride. To the best of the authors' knowledge, no previous PMWL assessments in the open literature have considered fighter pilots' IFR check rides.

Evaluation of the pilots' spare mental capacity requires measuring of PMWL for which task performance (see Chapter 3.2.1), subjective reports (see Chapter 3.2.2) and physiological metrics (see Chapter 3.2.3) are typically used (Boff et al. 1994). Subjective measures of PMWL, such as the NASA-TLX and the MCH scale, have been widely used in the aviation domain (Hart and Staveland 1988; Wierwille et al. 1985; Casali and Wierwille 1983). While the multidimensional scales, such as the NASA-TLX, have better reliability, diagnosticity and validity than the uni-dimensional scales, these types of subjective reports are too intrusive to be used during flight or simulated flight. Also, it should be noted that the subjective ratings can become dissociated with performance, especially if the task is resource limited (Yeh and Wickens 1988). In addition, the data for these measures are typically collected after the trial making them less capable of identifying sudden changes in PMWL. In the aviation domain even sudden, short term PMWL overload conditions may jeopardise flight safety and need to be therefore identified. The instantaneous self-assessment (ISA) technique was considered as a potential real-time subjective measure of PMWL. However, as the PMWL was measured during a real IFR check ride, the use of ISA had to be discarded due to potential primary task intrusion (Tattersall and Foord 1996). Furthermore, if PMWL is to be used as an additional criterion for an IFR check ride performance, possible pilot biases could compromise the reliability of the subjective measures.

Physiological measures (see Chapter 3.2.3) do not have the limitations mentioned above. Many physiological measures, however, are not suitable for a check ride use, mainly because they generate unacceptable pilot intrusion, lack pilot acceptance and disturb simulator and aircraft instruments. HR and HRV measures, although somewhat less sophisticated than some of the more recently developed physiological measures, have been widely employed in real and simulated aircraft environments, enjoy high face validity among the pilot population and generate little, if any, pilot intrusion (Dussault et al. 2004; Lee and Liu 2003; Hankins and Wilson 1998; Ylönen et al. 1997). For these reasons, this study used ECG based measures to measure task demand induced activation of ANS. From an ECG, the NN interval of the heart rhythm was identified. HR and HRV were derived from the NN interval and used as measures of PMWL. Before this study, HR and HRV have not been measured during a real F/A-18 IFR check ride.

Different components of HRV have been used as measures of ANS modulation. HR, although often associated with reactions to variations in the physical task demands, has also been associated with the changes in the piloting task's mental demands. Table 5 summarises the products of the NN interval used in this study. Also, Table 5 describes how HR and the components of HRV are affected by the increased PMWL.

Measure	Unit	Description	Expected change	References
MEANHR	[1/min]	The mean heart rate.	Increase	Vuksanović and Gal 2007; Wilson 2002; Roscoe 1993, 1975
MEANRR	[ms]	The mean of NN intervals.	Decrease	Sun et al. 2012, Terkelsen et al. 2005
SDNN	[ms]	The standard deviation of NN intervals.	Decrease	Tran et al. 2010; Terkelsen et al. 2005
RMSSD	[ms]	The square root of the mean squared differences between successive NN	Increase	Li et al. 2009; Orsila et al. 2008
NN50	[count]	The number of successive NN interval pairs that differ more than 50ms.	Decrease	Deepak et al. 2014
pNN50	[%]	The NN50 divided by the total number of NN intervals.	Decrease	Taelman et al. 2011
HRVTRI	[-]	The integral of the NN interval density distribution divided by the maximum of the distribution.	Decrease	Cinaz et al. 2013
LFnu	[-]	Normalized low frequency (0.04 - 0.15Hz) component of HRV.	Increase	Wu et al. 2011; Miyake et al. 2009
HFnu	[-]	Normalized high frequency (0.15 - 0.4 Hz) component of HRV.	Decrease	Wilson 2002
LF/HF	[-]	The ratio between the power of low frequency (LF) and high frequency (HF) components of HRV.	Increase	Skibniewski et al. 2015

Table 5. HR and HRV components and their expected change due to increased PMWL.

Several studies have shown HRV and HR to be relatively insensitive to changes in task demand, with HRV and HR being able to differentiate the task demand variations only between the task and rest conditions (Fallahi 2016; Wei et al. 2014; Veltman and Gaillard 1996; Jorna 1992; Wilson 1992). In a more recent study, Mansikka et al. (2016a) successfully used HR and HRV to identify different levels of task demands during simulated fighter missions when the task demand was intentionally and somewhat artificially varied from very modest to extremely high; the temporal demand of the repeated flying task varied from 6 min and 35 s to 2 min 20 s. In this study, the fighter pilots' performance and PMWL were measured during a real instrument check ride without artificial manipulation of task demand. The instrument check ride was carried out in a high fidelity simulator and comprised of clearly identifiable mission segments. Each mission segment consisted of a different piloting task and thus generated mission segment specific task demands. The pilots' PMWL measured with HR/HRV and performance variations between different mission segments was studied.

The aim of this study was to answer to the research questions three and four (see Chapter 1, Figure 1). It was hypothesised that HR and the HRV components presented in Table 5 could differentiate the task demand differences between the check ride's mission segments. Also, it was theorised that the PMWL measures could identify differences between the mission segments even when there were no significant performance differences between them. That is, even when the pilots could maintain their performance unchanged from mission segment to mission segment, there would be significant differences in their ANS responses to the changing task demands. Such a finding would support the use of both performance and PMWL measures in future check rides; the differences in the values of the PMWL measures could provide valuable insights about the PMWL's relation to performance and about the differences in the pilots' cognitive spare capacities during events of varying task demands. Ultimately, the level of PMWL could at some later stage be used as an additional IFR check ride criterion where the pilot would have to achieve a minimum performance score without exceeding the given level of PMWL (see page 41, Figure 13). This study was aimed at evaluating if HR and HRV have potential as such measures of PMWL.

6.2 Method

6.2.1 Participants

Data from 26 volunteer FinAF male F/A-18 pilots with a 1st class IFR qualification were collected. The pilots' average flying experience with the F/A-18 was 781 hours (SD=390). Relevant data concerning the pilots' activities for the 12 hours before the check ride were recorded (see Appendix 2). All pilots had passed an extensive aeromedical examination within the last 12 months and were fit to fly at the time of the study. A written, informed consent (see Appendix 1) was obtained from each subject. The study was reviewed and approved through the Coventry University's Ethical Review Process (see page iv).

6.2.2 Study Design

The data collection was undertaken during official F/A-18 1st class IFR check rides. A Boeing built WTSAT was used for the piloting task (see page 49, Figure 14). The WTSAT is used at the FinAF's fighter squadrons for basic and advanced F/A-18 pilot training. The WTSAT is a non-motion, high fidelity flying simulator, with a 135 degree field of view and a fully functional cockpit. The WTSAT replicates the F/A-18 flying characteristics with such a high accuracy that the FinAF F/A-18 pilots can use it to fly their annual instrument check rides. Each pilot's check ride was briefed, controlled, scored and debriefed by a qualified F/A-18 examiner pilot. A single examiner pilot was responsible for the check rides' scoring. The subjects' official IFR ratings were based on their performance score during the mission. It was therefore assumed that the subjects invested a high degree of mental effort on the task.

The mission comprised of seven recognisable segments: 'Takeoff and Ingress', 'Manoeuvering, 'Level Turns', 'Single Engine Manoeuvering (SEM)', 'VOR (VHF Omni Directional Radio Range) Approach', ILS Approach' and 'PAR (Precision Approach Radar) Approach'. The different segments were linked together to form a complete, logical flying mission. The ‘Takeoff and Ingress’ segment consisted of final checks before the takeoff, IFR takeoff and initial climb, turning climb as well as leveling at the designated altitude, speed and heading. The ‘Manoeuvering’ segment included basic aerobatic manoeuvres, recoveries from unusual attitudes and basic fighter manoeuvres. The ‘Level Turns’ segment contained a serial of steep turns with constant bank angle, altitude, load factor and airspeed. The ‘SEM’ segment included single engine emergency procedures and a simulated single-engine approach followed by a single engine go-around. The approach segments comprised of standard approaches with identifiable phases of initial approach, intermediate approach and final approach. The ‘VOR approach’ and the ‘ILS approach’ segments included also the missed approach phase. It was expected that the segments including instrument approaches would have the highest task loading as the pilots were not allowed to use any autopilot functions while the required control accuracy greatly increased as the pilots descended towards their approach specific minimum altitudes. On the other hand, the ‘Manoeuvering’ segment was expected to have the lowest task loading as this segment was closest to a ‘free flight’ condition where the pilots had numerous control input options, each providing an acceptable control accuracy.

The whole mission was flown in IMC. The cloud base was adjusted below the 1st class DH for the ILS approach and below the 1st class minimum descent altitude for the VOR approach, thus forcing the pilots to commence go-arounds after reaching their approach specific descent minimum. For the PAR approach, the cloud base was set at DH (60 m/200 ft) thus allowing a full stop landing. A moderate, variable and gusty wind was set for the mission. A typical IFR check ride lasted just over an hour from an engine start to the final landing. As the study utilised an existing proficiency test and the mechanisation of the HR/HRV data collection instrumentation had been tested during the previous study, there was no need to conduct a separate pilot study.

6.2.3 Procedure

Each mission segment was scored by the examiner pilot. For the purposes of analysis, the performance scores were retrieved and calculated as percentages of the maximum scores. For a pilot to achieve a 1st class IFR rating, s/he has to score at least 60% of the maximum score in each segment. Both the control accuracy and the smoothness of the aircraft control were assessed. To minimise the effects of the inter-rater variability, only the control accuracy scores were used for the analysis conducted in this study. The scoring of the ILS approach was based on deviations from the target airspeed, GS and LLZ. The VOR approaches were scored based on deviations from the target speed, step down fixes and the final approach course. A mission playback was used to increase the scoring accuracy of the approaches; the playback was stopped at every 0.5 NM (0.9 km) during the approaches. While stopped, the deviations were recorded and scored (see Appendix 4). The scoring of the PAR approach was not used for the analysis as different malfunctions were activated during the PAR approaches making them inconsistent between the pilots. To achieve a 100% ILS performance score, the maximum control error at 5 NM (9.3 km) was 60 ft (18.3 m) for the glideslope, 300 ft (91.4 m) for the localizer and 5 kts (9.3 km/h) for the airspeed. As the control accuracy requirement is increased towards the approach minima, the maximum allowable control error at DH was 10 ft (3.0 m) for the GS, 20 ft (6.1 m) for the LLZ and 5 kts (9.3 km/h) for the airspeed. The scoring of the VOR segment was similar to that of the ILS segment. As the VOR approach is a non-precision approach, its precision requirement was not as tight as it was for the ILS segment. The scoring of the other segments was based on variations of target flight parameters defined for different manoeuvres and reflected the control accuracy requirements of the ILS approach.

The ECG recording, manipulation and interpretation were done in accordance with the guidance in Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology (Camm et al. 1996). Before the mission, the subjects were equipped with Mind Media Nexus-10 MKII system for the ECG recording. Three electrodes were placed below the left (negative) and right (ground) clavicle and the left costal cartilage (positive), respectively (see page 55, Figure 16). ECG data were collected continuously during the whole mission. Five minute ECG samples were retrieved from each mission segment for further analysis. Data were first recorded using Biotrace+ software (version V2012C) from where the samples were exported to Kubios HRV 2.2 software for further analysis and NN interval artefact removal. A sampling rate of 1024 Hz was used for all samples and a 256 second window width with a 50% overlap was used for the fast Fourier transformation. Piecewise cubic spline interpolation was used to support artefact corrections; on preliminary inspection, all inter beat intervals 0.35 seconds longer or shorter than the local average, at HR of 60 beats per minute, were considered as artefacts. However, the artefacts were ultimately carefully edited using beat to beat visual checks and manual corrections (Tarvainen et al. 2014; Camm et al. 1996). Noisy data were excluded from the analysis. The values of HR and the following components of HRV were analysed from each subject: MEANRR, SDNN, RMSSD, NN50, pNN50, HRVTRI, LFnu, HFnu and LF/HF (see Table 5). Chapter 3.3.2 includes additional discussion about HR/HRV analysis and Kubios HRV 2.2 software.

6.3 Results

Each pilots' performance data from every mission segment were retrieved. In a similar fashion, the values of the pilots' HR and HRV components were collected from each mission segment. In general, ECG data were uncluttered with very few artefacts. However, ECG data from one subject were lost due to a software error. In addition, ECG data from one subject were corrupted and thus excluded from the analysis. As a result, the findings of this study were based on data from 24 subjects.

Data were analysed using IBM SPSS software (version 22). Normality of the distributions of the performance scores as well as the HR and the HRV components' values in each mission segment were verified using the Shapiro-Wilk test. The performance scores and the HR/HRV components' values were first analysed with the repeated measures ANOVA. Only after the ANOVA results proved to be significant, the results were further analysed using paired t-test for the subsequent pairwise comparisons.

The pilots were able to maintain high performance levels across all the mission segments; the 'SEM' segment had the highest mean performance score of 97.3% (SD=4.0) whereas the 'Manoeuvering' had the lowest mean performance score of 89.8% (SD=5.5). Table 6 presents the descriptive statistics for the performance scores.

Mission Segment	Performance Scores (% from maximum)			
	M	SD	Max	Min
Takeoff and Ingress	96.3	3.1	100.0	90.0
Manoeuvering	89.8	5.5	98.3	78.3
Level Turns	95.0	9.6	100.0	58.5
SEM	97.3	4.0	100.0	83.3
VOR Approach	94.4	6.3	100.0	73.6
ILS Approach	93.8	4.3	100.0	86.0

Table 6. Means (M), standard deviations (SD), maxima (Max) and minima (Min) of the mission segments' performance scores (N=24). SEM= Single Engine Manoeuvering, VOR= VHF Omni Directional Radio Range, ILS= Instrument Landing System.

The repeated measures ANOVA revealed significant differences in the performance scores between the mission segments; $F(5,115)=4.9$, $p<0.05$, partial $\eta^2=0.176$. In the pairwise comparisons, seven mission segment pairs had significant performance differences between them. The results of the pairwise comparisons are summarised in Table 7.

Mission Segment Pairs		M	SE	t	
Takeoff and Ingress	Manoeuvering	6.4	1.2	5.248	***
	Level Turns	1.3	2.2	0.577	
	SEM	-1.0	1.0	-0.957	
	VOR Approach	1.9	1.4	1.376	
	ILS Approach	2.5	1.0	2.427	*
Manoeuvering	Level Turns	-5.2	2.1	-2.419	*
	SEM	-7.4	1.3	-5.657	***
	VOR Approach	-4.5	1.5	-3.082	**
	ILS Approach	-4.0	1.3	-3.150	**
Level Turns	SEM	-2.2	2.2	-0.988	
	VOR Approach	0.7	2.4	0.273	
	ILS Approach	1.2	1.9	0.629	
SEM	VOR Approach	2.9	1.5	1.906	
	ILS Approach	3.4	1.1	3.085	**
VOR Approach	ILS Approach	0.6	1.5	0.389	

Table 7. Pairwise means (M) and standard errors (SE) of the performance scores as well as the corresponding test statistics (t) in the pairwise comparisons between the mission segments; ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$ (N=24).

While the pilots' performance remained relatively stable between the different mission segments, there were changes in HR and in the components of HRV. The descriptive statistics of the HR values and the HRV components' values for different mission segments are presented in Table 8.

		Takeoff and Ingress		Manoeuvering		Level Turns		SEM		VOR Approach		ILS Approach	
		M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
MEANRR	[ms]	677.6	115.3	676.0	107.3	689.4	126.8	686.7	114.0	661.8	103.6	666.2	105.1
MEANHR	[1/min]	91.7	14.1	91.7	13.2	90.2	15.4	90.2	13.7	93.5	13.5	92.7	13.4
SDNN	[ms]	64.7	27.0	67.8	33.3	53.4	19.6	61.2	22.5	64.4	25.3	58.0	20.5
RMSSD	[ms]	27.4	11.3	26.8	16.1	25.0	11.2	27.6	13.6	25.8	10.1	23.9	9.6
NN50	[count]	29.5	26.6	27.8	29.4	24.5	26.4	28.7	32.0	24.4	23.0	22.5	22.2
pNN50	[%]	7.2	7.1	7.0	8.2	6.4	7.6	7.3	8.8	5.7	6.0	5.4	5.9
HRVTRI	[-]	14.2	4.7	15.6	5.4	11.7	3.8	13.8	5.0	12.9	4.3	12.9	4.8
LFnu	[-]	77.0	10.0	78.3	12.7	79.8	11.1	80.2	9.3	83.0	9.2	82.0	9.4
HFnu	[-]	22.9	10.0	21.6	12.7	20.1	10.9	19.7	9.3	16.9	9.2	18.0	9.3
LF/HF	[-]	4.2	2.3	4.9	2.8	5.5	3.7	5.4	3.3	6.6	3.9	6.0	3.2

Table 8. Means (M) and standard deviations (SD) of the HR values and the HRV components' values for the mission segments (N=24).

Sphercity was assumed only for some HR/HRV measures. As a result, the degrees of freedom in ANOVAs vary between different HR/HRV measures. The repeated measures ANOVA revealed significant differences across the mission segments for: MEANRR $F(5,115)=3.15$, $p<0.05$, partial $\eta^2=0.120$; MEANHR $F(5,115)=2.78$, $p<0.05$, partial $\eta^2=0.108$; SDNN $F(3,71)=3.51$, $p<0.05$, partial $\eta^2=0.132$; HRVTRI $F(5,115)=7.79$, $p<0.05$, partial $\eta^2=0.253$; LF/HF$(5,115)=3.16$, $p<0.05$, partial $\eta^2=0.121$. ANOVA did not reveal significant differences for: LFnu $F(3,79)=2.33$, $p>0.05$, partial $\eta^2=0.092$; HFnu $F(3,79)=2.32$, $p>0.05$, partial $\eta^2=0.092$; RMSSD $F(3,74)=1.26$, $p>0.05$, partial $\eta^2=0.052$; NN50 $F(3,75)=1.65$, $p>0.05$, partial $\eta^2=0.067$; pNN50 $F(3,73)=2.06$, $p>0.05$, partial $\eta^2=0.082$. The measures with the significant ANOVA differences were further analysed with pairwise comparisons. These results are summarised in Table 9.

		MEANRR			MEANHR			SDNN			HRVTRI			LF/HF		
		M	SE	t	M	SE	t	M	SE	t	M	SE	t	M	SE	t
Takeoff and Ingress	Manoeuvering	1.6	6.4	0.256	0.0	0.8	-0.056	-3.1	3.2	-0.974	-1.4	0.5	-2.655 *	-0.7	0.6	-1.198
	Level Turns	-11.8	7.7	-1.528	1.4	1.1	1.279	11.4	2.7	4.225 ***	2.5	0.5	4.572 ***	-1.3	0.7	-1.954
	SEM	-9.1	7.5	-1.209	1.4	1.0	1.425	3.6	3.2	1.116	0.3	0.7	0.460	-1.2	0.6	-2.066
	VOR Approach	15.8	8.6	1.831	-1.9	1.2	-1.641	0.3	4.7	0.072	1.3	0.5	2.319 *	-2.4	0.8	-3.062 **
	ILS Approach	11.4	9.8	1.162	-1.1	1.2	-0.905	6.7	3.5	1.933	1.3	0.6	2.148 *	-1.7	0.6	-2.751 *
Manoeuvering	Level Turns	-13.4	8.4	-1.587	1.5	1.2	1.224	14.4	4.4	3.316 **	3.9	0.8	5.226 ***	-0.6	0.6	-0.971
	SEM	-10.7	7.9	-1.357	1.5	1.1	1.343	6.6	4.2	1.600	1.7	0.7	2.351 *	-0.5	0.7	-0.796
	VOR Approach	14.2	9.2	1.539	-1.9	1.3	-1.456	3.4	5.3	0.648	2.7	0.8	3.404 **	-1.7	0.8	-2.195 *
	ILS Approach	9.8	9.8	0.996	-1.1	1.3	-0.791	9.8	5.3	1.847	2.7	0.8	3.403 **	-1.0	0.6	-1.818
Level Turns	SEM	2.7	6.9	0.386	0.0	0.9	-0.150	-7.8	2.3	-3.325 **	-2.2	0.8	-2.644 *	0.1	0.6	0.134
	VOR Approach	27.6	10.9	2.534 *	-3.3	1.3	-2.538 *	-11.0	4.8	-2.320 *	-1.2	0.7	-1.848	-1.1	0.9	-1.311
	ILS Approach	23.2	10.9	2.126 *	-2.5	1.3	-1.904	-4.6	2.8	-1.637	-1.2	0.7	-1.875	-0.4	0.7	-0.638
SEM	VOR Approach	24.9	9.3	2.684 *	-3.3	1.1	-2.972 **	-3.2	4.0	-0.818	1.0	0.6	1.521	-1.2	0.6	-2.004
	ILS Approach	20.5	8.3	2.464 *	-2.5	1.0	-2.450 *	3.1	3.0	1.042	0.9	0.8	1.139	-0.5	0.7	-0.784
VOR	ILS Approach	-4.4	7.0	-0.636	0.8	0.9	0.905	6.4	4.4	1.443	0.0	0.5	-0.042	0.7	0.6	1.225

Table 9. Pairwise means (M) and standard errors (SE) of HR and the HRV components as well as the corresponding test statistics (t) in the pairwise comparisons between the mission segments; ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$ (N=24).

All mission segments that were differentiated by the pilots' performance scores were also differentiable by their HR/HRV responses. The 'Takeoff and Ingress' and 'SEM' mission segment pair and the "ILS Approach" and "VOR Approach" segment pair were neither differentiated by the performance scores nor by the HR/HRV responses. All other mission segments were differentiated by some of the HR/HRV measures. HR and the HRV components were able to differentiate six mission segment pairs that had non-significant performance differences. The mission segment pairs with the significant performance score differences and/or with the significant differences in the HR values or in the HRV components' values are summarised in Table 10.

	Takeoff and Ingress	Manoeuvering	Level Turns	SEM	VOR Approach
Manoeuvering	HRVTRI				
Level Turns	SDNN, HRVTRI	SDNN, HRVTRI			
SEM	-	HRVTRI	SDNN, HRVTRI		
VOR Approach	HRVTRI, LF/HF	HRVTRI, LF/HF	MEANRR, MEANHR, SDNN	MEANRR, MEANHR	
ILS Approach	HRVTRI, LF/HF	HRVTRI	MEANRR	MEANRR, MEANHR	-

Table 10. The mission segment pairs with the significant differences of the performance score and/or with the significant differences of HR and the HRV components (N=24). The mission segments pairs with the significant performance score differences are denoted by a shaded background.

6.4 Discussion

This study successfully utilised HR and HRV, as measures of PMWL, during real F/A-18 IFR check rides and was able to replicate the findings of the earlier mental workload related HR and HRV studies (Deepak et al. 2014; Cinaz et al. 2013; Sun et al. 2012; Taelman et al. 2011; Tran et al. 2010; Li et al. 2009; Orsila et al. 2008; Vuksanović and Gal 2007; Terkelsen et al. 2005; Svensson and Wilson 2002; Wilson 2002; Roscoe 1993, 1975). Unlike some earlier studies using HR and HRV (Fallahi 2016; Wei et al. 2014; Veltman and Gaillard 1996; Wilson 1992; Jorna 1992), this study was able to differentiate HR/HRV variations between mission segments instead of differentiating just the rest and trial conditions. In addition, whereas Mansikka et al. (2016a) successfully used HR and HRV to differentiate large task demand changes during a simulated flight (see Chapter 5), this study replicated these results with smaller task demand variations during a realistic, simulated flying task.

As discussed in Chapter 6.3, the experienced F/A-18 pilots were able to maintain high and mostly equal performance across the different segments of the IFR check ride (see Tables 6, 7 and 10). At the same time, the HR values and the HRV components' values indicated that their PMWL between the different mission segments were not equal (see Tables 8, 9 and 10). As summarised in Table 10, the 'Takeoff and Ingress' and 'SEM' mission segment pair and the "ILS Approach" and "VOR Approach" segment pair were neither differentiated by the performance scores nor by the HR/HRV responses. Out of the total of 15 mission segment pairs analysed, there were eight mission segment pairs which were not differentiated by the performance scores. However, of these eight mission segment pairs six were differentiated by one or more of the HR/HRV measures (see Table 10).

The pilot's performance in an instrument check ride affects his/her qualification to operate in adverse weather conditions. In addition, a failure to pass a 1st class instrument check ride can have a negative impact on a pilot's career progression – at least in the case of repeated failures. While the level of effort was not measured, it was assumed that the pilots were willing to invest a lot of effort to the flying task. However, especially in the case of the more experienced pilots, they were well aware of the required performance standards and their own, real-time performance in relation to those standards. As a result, it is reasonable to expect that some of those more experienced pilots did not invest more effort than they considered necessary.

From the HR/HRV data collection and analysis point of view, the ECG data of this study had far less artefacts than the data retrieved during the first study (see Chapter 5). The same instrumentation was used for the data collection and analysis in both studies. Also, the electrode placement was the same in both studies. It is likely that the almost total absence of artefacts in the second study resulted from the very minimal movement of the subjects' upper torsos; no inflight emergencies were introduced during the HR/HRV measuring phases of the mission and very few other activities requiring torso movement were present during the check ride. While the torso movement's impact on artefacts was not specifically studied, it can be expected that a similar electrode placement would not be practical on missions where the torso movement is excessive.

For the highly experienced pilots, an instrument check ride is a slow paced, routine mission with a few, if any, unexpected events. Therefore, an instrument check ride does not excessively challenge the pilots' perception or the central processing capacities. Nor does it challenge the pilots' ability to gain and maintain awareness of the mission requirements. If more complex, fast paced fighter missions are considered, it would probably be an oversimplification to expect that PMWL alone could explain the pilots' performance differences.

It is concluded that the differences in the pilots' PMWL between the check ride's mission segments can be differentiated by HR and HRV. HR and HRV were also capable of identifying the differences between the mission segments even when there were no significant performance differences between them. In other words, this study was able to answer the research questions three and four mentioned in Chapter 1 (see Figure 1). The utilisation of HR and HRV, as measures of PMWL, can improve the awareness of the pilots' potential to respond to high task demands and may support the assessment of the differences between their spare cognitive capacities. As illustrated in Figures 4, 12 and 13, an increase of PMWL will - at some point - degrade pilot performance, the evaluation of individual differences could reveal if some pilots are closer to the threshold of impaired performance than others with a similar performance. This can give valuable insights about the pilots' spare mental capacities during events of high task load, which in turn could be used to improve both the flight safety and the operational effectiveness.

The next study examines the fighter pilots' HRV and performance during a simulated mission consisting of instrument approaches and air combat. In addition, the study examines the pilots' awareness of the mission requirements; it was expected that the combination of low performance and low PMWL would be associated with the pilots' low awareness of the mission requirements.

7. FIGHTER PILOTS' MENTAL WORKLOAD AND PERFORMANCE: A COMPARISON OF SIMULATED INSTRUMENT APPROACHES AND AIR COMBAT

7.1 Introduction

While fighter aircraft have become easier to fly, their ever improving sensors, weapons and communications systems put increasing demands on pilots' cognitive processing capacity. As illustrated in Figures 4, 12 and 13, the task demands and cognitive stressors of an air combat have potential to degrade pilot performance to an unacceptable level (Ahmadi and Alireza 2007; Matthews et al. 2007; Noel et al. 2005; Paas and Van Merriënboer 1993). The performance degradation may, or may not, be associated with the high PMWL (Endsley 1996; Aasman et al. 1987). This study examined the relation of PMWL and pilots' performance during a simulated mission consisting of instrument approaches and air combat phases. In addition, the subjects' awareness of the mission requirements was studied to further explain the association between PMWL and performance. Doing so, this study attempted to answer the research questions five and six (see Chapter 1, Figure 1).

A flight simulator is a convenient environment to study the pilots' performance as the missions can be recorded, reviewed and paused for an accurate performance assessment. The performance assessment during the instrument approaches is a relatively straightforward task as the target parameters are well known and any deviations from them can be easily monitored (Mansikka et al. 2016a, 2016b). While air combat missions can be highly complex, their related tactics, techniques and procedures (TTPs) are typically highly standardised and provide the examiner or evaluator well defined target parameters of performance. Fighter pilots' performance is typically evaluated during proficiency checks or combat check rides. Different aspects of performance are tracked on different types of check rides, and different Air Forces and Squadrons have their own, and often classified, performance standards. As a result, no global fighter combat performance measures exist and performance comparisons between different check rides is difficult and also somewhat arbitrary. However, the performance measures provide a reliable discriminator of pilots' abilities in relation to their peers flying the same check ride.

Measuring PMWL during a combat check ride is more challenging as the measurement instrumentation has to be sensitive and reliable. In addition, the instrumentation may not disrupt the simulator's systems or distract the pilot's performance during the mission (Carmody 1994; O'Donnell et al. 1986). And finally, the instrumentation has to be approved by the aviation authorities. Psychophysiological measures are often the preferred method of measuring the pilots' ANS activation, or PMWL, in an aviation environment as they are capable of detecting the sudden changes during the flying mission (Carmody 1994). Also, should the level of PMWL be used as an additional pass-fail criterion on a check ride, the psychophysiological measures minimise the effects of possible pilot biases (Annett 2002a; Gopher and Donchin 1986). Among the variety of psychophysiological measures available, the HRV was selected as a measure of PMWL for this study. As discussed in Chapters 5 and 6, Mansikka et al. (2016a, 2016b) have successfully used MEANRR as a measure of PMWL during fighter pilots' instrument approaches and instrument proficiency check rides. Encouraged by the earlier results, MEANRR was selected as a measure of PMWL for this study as well. Before this study, the MEANRR's ability to capture fighter pilots' PMWL variations during a simulated combat check ride has not been reported.

If task demand is increased, the performance level can be maintained unchanged by investing more effort and cognitive resources on the task (Vicente et al. 1987). The increased allocation of the limited cognitive resources will eventually increase PMWL. While PMWL cannot be directly measured, the lowered MEANRR can be used as an indicator of an increased PMWL (Sun et al. 2012;

Terkelsen et al. 2005). Despite the pilots' willingness to invest more effort, the performance degradation will eventually occur if the task demands are increased beyond the point where there is no more mental capacity left to compensate it (Williges and Wierwille 1979). Evaluating PMWL during simulated air combat can improve the awareness of the pilots' potential to cope with the task demands higher than those experienced during the simulator mission itself (Camp et al. 2001). However, it would be an oversimplification to claim that impaired performance during a complex air combat would be automatically associated with an extremely high PMWL.

During a simulated air combat, the pilots are exposed to an overwhelming amount of data. From that data, the mission critical information has to be selected and integrated with past and real-time information to form a tactically sound understanding of the current situation and a prediction of future events (Endsley 1988). Depending on the pilots' awareness of the mission requirements they may, or may not, be capable of selecting the most appropriate tactical responses (Endsley 1995). For example, pilots unaware of their lowered awareness of the mission requirements may see no reason to invest more effort to improve his/her performance (Endsley 1990). This can lead to a 'not knowing what is not known'-situation characterised by a low actual awareness of the mission requirements, high perceived awareness, low PMWL and low performance (Matthews et al. 2011). Therefore, in order to gain a thorough understanding of the pilots' cognitive potential and performance on an air combat mission, the task performance and PMWL measures must be supplemented by a measure of the pilots' awareness of the mission requirements.

The aim of this study was to investigate if PMWL was associated with the changes in pilots' performance during the instrument approaches and air combat phases of the simulated mission (see research question five, Chapter 1, Figure 1). In addition, the subject's awareness of the mission requirements during the air combat phases was measured and used to further explain the association between PMWL and performance (see research question six, Chapter 1, Figure 1). When data from the pilots' awareness of the mission requirements are collected during a check ride- type setting, several requirements have to be satisfied. First, the data must be collected without pausing the mission or otherwise disturbing the mission execution. Second, the data has to be collected in a fashion that is not dependent on the pilots' ability to recall detailed events from the mission. Third, in fighter squadrons, the time available for the data collection is often very limited; the data collection has to be done very quickly, preferably in real-time. In this study, the data of the pilots' awareness of the mission requirements was collected by a weapons instructor, in real-time and without pausing the mission or disturbing the pilot. The pilots' awareness of the mission requirements was indirectly measured by monitoring their behaviour, i.e., the pilots' ability to receive and acknowledge the predefined tactical orders, and their overt actions to execute them. In summary, this study combined measures of PMWL, mission requirement awareness and performance during a simulated flying mission consisting instrument approaches and air combat. The intent was to examine if PMWL and performance were associated in a complex flying scenario and if the awareness of the mission requirements could explain some of the ANS responses.

7.2 Method

7.2.1 Participants

Thirty seven combat ready FinAF F/A-18 pilots volunteered. The subjects' background varied from a wingman to a weapons instructor. As a result, the subjects' mean experience with the F/A-18 was 686 flight hours (SD=329). Each pilot had passed an extensive aeronautical medical examination within the last 12 months and they were fit to fly at the time of the study. A written, informed consent was collected from each subject (see Appendix 1). A structured proforma was used to collect subjects' background data and information concerning their relevant activities for the 12 hours prior to participating (see Appendix 2). Data from five subjects were lost due to corrupted ECG samples and simulator malfunctions. As a result, the analysis was based on data from 32 subjects. The study was reviewed and approved through the Coventry University's Ethical Review Process (see page iv).

7.2.2 Study Design

WTSAT, a high fidelity F/A-18 simulator, was used for the flying mission (see page 49, Figure 14). Data collection was conducted during the OT&E of F/A-18 combat check ride. The content of the mission, like the content of most tactical check rides, was classified. The flying mission was built to capture the essential elements of a defensive counter-air mission. For the purposes of this study, data was retrieved from two ILS approaches and from two beyond-visual-range (BVR) attacks.

The first ILS approach (ILS-A) was flown to a primary landing field immediately after the mission's last combat phase and high speed egress. Both the egress and the ILS-A were flown in a two-ship radar trail with the subject flying the trailing aircraft. The lead aircraft reduced to approach speed just before intercepting the LLZ and maintained the approach speed throughout the ILS-A approach. The high speed egress, rapid deceleration before intercepting the LLZ and flying in the radar-trail generated additional task demands to subjects. While these additional tasks were simple, they had potential to compromise the ILS performance if the pilot's attention was not focused properly or the task management was poorly executed. The ILS-A performance scores reflect the consequences of mismanaged tasks and distracted attention. The second ILS approach (ILS-B) was flown to an alternative airfield after a diversion. The ILS-B was flown as a single-ship and subjects were free to choose their airspeed during the initial descent and the range at which they started to reduce speed to the approach speed. Both ILS approaches had almost identical, standard approach profiles.

Both BVR attacks (Commit-A and Commit-B) were flown in a four-ship formation with the subjects flying as a #4, i.e., as the wingman of the second element. While the Commit-A and Commit-B were similar type BVR attacks, they had different threat presentations and thus required different actions from the friendly four-ship. The mission was prepared to follow a predefined script and all objects, except the #4, were constructive simulation entities. The flight profiles of all simulated entities were preprogrammed; the enemy air followed predefined threat presentations whereas the simulated friendly aircraft manoeuvered based on the FinAF TTPs. The radio transmissions were prepared as an audio file which was time-synchronised with the flight profiles and activities of the simulated entities. The simulated radio communications included transmissions from three different air traffic controllers (tower and two radar controllers), two different fighter controllers and the transmissions of the #1, #2 and #3 of the friendly four-ship. All simulated communications were recorded using the real tower, radar, fighter controller and aircraft radios. To increase the sense of authenticity, the radios were operated by real air traffic controllers, fighter controllers and F/A-18 pilots. In addition to the normal radio traffic, the audio file included radio jamming and radio noise. The subjects were given directive and informative calls on a radio and they were expected to reply to them. The subjects' missile shots had no effect on the threat aircraft which were, or were not, removed from the aerial picture based on the written script. The subjects were not aware of their missiles' probability of kill being set to zero. As a result of this test setting, the simulated entities manoeuvered similarly during each subjects' trial and the only variations to the mission complexity and the four-ship's performance resulted from the subjects' own actions.

Both BVR attacks and ILS approaches were flown in IMC with 0 m (0 ft) IMC visibility. For the ILS-A approach, the cloud base was set below the landing minima, hence forcing a go-around. For the ILS-B approach, the cloud base was set to the DH and the RVR was adjusted to 700 m (2 296 ft) to allow full stop landing.

The flight profile of each simulated entity was prepared manually. The mission was flown numerous times as #4 during the development and pre-testing of the mission design. The purpose of the testing was to verify that the parameters of the simulated entities were realistic, that there was a reasonable chance for the #4 to stay with flow of the friendly four-ship and that the audio file was properly synchronised with the simulation. The final mission design was evaluated by the OT&E test pilots.

7.2.3 Procedure

Before the trial the subjects had 15 minutes to study the classified mission briefing material. The briefing material was given as a hard copy. No clarifications were given and no questions were answered during the briefing. After 15 minutes the simulation was activated.

The performance scores of the Commit-A and Commit-B were based on the subjects' reactions and the mission outcome. As long as the TTPs were followed and the mission's outcome met the mission objectives, the subjects were free to use the manoeuvres they considered most appropriate and they had the liberty to use the sensors and weapons as they considered proper. Therefore, only the kind of tactical items that were clearly observable and that could be unambiguously defined as correct or incorrect were scored. Tactical reactions were scored based on whether they were conducted correctly and safely. Unsafe, incorrect or missed tactical reactions were given a score of '0' whereas each safe and correct response was given a score of '1'. The subjects' performance in each Commit was formed by summing up the respective performance scores. The subjects' performance scores in the Commit-A and Commit-B were communicated as percentage values of each Commit's maximum performance score. The different tactical reactions were not weighted by their importance.

The ILS rating was based on deviations from the target speed along with the LLZ and GS errors. Each error component was scored independently and their values ranged from 5 (best performance) to 0 (worst performance). The deviations from the target values were recorded and scored every 0.5 NM (0.9 km) between the glide slope interception range and DH. Each subject's ILS-A and ILS-B performance scores were generated by calculating the mean of the performance scores from each data collection point (see Appendix 4). Both ILS approaches were scored using an official FinAF instrument check ride rating scale. The ILS performance scores were communicated as percentage values of each approach's maximum performance score. Arcsine transformation was not used, as it was likely that the shape of the distribution of the data in the sample reflected the shape of the distribution of data in the population.

The subjects' ability to execute targeting during the Commit-A and Commit-B was evaluated and used as a behavioural measure of subjects' awareness of the mission requirements (Salmon et al. 2009; Gugerty 1997; Endsley 1996; Vidulich et al. 1994; Hansman et al. 1993). The targeting data was not used as part of the Commits' performance criteria. Targeting commands were given as a radio calls. Using the targeting command's information, the subject's duty was to locate the specified target from the air picture and allocate onboard sensors and weapons against it, and to eventually commit against the specified target. Properly executed targeting required complex mental processing as the radio frequency used for the targeting communication was often cluttered and required well focused attention. Processing of the targeting command generated high demands on recipient's WM. The command was typically not repeated, and once given, the radio frequency was filled with other tactical communication making the auditory rehearsal process (see Chapter 2) difficult or impossible. The target's location was typically given as a bearing, range and altitude from a common reference point. Coupling the target's location to a constantly changing aerial picture visible in the cockpit displays required further mental processing. As a result, the properly executed targeting was treated as a behavioural measure which reflected the pilots' awareness of the mission requirements (Vidulich et al. 1994). The targeting success was given a score of either 1 (successful) or 0 (unsuccessful). The targeting success was communicated as a percentage value of the maximum targeting, or behavioural score. The subjects' performance and targeting success were scored by a qualified F/A-18 weapons instructor.

MEANRR was retrieved from both ILS approaches and from both Commits. In addition, a five minute MEANRR sample was recorded after the check ride and used as a MEANRR rest baseline. During the rest baseline recording the subjects sat still in the simulator.

As the purpose of the combat check ride was to differentiate the performance of members of a highly skilled subject population, the mission was designed to be highly intensive and demanding. In order to complete the BVR attacks successfully, the subjects had to manoeuvre according to the TTPs while complying with the platform and safety limitations. In addition, the subjects had to select and perceive the relevant information provided by the audio file and the simulator's sensors. From the information provided, the subjects had to generate a mental model of the aerial picture and to select their responses accordingly. A missed or misunderstood piece of relevant information, improper understanding of the relative positions and actions of the threat and friendly entities were each likely to generate sub-optimal pilot responses. The subjects' ability to comply with the rules of engagement and their adherence of the TTPs and the directives given by the air traffic controllers, ground controllers and element leads (#3 and #1 in the formation) was assessed. The safety of flight was assessed as a separate item. During the mission briefing, the importance of the subjects' mental effort during the trial was emphasised. All subjects reported their willingness to invest as much effort on the task as possible. Therefore, it was assumed that the results were not biased by the lack level of mental effort.

7.3 Results

The performance and the ECG data were analysed using IBM SPSS Statistics software (version 22). Normality of the distributions of the data was verified using the Shapiro-Wilk test. The data were further analysed with the paired t-test. The differences between the ILS approaches and commits were not analysed.

The subjects' mean performance on the ILS-A was 69.1% (SD=17.4) whereas the mean performance on the ILS-B was 82.1% (SD=8.3). The minimum performance on the ILS-A was 4.0%. The subjects' mean performance on the Commit-A was 57.0% (SD=10.4) and 44.8% (SD=8.0) on Commit-B. Table 11 presents the descriptive statistics of the performance scores. The pairwise comparisons indicated significant differences between the ILS-A and ILS-B (Mean difference = -13.0; Standard error 3.3; t= -3.886, $p<0.01$) and between the Commit-A and Commit-B (Mean difference = 12.3; Standard error= 2.2, t= 5.561, $p<0.001$).

	Min	Max	M	SD
Commit-A	40.0	75.0	57.0	10.4
Commi-B	32.0	68.0	44.8	8.0
ILS-A	4.0	87.1	69.1	17.4
ILS-B	62.4	93.9	82.1	8.3

Table 11. Means (M), standard deviations (SD), maxima (Max) and minima (Min) of the performance scores as percentages from the maximum score (N=32).

The mean of the subjects' MEANRR during the ILS-A was 644.1 ms (SD=108.0) while in the ILS-B it was 659.3 ms (SD=112.0). The Commit-A had a mean MEANRR of 649.8 ms (SD=110.7) and Commit-B 680.2 ms (SD=118.5), respectively. The mean rest baseline MEANRR was 768.0 ms (SD=130.1). Table 12 presents the descriptive statistics of the MEANRR values. The pairwise comparisons revealed significant MEANRR differences both between the ILS-A and ILS-B approaches (Mean difference=-15.2; Standard error=7.2; t=-2.099; p<0.05) and between the Commit-A and Commit-B (Mean difference=-30.4; Standard error=4.1, t=-7.356, p<0.001).

	Min	Max	M	SD
Commit-A	435.2	830.3	649.8	110.7
Commi-B	460.2	864.2	680.2	118.5
ILS-A	446.4	787.1	644.1	108.0
ILS-B	448.2	848.1	659.3	112.0
Rest	514.3	1004.6	768.0	130.1

Table 12. Means (M), standard deviations (SD), maxima (Max) and minima (Min) of MEANRR values (ms), (N=32).

The mean of the subjects' targeting, or behavioural, scores in the Commit-A was 43.8% (SD=50.4, Min=0%, Max=100%) while in Commit-B it was 12.5% (SD=33.6, Min=0%, Max=100%). A pairwise comparison revealed significant difference in the behavioural scores between the Commit-A and Commit-B (Mean difference=31.3; Standard error=11.4; t=2.743; p<0.05). The performance scores and MEANRR values are summarised in Figure 18. The shaded boxes in Figure 18 indicate the Commit with a low awareness of the mission requirements.

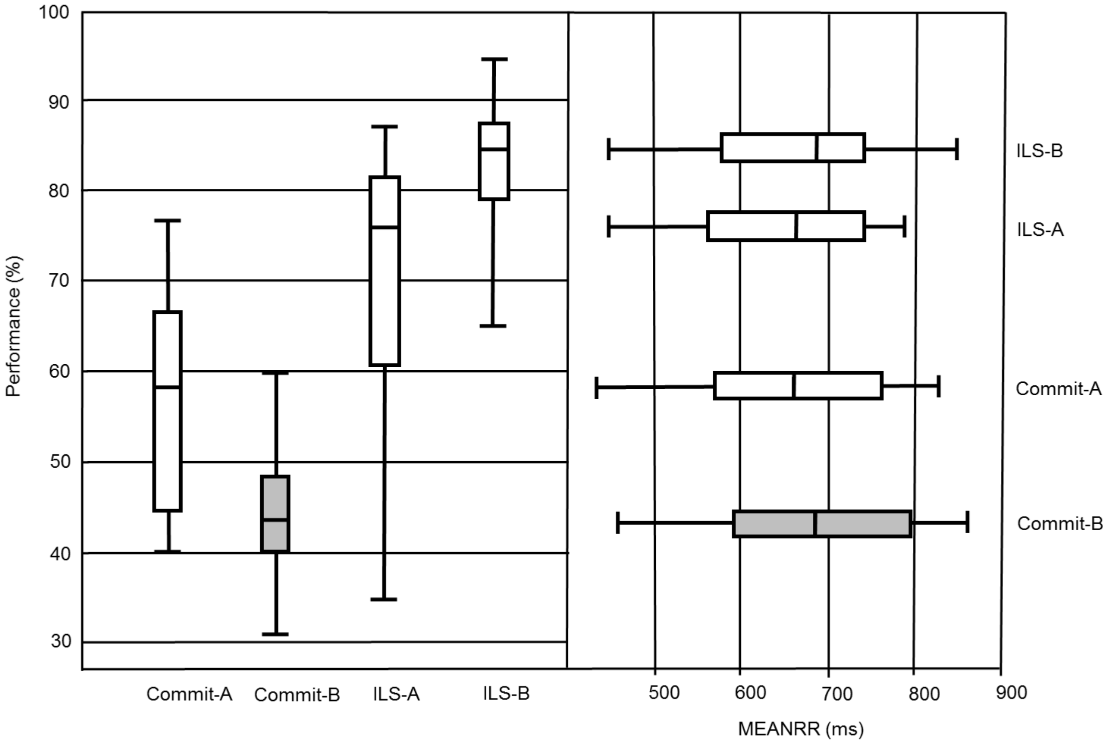


Figure 18. The performance (%) and MEANRR (ms) for Commit-A, Commit-B, ILS-A and ILS-B. The shaded boxes indicate the Commit with a low awareness of mission requirements

A Pearson product-moment correlation was run to determine the relationship between MEANRR and performance. A significant negative correlation was found between ILS-B performance and ILS-B MEANRR ($r=-0.412$, $n=32$, $p<0.05$). There were no correlations between ILS-A ($r=-0.004$, $n=32$, $p>0.05$), Commit-A ($r=0.311$, $n=32$, $p>0.05$) or Commit-B ($r=-0.181$, $n=32$, $p>0.05$) and their respective MEANRR values.

.

7.4 Discussion

This study examined the pilots' performance and PMWL during simulated ILS approaches and air combat and answered the research questions five and six (see Chapter 1, Figure 1). The performance scores were analysed against the pilots' ANS responses and their awareness of the mission requirements.

When HRV was studied during the BVR commits, the results from the earlier studies (Chapters 5.3 and 6.3) gave reason to expect that the lower performance scores would be associated with a lower HRV (see Figure 17, Table 6 and Table 8). The subjects were able to get significantly higher performance scores in the ILS-B compared to those in the ILS-A (see Chapter 7.3). As also the MEANRR values in the ILS-A were significantly lower than in the ILS-B, it was concluded that the ILS-A had significantly higher task demand than the ILS-B. MEANRR values and the performance scores during the ILS approaches followed the expected pattern and replicated the findings discussed in Chapters 5 and 6; an increased task demand was reflected in increased PMWL as indicated by the lowered values of the MEANRR (see Figure 17, Table 6 and Table 8). When the subjects' performance scores in the Commit-A and Commit-B were compared, the Commit-A had a significantly higher performance score average than the Commit-B (see Chapter 7.3). The MEANRR values during the Commit-B were significantly higher than during the Commit-A (see Chapter 7.3 and Figure 18).

By evaluating the performance and MEANRR alone, the results of the third study would had made little sense. But as discussed in Chapter 2, only after the level of effort is increased, will the mental resources be expended. For the pilots to increase their effort, they need to be aware of what is required for successful task accomplishment. More specifically, pilots need to be aware of the mission requirements - or situation awareness (SA). SA has an open ended nature; there are no such things as "zero SA" or "full SA". SA is also nominal in nature; it cannot be said that a SA of "4" would be twice as good as a SA of "2". Depending on the pilots' duties or role, different levels of SA may be considered acceptable (Endsley and Bolstad 1994). An acceptable level of SA can be considered as an awareness that an operator needs to possess in order to perform his/her duties safely and effectively. For example, if the subjects had been flying as leaders of the four aircraft formation (instead of flying as wingmen), their required SA would had been completely different. When SA is measured during an air combat engagement, it should be carefully considered what level of SA is considered acceptable and what kind of (behavioural) indicators can be used to reflect it. When SA is measured during proficiency checks, the SA measures should be objective and non-intrusive. Unfortunately, these requirements greatly narrow down the available SA measuring techniques.

It is concluded that during the complex air combat tasks, the association between the performance and PMWL was related to the pilot's perception of the task demands, not actual, or objective, task demands. When the pilot's awareness of the mission requirements was low, a combination of low performance and low PMWL occurred.

8. FINAL DISCUSSION

Flying a fighter aircraft can be a challenging task. The possible variations of the expected tasks and the ever present possibility of additional, unexpected tasks generate a potential for the task demands to increase dramatically. Despite the careful selection process and extensive training programs, the pilots are still sometimes unable to cope with these demands and to maintain the required performance. An unfavorable mismatch between the task demands and the mental capacity to cope with them can generate a serious risk to the pilot performance and flight safety. In addition, long term exposure to mentally demanding environment is likely to generate both short and long term stress and fatigue. The mental health effects of the extended exposure to stressful work environment, although not the focus of this thesis, should not be taken lightly. If the health effects are set aside, the main concern regarding the excessive task demands in fighter aviation is the impact these adverse task conditions have on flight safety and mission success.

As described by Rasmussen and Jensen (1974) the mental information processing of pilots comprises of sequenced, simple and quick decisions. For the serialised decisions to be effective, an operator needs to have a sufficient WM capacity; s/he needs to be able to swiftly select the relevant information, to process it, to execute appropriate responses and to predict likely system responses (see Figure 2). Although this may seem an economically inefficient procedure, it allows an operator to plan his/her actions and partition the complex task into a series of simpler tasks. When the decision cycle is kept fast enough, the simultaneous resource demands on WM can be managed and high system performance can be maintained. In a modern fighter aircraft environment, the mental information processing is challenged by the vast amount of unfiltered information provided to the pilot. In order to maintain the required pace of the decision cycle, the pilot has to rapidly differentiate between the relevant and the irrelevant information. “Being behind the jet”, a common saying in fighter aviation, is used to describe the operator's inability to maintain the required speed of this cyclical mental processing. Should any of the phases of this process fail, the pilot's performance has a tendency to drop. Unfortunately, these kind of information processing failures are self-reinforcing: First, the mismanaged control inputs complicate perception and situation analysis. Second, improperly perceived or analysed flight data results in less than optimal control input decisions. Finally, a lowered rate of the cyclical process creates high demands for the WM – mainly due to a slower update rate of the memorised items and the increased complexity of the required analysis and response selection. When the task demands do not overstress the perception or response stages of the information processing (see Figure 3), it is likely that the pilots' performance decrements result from the limitations of the central processing capacity in general, and WM limitations in particular. The pilots are able to adjust to the increased task demands only to a certain extent. Once the pilots have no more excess mental capacity left, any increase in the task demand will gradually degrade the pilots' performance (see Figure 12) (Noel et al. 2005; Paas and Van Merriënboer 1993; Williges and Wierwille 1979). However, PMWL and the pilot performance are not always correlated (see Chapter 7.3). As a result, it is not possible to draw conclusions about one by measuring the other; pilots with overtly similar performances may be exposed to significantly different PMWLs. It can also be assumed that pilots with similar PMWL may have remarkably different performances.

Check rides, along with other safety mechanisms, are in place to ensure the pilots' ability to cope with the task demands of their flying duties (Mavin and Roth 2014). Pilot proficiency checks have for a long time been based on scripted missions which remain unchanged from one proficiency check to another. Standardised, scripted proficiency checks are practical when the test is purposed to evaluate the pilots' technical, 'stick and rudder' flying skills; the desired manoeuvres can be introduced in a controlled fashion and an experienced examiner can perform the real-time performance assessment without difficulty.

On the other hand, the problem with the scripted proficiency tests is that while they test the pilots' basic aircraft handling abilities, they are unlikely to reveal the pilots' competency deficiencies in such areas as buildup and maintenance of SA, decision making and problem solving. These higher level cognitive functions – or competencies – are the ones that are likely to differentiate the pilots' performance in a modern air combat environment (see Chapter 7).

As discussed in Chapters 1 and 3, the flight safety and mission success can be compromised if the pilots' mental capacity is exceeded during flight. The unexpected events during a live flying mission can exceed the task demands experienced during a simulator check ride. If a pilot is already at the upper limit of his/her cognitive capacity during a simulator check ride, an increased task demand during a live flying mission have a potential to exceed the pilot's mental capacity and impair his/her performance (see Figures 12). PMWL assessment during check rides, if conducted at all, has traditionally relied on primary task measures (Hyland et al. 1994; Ruffner et al. 1984; Childs 1979; Weitzman et al. 1979; Woodruff et al. 1976; Caro and Isley 1966). While the unexpected task demands of flying and the limitations of the human mental capacity are well known, it is somewhat concerning that the PMWL measurement during proficiency checks is still in its infancy. Measuring pilots' PMWL and performance during check rides can give valuable insights about pilots' ability to cope with the high task demands (see Chapter 3.4 and Figure 13).

In this thesis, the utility of HR/HRV as an adjunct to pilot proficiency checks were examined. While the multidimensional nature of PMWL is acknowledged, there are many studies that have been able to utilise HR and HRV as measures of PMWL (Lahtinen et al. 2007; Magnusson 2002; Veltman 2002; Svensson et al. 1997). That being said, a single measure is typically not enough to capture all the different dimensions of mental workload. When PMWL is considered, the number of usable measures of pilots' psychophysiological responses is highly limited, mainly because of the limitations and restrictions a simulator environment puts on the measuring instrumentation. A real flying environment adds more restrictions to the available measures, as the instrumentation has to be flight-worthy and it may not compromise the flight safety and mission success. If pilot responses to varying (mental) task demands are measured during a real flight, the physical demands of fighter aviation generate remarkable source of error. It is fortunate that there are certain check rides that are flown in a simulator and can therefore be considered as ‘operational missions’. But even if multiple physiological measures could be utilised in a simulator environment, they may still fail to fully explain the relationship between mental task demand, PWML and performance as there are other constructs, such as SA, which may influence PMWL and performance. These concepts, however, were not the emphasis of this thesis. While the ergonomics community waits for the more sophisticated objective measures of task demand, PMWL and SA to become available also for the flying environment, it has to settle to those - maybe less sophisticated - measures accepted (on a case by case basis) by the aviation authorities. In that sense, HR and HRV have both justified their place among the usable measures of task demand in aviation domain (see for example: Magnusson 2002, Ylönen et al. 1997, Jorna 1993).

HR/HRV demonstrated to be useful measures of PMWL in a flight simulator environment. The method was accepted by the pilots as it was not intrusive to their task and it appeared to be objective. In addition, the implementation requirements were by no means excessive; the electrode placement was not time consuming, the installation of the Biotrace Nexus-10 instrumentation into the simulator was simple and the Kubios HRV Software proved to be a powerful and user friendly tool to analyse HR/HRV. Considering the objectives of this thesis, the low diagnosticity of HR/HRV was not seen as a problem. Finally, HR/HRV proved to be sensitive measures of varying task demands in a flight simulator – especially when measured together with the pilots’ awareness of the mission requirements. For an additional discussion regarding the PMWL measurement criteria, see Chapter 3.2.

The first study of this thesis (see Chapter 5) extended the findings of earlier studies by investigating the associations between the PMWL and performance. The results can be used to support the identification of the PMWL redline in a realistic, or realistic simulated environment. To this end, the association between the HR/HRV components and the pilot's ILS performance was studied. With careful selection of PMWL metrics and thorough task analysis, it was possible to replicate the findings of the earlier mental workload related HR/HRV studies (see, e.g., Saperova and Dimitriev 2014; Monge et al. 2014; Perusse-Lachance et al., 2012; Sun et al. 2012; Taelman et al. 2011; Tran et al. 2010; Li et al. 2009; Orsila et al. 2008; Terkelsen et al. 2005).

The results of the first study (see Chapter 5.3) clearly indicated that HR and HRV are sensitive to varying ratios between the time available and the time required for completing the tasks in the fighter aviation environment. Also, ECG monitoring was a relatively cheap method to assess HR and HRV, it had a high face validity and it did not generate intrusion of any kind. As this study had a within-subject, repeated measures design, most of the issues related to variations in skill and experience and the idiosyncratic heart rate responses were avoided. Furthermore, because of the way the three categories of the ILS performance were defined, every pilot effectively set their own datum.

The MEANRR was able to differentiate the high performance and the sub-standard performance categories, i.e., their MEANRR averages were significantly different across the pilot population (see Figure 17). This finding, however, did not represent the redline of the subject or the subject population but rather suggested that such a redline could be found. The sub-standard performance category and the low performance category were not differentiated by the HR/HRV components used (see Table 4). Some subjects may have found the low performance category trial impossible and have eventually given up, i.e., they have invested less effort for the low performance category trial than they did for the sub-standard performance category trial. As a result, their ILS performance was extremely poor while the HR/HRV response to increased task demand suggested lower PMWL. The first study of this thesis showed that the MEANRR component of the HRV is a strong candidate when different measures of PMWL are being considered for a fighter aviation environment.

The second study of this thesis (see Chapter 6) successfully utilised HR and HRV, as measures of PMWL, during real F/A-18 IFR check rides and was able to replicate the findings of the earlier mental workload related HR and HRV studies (Deepak et al. 2014; Cinaz et al. 2013; Sun et al. 2012; Taelman et al. 2011; Tran et al. 2010; Li et al. 2009; Orsila et al. 2008; Vuksanović and Gal 2007; Terkelsen et al. 2005; Svensson and Wilson 2002; Wilson 2002; Roscoe 1993, 1975). Unlike some earlier studies using HR and HRV (Fallahi 2016; Wei et al. 2014; Veltman and Gaillard 1996; Wilson 1992; Jorna 1992), the second study was able to differentiate HR/HRV variations between mission segments instead of differentiating just the rest and trial conditions. In addition, whereas Mansikka et al. (2016a) successfully used HR and HRV to differentiate large task demand changes during a simulated flight (see Chapter 5), the second study of this thesis replicated these results with smaller task demand variations during a realistic, simulated flying task. The flying mission consisted of separate mission segments. Each mission segment exposed pilots to different task demands as each mission segment tested different aspects of pilots' IMC flying abilities. Both the performance data and the ECG data were retrieved from each mission segment. As a result, HR and HRV data provided an adjunct and sensitive measure of PMWL and could, at some later stage, be used to support the evaluation of the pilots' spare mental capacities. The measured differences in the pilots' spare mental capacities can give valuable information about their ability to maintain the desired performance during events of high task demand.

As discussed in Chapter 6.3, the experienced F/A-18 pilots were able to maintain high and mostly equal performance across the different segments of the IFR check ride (see Tables 6 and 7). At the same time, the HR values and the HRV components' values indicated that their PMWL between the different mission segments was not equal (see Tables 8 and 9). As summarised in Table 10, the 'Takeoff and Ingress' and 'SEM' mission segment pair and the "ILS Approach" and "VOR Approach" segment pair were neither differentiated by the performance scores nor by the HR/HRV responses. Thus, it can be concluded that the task demand and the resulting PMWL of these mission segment pairs were very similar. There were seven mission segment pairs ('Takeoff and Ingress' and 'Manoeuvering'; 'Takeoff and Ingress' and 'ILS Approach'; 'Manoeuvering' and 'Level Turns'; 'Manoeuvering' and 'SEM'; 'Manoeuvering' and 'VOR Approach'; 'Manoeuvering' and 'ILS Approach'; 'SEM' and 'ILS Approach') that were differentiated both by the performance scores and by one or more of the HR/HRV measures. It was concluded that these segments were different in their task demands and also generated different ANS responses as revealed by the changes in HR/HRV.

Out of the total of 15 mission segment pairs analysed, there were eight mission segment pairs which were not differentiated by the performance scores. However, of these eight mission segment pairs six were differentiated by one or more of the HR/HRV measures (see Table 10). In other words, there were PMWL differences in 75% of those mission segment pairs that could not reveal differences in performance, i.e., PMWL measured with HR/HRV was more sensitive than the performance score when the mission segments and their task demands were differentiated. Although the overt, traditional performance measures suggest otherwise, the subjects' average potential to operate effectively and safely varied between the mission segments. HR/HRV proved to be potential measures of PMWL should the individual PMWL and performance differences between the pilots be evaluated and used as an additional IFR check ride criterion, i.e., the pilot would have to achieve a minimum performance score without exceeding the given level of PMWL. With a slightly modified test design, it would be possible to study if the pilots could be differentiated by their performance and PMWL; insignificant and non-significant differences in pilots' performance coupled with significant differences in their PMWL could be used to reflect the pilots' different mental spare capacities.

Pilots are usually well aware of the task requirements of the instrument proficiency check ride. The information required for decision making and response selection is continuously available and the required physical movements of the throttle and stick are easy to perform. In addition, the physical demands during an instrument check ride are almost non-existent. From an information processing perspective very simple, sequenced decisions and responses are required from the pilot. Most of the information required for successful response selection stresses the visual, and more specifically the focal element of the visual modality. However, the response selection is seldom based on a single cockpit instrument reading only. Therefore, a pilot needs to maintain a rapid and continuous visual scan of all relevant instruments to be able to gain a proper awareness of the aircraft's state and the required responses. Once the responses have been selected and executed, the instrument readings change and most likely require another cycle of decision making, response selection and response execution. For an averagely skilled pilot, the perception and response selection are unlikely to become limited during an IFR check ride. However, the requirement to memorise, even for a very short time, the readings of the different instruments may place high demands on WM, especially during the mission phases where the aircraft state changes aggressively. If, for any reason, the cyclical process of perception, decision making, response selection and responding slows down, the update rate of the perceived data is equally lowered and the unnoticed changes in the aircraft's state are more likely to occur. In addition, due to the lower update rate the unnoticed changes in the aircraft's state may easily become greater, requiring larger simultaneous corrections. Before corrections can be made, the combined effect of the affected instruments has to be analysed. Mis-analysed control input requirements result in inadequate or inappropriate changes in the aircraft's state. As a result, the follow-on analysis of the instruments, aircraft's state and required follow-on corrections become even more difficult. Hard, over-controlled stick and throttle inputs are probably the most commonly known indication of an approaching breakdown of the pilot's information processing cycle. For an additional discussion about pilots' information processing, see Chapter 2.

In the third study of this thesis (see Chapter 7), the MEANRR and the performance of pilots were measured during two ILS approaches and two BVR attacks. The subjects were able to get significantly higher performance scores in the ILS-B compared to those in the ILS-A (see Chapter 7.3 and Table 11). As also the MEANRR values in the ILS-A were significantly lower than in the ILS-B, it was concluded that the ILS-A had significantly higher task demand than the ILS-B (see Chapter 7.3 and Table 12). The tactical, high speed radar trail egress combined with the tactical air combat tasks preceding the ILS-A approach increased the temporal demand during the approach preparations, increased the amount of distracting information and generated potential for false attention focus during the initial approach. In addition, the radar trail approach itself increased the subjects' task loading; the onboard radar had to be adjusted to maintain a radar track of the preceding aircraft and the radar trail distance had to be kept within acceptable limits. It is assumed that these factors were enough to generate a significant task demand difference between otherwise similar approaches. Despite the distracting elements during the initial approach of ILS-A, the pilots were highly familiar with the approaches; they knew the approach profiles and procedures, they knew what information was needed to execute the approaches and they knew where the information could be acquired. It can be thus assumed that any observed performance degradations were not caused by the pilots' lack of understanding of the mission requirements. As a result, the MEANRR values and the performance scores during the ILS approaches followed the expected pattern and replicated the findings discussed in Chapters 5 and 6; an increased task demand was reflected in increased PMWL as indicated by the lowered values of the MEANRR.

When the subjects' performance scores in the Commit-A and Commit-B were compared, the Commit-A had a significantly higher performance score average than the Commit-B (see Table 11). However, the MEANRR values of the Commit-A and Commit-B did not follow the same pattern as they did during the ILS approaches (see Table 12). The MEANRR values during the Commit-B were significantly higher than during the Commit-A (see Chapter 7.3 and Figure 18). In other words, the Commit-B had lower performance scores and lower PMWL. Unlike the ILS approaches, the Commit-A and Commit-B were highly complex tasks with rapidly and unexpectedly changing tactical situation. These conditions generated extremely high demands on subjects' ability to collect and process relevant information, to build a coherent mental picture of current and future events as well as to select their responses accordingly.

When the subjects failed to gain and maintain a proper awareness of the mission requirements during the Commit-A and Commit-B, it was likely that their responses were less than optimal and their performance scores were compromised. Also, when the subjects failed to perceive the necessary information related to the tactical situation, they had no basis to conduct the higher level mental processing required for the establishment of the mental picture of current and future events (Matthews et al. 2011; Endsley 1996, 1995, 1990, 1988; Aasman et al. 1987). In other words, when the subjects were unaware of the required tasks, they did not excessively consume their mental processing capacity and their PMWL remained low.

When the association of PMWL and the pilot performance is being evaluated during a complex air combat mission, a pilot with a low PMWL and a low performance might be inferred as having a high cognitive potential neglected by the lack of effort. But when the pilot's awareness of the mission requirements is considered as an additional measure, the conclusion may be very different; a low performance could result from the pilot working at the upper limits of his/her cognitive resources thus being unable to even perceive and process the mission critical information.

The aim of this thesis was to investigate if HR and HRV are sensitive measures of PMWL in a simulated fighter aviation environment. HR and HRV proved to be associated with the changes in task demands and pilots' performance during the simulated instrument approaches and air combat. The research questions are summarised in Figure 19.

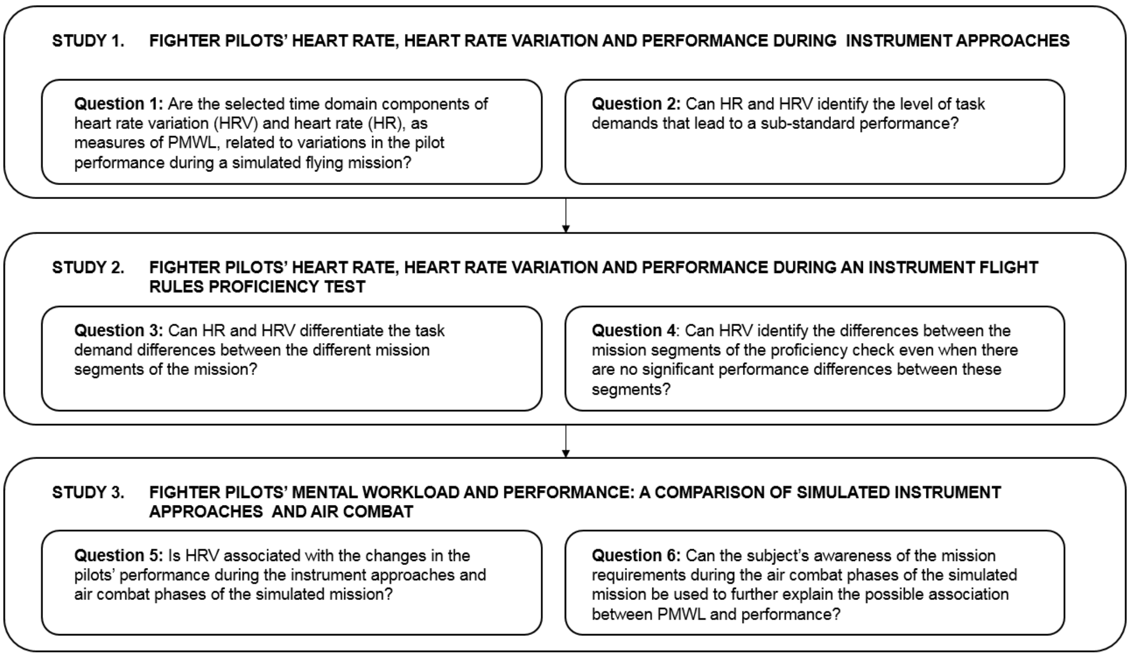


Figure 19. Summary of the research questions.

The three studies reported in this thesis provided answers for all research questions. More specifically:

- The first study showed that HR and HRV were able to identify the level of task demands that led to a sub-standard performance during instrument approaches. This finding provided answers for the research questions one and two.

- The second study showed that HR and HRV were able to differentiate varying levels of task demands during a real instrument flight rules proficiency check. In addition, the HR and HRV were able to identify the pilots' PMWL differences even in situations where the performance differences between pilots were not significant. This finding provided answers for the research questions three and four.

- The third study showed that HRV was associated with the varying task demands and performance during simulated air combat. It became obvious that during the air combat phases of the mission the analysis of HRV and performance benefits greatly of the simultaneous assessment of pilots' awareness of the mission requirements. This finding provided answers for the research questions five and six.

While the results of this thesis are highly promising, further research is still required before PMWL can be used as an additional pass-fail criterion for a check-ride. The results of this thesis suggest that measuring just PMWL and performance is not sufficient – especially if the task of interest is complex and dynamic. To fully understand the pilot performance in such environment, the relationship between SA, workload and performance needs to be untangled. Also, the measures of the awareness of the mission requirements, or SA, require further research and testing; whatever measuring technique is selected, it must be objective and it may not disturb the mission accomplishment.

Whereas this thesis concentrated on the pilots' performance and PMWL, the test settings described in the three studies of this thesis could be used for various research purposes. More specifically, similar test settings – with only slight modifications – could be used to study learning effects, pilots' decision making strategies, skill decay and the demographic and experience related factors influencing them. It is concluded that this thesis was successful in achieving its aims and serves as an encouraging basis for future research aiming to improve the effectiveness and safety of current and future fighter pilots.

References

Aasman, J., Mulder, G. and Mulder, L. 1987. Operator Effort and the Measurement of Heart-Rate Variability. Human Factors 29 (2): 161-170.

Ahmadi, K. and Alireza, K. 2007. Stress and Job Satisfaction among Air Force Military Pilots. Journal of Social Sciences 3 (3): 159.

Anderson, P. and Toivanen, M. 1970. Effects of Varying Levels of Autopilot Assistance and Workload on Pilot Performance in the Helicopter Formation Flight Mode. Joint Army-Navy Aircraft Instrumentation Research (JANAIR) Report 680610. Honeywell Inc., Clearinghouse, Springfield, VA.

Annett, J. 2002a. Subjective Rating Scales in Ergonomics: A Reply. Ergonomics 45 (14): 1042-1046.

Annett, J. 2002b. Subjective Rating Scales: Science Or Art? Ergonomics 45 (14): 966-987.

Atkinson, R. and Shiffrin, R. 1968. Human Memory: A Proposed System and its Control Processes. The Psychology of Learning and Motivation 2: 89-195.

Ayaz, H., Willems, B., Bunce, B., Shewokis, P., Izzetoglu, K., Hah, S., Deshmukh, A. and Onaral, B. 2010. Cognitive Workload Assessment of Air Traffic Controllers using Optical Brain Imaging Sensors. Advances in Understanding Human Performance: Neuroergonomics, Human Factors Design and Special Populations: 21-31.

Baber, C. 2002. Subjective evaluation of usability. Ergonomics 45 (14): 1021-1025.

Bach, R., 1964. Stranger to the Ground, Dell Publishing, New York, NY.

Baddeley, A. 2012. Working memory: theories, models, and controversies. Annual Review of Psychology 63: 1-29.

Baddeley, A. 2000. The Episodic Buffer: A New Component of Working Memory? Trends in Cognitive Sciences 4 (11): 417-423.

Baddeley, A. and Hitch, G. 2000. Development of Working Memory: Should the Pascual-Leone and the Baddeley and Hitch Models be Merged? Journal of Experimental Child Psychology 77 (2): 128-137.

Baddeley, A. 1998. Recent Developments in Working Memory. Current Opinion in Neurobiology 8 (2): 234-238.

Baddeley, A. 1996a. Exploring the Central Executive. The Quarterly Journal of Experimental Psychology: Section A 49 (1): 5-28.

Baddeley, A. 1996b. The Fractionation of Working Memory. Proceedings of the National Academy of Sciences of the United States of America 93 (24): 13468-13472.

Baddeley, A. 1992. Working Memory. Science 255 (5044): 556-559.

Baddeley, A., Lewis, V. and Vallar, G. 1984. Exploring the articulatory loop. The Quarterly journal of experimental psychology, 36 (2): 233-252.

Baddeley, A., Thomson, N. and Buchanan, M. 1975. Word Length and the Structure of Short-Term Memory. Journal of Verbal Learning and Verbal Behavior 14 (6): 575-589.

Baddeley, A. and Hitch, G. 1974. Working memory. Psychology of learning and motivation, 8: 47-89.

Bailón, R., Laguna, P., Mainardi, L. and Sornmo, L. 2007. Analysis of heart rate variability using time-varying frequency bands based on respiratory frequency. Paper presented at the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society: 6674-6677.

Battiste, V. and Bortolussi, M. 1988. Transport Pilot Workload: A Comparison of Two Subjective Techniques. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 32 (2): 150-154.

Beatty, J. 1982. Task-evoked pupillary responses, processing load, and the structure of processing resources. Psychological bulletin 91 (2): 276.

Berka, C., Levendowski, D., Lumicao, M., Yau, A., Davis, G., Zivkovic, V., Olmstead, R., Tremoulet, P. and Craven, P. 2007. EEG Correlates of Task Engagement and Mental Workload in Vigilance, Learning and Memory Tasks. Aviation, Space and Environmental Medicine 78 (Supplement 1): B231-B244.

Berkun, M. 1964. Performance Decrement under Psychological Stress. Human Factors 6 (1): 21-30.

Billings, C. and Reynard, W. 1984. Human factors in aircraft incidents: results of a 7-year study. Aviation, Space, and Environmental Medicine 55 (10): 960-965.

Boff, K., Kaufman, L. and Thomas, J. 1994. Handbook of Perception and Human Performance. Vol. 2. Cognitive Processes and Performance. Wiley-Interscience, New York, N.Y.

Brookhuis, K., De Waard, D. and Fairclough, S. 2003. Criteria for Driver Impairment. Ergonomics 46 (5): 433-445.

Brown, I. 1978. Dual task methods of assessing workload. Ergonomics 21 (3): 221-224.

Camm, A., Malik, M., Bigger, J., Breithardt, G., Cerutti, S., Cohen, R., Coumel, P., Fallen, E., Kennedy, H. and Kleiger, R. 1996. Heart Rate Variability: Standards of Measurement, Physiological Interpretation and Clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Circulation 93 (5): 1043-1065.

Camp, G., Paas, F., Rikers, R. and van Merriënboer, J. 2001. Dynamic Problem Selection in Air Traffic Control Training: A Comparison between Performance, Mental Effort and Mental Efficiency. Computers in Human Behavior 17 (5): 575-595.

Carmody, M. 1994. Current Issues in the Measurement of Military Aircrew Performance: A Consideration of the Relationship between Available Metrics and Operational Concerns. Report No. NAWCADWAR-94139-60, Naval Air Warfare, Center Aircraft Division, Warminster, PA.

Caro, P. and Isley, R. 1966. Helicopter trainee performance following synthetic flight training. Journal of the American Helicopter Society 11 (3): 38-44.

Casali, J. and Wierwille, W. 1984. On the Measurement of Pilot Perceptual Workload: A Comparison of Assessment Techniques Addressing Sensitivity and Intrusion Issues. Ergonomics 27 (10): 1033-1050.

Casali, J. and Wierwille, W. 1983. A Comparison of Rating Scale, Secondary-Task, Physiological and Primary-Task Workload Estimation Techniques in a Simulated Flight Task Emphasizing Communications Load. Human Factors 25 (6): 623-641.

Casner, S. and Gore, B. 2010. Measuring and Evaluating Workload: A Primer. NASA Technical Memorandum 2010-216395. Ames Research Center, San Jose, CA.

Casner, S. 2009. Perceived vs. Measured Effects of Advanced Cockpit Systems on Pilot Workload and Error: Are Pilots' Beliefs Misaligned with Reality? Applied Ergonomics 40 (3): 448-456.

Childs, J. 1979. The development of objective inflight performance assessment procedures. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 23 (1): 329-333.

ChuDuc, H., NguyenPhan, K. and NguenViet, D. 2013. A Review of Heart Rate Variability and its Applications. APCBEE Procedia 7: 80-85.

Cinaz, B., Arnrich, B., La Marca, R. and Tröster, G. 2013. Monitoring of Mental Workload Levels during an Everyday Life Office-Work Scenario. Personal and Ubiquitous Computing 17 (2): 229-239.

Cocchini, G., Logie, R., Della, S., MacPherson, S. and Baddeley, A. 2002. Concurrent performance of two memory tasks: Evidence for domain-specific working memory systems. Memory & Cognition 30 (7): 1086-1095.

Collet, C., Salvia, E. and Petit-Boulanger, C. 2014. Measuring Workload with Electrodermal Activity during Common Braking Actions. Ergonomics 57 (6): 886-896.

Cowan, N., Wood, N., Wood, P., Keller, T., Nugent, L. and Keller, C. 1998. Two separate verbal processing rates contributing to short-term memory span. Journal of experimental psychology: General, 127 (2): 141.

Dahlstrom, N. and Nahlinder, S. 2009. Mental Workload in Aircraft and Simulator during Basic Civil Aviation Training. The International Journal of Aviation Psychology 19 (4): 309-325.

Deepak, A., Deepak, A. N, Nallulwar, S. and Khode, V. 2014. Time Domain Measures of Heart Rate Variability during Acute Mental Stress in Type 2 Diabetics-A Case Control Study. National Journal of Physiology, Pharmacy and Pharmacology 4 (1): 34-38.

Diehl, A. 1991. Human performance and systems safety considerations in aviation mishaps. The International Journal of Aviation Psychology 1 (2): 97-106.

Driskell, J. and Salas, E. 1991. Group Decision Making Under Stress. Journal of Applied Psychology 76 (3): 473.

Duncan, J., Martens, S. and Ward, R. 1997. Within but not between sensory modalities. Nature 387 (809): 1000-1038.

Dussault, C., Jouanin, J.-E. and Guezennec, C.-Y. 2004. EEG and ECG Changes during Selected Flight Sequences. Aviation, Space and Environmental Medicine 75 (10): 889-897.

Eggemeier, F. 1988. Properties of workload assessment techniques. Advances in Psychology 52: 41-62.

Endsley, M. 2001. Designing for situation awareness in complex systems. Proceedings of the Second International Workshop on symbiosis of humans, artifacts and environment: 1-14.

Endsley, M. 1999. Situation awareness in aviation systems. In Garland, D., Wise, J. and Hopkin, V. (Eds.), Handbook of aviation human factors: 257-276. Lawrence Erlbaum Associates, Mahwah, NJ.

Endsley, M. 1996. Automation and Situation Awareness. Automation and Human Performance: Theory and Applications: 163-181.

Endsley, M. 1995. Toward a Theory of Situation Awareness in Dynamic Systems. Human Factors 37 (1): 32-64.

Endsley, M. and Bolstad, C. 1994. Individual Differences in Pilot Situation Awareness. The International Journal of Aviation Psychology 4 (3): 241-264.

Endsley, M. 1990. Predictive Utility of an Objective Measure of Situation Awareness. Proceedings of the Human Factors Society 34th Annual Meeting 34 (1): 41-45.

Endsley, M. 1988. Situation Awareness Global Assessment Technique (SAGAT). Proceedings of the Aerospace and Electronics Conference 34: 789-795.

Fallahi, M., Motamedzade, M., Heidarimoghadam, R., Soltanian, A. and Miyake, S. 2016. Effects of Mental Workload on Physiological and Subjective Responses during Traffic Density Monitoring: A Field Study. Applied Ergonomics 52: 95-103.

Fox, S. 1996. Human Physiology, 9th edition. McGraw-Hill Education, Boston, MA.

Glanzer, M. and Cunitz, A. 1966. Two Storage Mechanisms in Free Recall. Journal of Verbal Learning and Verbal Behavior 5 (4): 351-360.

Gomer, F., Spicuzza, R. and O'Donnell, R. 1976. Evoked Potential Correlates of Visual Item Recognition during Memory-Scanning Tasks. Physiological Psychology 4 (1): 61-65.

Gopher, D. and Donchin, E. 1986. Workload – an Examination of the Concept. In Boff, K., Kaufman, L. and Thomas (Eds.), Handbook of Perception and Human Performance. Cognitive Processes and Performance Vol. 2: 41, 1-49. Wiley-Interscience, New York, N.Y.

Grandjean, E. and Kroemer, K. 1997. Fitting the Task to the Human: A Textbook of Occupational Ergonomics.Taylor and Francis, Boca Raton, FL.

Green, R. and Flux, R. (1977). Auditory Communication and Workload. Paper presented at the ASMP Meeting, Studies of Pilot Workload.

Gugerty, L. 1997. Situation Awareness during Driving: Explicit and Implicit Knowledge in Dynamic Spatial Memory. Journal of Experimental Psychology: Applied 3 (1): 42.

Hall, J. 2015. Guyton and Hall Textbook of Medical Physiology. Elsevier Health Sciences. Elsevier, Philadelphia, PA.

Hancock, P. and Warm, J. 1989. A Dynamic Model of Stress and Sustained Attention. Human Factors 31 (5): 519-537.

Hankins, T. and Wilson, G. 1998. A Comparison of Heart Rate, Eye Activity, EEG and Subjective Measures of Pilot Mental Workload during Flight. Aviation, Space and Environmental Medicine 69 (4): 360-367.

Hansman, R., Wanke, C., Kuchar, J., Mykityshyn, M., Hahn, E. and Midkiff, A. 1993. Hazard Alerting and Situational Awareness in Advanced Air Transport Cockpits. Paper presented at the 18th ICAS Congress, Beijing, China.

Harris, D., 2011. Human Performance on the Flight Deck, Ashgate, Farnham, Surrey.

Hart, S. and Staveland, L. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. Advances in Psychology 52: 139-183.

Hart, S. and McPherson, D. 1976. Airline Pilot Time Estimation during Concurrent Activity Including Simulated Flight. Paper presented at the 47th Annual Meeting of the Aerospace Medical Association, Bar Harbour, Florida, FL.

Hayward, C., Patel, H., Manohar, S., Waton, J., Middleton, L., Lyon, A., Sutton, R. and Rosen, S. 2014. 39 Short Time Window Analysis of Heart Rate Variability during Tilt Table Testing Provides Insights in to the Timing of Changes in the Autonomic Nervous System. Europace 16 (suppl 3): iii16-iii16.

Hecker, R. and Mapperson, B. 1997. Dissociation of visual and spatial processing in working memory. Neuropsychologia 35 (5): 599-603.

Hendrick, C. and Costantini, A. 1970. Effects of varying trait inconsistency and response requirements on the primacy effect in impression formation. Journal of Personality and Social Psychology 15 (2): 158.

Hill, S., Iavecchia, H., Byers, J., Bittner, A., Zaklade, A., and Christ, R. 1992. Comparison of four subjective workload rating scales. Human Factors 34 (4): 429-439.

Horwitz, L., Atkins, J. and Leshin, S. 1972. Role of the Frank-Starling mechanism in exercise. Circulation research 31 (6): 868-875.

Horrey, W. and Wickens, C. 2003. Multiple resource modeling of task interference in vehicle control, hazard awareness and in-vehicle task performance. Proceedings of the Second International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design: 7-12.

Houghton, A., and Gray, D. 2014. Making sense of the ECG: a hands-on guide. CRC Press, Boca Raton, FL.

Hyland, D., Kay, E. and Deimler, J. 1994. Age 60 study, part 4: Experimental evaluation of pilot performance. Final Report DOT/FAA/AM-94/23. U.S. Department of Transportation, Federal Aviation Administration. Washington, D.C.

Huddleston, H. and Wilson, R. 1971. An Evaluation of the Usefulness of Four Secondary Tasks in Assessing the Effect of a Lag in Simulated Aircraft Dynamics. Ergonomics 14 (3): 371-380.

Iani, C., Gopher, D., Grunwald, A. and Lavie, P. 2007. Peripheral Arterial Tone as an on-Line Measure of Load in a Simulated Flight Task. Ergonomics 50 (7): 1026-1035.

Isreal, J., Wickens, C., Chesney, G. and Donchin, E. 1980. The event-related brain potential as an index of display-monitoring workload. Human Factors 22 (2), 211-224.

Ivanov, P., Amaral, L., Goldberger, A., Havlin, S., Rosenblum, M., Struzik, Z. and Stanley, H. 1999. Multifractality in human heartbeat dynamics. Nature 399 (6735): 461-465.

Johanssen, G., Moray, N., Pew, R., Rasmussen, J., Sanders, A. and Wickens, C. 1979. Final report of experimental psychology group. In Mental Workload Vol. 8 of the NATO Conference Series: 101-114, Springer, New York, NY.

Jorna, P. 1993. Heart Rate and Workload Variations in Actual and Simulated Flight. Ergonomics 36 (9): 1043-1054.

Jorna, P. 1992. Spectral Analysis of Heart Rate and Psychological State: A Review of its Validity as a Workload Index. Biological Psychology 34 (2): 237-257.

Kaber, D. and Endsley, M. 2004. The Effects of Level of Automation and Adaptive Automation on Human Performance, Situation Awareness and Workload in a Dynamic Control Task. Theoretical Issues in Ergonomics Science 5 (2): 113-153.

Kahneman, D. 1973. Attention and effort. Prentice Hall, Englewood Cliffs, NJ.

Kavanagh, J. 2005. Stress and Performance. A Review of the Literature and its Applicability to the Military. RAND Corporation, Santa Monica, CA.

Katz, A. 1977. Physiology of the Heart. Raven Press, New York, NY.

Kerr, B. 1973. Processing demands during mental operations. Memory & Cognition 1 (4): 401-412.

Klauer, K., and Zhao, Z. 2004. Double dissociations in visual and spatial short-term memory. Journal of experimental psychology: General 133 (3): 355.

Kramer, A., Sirevaag, E. and Braune, R. 1987. A Psychophysiological Assessment of Operator Workload during Simulated Flight Missions. Human Factors 29 (2): 145-160.

Lahtinen, T., Koskelo, J-P., Laitinen, T. and Leino, T. 2007. Heart Rate and Performance during Combat Missions in a Flight Simulator. Aviation, Space and Environmental Medicine 78 (4): 387-391.

Lee, Y.-H. and Liu, B-S. 2003. Inflight Workload Assessment: Comparison of Subjective and Physiological Measurements. Aviation, Space and Environmental Medicine 74 (10): 1078-1084.

Levy, M. 1971. Sympathetic-parasympathetic interactions in the heart. Circulation research 29 (5): 437-445.

Li, Z., Snieder, H., Su, S., Ding, X., Thayer, J., Treiber, F. and Wang, X. 2009. A Longitudinal Study in Youth of Heart Rate Variability at Rest and in Response to Stress. International Journal of Psychophysiology 73 (3): 212-217.

Lieberman, H., Bathalon, G., Falco, C., Morgan, C., Niro, P. and Tharion, W. 2005. The Fog of War: Decrements in Cognitive Performance and Mood Associated with Combat-Like Stress. Aviation, Space and Environmental Medicine 76 (Supplement 1): C7-C14.

Logie, R. 2014. Visuo-spatial working memory. Lawrence Erlbaum Associates, Hove, East Sussex.

Logie, R. and Marchetti, C. 1991. Visuo-spatial working memory: Visual, spatial or central executive? Advances in Psychology 80: 105-115.

Lutfi, M., and Sukkar, M. 2011. Effect of blood pressure on heart rate variability. Khartoum Medical Journal 4 (1): 548-553.

Lysaght, R., Hill, S., Dick, A., Plamondon, B. and Linton, P. 1989.Operator Workload: Comprehensive Review and Evaluation of Operator Workload Methodologies. Technical Report 851. U.S. Army Institute for Behavioral and Social Sciences, Alexandria, VA.

Magnusson, S. 2002. Similarities and Differences in Psychophysiological Reactions between Simulated and Real Air-to-Ground Missions. The International Journal of Aviation Psychology 12 (1): 49-61.

Mallis, M. and Dinges, D. 2004. Monitoring Alertness by Eyelid Closure. In The Handbook of Human Factors and Ergonomics Methods, CRC Press, New York, NY.

Mansikka, H., Simola, P., Virtanen, K., Harris, D. and Oksama, L. 2016a. Fighter Pilots' Heart Rate, Heart Rate Variation and Performance during Instrument Approaches. Ergonomics: 1-9.

Mansikka, H., Virtanen, K., Harris, D. and Simola, P. 2016b. Fighter Pilots' Heart Rate, Heart Rate Variation and Performance during an Instrument Flight Rules Proficiency Test. Applied Ergonomics 56: 213-219.

Mansikka, H., Virtanen, K., Harris, D. and Simola, P. 2016c. Fighter Pilots' Mental Workload and Performance: A Comparison of Simulated Instrument Approaches and Air Combat. Submitted for Publication.

Matthews, G., Warm, J. and Washburn, D. 2007. Diagnostic Methods for Predicting Performance Impairment Associated with Combat Stress. Final Report. University of Cincinnati, Cincinnati, OH.

Matthews, M., Eid, J., Johnsen, B. and Boe, O. 2011. A Comparison of Expert Ratings and Self-Assessments of Situation Awareness during a Combat Fatigue Course. Military Psychology 23 (2): 125.

Mavin, T. and Roth, W.-M. 2014. A Holistic View of Cockpit Performance: An Analysis of the Assessment Discourse of Flight Examiners. The International Journal of Aviation Psychology 24 (3): 210-227.

Mehler, B., Reimer, B. and Wang, Y. 2011. A Comparison of Heart Rate and Heart Rate Variability Indices in Distinguishing Single Task Driving and Driving Under Secondary Cognitive Workload. Proceedings of the Sixth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design: 590-597.

Meshkati, N. and Hancock, P. 2011. Human mental workload, Vol. 52. Elsevier, Amsterdam.

Michell, J. 1997. Quantitative Science and the Definition of Measurement in Psychology. British Journal of Psychology 88 (3): 355-383.

Michon, J. 1966. Tapping Regularity as a Measure of Perceptual Motor Load. Ergonomics 9 (5): 401-412.

Miller, G. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological review 63 (2): 81-97.

Miyake, S., Yamada, S., Shoji, T., Takae, Y., Kuge, N. and Yamamura, T. 2009. Physiological Responses to Workload Change. A test/retest Examination. Journal of Applied Ergonomics 40 (6): 987-996.

Monge, M., Gomez, I. and Molina, A. 2014. Robustness of parameters from heart rate for mental stress detection. Experimental & Clinical Cardiology 20 (8): 3509-3517.

Morris, C. and Leung, Y. 2006. Pilot Mental Workload: How Well do Pilots Really Perform? Ergonomics 49 (15): 1581-1596.

Mulder, G. 1986. The concept and measurement of mental effort. In Hockey, G. (Eds.), Energetics and human information processing. Springer, Dordrecht.

Murdock Jr, B. 1962. The Serial Position Effect of Free Recall. Journal of Experimental Psychology 64 (5): 482.

Muse, L., Harris, S. and Field, H. 2003. Has the Inverted-U Theory of Stress and Job Performance had a Fair Test? Human Performance 16 (4): 349-364.

Navon, D. and Gopher, D. 1979. On the Economy of the Human-Processing System. Psychological Review 86 (3): 214.

Niskanen, J.-P., Tarvainen, M., Ranta-Aho, P., and Karjalainen, P. 2004. Software for advanced HRV analysis. Computer methods and programs in biomedicine 76 (1): 73-81.

Noel, J., Bauer Jr, K. and Lanning, J. 2005. Improving Pilot Mental Workload Classification through Feature Exploitation and Combination: A Feasibility Study. Computers & Operations Research 32 (10): 2713-2730.

Norman, D. 1976. Memory and attention: introduction to human information processing, 2nd edition. John Wiley and Sons, New York, NY.

Norman, D. and Bobrow, D. 1975. On Data-Limited and Resource-Limited Processes. Cognitive Psychology 7 (1): 44-64.

O’Donnell, R., Eggemeier, F. and Thomas, F. 1986. Workload Assessment Methodology. In Boff, K., Kaufman, L. and Thomas, J. (Eds.), Handbook of Perception and Human Performance Vol. 2: 42, 1-49. Wiley-Interscience, New York, N.Y.

Ogden, G., Levine, J. and Eisner, J. 1979. Measurement of Workload by Secondary Tasks. Human Factors 21 (5): 529-548.

O'hare, D. 2000. The ‘Wheel of Misfortune’: A Taxonomic Approach to Human Factors in Accident Investigation and Analysis in Aviation and Other Complex Systems. Ergonomics 43 (12): 2001-2019.

O'Hare, D., Wiggins, M., Batt, R. and Morrison, D. 1994. Cognitive Failure Analysis for Aircraft Accident Investigation. Ergonomics 37 (11): 1855-1869.

Opie, L. 2004. Heart physiology: from cell to circulation. Williams & Wilkings, Baltimore, MD.

Opmeer, C. 1973. The Information Content of Successive RR-Interval Times in the ECG. Preliminary Results using Factor Analysis and Frequency Analysis. Ergonomics 16 (1): 105-112.

Orsila, R., Virtanen, M., Luukkaala, T., Tarvainen, M., Karjalainen, P., Viik, J., Savinainen, M. and Nygard, C.-H. 2008. Perceived Mental Stress and Reactions in Heart Rate Variability–a Pilot Study among Employees of an Electronics Company. International Journal of Occupational Safety and Ergonomics (JOSE) 14 (3): 275-283.

Paas, F., Tuovinen, J., Tabbers, H. and Van Gerven, P. 2003. Cognitive load measurement as a means to advance cognitive load theory. Educational psychologist 38 (1): 63-71.

Paas, F. and Van Merriënboer, J. 1993. The Efficiency of Instructional Conditions: An Approach to Combine Mental Effort and Performance Measures. Human Factors 35 (4): 737-743.

Parasuraman, R., Sheridan, T. and Wickens, C. 2008. Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. Journal of Cognitive Engineering and Decision Making 2 (2): 140-160.

Pashler, H. 1994. Dual-task interference in simple tasks: data and theory. Psychological bulletin 116 (2): 220.

Pérusse-Lachance, E., Tremblay, A., Chaput, J.-P., Poirier, P., Teasdale, N., Drapeau, V., Sénécal, C. and Brassard, P. 2012. Mental Work Stimulates Cardiovascular Responses through a Reduction in Cardiac Parasympathetic Modulation in Men and Women. Bioenergetics S1 1: 1-6.

Prophet, W. 1976. Long-Term Retention of Flying Skills: A Review of the Literature. Final Report 76-35. Human Resources Research Organization, Virginia, VA.

Rasmussen, J. and Jensen, A. 1974. Mental Procedures in Real-Life Tasks: A Case Study of Electronic Trouble Shooting. Ergonomics 17 (3): 293-307.

Reid, G. and Nygren, T. 1988. The Subjective Workload Assessment Technique: A Scaling Procedure for Measuring Mental Workload. Advances in Psychology 52: 185-218.

Reimer, B. and Mehler, B. 2011. The Impact of Cognitive Workload on Physiological Arousal in Young Adult Drivers: A Field Study and Simulation Validation. Ergonomics 54 (10): 932-942.

Robert, G. and Hockey, J. 1997. Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. Biological psychology 45 (1): 73-93.

Roman-Liu, D., Grabarek, I., Bartuzi, P. and Choromański, W. 2013. The Influence of Mental Load on Muscle Tension. Ergonomics 56 (7): 1125-1133.

Roscoe, A. 1993. Heart Rate as a Psychophysiological Measure for in-Flight Workload Assessment. Ergonomics 36 (9): 1055-1062.

Roscoe, A. 1992. Assessing Pilot Workload. Why Measure Heart Rate, HRV and Respiration? Biological Psychology 34 (2): 259-287.

Roscoe, A. and Ellis, G. 1990. A Subjective Rating Scale for Assessing Pilot Workload in Flight: A Decade of Practical Use. Technical Report TR 90010. Royal Aerospace Establishment, Farnborough, Hampshire.

Roscoe, A. 1975. Heart Rate Monitoring of Pilots during Steep-Gradient Approaches. Aviation, Space and Environmental Medicine 46 (11): 1410-1413.

Ruffner, J., Wick, D. and Bickley, W. 1984. Retention of Helicopter Flight Skills: Is there a “Critical Period” for Proficiency Loss? Proceedings of the Human Factors and Ergonomics Society Annual Meeting 28 (4), 370-374.

Ryu, K. and Myung, R. 2005. Evaluation of Mental Workload with a Combined Measure Based on Physiological Indices during a Dual Task of Tracking and Mental Arithmetic. International Journal of Industrial Ergonomics 35 (11): 991-1009.

Saladin, K. 1998. Anatomy & physiology. McGraw Hill, Boston, MA.

Salmon, P., Stanton, N., Walker, G., Jenkins, D., Ladva, D., Rafferty, L. and Young, M. 2009. Measuring Situation Awareness in Complex Systems: Comparison of Measures Study. International Journal of Industrial Ergonomics 39 (3): 490-500.

Salvia, E., Delpuech, C., Daligault, S., Martin, R. and Collet, C. 2015. Mental Strain while Driving on a Driving-Simulator: Potential Effect on Central and Autonomic Responses. International Journal of Ergonomics 5 (2), 26-43.

Sampson, M. and McGrath, A. 2015. Understanding the ECG Part 2: ECG basics. British Journal of Cardiac Nursing 10 (12): 588-594.

Saperova, E. and Dimitriev, D. 2014. The Effects of Mood State and Psychological Well-being on Heart Rate Variability during Mental Stress. International Journal of Psychophysiology 94 (2): 232.

Sarter, N., Mumaw, R. and Wickens, C., 2007. Pilots' Monitoring Strategies and Performance on Automated Flight Decks: An Empirical Study Combining Behavioral and Eye-Tracking Data. Human Factors 49 (3): 347-357.

Scholl, B. and Xu, Y. 2001. The magical number 4 in vision. Behavioral and Brain Sciences 24 (1): 145-146.

Seck, M., Frydman, C., Giambiasi, N., Oren, T. and Yilmaz, L. 2005. Use of a Dynamic Personality Filter in Discrete Event Simulation of Human Behavior under Stress and Fatigue. Paper presented at the1st International Conference on Augmented Cognition, Las Vegas, NV.

Selvaraj, N., Jaryal, A., Santhosh, J., Deepak, K. and Anand, S. 2008. Assessment of heart rate variability derived from finger-tip photoplethysmography as compared to electrocardiography. Journal of medical engineering & technology 32 (6): 479-484.

Setz, C., Arnrich, B., Schumm, J., La Marca, R., Troster, G. and Ehlert, U. 2010. Discriminating Stress from Cognitive Load using a Wearable EDA Device. Information Technology in Biomedicine 14 (2): 410-417.

Shappell, S. and Wiegmann, D. 2001. Applying reason: The human factors analysis and classification system (HFACS). Human Factors and Aerospace Safety 1 (1): 59-86.

Shappell, S. and Wiegmann, D. 1997. A Human Error Approach to Accident Investigation: The Taxonomy of Unsafe Operations. The International Journal of Aviation Psychology 7 (4): 269-291.

Shaw, T., Satterfield, K., Ramirez, R. and Finomore, V. 2013. Using Cerebral Hemovelocity to Measure Workload during a Spatialised Auditory Vigilance Task in Novice and Experienced Observers. Ergonomics 56 (8): 1251-1263.

Sheridan, T. and Simpson, R. 1979. Toward the Definition and Measurement of the Mental Workload of Transport Pilots. Final Report R-79-4. Flight Transportation Laboratory, Cambridge, MA.

Shiffrin, R. and Nosofsky, R. 1994. Seven plus or minus two: A commentary on capacity limitations. Psychological Review 101 (2): 357-361.

Silverthorn, D., Ober, W., Garrison, C., Silverthorn, A. and Johnson, B. 2009. Human physiology: an integrated approach. Benjamin Cummings, San Francisco, CA.

Skibniewski, F., Dziuda, L., Baran P., Krej, M., Guzowski, S., Piotrowski, M. and Truszczyński, O. 2015. Preliminary Results of the LF/HF Ratio as an Indicator for Estimating Difficulty Level of Flight Tasks. Aerospace medicine and human performance 86 (6): 518-523.

Spyker, D., Stackhouse, S., Khalafalla, A. and McLane, R. 1971. Development of Techniques for Measuring Pilot Workload. Technical Report ID 19720004411, NASA, Systems and Research Centre, Saint Pu, MN.

Stauss, H. 2003. Heart rate variability. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology 285 (5): R927-R931.

Stuiver, A., Brookhuis, K., de Waard, D. and Mulder, B. 2014. Short-Term Cardiovascular Measures for Driver Support: Increasing Sensitivity for Detecting Changes in Mental Workload. International Journal of Psychophysiology 92 (1): 35-41.

Sun, F.-T., Kuo, C., Cheng, H.-T., Buthpitiya, S., Collins, P. and Griss, M. 2012. Activity-Aware Mental Stress Detection using Physiological Sensors. In Mobile Computing, Applications and Services: 211-230. Springer, Berlin.

Svensson, E. and Wilson, G. 2002. Psychological and Psychophysiological Models of Pilot Performance for Systems Development and Mission Evaluation. The International Journal of Aviation Psychology 12 (1): 95-110.

Svensson, E., Angelborg-Thanderz, M. and Wilson, G. 1999. Models of Pilot Performance for Systems and Mission Evaluation-Psychological and Psychophysiological Aspects. Interim Report. Human Effectiveness Directorate, Wright-Patterson Air Force Base, Dayton, OH.

Svensson, E., Angelborg-Thanderez, M., Sjöberg, L. and Olsson, S. 1997. Information Complexity-Mental Workload and Performance in Combat Aircraft. Ergonomics 40 (3): 362-380.

Taelman, J., Vandeput, S., Vlemincx, E., Spaepen, A. and van Huffel, S. 2011. Instantaneous Changes in Heart Rate Regulation due to Mental Load in Simulated Office Work. European Journal of Applied Physiology 111 (7): 1497-1505.

Tan, G., Dao, T., Farmer, L., Sutherland, R. and Gevirtz, R. 2011. Heart rate variability (HRV) and posttraumatic stress disorder (PTSD): A pilot study. Applied psychophysiology and biofeedback 36 (1): 27-35.

Tarvainen, M., Niskanen, J.-P., Lipponen, J., Ranta-Aho, P. and Karjalainen, P. 2014. Kubios HRV – Heart Rate Variability Analysis Software. Computer Methods and Programs in Biomedicine 113 (1): 210-220.

Tattersall, A. and Foord, P. 1996. An Experimental Evaluation of Instantaneous Self-Assessment as a Measure of Workload. Ergonomics 39 (5): 740-748.

Terkelsen, A., Mølgaard, H., Hansen, J., Andersen, O. and Jensen, T. 2005. Acute Pain Increases Heart Rate: Differential Mechanisms during Rest and Mental Stress. Autonomic Neuroscience 121 (1–2): 101-109.

Tran, B., Papoiu, A., Russoniello, C., Wang, H., Patel, T., Chan, Y.-H. and Yosipovitch, G. 2010. Effect of Itch, Scratching and Mental Stress on Autonomic Nervous System Function in Atopic Dermatitis. Acta Dermato-Venereologica 90 (4): 354-361.

Tresch, M., Sinnamon, H. and Seamon, J. 1993. Double dissociation of spatial and object visual memory: Evidence from selective interference in intact human subjects. Neuropsychologia 31 (3): 211-219.

Ursin, H. and Ursin, R. 1979. Physiological indicators of mental workload Mental Workload. Plenum, New York, NY.

Vallar, G. and Papagno, C. 1986. Phonological short-term store and the nature of the recency effect: Evidence from neuropsychology. Brain and Cognition 5 (4): 428-442.

Valverde, H., 1973. A Review of Flight Simulator Transfer of Training Studies. Human Factors 15 (6): 510-522.

Veltman, J. 2002. A Comparative Study of Psychophysiological Reactions during Simulator and Real Flight. The International Journal of Aviation Psychology 12 (1): 33-48.

Veltman, J. and Gaillard, A. 1998. Physiological Workload Reactions to Increasing Levels of Task Difficulty. Ergonomics 41 (5): 656-669.

Veltman, J. and Gaillard, A. 1996. Physiological Indices of Workload in a Simulated Flight Task. Biological Psychology 42 (3): 323-342.

Vergauwe, E., Barrouillet, P. and Camos, V. 2010. Do mental processes share a domain-general resource? Psychological Science 21 (3): 384-390.

Verwey, W. and Veltman, H. 1996. Detecting short periods of elevated workload: A comparison of nine workload assessment techniques. Journal of Experimental Psychology: Applied 2 (3): 270-285.

Vicente, K., Thornton, D. and Moray, N. 1987. Spectral Analysis of Sinus Arrhythmia: A Measure of Mental Effort. Human Factors 29 (2): 171-182.

Vidulich, M. and Tsang, P. 1986. Techniques of Subjective Workload Assessment: A Comparison of SWAT and the NASA-Bipolar Methods. Ergonomics 29 (11): 1385-1398.

Vidulich, M., Stratton, M., Crabtree, M. and Wilson, G. 1994. Performance-Based and Physiological Measures of Situational Awareness. Aviation, Space and Environmental Medicine 65 (5), Supplement: A7-A12.

Vitense, H., Jacko, J. and Emery, K. 2003. Multimodal Feedback: An Assessment of Performance and Mental Workload. Ergonomics 46 (1-3): 68-87.

Vuksanović, V. and Gal, V. 2007. Heart Rate Variability in Mental Stress Aloud. Medical Engineering & Physics 29 (3): 344-349.

Wei, Z., Zhuang, D., Wanyan, X., Liu, C. and Zhuang, H. 2014. A Model for Discrimination and Prediction of Mental Workload of Aircraft Cockpit Display Interface. Chinese Journal of Aeronautics 27 (5): 1070-1077.

Weitzman, D., Fineberg, M., Gade, P. and Compton, G.1979. Proficiency Maintenance and Assessment in an Instrument Flight Simulator. Human Factors 21 (6): 701-710.

Wickens, C. 2008. Multiple Resources and Mental Workload. Human Factors 50 (3): 449-455.

Wickens, C. 2002a. Multiple Resources and Performance Prediction. Theoretical Issues in Ergonomics Science 3 (2): 159-177.

Wickens, C. 2002b. Situation awareness and workload in aviation. Current directions in psychological science 11 (4): 128-133.

Wickens, C. and Huey, B. 1993. Workload Transition: Implications for Individual and Team Performance. National Academies Press, Washington, DC.

Wickens, C. 1991. Processing Resources and Attention. In Damos, D. (Ed.), Multiple-Task Performance: 3-34. Taylor and Francis, London.

Wickens, C. and Liu, Y. 1988. Codes and modalities in multiple resources: a success and a qualification. Human factors 30 (5): 599-616.

Wickens, C., Sandry, D. and Vidulich, M. 1983. Compatibility and resource competition between modalities of input, central processing, and output. Human Factors 25 (2): 227-248.

Wickens, C. (1980). The structure of attentional resources. In Nickerson (Ed.), Attention and Performance VIII. Erlbaum, Hillside, NJ.

Wierwille, W. and Eggemeier, F. 1993. Recommendations for Mental Workload Measurement in a Test and Evaluation Environment. Human Factors 35 (2): 263-281.

Wierwille, W., Rahimi, M. and Casali, J. 1985. Evaluation of 16 Measures of Mental Workload using a Simulated Flight Task Emphasizing Mediational Activity. Human Factors 27 (5): 489-502.

Wierwille, W. and Casali, J. 1983. A Validated Rating Scale for Global Mental Workload Measurement Applications. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 27 (2): 129-133

Wierwille, W. 1979. Physiological measures of aircrew mental workload. Human Factors 21 (5): 575-593.

Williges, R. and Wierwille, W. 1979. Behavioral Measures of Aircrew Mental Workload. Human Factors 21 (5): 549-574.

Wilson, G. and Russell, C. 2003. Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. Human Factors 45 (4): 635-644.

Wilson, G. 2002. An Analysis of Mental Workload in Pilots during Flight using Multiple Psychophysiological Measures. The International Journal of Aviation Psychology 12 (1): 3-18.

Wilson, G. 1993. Air-to-Ground Training Missions: A Psychophysiological Workload Analysis. Ergonomics 36 (9): 1071-1087.

Wilson, G. 1992. Applied use of Cardiac and Respiration Measures: Practical Considerations and Precautions. Biological Psychology 34 (2): 163-178.

Woodruff, R., Smith, J., Fuller, J. and Weyer, D. 1976. Full mission simulation in undergraduate pilot training: An exploratory study. Technical Repoprt AFHRL-TR-76-84. Air Force Human Resources Laboratory, Brooks Air Force Base, TX.

Wu, B., Hou, F., Yao, Z., Niu, J., Huang, W. 2011. Using Physiological Parameters to Evaluate operator's Workload in Manual Controlled Rendezvous and Docking (RVD). In Duffy, V. (Ed.), Digital Human Modeling: 426-435. Springer, Berlin.

Xhyheri, B., Manfrini, O., Mazzolini, M., Pizzi, C. and Bugiardini, R. 2012. Heart Rate Variability Today. Progress in Cardiovascular Diseases 55 (3): 321-331.

Yeh, Y.-Y. and Wickens, C. 1988. Dissociation of Performance and Subjective Measures of Workload. Human Factors 30 (1): 111-120.

Yerkes, R. and Dodson, J. 1908. The Relation of Strength of Stimulus to Rapidity of Habit-formation. Journal of Comparative Neurology and Psychology 18 (5): 459-482.

Ylönen, H., Lyytinen, H., Leino, T., Leppäluoto, J. and Kuronen, P. 1997. Heart Rate Responses to Real and Simulated BA Hawk MK 51 Flight. Aviation, Space and Environmental Medicine 68 (7): 601-605.

Young, M., Brookhuis, K., Wickens, C. and Hancock, P. 2014. State of Science: Mental Workload in Ergonomics. Ergonomics 58 (1): 1-17.

Young, M. and Stanton, N. 2002a. Attention and automation: new perspectives on mental underload and performance. Theoretical Issues in Ergonomics Science 3 (2): 178-194.

Young, M. and Stanton, N. 2002b. It's all relative: defining mental workload in the light of Annett's paper. Ergonomics 45 (14): 1018-1020.

Zakay, D. and Shub, J. 1998. Concurrent Duration Production as a Workload Measure. Ergonomics 41 (8): 1115-1128.

Web References

https://easa.europa.eu/system/files/dfu/AMC%20and%20GM%20to%20Part-FCL.pdf. Accessed 26.08.2015.

APPENDIX 1: WRITTEN, INFORMED CONSENT, STUDIES 1-3

I volunteer to participate in Heikki Mansikka's PhD research and agree to be a subject in Study 1 / 2 / 3 (select appropriate). The studies are aimed at understanding the relationship between the fighter pilots' mental workload and performance during a simulated flying mission. Subject shall be exposed to varying levels of task demands and mental workload. During trials, the mental workload may be extremely high.

Heart rate (HR) and heart rate variation (HRV) will be used as indexes of subject's mental workload. Nexus-10 Mk II biofeedback device will be used to detect subject's HR and HRV. Weapon Tactics and Situation Awareness Trainer will be used for the flying missions. Additional data will be collected using a structured proforma. Historical flight data will be retrieved from LSSJ database.

Subject holds a right to view his/her personal data at any time. Identifiable personal data shall be handled confidentially. Only Heikki Mansikka and those directly involved in data collection may access the identifiable personal data. All results shall be presented in a statistical, unidentifiable form. Final research reports and publications are unclassified and releasable for public. Subject may withdraw from the research at any point without a need to provide a reason. Withdrawal will have no negative impacts on a subject. Until the results are made public, subject may deny the use of any personal data collected during the research. During the research, subject shall be covered by the employer's insurance. This consent is made in two copies.

____________________________	________________________________
Time, Place	Name
____________________________	________________________________
Time, Place	Heikki Mansikka

APPENDIX 2: PROFORMA, STUDIES 1-3

WTSAT 14171A PROFORMA   TEST DATE ________

SUBJECT NAME ________________   SUBJECT ID ________

		Select appropriate answer	
Do you use snuff?		Y	N
If you do, how much snuff do you use in a day?	________________		
Do you smoke?		Y	N
If you do, how much do you smoke in a day?	________________		
Do you use electronic tobacco or other nicotine products?		Y	N
If you do, please specify product and daily consumption.	________________		
Do you use products with high caffeine content (coffee, tee etc)?		Y	N
If you do, please specify product and daily consumption.	________________		
At what time did you go to bed last night?	________________		
At what time did you woke up this morning?	________________		
Was the amount and quality of sleep appropriate?		Y	N
Do you feel hungry of dehydrated?		Y	N
When was the last time you ate (specify hours passed)?	________________		
Are you currently fit to fly?		Y	N
Do you have any limitations for flying duties?		Y	N
Do you use any medication?		Y	N
Have you used high-caffeine or nicotine products during the last 12 hours?		Y	N
If you have, please specify what, when and how much.	________________		

Height	
Weight	
Total flight hours	
F/A-18 flight hours	

APPENDIX 3: ILS PERFORMANCE COLLECTION FORM, STUDY 1

	Clear ILS			15 nm			14 nm			13 nm		
Saved												
Trial#												
	GS	LLZ	AOA	GS	LLZ	AOA	GS	LLZ	AOA	GS	LLZ	AOA
5.0												
4.5												
4.0												
3.5												
3.0												
2.5												
2.0												
1.5												
1.0												
0.5												
0.0												
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00			**0.00**			**0.00**			**0.00**	

	12 nm			11 nm			10 nm			9 nm		
Saved												
Trial#												
	GS	LLZ	AOA	GS	LLZ	AOA	GS	LLZ	AOA	GS	LLZ	AOA
5.0												
4.5												
4.0												
3.5												
3.0												
2.5												
2.0												
1.5												
1.0												
0.5												
0.0												
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00			**0.00**			**0.00**			**0.00**	

	8 nm			7 nm			6 nm			5 nm		
Saved												
Trial#												
	GS	LLZ	AOA	GS	LLZ	AOA	GS	LLZ	AOA	GS	LLZ	AOA
5.0												
4.5												
4.0												
3.5												
3.0												
2.5												
2.0												
1.5												
1.0												
0.5												
0.0												
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00			**0.00**			**0.00**			**0.00**	

APPENDIX 4: INSTRUMENT APPROACH PERFORMANCE COLLECTION FORM, STUDIES 2 AND 3

	ILS (A)				VOR			PAR				ILS (B)		
	GS	LLZ	AOA		GS	LLZ	AOA		KORJ.	AOA		GS	LLZ	AOA
5.0								5.0			5.0			
4.5								4.5			4.5			
4.0								4.0			4.0			
3.5								3.5			3.5			
				8.0										
3.0								3.0			3.0			
				7.0										
2.5								2.5			2.5			
				6.0										
2.0								2.0			2.0			
				5.0										
1.5								1.5			1.5			
				4.1										
1.0								1.0			1.0			
				Minima										
0.5								0.5			0.5			
				Go-Around										
Go-Around											Go-Around			
TOTAL	0.00	0.00	0.00	TOTAL	0.00		0.00		0.00	0.00	TOTAL	0.00	0.00	0.00
AVERAGE		0.00		AVERAGE		0.00			0.00	0.00	AVERAGE		0.00	