UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

MULTIMODAL ANALYSIS OF PILOT'S FATIGUE DURING A MULTI-PHASE FLIGHT MISSION

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

SALEM MAHMOUD NAEERI
Norman, Oklahoma
2020

MULTIMODAL ANALYSIS OF PILOT'S FATIGUE DURING A MULTI-PHASE FLIGHT MISSION

A DISSERTATION APPROVED FOR THE
SCHOOL OF INDUSTRIAL AND SYSTEMS ENGINEERING

BY

Dr. Ziho Kang, Chair

Dr. Robert Terry

Dr. Randa Shehab

Dr. Shivakumar Raman

Dr. Theodore Trafalis

© Copyright by SALEM MAHMOUD NAEERI 2020
All Rights Reserved

To my father, sisters, brothers, and mother.

Acknowledgements

I would express my gratitude to my supervisor, Dr. Ziho Kang for giving me this opportunity and trusting in my ability to complete this research. His guidance ultimately changed my way of thinking. I am sincerely grateful to my committee members, Dr. Randa Shehab, Dr. Shivakumar Raman, Dr.Terry Robert, and Dr.Theodore Trafalis, for their valuable feedback and guidance. I also wish to express my deepest thanks to the staff members of Industrial and Systems Engineering department for their support. I am so grateful to all my friends for their encouragement, support, and good wishes, especially Benmadi. I'm also grateful for my colleague Saptarshi for sharing with me his wisdom and experiences throughout my entire research work. My greatest thanks and appreciation to all my family members; my parents; my lovely wife, Amira, and my sons; Mahmuod; Munder and Yazan, for their entire support on this long endeavor.

Table of Contents

List of Tables

List of Figures

Abstract

One troubling threat to successful flight missions is attributed to fatigue induced and errors. Therefore, discovering effective methods to assess fatigue has been a major topic discussed by professional pilots and aviation experts. Fatigue is a major human factor related issue in aviation and currently subject to increased discussion by aviation administrations and professional pilots. Therefore, effective assessment of fatigue will provide opportunities to reduce the risk of fatigue-induced errors. Currently available subjective measures that assess fatigue can be somewhat affected by external and internal factors, that might cause biased judgment. Therefore, Psychomotor Vigilance Task (PVT), which provides objective measures, can be a viable approach to measure fatigue. In addition, eye movement analysis might augment the fatigue assessment, because eye movement analysis is an unobtrusive approach that does not require direct contact with the participant and can be measured for a long duration. However, it is unknown how eye movement characteristics are correlated with fatigue.

In this research, a multi-modal fatigue measurement framework was developed by combining the PVT analysis with eye movement analysis. In detail, PVT measures (i.e., reaction time, lapses & false starts) and eye movement characteristics (i.e., eye fixation duration, pupil size, number of eye fixations, gaze entropy) were measured to determine pilots' fatigue level under different flight conditions. The results show that significant correlations exist among the eye movement characteristics and the PVTs measures. The proposed multi-modal approach show promise on evaluating pilot fatigue in near real time, which in turn might enable timely recovery interventions.

Chapter 1: Introduction

1.1. Introduction

Although airplanes have become more advanced and reliable, pilot errors have still been one of the main causes of aviation accidents (Avers & Johnson, 2011). Pilot error has long been recognized as a primary factor in air accidents. A recent review on major airline crashes suggested that about 48% of the aviation crashes were attributed to pilot errors, where 15-20% of these errors are associated with pilot fatigue (Li, Baker, Grabowski, & Rebok 2001; Oster, Strong & Zorn, 2013). Fatigue and stress, negative factors associated with human cognition, can severely impact pilots' and potentially impair their ability to safely perform their duties (Dismukes et al., 2015).

Especially, it becomes important to understand the conditions that can impair pilot's ability to safely operate an aircraft and perform safety related tasks and evaluate the impact of fatigue for some scenarios. The Federal Aviation Administration (FAA) defines fatigue as "a condition characterized by increased discomfort with lessened capacity for work, reduced efficiency of accomplishment, loss of power or capacity to respond to stimulation, and is usually accompanied by a feeling of weariness and tiredness," (Salazar, 2007). The International Civil Aviation Organization (ICAO) defines fatigue as "a state of reduced mental or physical ability resulting from extended circadian phase, workload, or wakefulness that can harm a crew member's ability to perform tasks," (ICAO, 2010). Pilot fatigue generally arises from long distance flights and prolonged engagement in operational tasks, long wakefulness, insufficient sleep, and/or sleep disorders (Bourgeois, Carbon, Gounelle, Mollard, & Coblentz, 2003).

Various researchers found that high levels of fatigue severely impact pilots' ability to interpret complex information, detect safety threats, and solve inflight emerging problems by providing timely responses (Dismukes, Goldsmith, & Kochan, 2015; Hartzler, 2014; Lee, & Kim, 2018). The onset of pilot fatigue during aircraft operation is a severe safety concern as it impacts the pilot's ability to stay alert and be attentive (Caldwell et al., 2009).

The National Transportation Safety Board (NTSB) investigations, completed between 2001 and 2012, cited fatigue as one of the possible causes of pilot errors and potential fatalities (Marcus & Rosekind, 2017). The European Cockpit Association (2012) reported that that a fatigued pilot is likely to experience short-term memory loss, longer reaction times, decreased visual perception, poor judgment and to make errors in emergency situations. Many other studies show that fatigue negatively impairs a pilot's capacity to analyze complex processes, to concentrate, solve the emerging problem and make the right decision while flying (Gregory, Winn, Johnson, & Rosekind, 2010; Avers & Johnson, 2011; Hartzler, 2014).

Pilot fatigue becomes one of the foremost concerning factors in flight operations, because exceeding fatigue level could impact the pilot's ability to stay alert and attentive to the demands of controlling a flight safely (Civil Aviation Safety Authority of Australia, 2014; Caldwell & Caldwell, 2016; Velazquez, 2018). The troubling threats of fatigue in the aviation world is that a pilot might be fatigued when they operate the aircraft, fatigue builds up over time, and fatigue can harmfully affect a pilot's ability to perform tasks (Phillips, 2014). Therefore, understanding the conditions that predispose a pilot toward errors/accidents could improve safety, profitability and customer satisfaction in the aviation industry.

1.2. Statement of the Problem

Currently, it is difficult to accurately and objectively assess the pilots' fatigue and how the fatigue level changes over a prolonged period. For example, subjective measures help evaluate how fatigued the person feels, but the results obtained from these measures suffer from the chances of being influenced by biased judgment (Goker, 2018).In addition, Bio-mathematical models help in comparing the merits of work schedules to develop regulations for duty time limitations but cannot definitively answer the question of whether the work schedule is acceptable and safe (Williamson, Lombardi, Folkard, Stutts, Courtney & Connor, 2011).

Unlike these, an objective measure such as Psychomotor Vigilance task (PVT) can detect sensitive changes in the fatigue levels, which allows for an unbiased judgment on the amount of fatigue arising due to different tasks. In addition, non-intrusive eye movement analysis might augment the effective analysis of the pilots' fatigue level; however, currently it is unclear how various eye movement characteristics are correlated with fatigue, especially during a prolonged multi-flight task.

1.3. Research Questions

The recent research has demonstrated that pilot error is associated with a higher prevalence of fatigue. However, several crucial questions remain. This research examines the impact of fatigue on pilots using various factors such as flight conditions, expertise, prolonged multi-flight task. In addition, this research examines whether and how the fatigue level is correlated with eye movement characteristics. This research is organized around the following questions:

Q1. How does flight conditions impact the pilot when the pilot is observing the cockpit interface?

Q2. How does expertise level impact the fatigued pilot?

Q3. How does a prolonged flight mission (based on sequential multi-flight tasks) impact the pilot fatigue level?

Q4. Can we use eye movement characteristics to identify fatigue level?

1.4. Hypotheses

Four main hypotheses are investigated based on my preliminary research:

Hypothesis 1: Extreme flight conditions (severe whether & engine failures) will negatively impact pilot eye-scan behavior (increase the eye fixation duration & decrease the number of fixations)

Hypothesis 2: Effect of fatigue will be higher on novice pilots than expert pilots.

Hypothesis 3: As the number of flights increase (i.e. first, second, third, fourth, and fifth consecutive flights), fatigue level will increase.

Hypothesis 4: Eye movement characteristics (i.e., eye fixation durations, eye fixation numbers & gaze entropy) will be highly correlated with fatigue level.

1.5. Proposed analysis approach

It is important to implement reliable pilot fatigue measures that can be observed regularly to identify potential threats and the impact of fatigue on a pilot. with recent developments, eye tracking technology (being non-intrusive, portable and cheap) has shown to be a powerful tool to observe visual search and information processing strategies (Ziv, 2016). It measures various eye movement characteristics, such as where and for how long observers are focusing at a certain location, and how the observers shift their visual attention among various regions on the display (Mandal, Kang, & Millan, 2016; Naeeri & Kang, 2018; Naeeri et al., 2018, Mandal & Kang, 2018). Eye tracking technologies make it possible to monitor human visual behavior reliably and conveniently without being intrusive in nature (Eckstein, Guerra-Carrillo, Singley & Bunge, 2017; Peibl, Wickens, & Baruah, 2018). Thus, it is less complicated to use them for fatigue measurement purposes.

Glaholt (2014) found a strong correlation between cockpit crewmembers' eye movement attributes (e.g., eye fixation numbers, durations and fixation transitions, pupil size and blink rate) and their cognitive workload. However, the primary metrics alone, e.g. eye fixation number and duration data, only cannot provide us with whole picture of the pilot visual dispersion. Furthermore, blink rate data has been found to be noisy and might not be able to track fatigue accurately (Thomas, Gast, Grube & Craig, 2015).

Eye movements can be further characterized into visual scanpaths (i.e. sequence of eye fixations and saccades) that allow us to investigate the visual entropy (i.e. a single quantitative value that represents the degree of randomness in eye movements; as a result, it is easy to aggregate and compare across many observers and scenarios (Krejtz et al., 2015). Krejtz et al (2015), used gaze transitions entropy measures to compare gaze transitions from viewing different style of painting. Similarly, the gaze entropy was also used by Shiferaw et al. (2018) to predict the lane departure of sleep deprived drivers.

Di Nocera, Camilli, and Terenzi (2007) applied the entropy metrics to observe the pilot visual dispersion over different simulated flight phases. However, the effects of pilots' fatigue on gaze entropy are yet to be explored, especially for long haul flight missions associated with multiple landing and take-off phases. Furthermore, how fatigue impacts pilots with different expertise levels has also not been explored. There are some questions that demand further investigation, including whether the increase in fatigue level follows a linear trend or not and if it follows the same trend among both novices and experts.

To address these issues and in conjunction with the limitations of the various prevalent fatigue measurement approaches, we require a multi-modal approach that is non-intrusive, easier to implement, and easier to analyze. Based on the related work described in the previous sections, our fatigue detection model focuses on two types of measures: (i) PVT measures and (ii) eye tracking measures. The set of measures within PVT were used previously for mental fatigue assessment during cognitive tasks (Roma, Mallis, Hursh, Mead & Nesthus, 2010). The various eye tracking measures have been used for characterizing eye movements in other contextual scenarios, e.g. inferring brain related diseases (Lorist et al., 2000; Hopstaken, Van Der Linden, Bakker & Kompier, 2015). Furthermore, our proposed model also investigates the changes in pilot fatigue level using their gaze entropy values (see Figure 1.1).

To validate our approach, we conducted a moderate fidelity simulated experiment involving a long duration flight operation which consists of multiple takeoff and landing phases, involving two groups of pilots (experts and novices) to analyze how fatigue level affects pilot behavior. Both groups of pilots underwent four standardized prolonged flight missions (flying under instrument flight rules). Subsequently, we studied the relationship between the eye movement characteristics and PVT fatigue measures to investigate whether those measures might show mutually supporting or possibly conflicting evidence in measuring pilots’ fatigue.

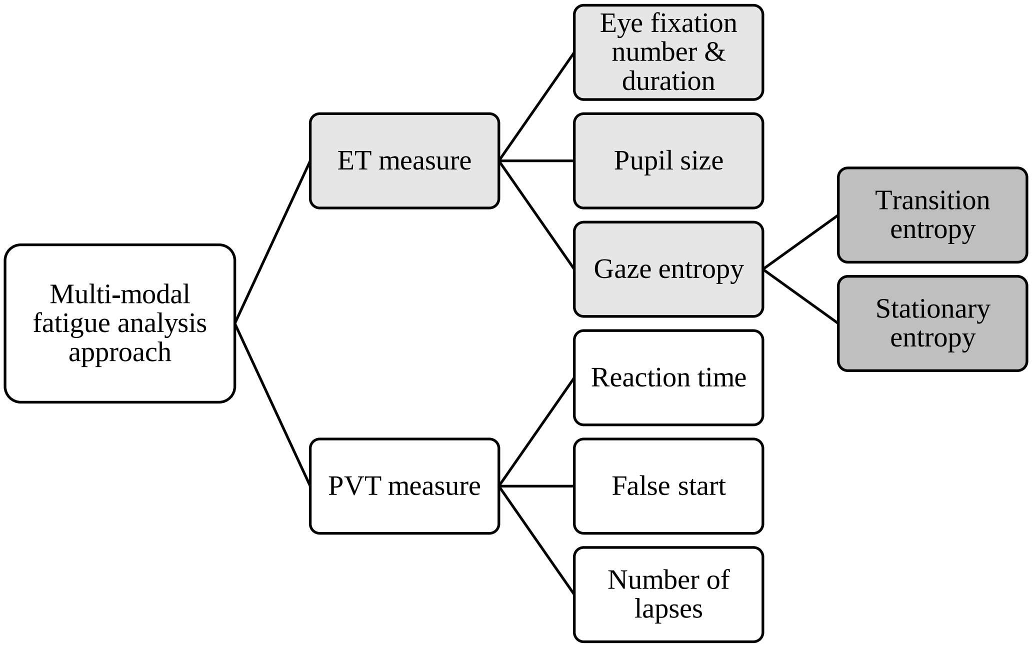


Figure 1.1. The proposed multi-modal fatigue detection framework combining PVT and eye tracking measures.

1.6. Contributions

The main contribution of this research is to develop a multi-modal fatigue measurement framework combining traditional methods and eye tracking measures (see Figure 1.1.). The multi-modal framework integrates psychomotor vigilance task (PVT) metrics and eye tracking metrics to identify physiological indications of fatigue, and then to determine a relationship, if any, between fatigue effects and pilot eye-scan behavior. Moreover, we conducted several studies to examine the effect of fatigue on pilot under different flight scenarios and applied various measures, including 1] eye movement measures such as eye movement network (using the Graph theory), eye fixation durations, eye fixation numbers, pupil sizes, and visual entropy, and 2] PVT measures such as reaction time, false start, and number of lapses, and 3] a subjective measure called the Samn-Perelli Fatigue rating

The contributions are provided in detail within Chapters 3, 4, and 5. A short summary is as follows: In Chapter 3, we were able to discover how the pilots' eye-scan behavior impacted under normal and extreme flight conditions (i.e. under extreme weather condition and engine failures). In Chapter 4, we compared the effect of fatigue on pilot during prolonged flight missions composed of either single or multiple takeoffs and landings. In Chapter 5, the proposed multi-model framework was developed and applied to analyze pilot fatigue in a prolonged flight mission composed of multiple takeoffs and landing which will be called a multi-phase prolonged flight mission). The results from the multiple experiments provided in Chapters 3, 4, and 5 showed that experts showed shorter eye fixation durations, more eye fixation numbers, faster and wider scanning frequencies and wider scan areas compared to those of the novices. In addition, significant correlations were found among the PVT measures and the eye movement measures.

Chapter 2: Literature Review

2.1. Introduction

Human factors approaches service switched from an offensive position instead of a defensive one (see Figure 2.1), based on a sound analysis of needs in the early stages to avoid reactive situations with late damage control (Russ et al., 2013). As aviation machines have evolved in line with the technological advances in computer, the importance of human factors in aviation safety also further increases (Wickens, Gordon, & Liu, 1998). The human factors proactive approach can help prevent system failures and downtime and reduce human error (Bruseberg, 2009; Bris & Soares, 2009).

Human error in aviation has long been recognized one of the main causes of accidents (Baker et al., 2001). Overall, about 70 % of aviation accidents are attributed to human error, mostly pilot error, even though controller error and maintenance error are also sometimes cited (Shappell & Wiegmann, 2004). Beyond safety issues, considerable financial losses in the aviation industry in the form of ramp damage, equipment damaged, flights delayed, and fuel costs are attributed to human error. For example, around 79% of U.S. fatal accidents in 2006 were attributed to pilot error (Krey, 2007; Lacagnina, 2007).

Human errors occur when an operator lose the required information, has limited cognitive abilities or is unable to make the right decisions timely. In such cases, errors may result in a system failure, injury, or accidents, which significantly influences the efficiency and safety, and most of these accidents exhibited a common underlying set of circumstances: a series of human errors, miscommunications or mis-assessment of the situation (Sexton, Thomas & Helmreich, 2000; Shappell et al., 2006).

The aviation system can be made safer by understanding the conditions that predispose an operator to errors/accidents recognizing the potential for error, and by developing systems and strategies to learn from mistakes, to minimize their occurrence and effects (Dhillon, 2009). A modern, scientific approach demonstrates that human errors are not the main cause of accidents, but symptoms of the defects and limitations of the overall system in which they work. They emphasized that causes of errors relate to the cognitive mechanisms (Dismukes, 2017; Perrow, 2011; Rasmussen, 1990). However, studies on human error has been limited mainly to address issues associated with the operational and behavioral acts that contribute to aviation accidents and to classify those acts according to a set of taxonomies (Miranda, 2018). Although this research has helped to outline and classify the patterns of human error, it provides little information about the causes of error. The lack of required information on contributing factors underlying human errors has made it difficult to develop effective intervention strategies.

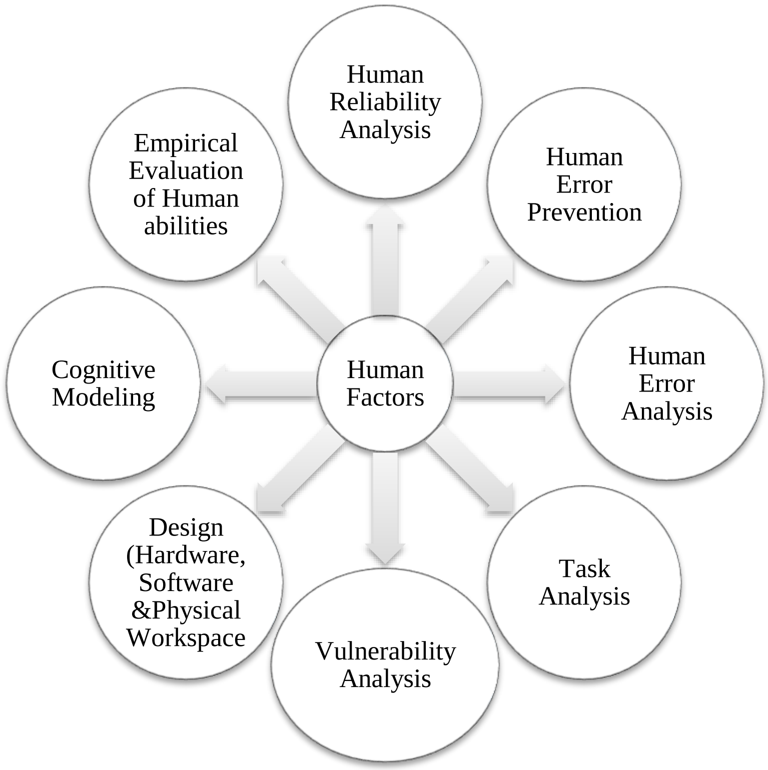


Figure 2.1. Human factors systemic perspective (Wenner, 2011).

2.2. Human factors concepts in aviation

Human factors is a process of re-designing the physical and social environments to better suit the workers' abilities (see Figure 2.2) (Dul, et al., 2012). It is re-engineering the relationship between the worker and machines, procedures and environment about them and with other workers. The better the match between human characteristics and other system components, the better the productivity of the system. The knowledge that human factors offers has applications to manage issues, such as: human error, visual scanning, collision avoidance, and cockpit display design. There are two approaches of managing human error from the human factors perspective. The first approach is to reduce the errors. This can be achieved by improving the human-machine interface. The second approach is to reduce the impact of errors once committed. This can be achieved by establishing error-tolerant systems (Shappell, et al., 2017).

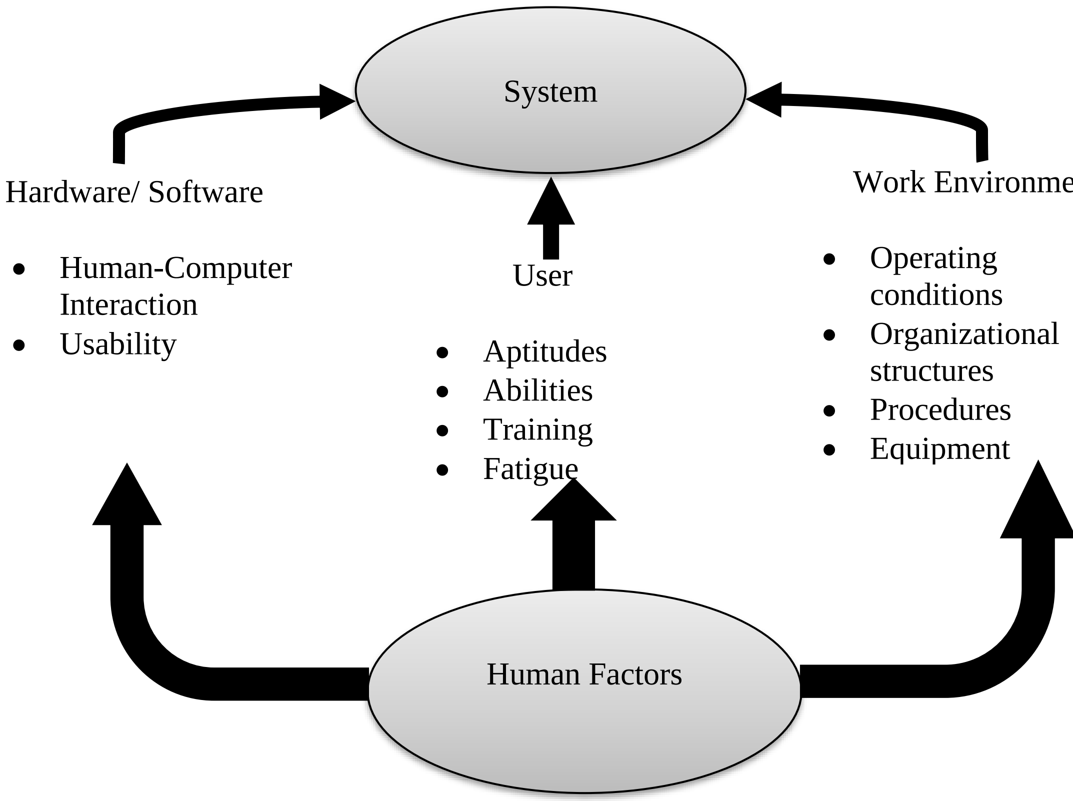


Figure 2.2. Human factors related elements (Dul, et al.,2012).

2.3.Human factors models

Various taxonomies have been developed to classify errors into categories, many of which have been offered quite diverse perspectives. In terms of unsafe factors, Edwards (1972) developed software-hardware-environment-liveware (SHELL) model. Based on the SHELL model philosophy, the discrepancy between liveware and one of the other elements of the model contributes to the human error. In terms of classifying accident and contributing factors, Shappell and Wiegmann. (2004), had introduced the human factors analysis and classification system

(HFACS). The goal of HFACS is to understand the causal factors that contribute to an accident (Wiegmann & Shappell, 2017).

One of the other taxonomies that has recently received considerable attention and has been applied in aviation domain: the accident tree model (Acci Tree), developed for accident modeling and safety defense identification. The philosophy of the accident tree taxonomy drew upon theoretical perspective that regulation and legislation, organizational influence and unsafe supervision interact to produce accidents (Gong, Zhang, Tang & Lu, 2014). Acci Tree describes human error at four layers: regulation and legislation, organizational influence, unsafe supervision, and unsafe acts by operators (see Figure 2.3 below). Each of these four levels is further divided into several categories; for example, regulation and legislation are divided into legislation inefficiency and regulatory failure; organizational influence is divided into organizational influence levels; unsafe supervision is divided into three supervisory failure levels; Y-shaped lyre in turn split into three elements; environment, human, and aircraft respectively. Reaction of a Y-Shaped element can be prompted by multiple events; in the meantime, an event may prompt reactions of multiple elements. These reactions are represented by reaction chains, and the direct process is represented by a coil reaction network.

The starting event of the network is defined as an initial state of related personnel, system, or environment, which leads directly to the first abnormality in the accident direct process. Events in the reaction network may also involve other consequences which are not shown in the accident but might have other potential side risks. This model identifies two patterns of accidents in aviation: human conditions (e.g., mental, physical states, personalities, training factors) and reaction preconditions (e.g., responses of systems) (Salmon, Cornelissen & Trotter, 2012). The concern with taxonomies is that taxonomies do not explain human error and the ways in which errors contribute to accidents, they just provide a method to track of what errors occur, allowing comparison among different types of operation trends (Dekker, 2002; Woods & Cook, 2003).

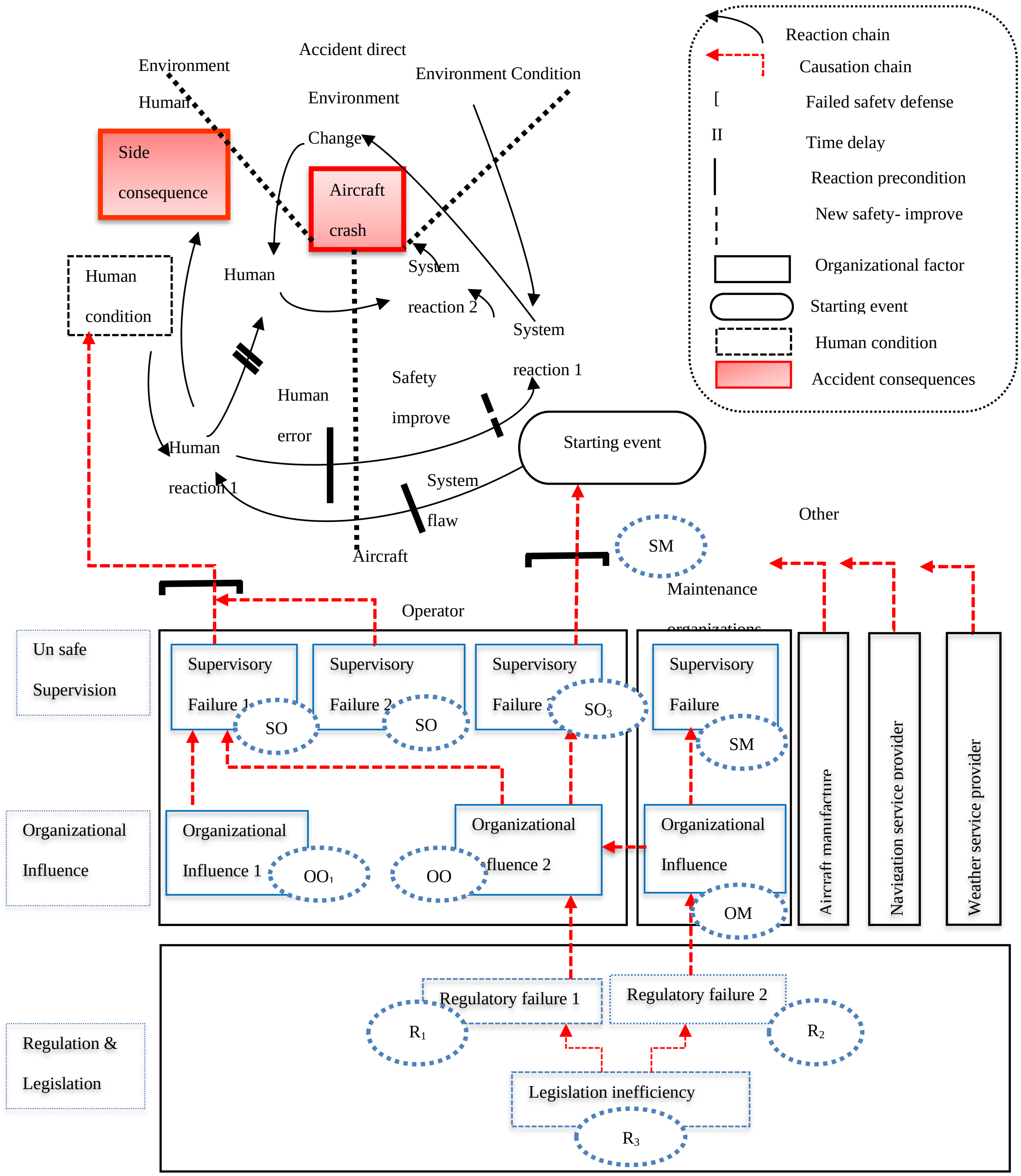


Figure 2.3. Accident tree model (Gong, Zhang, Tang and Lu, 2014).

2.4.Human error classifications

Human error can be loosely classified into two types: an unintentional error and intentional error. An unintentional error includes an error in human action (a slip), opinion, insufficient knowledge (a mistake) or judgment caused by poor reasoning. An example of unintentional error is when someone transposes the number 78 to 87 (Donaldson, Corrigan & Kohn, 2000). An intentional error (violation) is when someone knowingly is deviated from safe practices, procedures, standards, or regulations (see Figure 2.4) (Maurino, Reason, Johnston, & Lee, 2017).

Rasmussen (1983) developed an approach, known as the Skill, Rule, and Knowledge-Based Classification. This approach provides a useful basis for identifying the types of error (see Table 2.1). The pattern of classification provides a useful technique for identifying errors possible to occur in different industrial tasks. The term knowledge-based mode refers to the deliberate control level practiced by the operator over the performed activity. In knowledge-based mode, the worker performs a task in a fully conscious manner. This would occur where a task is new. In this case, the individual would have to apply some mental efforts to evaluate the situation and make the decision. Mistakes occur in this mode. The skill-based mode refers to the physical actions in which there is no conscious thought. Skill-based mode is generally a reaction to an event. Slips occur in this mode. In Table 2.2 the knowledge-based and the skill-based mode distinction is further explained by relating it to the source of error (Embrey, 2005).

Table 2.1. Knowledge- and skill-based behavior features and sources of error distinction (Reason, J, 1990).

Knowledge-based mode	Skill-based mode
Features	**Features**
New to the work environment	Familiar environment
Slow response	Quick response
Requires big efforts	Requires small effort
Requires feedback	Requires little to no feedback
Lacks training	Highly trained
Sources of error	**Sources of error**
Work overload	Inappropriate use of rules
Process variability	Situation changes
Lack of knowledge and awareness of work methods and consequences	Routine action

Table 2.2. Locating the levels in the activity space (Dismukes, 2017).

Situations	Control modes: Mainly conscious	Control modes: Conscious and automatic	Control modes: Mainly automatic
Routine expected			Skill-based
Trained for problems		Rule-based	
New or difficult problems	Knowledge based		

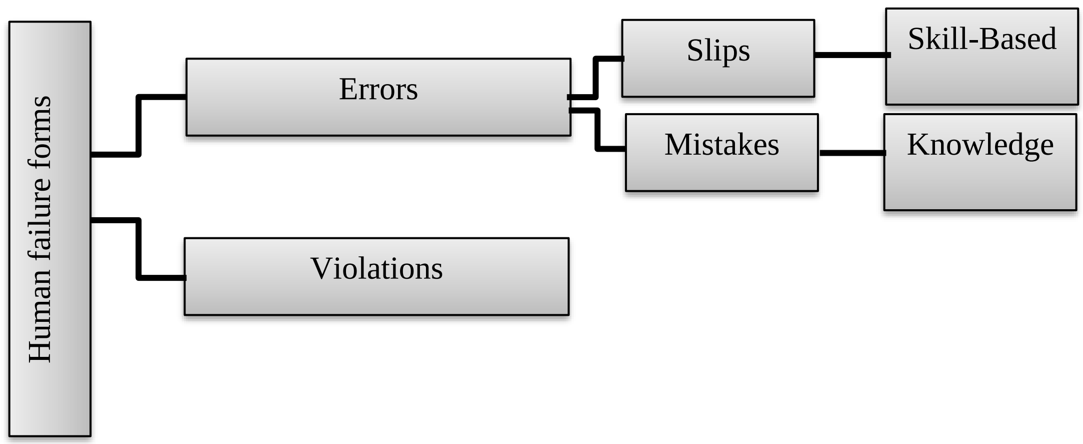


Figure 2.4. Classification of human errors (Maurino, Reason, Johnston, and Lee, 2017).

2.5. Contributory factors and causes of errors

An understanding of human error in the aviation domain, causes and mechanisms, and their relation to different factors, is a critical part of human factors. The Aviation Safety Reporting System (ASRS) reports all civilian aviation crashes using a set of standard forms. SARS factual report provides detailed information about the accident circumstances, causes, factors related to the aircraft: light, flight plan, flight phase and weather. For the purposes of analyzing the most common causes of aviation accidents, we reviewed SARS crashes that occurred during the last 18 years 2001 to 2018. The data included the U.S. fleets of commercial jet airplanes under regulations 14 FAR 121 (major airlines) and 14 FAR 91 (general aviation) (Aviation safety reporting system, 2019).

Accident rates were computed based on the following factors: light, flight plan, flight phase and weather. Since this example included only categories of civil aviation that are operated under regulations 14 FAR 121 and 14 FAR 91, data were analyzed and presented separately for major airline crashes and general aviation crashes. Table 2.3 shows data resulting from an examination of 7048 incidents that were reported over an 18-year period. Years of crashes were categorized based on the interval for each of the two aviation groups, with interval 1 representing the period from 2001-2006, interval 2 representing the period from 2007-2012 and interval 3 representing the period from 2012-2018.

Overall, human error was listed as a probable cause for 64 % of the 7048 Major airlines (FAR part 121) crashes, and 67% of the 1898 general aviation (FAR part 91) crashes. The trend of human error was unsteady over the 18 years period, particularly in major airline operating under FAR part 121 (see Table 2.3). Crashes involving human error in major airline that occurred during 2007-2012 appeared to be less than those occurring during 2013-2018 (56% vs. 67%) (see Table 2.3). Of the major airline crashes involving flying under visual flight rules (VFR), 68% were attributed to human error, compared to 64% crashes involving flying under instrument flight rules (IFR). The prevalence rates of human error related to the light was almost the same for General aviation during the four-light period, while it was higher during dawn period than other periods for major airline. Human error crashes rate was also categorized based on the flight phase (e.g., takeoff, climb, cruise, descent and landing). A higher prevalence of crashes involving human errors was observed in descent and landing phases, as compared with other three phases (e.g., takeoff, climb and cruise phases for major airlines and general aviation (see Table 2.3).

Table 2.3. Crashes rates of human error by aviation category CFR 91 and 121 (Aviation safety reporting system, 2019).

	Major airline operating under FAR part 121			General aviation operating under FAR part 91 and 135		
	Crashes involving human errors			Crashes involving human errors		
	Total crashes	Yes	%	Total crashes	Yes	%
Year of crash						
2001-2006	441	360	0.82	79	54	0.68
2007-2012	2597	1463	0.56	774	565	0.73
2013-2018	4010	2703	0.67	1045	657	0.63
Total	7048	4526	0.64	1898	1276	0.67
Light						
Dawn	278	198	0.71	33	23	0.70
Daylight	4702	2976	0.63	1500	994	0.66
Dusk	355	238	0.67	89	62	0.70
Night	1713	1114	0.65	276	197	0.71
Total	7048	4526	00.64	1898	1276	0.67
Flight Plan						
IFR	7026	4511	0.64	1745	1165	0.67
VFR	22	15	0.68	153	111	0.73
Total	7048	4526	0.64	1898	1276	0.67
Flight Phase						
Takeoff	1428	979	0.69	289	210	0.73
Climb	1073	521	0.49	517	309	0.60
Cruise	926	376	0.41	292	142	0.49
Descent	2774	1983	0.71	627	497	0.79
Landing	847	667	0.79	173	118	0.68
Total	7048	4526	0.64	1898	1276	0.67

2.6. Common causes of human error in aviation

Air travel has proved to be one of the safest means of transportation. However, when an accident occurs, the results are often catastrophic. Major advancements in aircraft design and maintenance have reduced mechanical errors and greatly improved the overall safety of aircraft operation. Thus, more attention needs be given to the factors that have been shown to impact aviation personnel over time (see Figure 2.5) (Shappell & Wiegmann, 2000). There are many contributing factors that were identified by the aviation industry as a straightforward means to frame human error (see Table 2.4). These factors are classified into three groups: organizational, work group, and individual factors (Wiegmann & Shappell, 2017). Understanding the interaction between these factors may help operators to prevent accidents and incidents proactively. This framework was developed by DuPont (1997). DuPont identified the 12 of the most common causes of the human error (see Table 2.4), ignoring any of these factors can lead to a catastrophic accident.

Therefore, improving the human-system interface can mitigate the risks of accidents and incidents and help to specify suitable efforts, generate data, and support the planning related to potential problems. Safety-critical workers in aviation, such as pilot and air traffic controllers, also can degrade flight safety if they are not fully rested. Fatigue takes different forms, including mental and physical fatigue depending on the nature of its causes. In the last couple of years fatigue has been cited as a significant factor in a series of accidents and incidents (Marcus & Rosekind, 2017). Fatigue has different levels and different negative effects. Moreover, the effects of fatigue can vary considerably from one pilot to the other based on the causes (Williamson et al., 2011).

Table 2.4. Common causes of human error in aviation (Dupont, 1997).

Organizational factors	Work group factors	Individual factors
Lack of Communication Lack of Resources Norms	Lack of Teamwork Pressure Complacency	Lack of Knowledge Distraction Fatigue Lack of Assertiveness Stress Lack of Awareness

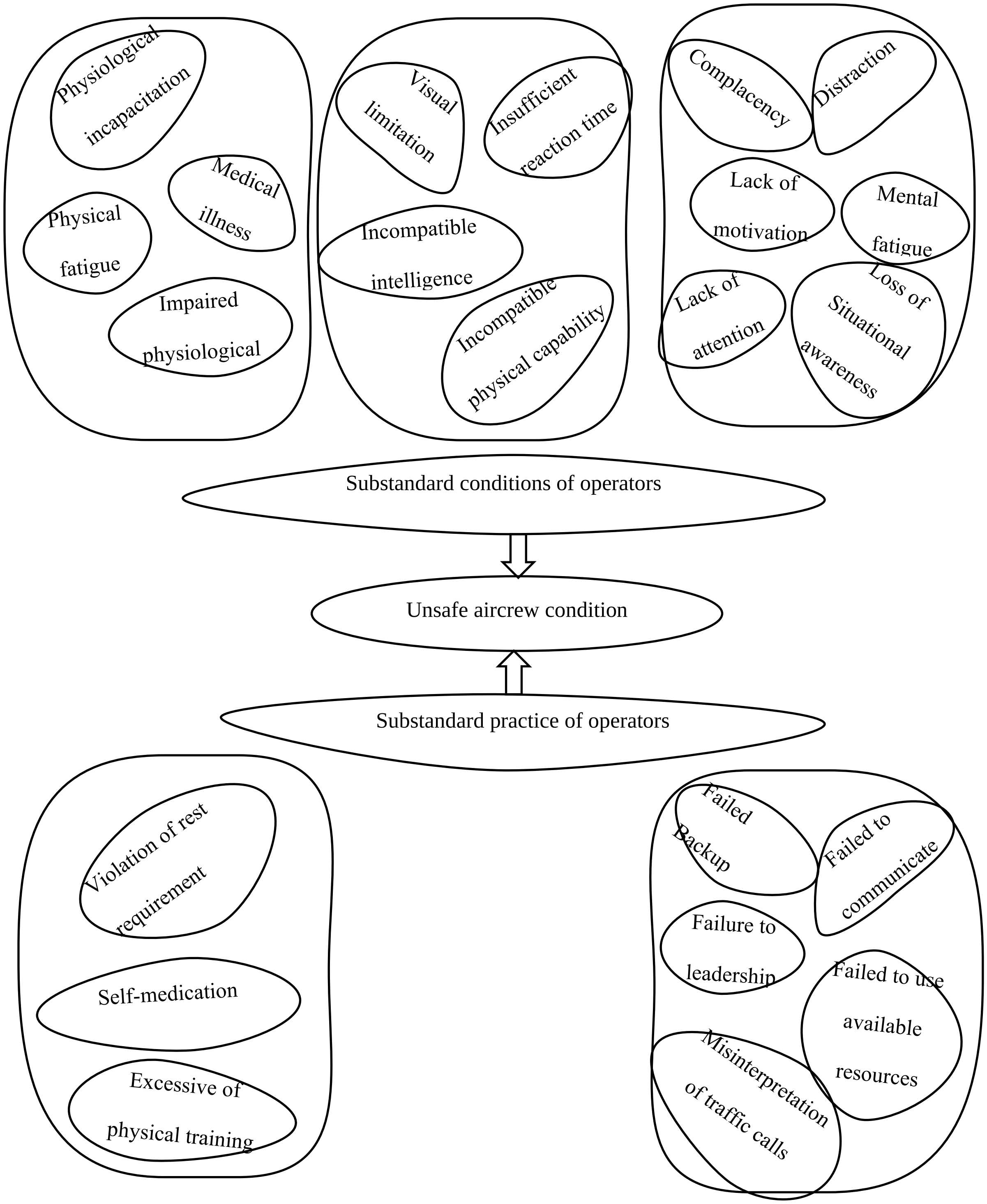


Figure 2.5. Unsafe aircrew conditions (Shappell & Wiegmann, 2000).

2.7. Safety challenges

Safety is basically defined as the system quality that is necessary to ensure that all safety risks and harmful events have been identified, evaluated, mitigated and controlled to an acceptable level (ICAO, 2013). Safety management is a combination of the two perspectives: reactive and proactive. The reactive approach is based on the strategy that includes reporting safety occurrences and making the required adjustments if it deviated from what it was supposed to be (Salas et al., 2010). This approach focuses on meeting the minimum safety limits. The limitation of this approach is that it fails to identify events may cause harm to the system. The proactive approach focuses on a strategy that identifying the safety risks and applying the required actions before they evolve into accidents (Salas et al., 2010).

The proactive approach focuses on patterns and relations across events to identify systemic safety deficiencies and eliminate problem areas. The main characteristics of the proactive approach are that it enables an organization to create a sense of alertness; it invests efforts up front and makes a contingency plan to respond to unsafe acts and conditions (Hollnagel, 2018). In Table 2.5, the reactive and proactive safety management distinction is further explained by relating it to the main characteristics (Hudson, 2001).

Table 2.5. Differences between reactive approach and proactive approach (Oster Jr, Strong, & Zorn, 2013).

Characteristics	Reactive approach	Proactive approach
Philosophy	No advanced action to maintain the system.	Real-time measurements that detect the system deterioration, predict of the failure.
Plan	No contingency plan in place	Contingency plan in place
System downtime	High system downtime	Low system downtime
Signs	People-dependent	System dependent

2.7.1. Aviation safety and legislations

There are regulations to govern the working environments of aviation personnel: the pilots who fly the aircraft, the air traffic controllers who direct and arrange flights and the aviation maintenance engineers who maintain the aircraft and keep flying (International Civil Aviation Organization, 2004). The most important role of the International Civil Aviation Organization (ICAO) is to govern the aviation international and develop standards and regulations that keeps the civil aviation industry safe, efficient, secure, environmentally responsible and economically sustainable (Guimarans, Tomasella, & Wu, 2019).

ICAO focuses on three inter-related areas simultaneously: improving the capability of national civil aviation authorities to address and treat security system deficiencies; developing standards and recommended practices; and promoting their implementation (Abeyratne, 2012). In addition to ICAO there are two large aviation agencies: the European Aviation Safety Association (EASA) in Europe and the Federal Aviation Administration (FAA) in the U.S. The primary responsibilities of the EASA and FAA include monitoring the application of common aviation safety requirements, regulating flight operations, collecting and analyzing safety data and issuing acceptable means of compliance, guidance material and certification specifications (Pierre & Peters, 2009; Stolzer, 2017).

2.7.2. Perceived fatigue in aviation

The importance of fatigue comes from the fact that it may interfere with the high efficiency demand in many work-related tasks. Fatigue may result in work-related problems such as omission of details, poor judgement (Ahsberg, 1998). Fatigue has been classified into three types: mental, physical and total fatigue. Mental fatigue refers to a psychophysiological state that results from increasing cognitive task demands and leads to a perceived sense of reduced alertness, reduced motivation, and decreased mental capabilities. While, physical fatigue refers to a state that impairs the whole body and results in feelings of tiredness and a reduced capacity to generate force or power (Sadeghniiat-Haghighi & Yazdi, 2015). On the other hand, total fatigue refers to a state that arises in workers who are exposed to excessive mental, physical, and emotional demands through their work tasks and schedules, and overtime (Van et al., 2017).

According to Boksem & Tops, 2008, there are many factors that affect the general state of the workers. Mental load refers to the impairment in cognitive ability, reaction time, and reduces a person’s ability to focus and hold attention on the task being performed. Sensory load refers to conditions which put demands on the visual system. Psychological working conditions refer to conditions such presence of a goal, probability of success, and a person's motivation. Physical environment refers to conditions such as heat, cold, noise, and the time of the day when the work is being performed.

Fatigue can be described as: (i) acute Fatigue, which can be due to physical effort, circadian rhythm effects or the lack of quality of sleep before a task, (ii) chronic fatigue, which can be due to sleep debt accumulated over a period more than 48 hours, and (iii) cumulative Fatigue builds up across work and long duty periods (Cheung, Vartanian, Hofer & Bouak, 2010).

2.7.3. Pilot fatigue

Pilot fatigue comes from a variety of causes, ranging from mental or emotional stress, physical exertion and lack of quality sleep, and monotonous tasks (Caldwell, 2012). According to Caldwell et al. (2009), the troubling threat of fatigue in the aviation world is that a pilot might be fatigued when they operate the aircraft. Caldwell et al. (2009) emphasized that the pilots could experience fatigue any time while flying. Therefore, the reduction of pilot fatigue is necessary to improve a pilot's ability to quickly interpret complex information and execute the appropriate response to minimize errors.

The scientific literature shows proof of the negative impact of fatigue on performing safety related activities (Lee & Kim, 2018). Hartzler (2014) found that a pilot is likely to have decreased visual perception, impaired judgment, short-term memory loss, poor decision-making, and increased reaction times, as he gets fatigued. Many other studies show that fatigue affects cognition and leads to impairment of the pilot's ability to detect safety threats, solve the emerging problem and make the right decision while flying.

Even after years of numerous improvements in available countermeasures, the costs associated with fatigue in aviation are still tremendous. However, fatigue is still one of the foremost concerning factors in aviation domain (Caldwell & Caldwell, 2016). The NTSB has identified fatigue as a probable cause of pilot errors (Brown & Whitehurst, 2012). Accident statistics, operational flight and reports from pilots themselves showed that exhaustion and fatigue are a major problem for 90% of the pilots questioned (Lee & Kim, 2018).

The European Cockpit Association (2012) found that most of the 6,000 European commercial airline pilots, who have participated in eight-fatigue surveys between 2010 and 2012, had experienced different levels of fatigue while flying. The surveys' results also showed that on average, more than three out of five pilots have acknowledged having made mistakes due to fatigue and that fatigue played a crucial role in accidents or incidents they were engaged in. A significant majority (68%) of the participating pilots had experienced fatigue at least once (Lee & Kim, 2018).

Gregory et al. (2010) found that more than 84% out of 697 surveyed pilots had examined fatigue while flying the aircraft. Almost half (46%) of the surveyed pilots reported indicated that both their ability to concentrate was degraded. This situation could occur with pilots who are involved in an ongoing physical or mental activity with insufficient rest or rest opportunities that may not coincide with the times that they are most fatigued. Detwiler et al. (2006) found that pilots who do not get enough sleep over a long period or pilots who are involved in an ongoing physical or mental activity with insufficient rest are likely to have decreased visual perception, short-term memory loss, increased reaction times, impaired judgment, poor decision making (Hartzler, 2014; Lee & Kim, 2018).

The findings from this review suggest that fatigue may happen in a relatively short time after some significant physical or mental activity, and it may occur gradually over time. Thus, mitigate the risks of pilot's fatigue require understanding the conditions that can impair pilot's ability to safely operate an aircraft and perform safety related tasks

2.7.4. Pilot error and fatigue

Pilot error has long been recognized as a primary factor in air accidents and incidents, with about 48% of commercial flight accidents being attributed to pilot errors (Oster et al., 2013). Several negative factors associated with human cognition, such as fatigue and stress, can potentially impair their ability to safely perform their duties (Dismukes et al., 2015). An estimated

15 - 20 % of all fatal aviation accidents can be attributed to pilot fatigue (Dillard, Orhan & Letsu-Dake, 2016).

Prolonged wakefulness, insufficient sleep, and/or sleep disorders are the main causes of pilot fatigue. Pilot fatigue can also arise from long distance flights and prolonged engagement in operational tasks (Bourgeois-Bougrine, Carbon, Gounelle, Mollard & Coblentz, 2003). According to Caldwell (2012) pilot fatigue comes from a mental or emotional stress, physical exertion and lack of quality sleep, and monotonous tasks.

Fatigue contributes to aviation accidents since mental or physical capabilities are impaired (Velazquez, 2018). For the average individual, the impact of fatigue is little discomfort, and can be resolved by suspending whatever activity that led it on without any significant implications. However, if that person is engaged in safety-related activities as flying an airplane the impacts of fatigue can be catastrophic (Caldwell, 2005).

2.7.5. Fatigue Measures

It is important to determine which measurement tools could estimate pilot fatigue in a meaningful method. Reliable pilot fatigue measures that can be observed regularly to identify potential threats are needed. There is no single approach to measure fatigue. A wide variety of techniques were developed to measure fatigue (Powell, Spencer, & Petrie, 2014). These techniques are divided into four major categories: (a) self-report or subjective measures (Gander et al., 2013), (b) bio mathematical models (Powell et al., 2014) (c) objective-based measures (Gander et al., 2013) and (d) physiological measures (Li, Lin, Braithwaite & Greaves, 2016; Marandi, Madeleine, Omland, Vuillerme, & Samani, 2018).

Subjective or self-assessment of fatigue (e.g. Samn-Perelli Fatigue rating, Karolinska Sleepiness Scale, 20- item Checklist Individual Strength and Jenkins Sleep Scale) depends on the perception of the individual who is tested. It helps identify how the person's perceived fatigue impact them rather than showing the negative effect of fatigue (Gander et al., 2013).

In addition to the self-reporting, some other measures are generated from self-rating such as quality, amount of sleep and work schedule (i.e., biomathematical models). Bio mathematical models use precise inputs, such as timing and duration of actual sleep, duty shifts, etc. Bio-mathematical measures include Boeing Alertness Model (BAM), Circadian Alertness Simulator (CAS), Fatigue Risk Index (FRI), Fatigue Assessment Tool by Inter-dynamics, and "Sleep, activity, fatigue, and task effectiveness" (SAFTE). Bio-mathematical models help in comparing the merits of work schedules to develop regulations for duty time limitations but cannot definitively answer the question of whether the work schedule is acceptable and safe (Williamson et al., 2011; Civil Aviation Safety Authority of Australia, 2014).

Objective measures of fatigue include Electroencephalogram (EEG), Electrooculogram (EOG), psychomotor vigilance tasks (PVT), and eye movement measures. Objective measures are more sensitive to changes in fatigue levels and allow for unbiased interpretation of the data. Objective tools have specific aspects to assess fatigue levels in a more practical way. Objective techniques can identify the capacity of the worker to achieve his/her task in a more precise way, using quantitative indicators, such as reaction time and number of lapses (Dinges, Mallis, Maislin & Powell, 1998).

2.7.5.1. Self-report or subjective measures

Subjective measures are classified into: Unidimensional and multidimensional measures of fatigue. Unidimensional measures such as visual analog scales (VAS) can only measure one dimension of fatigue. The VAS is a line that is anchored at both ends with words that are descriptive of the extreme level (e.g., "No fatigue" to "Total exhaustion"). Participants are asked to draw a mark on the line to indicate the level of the fatigue, thereby providing an objective representation of fatigue level. Multidimensional assessment of fatigue (MAF) considers all dimensions of fatigue: physical fatigue, activity, motivation, and mental fatigue. An example of MAF is Swedish occupational fatigue inventory (SOFI), which can measure all dimensions of work-related perceived fatigue (Ahsberg, 1998). There are many specific self-assessment questionnaires designed to measure cognitive abilities. The following methods such as chalder fatigue questionnaire, checklist individual strength fatigue severity scale have been shown to have good assessment capabilities.

2.7.5.1.1. Chalder fatigue questionnaire (CFQ)

The CFQ was developed to assess fatigue severity in a community population. The CFQ is a scale rather than a measurement of fatigue impact (Jing et al., 2016). The scale includes eleven items to produce fatigue score. The response scale for each item has four options ("Less than usual," "No more than usual," "More than usual," and "Much more than usual"). The interpretation of this scale is that the higher scores obtained, the greater the severity of fatigue. CFQ has shown high sensitivity to change in physical and mental fatigue level (Jing et al., 2016).

2.7.5.1.2. Checklist individual strength (CIS)

The CIS was developed to measure several dimensions of fatigue in chronic work environments. CIS includes four subscales: fatigue experience subscale with eight items, concentration subscale with five items, motivation subscale with four items, and physical activity subscale with three items. The scale of response options for each item ranges from, "Yes that is true," to, "No that is not true." The interpretation of this scale is that the higher the scores of subjects, the higher the fatigue level (Beurskens et al., 2000).

2.7.5.1.3. Fatigue severity scale (FSS)

The FSS was developed to assess all dimensions of fatigue (total fatigue), including physical, social, and cognitive. The response options for each of the nine items ranges from “Strongly disagree” to “Strongly agree.” Total scores of the nine items are summed to produce global scores, where higher scores reflect greater fatigue level (Valko, Bassetti, Bloch, Held, & Baumann, 2008).

2.7.5.1.4. Functional assessment chronic illness therapy (FATIGUE) (FACIT-F)

The FACIT-F was developed to assess chronic illnesses. This method covers four domains: physical fatigue, functional fatigue, emotional fatigue, and social consequences of fatigue. FACIT-F scale consists of thirteen items. The response options for each item scores between zero and four; where zero refers to “Not at all” and four refers to “Very much.” The result is summed and multiplied by thirteen, and then divided by the number of items answered. The FACIT-F method needs at least seven items to be answered to acquire clear results. The interpretation of the result is that the higher scores the subject obtains, the greater the fatigue level (Webster, Cella, & Yost, 2003).

2.7.5.1.5. Multi-dimensional assessment of fatigue (MAF)

The MAF was developed to measure multiple dimensions of fatigue. MAF consists of sixteen elements. The response options for each element provides global scores that contribute to global fatigue index (GFI). Each element ranges from one score where no fatigue to 50 scores where the individual is fully fatigued. Item sixteen scores range from one (no fatigue) to four (fully fatigued). Higher scores represent greater fatigue severity. Time to complete is likely to be ten minutes (Neuberger, 2003).

2.7.5.1.6. Multi-dimensional fatigue inventory (MFI)

The MFI was developed to measure fatigue using a multidimensional short questionnaire. MFI comprises all fatigue dimensions: mental fatigue, physical fatigue, activity and motivation. MFI consists of five subscales of four items each. Score options of each item ranges from zero (yes that is true) to twenty (no that is not true). Subscale items scores are summed to obtain the final scores for relating dimension. The overall scores reflect the total fatigue level. Twenty is the highest score which reflects greater severity of fatigue. Time to complete and report scores is about 10 minutes (Smets, Garssen, Cull, & De Haes, 1996).

2.7.5.1.7. Visual analog scales (VAS)

VAS is a unidimensional fatigue measure that was developed to measure the severity or intensity of fatigue. Typical fatigue of VAS comprises of a ten-cm horizontal line, anchored by 2 statements on the ends of the horizontal line ranging from “No fatigue” to “Total exhaustion.” Fatigue scores are calculated based on the distance from the left-hand anchor to the respondent’s mark on the line. Ten is the higher score on this scale and represents a severe fatigue. The VAS scale takes less than one minute to complete (Wewers, Rachfal, & Ahijevych, 1990).

2.7.5.1.8. Swedish occupational fatigue inventory (SOFI)

The Swedish occupational fatigue inventory (SOFI) is a multi-dimensional technique that was developed to measure all dimensions of work-related perceived fatigue. SOFI is categorized into five dimensions. Physical exertion and physical discomfort dimensions are considered as physical factors, while lack of motivation, Lack of energy and sleepiness are considered as a mental factor. Figure 2.6 shows the relationship between the work domain and the perceived fatigue: Physical fatigue is described as lack of energy, physical exertion, and physical discomfort. Mental fatigue after mental work is described as sleepiness, lack of energy, and lack of motivation.

Fatigue after shift work is described as sleepiness, lack of energy, and lack of motivation. The SOFI consists of 25 items, and each one assesses by numbers ranging from zero (not at all) to ten (very high degree). The higher the score, the greater the fatigue level (Ahsberg, 1998).

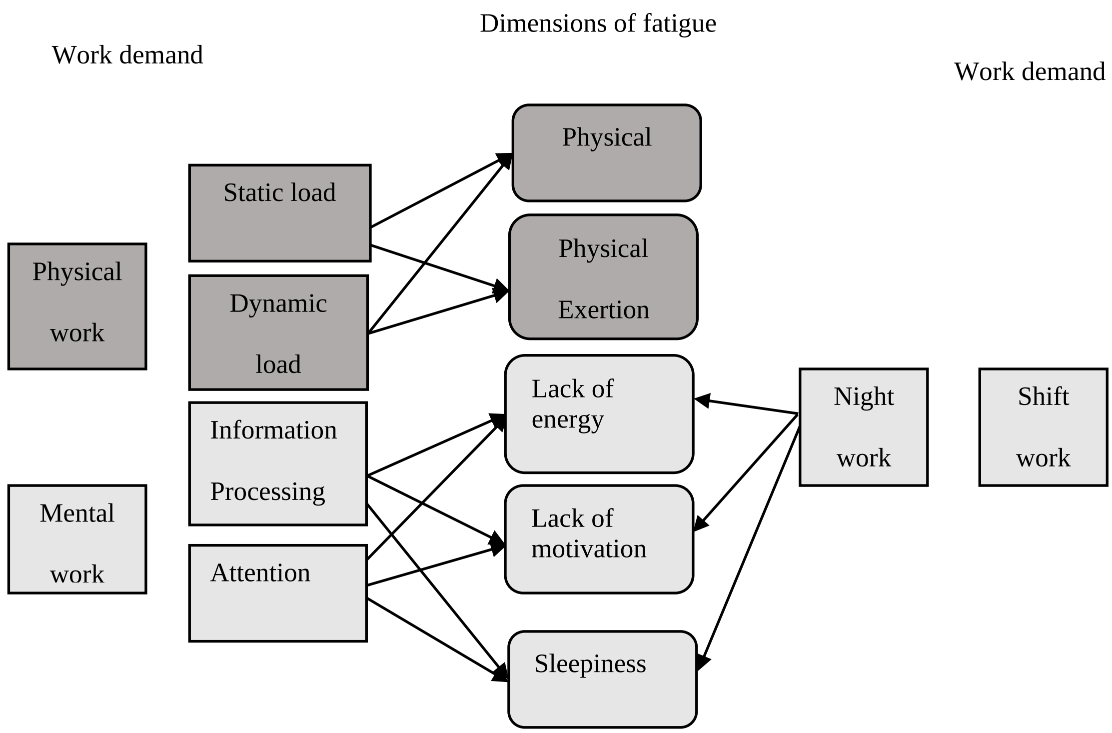


Figure 2.6. The relations between workload and perceived fatigue (Ahsberg, 1998).

2.7.5.2. Bio mathematical models

A bio mathematical model is defined as set of equations that are used to quantitatively predict human capabilities based on known sets of inputs into the system (e.g., on duty, off duty, and rest length, etc.) (Peng & Bouak, 2015). The bio mathematical models can help in developing the best countermeasures to improve human capabilities when interacting with the system interface (Civil Aviation Authority of Australia, 2014). In the aviation domain, bio mathematical models are mainly used to predict the average level of crew member fatigue that are expected during an operation (Stokes & Kite, 2017).

The purpose of these models is to provide fatigue risk metrics based on a set of inputs: circadian factors, shift lengths, time of shift, and previous duty periods. Bio-mathematical is a useful tool for assessing safety of work schedules and rosters. Most of these models are two-process models (Borbely, 1982). Two-process models involve: a homeostatic process which explains the degradation of alertness with time awake and its recovery with time asleep; a circadian process which explains the circadian variation in alertness (Goel, Basner, Rao & Dinges, 2013). In the those models the sleep homeostat and the circadian process interact only at distinct events. The Homeostatic pressure (known as process sleep (S)) which explains the downturn of alertness with time awake and recovery with time asleep. Wake-up happens when S reaches the lower threshold and sleep happens when the homeostatic Process (S) reaches the high threshold.

The Circadian clock (known as process (C)) describes the changes in the level of fatigue and the tendency to sleep during day and night times. Process (C) functions independently of the time awake programed by the homeostatic process (S). In aviation, process (C) is highly influenced by external factors (e.g., time zone, light, dark) (Goel et al., 2013). Akerstedt, Folkard, & Portin, (2004) developed three processes model to include sleep inertia as a third process. Sleep inertia (known as process waking (w)), refers to the temporary drop in alertness, and increase in sleep tendency after waking up from deep sleep.

Bio-mathematical fatigue models have some limitations, a major limitation of bio-mathematical fatigue models is that they may underestimate risk, simply because they predict risk probabilities for a population average rather than instantaneous fatigue levels of a specific individual; prediction is often based on criteria that exclude other factors rather than definitively identifying the contribution of fatigue (Civil Aviation Safety Authority of Australia (CASA), 2010). Williamson et al. (2011) demonstrated that bio-mathematical models can be used effectively to compare the relative merits of two or more work schedules but cannot definitively answer the question as to whether a work schedule is acceptable or safe. The most frequently used bio mathematical models are explained below.

2.7.5.2.1. Sleep, activity, fatigue, and task effectiveness (SAFTE)

The SAFTE model has been developed to mimic the underlying physiological causes which affect people's cognitive abailities (Gundel, Marsalek & Thoren, 2007). SAFTE is widely used in aviation sector to predict and administrate the expected fatigue risk. Based on the prediction results, a set of measures could be implemented proactively to mitigate excessive fatigue risk. The outputs of this model include metrics (e.g., response speed, number of lapses, reaction time), and sleep-wake metrics (e.g., circadian phase and sleep reservoir) (Gundel et al., 2007).

2.7.5.2.2. Sleep and wake predictor (SWP)

The Sleep and wake predictor model were developed to predict alertness by determining the level of sleepiness that is associated with changes in circadian rhythms and time awake or asleep (Gundel et al., 2007). SWP model is mainly used to evaluate the effects of work schedules on alertness. SWP is widely applied in navy, aviation, rail, and nuclear power sectors. These sectors apply the SWP to help schedulers and planning team assessing the effects of a given work schedule on human capabilities (Gundel et al., 2007).

2.7.5.2.3. Fatigue audit inter-dyne (FAID)

The FAID model was developed to estimate work-related fatigue that is associated with hours of work. The FAID model provides fatigue risk metrics based the following set of inputs: circadian factors, shift lengths, time of shift, and previous duty periods (Mallis, Mejdal, Nguyen, & Dinges, 2004). This model is applied to analyze the fatigue impact factors, such as time zone changes, start, and finish times. Applying FAID can assist in defining the highest risk times within shifts and supporting risk-based decisions.

2.7.5.2.4. Fatigue risk index (FRI)

The Fatigue and Risk Index (FRI) model was designed to compare different work schedules. The main applications of FRI are to compare work schedules and examine the impact of changes in each work schedule based on worker capabilities (Civil Aviation Authority of Australia, 2014). In addition, FRI can be used to identify the impact of fatigue or risk results from any change within a given work schedule. This model provides two metrics; the fatigue Index and the risk Index. The fatigue index provides an estimation of average probability outputs that are multiplied by one hundred while the risk index provides an estimation of the relative risk of making an error that could lead to an incident or accident (Greubel & Nachreiner, 2013).

2.7.5.2.5. System for aircrew fatigue evaluation (SAFE)

SAFE was designed by the UK Civil Aviation Authority for the aviation industry. SAFE is broadly applied in aviation domain to predict the expected fatigue and sleep patterns experienced by pilots for a given work schedule. Simply, SAFE produces Samn Perelli (SP) fatigue scores every fifteen minutes. Based on these predictions, the risks associated with duty schedules can be assessed and eliminated (Mellor, 2012).

2.7.5.2.6. Bio mathematical model's comparison

Table 2.6 shows detailed comparisons between the above discussed five models. The table summaries advantages, applications, and the outputs of the five bio mathematical models. The comparisons show that each model has its strengths and weaknesses. The lowest powerful model is the SWP because it does not have the enough capabilities to deal with large scale applications.

It also does not consider sleep data and workload data in its predictions. Table 2.6 shows the advantages of these models.

Table 2.6. Key characteristics of biomathematical models (Mallis, Mejdal, Nguyen & Dinges, 2004).

Model	Advantages	Applications		The output
		Consider actual sleep data	Consider workload	
SAFTE	Large-scale all industrial settings and transport Model outputs (work related and cognitive fatigue)	Yes	Yes	Estimated sleep / wake times Fatigue-related risk of operational
SWP	Few inputs to make predictions. Small-scale	Not	Not	Estimated sleep / wake times Fatigue-related risk of operational
FAID	Large-scale Provides KPIs	Yes	Yes	Estimated sleep / wake times Fatigue-related risk of operational
FRI	Estimates of both fatigue and risk. Requires few inputs to predict	Not	Yes	Subjective alertness Fatigue-related risk of operational
SAFE	Provides rapid assessment Effective in all types of aviation operations	Yes	Yes	Subjective alertness Estimated sleep / wake times Fatigue-related risk of operational

2.7.5.3. Objective measures

The widely used tool for this purpose is the psychomotor vigilance task (PVT). In addition, eye tracking technology is one of the tools to assess human capabilities based on eye activity. Eye tracking technology is used to monitor, measure, and quantify eye activities in real time. In addition, an electrooculography (EOG) measures the electrooculogram signal differences between the human front of the eye “cornea” and the back of the eye “retina”. In addition, electroencephalography (EEG) uses neuronal information to represent mental fatigue because mental fatigue can be strongly correlated with brain activities and is widely used in the healthcare sector.

Other commonly utilized physiological measures in fatigue studies include peripheral nervous system (i.e., visual activity) and the central nervous system (i.e., brain activity) (Wilson, Caldwell, & Russell, 2007). Physiological measures of fatigue assume that physiological responses correspond to the processes activated.

The advantage of the physiological methods is that they do not require any explicit response from the subject and allow fatigue to be assessed directly and continuously (Borghini, Astolfi, Vecchiato, Mattia, & Babiloni, 2014). In the last decades, various physiological fatigue detection methods have been proposed. The most frequently used physiological method for the assessment of fatigue are the EEG, EOG, and eye tracker system. The EEG provides an objective, functional mapping of brain activity on the scale of seconds (Gao et al., 2014). The EOG provides the percentage of eye closure (PERCLOS) as a measure of fatigue (Golz et al., 2010). The eye tracking system provides a natural and efficient way to observe visual search and information processing (Rayner, 2009).

A major limitation of EOG measure is that the existing methods for calculating Per-close are not robust in practical applications due to the complexity of eye detection. Per-close is usually calculated with a 60-second window, which makes it not suitable for real-time flight fatigue detection (Muller, Wendt, Kollmeier, & Brand, 2016).

The EEG provides an objective, functional mapping of brain activity on the scale of seconds. Some studies have established the sensitivity of EEG activity to stressors in the drive.

However, using EEG data to assess fatigue has shown some limitations: 1] EEG theta increases and EEG alpha decreases as a function of heightened flying demands, and 2] the collection and analysis method of data are somewhat difficult (Cao, Wan, Wong, da Cruz & Hu, 2014; Müller, Wendt, Kollmeier, & Brand, 2016).

2.7.5.3.1. The Psychomotor Vigilance task (PVT)

PVT is used to detect fatigue and to assess the human's behavioral vigilance (Lamond et al., 2008). PVT is a visual test which is used to assess individuals' reaction times to the visual stimuli at random times (e.g. a red dot against a black background screen) (Basner, Mollicone & Dinges, 2011). PVT is a simple test where participants need only to press the keyboard space bar key as soon as a red dot appears on the screen. The purpose of this test is to measure the participants' attention to provide numerical measurements of participant's attention. The main advantages of the use of PVT is that PVT does not require training and the test time is short (either five or ten minutes) (Lamond et al., 2008). The outcome measurements of PVT, that are highly sensitive to fatigue are mean reaction time (RT), median reaction time, false starts (responding with no visual stimulus is presented on screen), and number of lapses (failure of the subject to response within 500 MS) (Basner et al., 2011).

2.7.5.3.2. Electroencephalography (EEG) method

The EEG method has been widely applied in the healthcare domains among others. The EEG uses neuronal information to represent mental fatigue because mental fatigue is strongly correlated with brain activities. In other words, electroencephalography measures the voltage differences between different parts of the brain (Farnsworth, 2018). It uses electrodes placed on the scalp to record brain waves. The EEG has capabilities to detect changes in electrical activity in the brain on a millisecond-level (Farnsworth, 2018). The EEG mechanism uses four frequency bands to analyze the collected data about the brain waves. Table 2.7 shows that bands shift from fast and low amplitude waves once person becomes fatigued. More specifically, the alpha power peak frequency decreases with mental workload and increases when resting or relaxing. Beta frequency band activity reflects emotional and cognitive processes. Delta and theta increase largely with increasing mental fatigue. Alpha and theta power could provide an appropriate index to the level of fatigue (Boksem, Meijman, & Lorist, 2005).

Table 2.7. EEG records and pattern of brain electrical activities (Boksem, Meijman, & Lorist, 2005).

Rhythm	Frequency (Hz)	Pattern
Beta(β)	14 - 30	Excitement
Alpha(α)	8 – 13	Relaxed Awake
Theta(θ)	5 – 7	Drowsy, Sleep
Delta(δ)	2 – 4	Normal Sleep Rhythm

2.7.5.3.3. Electrooculography (EOG) method

The EOG measures the electrooculogram potential signal differences between the human front of the eye "cornea" and the back of the eye "retina" (see Figure 2.7) (Muller et al., 2016). In general, the measurements of electrooculography are similar to the eye movement recordings system. However, the EOG work mechanism is quite different. EOG is based on measuring the metabolic activity potential differences that exists between the front cornea and the back of the human eye retina (Deng, Hsu, Lin, Tuan, & Chang, 2010). Figure 2.7 shows that the electrooculography uses five electrodes to collect the signals: one electrode places on the nose; the second directly across from this on the edge of the right eye socket; the third electrode places above the right eyebrow; the fourth on the lower edge of the right eye socket; and the fifth electrode is a signal reference placed in the middle of the forehead. These electrodes, which are located above and below the eye (left or right), are used to measure vertical eye movement. The other two electrodes are located on the right and the left of the head to measure horizontal eye movement.

The electrical signal from the electrodes is amplified using direct-current to provide more reliable measurement of eye position during visual fixation (Zhang, Gao, Zhu, Zheng, & Lu, 2015). When eyes move from the center position towards one of the two electrodes (left & right), one electrode captures the data from positive side of the retina and the opposite electrode captures the data from the negative side of the retina (Zhang et al., 2015). The gaze is measured by recording the potential difference between electrical field generated by the movements the eyes. The signal from vertically placed electrodes record vertical eye-movement (eyelid movement, muscular activity near the eyebrow, and eye blinks) (Zhang et al., 2015).

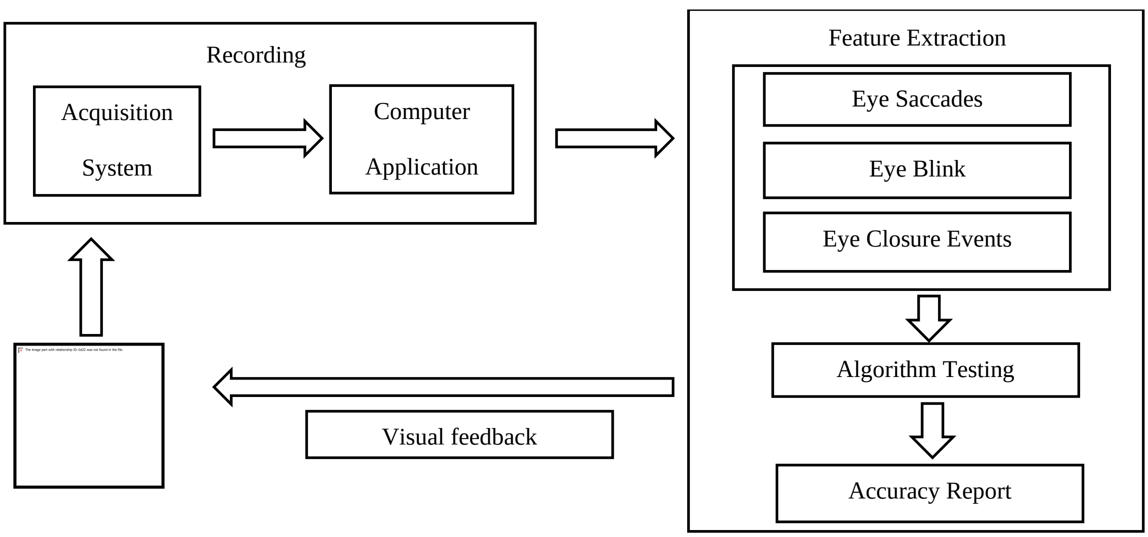


Figure 2.7. System architecture and EOG (Postelnicu, Girbacia, & Talaba, 2012)

2.7.5.3.4. Eye tracking method

Eye tracking technology is an advanced tool in human behavior research. It is used in different research areas to monitor, measure, and quantify eye activities in real time (Blascheck et al., 2014). Eye tracking technology utilizes two main components: light sources and camera. The light source is directed to cornea through the eye pupil to cause light reflections, while the camera embedded in the computer, tracks and records the reflections (Al-Rahayfeh & Faezipour, 2013). Eye tracking uses either a remote or head-mounted device connected to a computer. Remote eye tracker is mounted below computer screen to record eye activity within certain distance. Head mounted eye tracker uses eye glass frame to record the eye movement. Both techniques can deliver high accuracy data if they are calibrated properly (Richardson and Spivey, 2004). The eye tracking technology makes it possible to monitor human cognition reliably and comfortably (Eckstein et al., 2017; Lupu & Ungureanu, 2013). The eye movement terms include:

Eye fixation: Maintaining of the visual gaze on location of a stimulus for a period (150-300 MS). The relation between cognitive processes and eye fixations depends on two basic assumptions: the eye-mind assumption holds that; an individual pays her/his attention to the until she comprehends it. It directly reflects the complexity of the underlying cognitive processes: the immediacy assumption holds that, as soon as an individual fixate on a stimulus tries to interpret it. Saccade: rapid and continuous eye movements among fixations (between 40-100 MS) to visually scan the scene. Areas of interest: a defined area on the tested field of view which you can select in order to calculate detailed statistics and measurements on this area. The location and the size of area of interest depends on the purpose of the task being tested. Gaze plot: a map of the eyesight over the course of the tested areas. It reveals information about the time sequence of looking. Scan path: A series of fixations and saccades that represents a participant's order and location of each fixation (pattern of eye movements) (Poole & Ball, 2005). Heatmap: represents the distribution of eye fixations and gaze points on the areas of interest. Opacity map: the most attractive areas in course of the test are lighted up while less attractive areas are dimmed down.

One of the benefits of eye tracking based methods is that they are grounded in data that relates directly to cognitive processing for how humans understand the visualizations across task contexts, rather than subjective methods where the analysis depends on human willingness and competency to describe how they feel when they are exposed to an event (Reis, Mestre, Canhão, Gradwell, & Paiva, 2016). Most recently, part of the scientific research put forward has a focus on eye movement of the human. The eye tracking system provides an accurate, robust and objective method for measuring fatigue, and many eye movement parameters related to fatigue can be recorded in real-time. Eye movement techniques have been widely applied in many applications such as human attention analysis, human emotional state analysis, interactive user interfaces, and human factor domain (Al-Rahayfeh & Faezipour, 2013). There are many approaches to implementing eye detection and tracking systems as a robust method in a wide range of applications (Hansen & Ji, 2009). Kang and Landry (2014) have applied an eye tracking method to understand and characterize the ATCs' eye tracking data. Additionally, Mandal, Kang, and Millan (2016) have developed a new methodology to represent the eye movement data of air traffic controllers. They used the DWN to visualize the eye movement data of ATCs. Blascheck et al. (2014) have classified eye-tracking techniques into two approaches: scan path (order of eye fixations) based and areas of interest (AOI) based.

The eye tracking information provides many parameters, possibly related to human fatigue, such as fixation duration and fixation number. Through the DWN visualization techniques, we would be able to obtain information about pilot cognitive processes associated with the eye-tracking data (i.e., how often attention changed between two AOIs). Utilizing the advantages of both measures could help us: (a) predict the impact of pilot fatigue over the course of the flight and (b) better model fatigue states of pilot under different flight situation.

I. Mechanism of Visual Search

Visual search is an unsystematic process involves an active scan of the visual environment searching for a specific target among a complex array of stimuli (Wolfe, Alvarez, & Horowitz, 2000). Visual search can occur with or without eye movements. The efficiency of visual search depends on the reaction time to presented stimuli. The search is efficient if the attention is directed to the most activated target. There are two ways in which attention can be directed: top-down way or user driven; bottom-up or stimulus driven. The efficiency of visual search is proportional to the reaction time of identifying stimuli. Reder, Weber, Shang, & Vanyukov, (2003) have demonstrated that the visual search efficiency improves with expertise, requiring less search time and fewer visual scans to find the target.

II. visual scanning

Visual scanning is defined as a number of fixated points that occur from the time a search stimulus is presented until detects and recognizes it (Foulsham & Underwood, 2008). It is the individual ability to efficiently, quickly, and actively use vision to search in a systematic manner, such as left to right and top to bottom. The mechanism of visual scan includes a series of interrelated processes: Focused Attention, refers to the ability of the operator to focus her/his attention on a stimulus; Selective attention, refers to the ability to pay attention to a single stimulus when there are other stimuli present; cognitive shifting, refers to the brain's ability to adapt behavior and thoughts to new, changing, or emerged events. In other words, shifting is the ability to make the appropriate changes to adapt to new situations; Visual Perception, refers to the ability to distinguish, identify, and interpret, lights, shapes and colors. Recognition: refers to the ability to compare the visual information one receives to determine whether he has prior experience with this information or not. Visual scanning is an important skill for a pilot to efficiently carry out several different tasks. A pilot needs to use visual scanning to avoid obstacles when navigating their cockpit display.

III. Eye-Tracking Measures

Eye trackers help evaluate a subject visual attention by recording eye movements (Rayner, 1998; Duchowski,2007), which show where a subject being tested is looking (see Figure 2.8). The eye tracking system provides many different metrics including visualizations (i.e., heatmaps, gaze plots) and more comprehensive statistical measures (see Table 2.8), which gained acceptance and began to be used as reliable measures of the fatigue (Wilson et al., 2007).

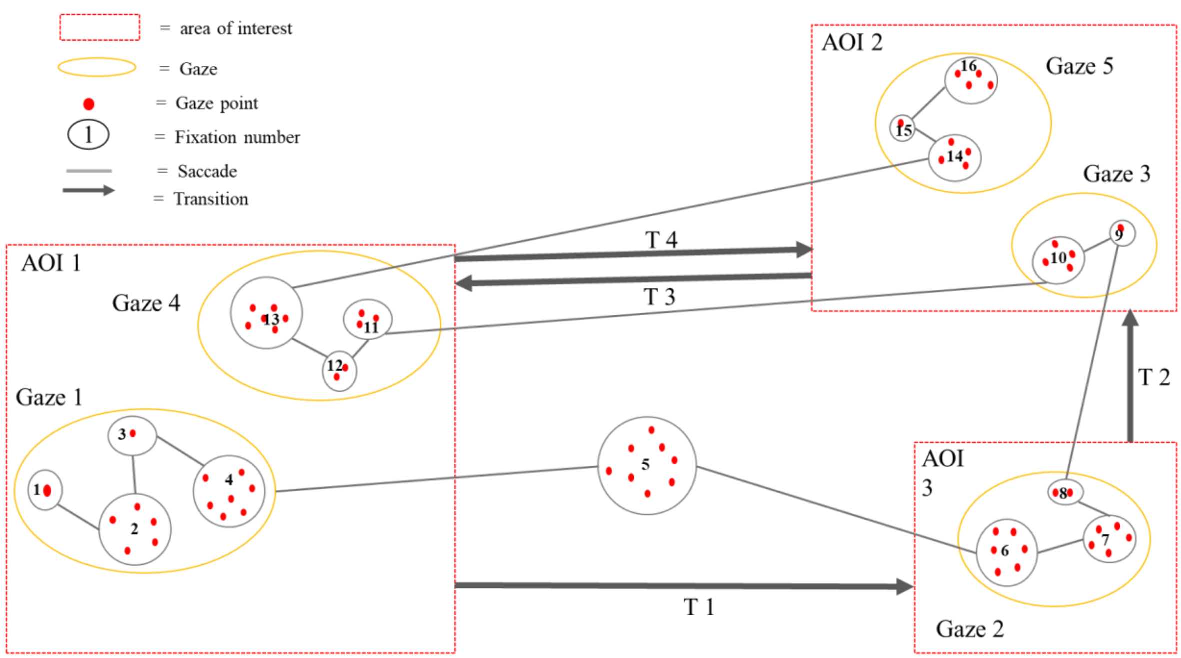


Figure 2.8. Eye tracking different data types.

Table 2.8. Eye tracking system metrics ((Sharafi, Shaffer, Sharif, & Guéhéneuc, 2015).

	Metrics based on the number of fixations
Fixation count	A metric which shows how many a participant fixated her/his gaze on an area of interest to observe an object. Several studies report that a higher number of fixations devoted to a stimulus shows inefficient search for finding the relevant information
Ratio of Fixation Count	The ratio of the total number of fixations in one AOI to all fixations in the area of glance, an AOG can be the entire stimulus or a set of AOIs. A higher ratio shows more efficient search: participants spend low effort to find the relevant AOI.
	Metrics based on fixations duration
Time to the first fixation	Shows the time it took an individual to locate his gaze on the specified area of interest.
Fixations before	A metric which shows where a participant looked before focusing on the AOI.
Fixation duration	Shows fixation time on a specified area of interest.
	Metrics based on scan paths
Attention switching frequency	The total number of switches among a set of AOIs per minute
Transitional matrix	A matrix represents the number of transitions between AOIs, calculated using the fixations number in each cell. Higher density cell shows that the AOI is more visited.
	Pupil size and blink rate
Blink rate	A metric which shows the level of fatigue and cognitive workload. Higher blink rates indicate higher level of fatigue.
Pupil size	A metric which shows the level of fatigue.

Chapter 3: Analysis of pilot's visual scanning characteristics under normal and extreme flight conditions

Author's Note: The contents of this chapter was published as a conference paper, titled "Analysis of pilot's visual scanning characteristics under normal and extreme flight conditions", In Proceedings of the 6th Annual World Conference of the Society for Industrial and Systems Engineering (See Naeeri and Kang (2017)).

3.1. Introduction

To study why some pilots, perform better than others while flying aircraft is to describe how they perform and behave under different flight conditions and how they tackle the emergent problems differently. However, mental processes during flying aircraft are usually not directly observable. Researchers have used various methodologies to measure cognitive processes using techniques such as eye tracking, EOG, and EEG automatically recorded for simulated tasks. In this chapter, we use one of these techniques, eye tracking, to study pilots' eye-movement patterns during fly aircraft under normal and adverse flight conditions.

We investigated the transitions among different areas of interests that pilots look at during normal and emergent situations. While transitions among AOIs have long been studied in eye-tracking research (Holmqvist et al., 2011). We applied different techniques for analyzing transitions of pilots' eye movements. Eye movement data were collected from six pilots. We analyzed the transition patterns of all pilots and compared the differences between expert and novice pilots. This study introduces the network visualization, network centrality measures and statistical related results to compare pilots with different experiences levels. The goal of the experiment was to identify if there are significant differences in the eye movement characteristics between the novice and the experienced pilots during normal and adverse (i.e. severe whether accompanied by engine failures) flight conditions, especially when utilizing complex interfaces found on military aircraft such as B-52. Comparing the visual scanning characteristics between experts and novices can help to find the effective and practical models of experts in training novices.

3.2. Methodology

3.2.1. Participants

Six subjects (i.e. three novices and three experiences pilots) were recruited from the Department of Aviation at the University of Oklahoma. Those with flight experiences that were less than 18 months were defined as novice pilots, and those with more than 18 months were defined as experienced pilots. The age of participants ranged from 21 to 30 years with a mean of 25.4 years and S.D. 3.65 years. All had normal vision.

3.2.2. Apparatus

Experimental measurement systems included Microsoft flight simulator software (FSX) with Logitech extreme 3D Pro Joystick and eye movement measurement system. The prototype of the simulator was a Boeing B-52H model. Eye movement measurement system was remote eye tracking system Tobii TX300 with sampling frequency of 60 Hz. The eye tracker average gaze error was less than 0.5. Tobii TX300 system could measure eyes numerical data, fixation position, and the sequence to scan instrument of subjects. Many eye-movement data could be obtained, such as the fixation numbers, duration, pupil diameters and so on. The collected data was analyzed using R programming and SAS software version 9.3.

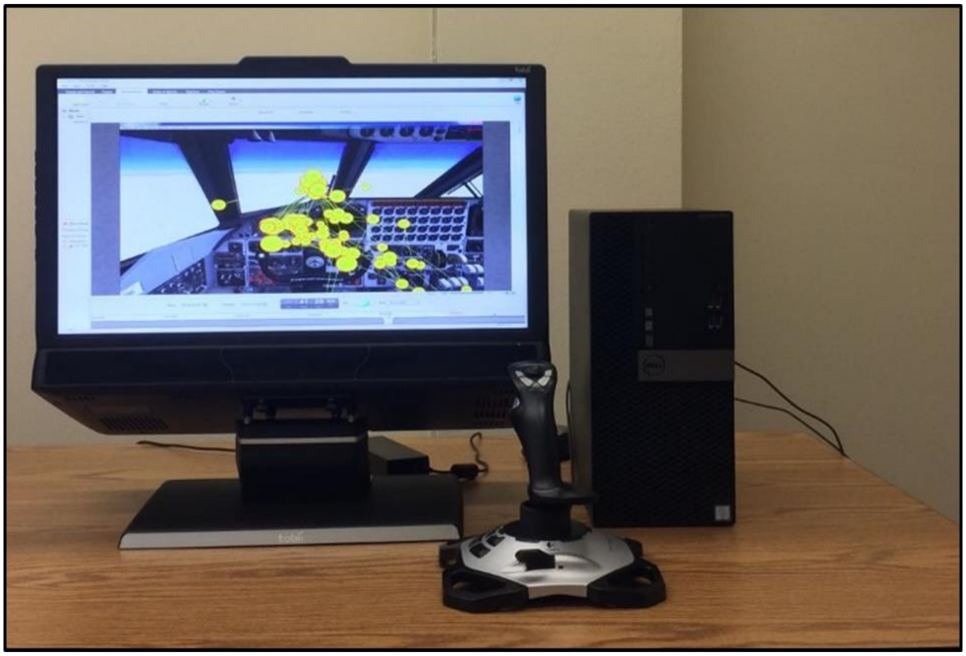


Figure 3.1. Experimental set up. Windows workstation with a Tobii TX300 eye-tracker in the Microsoft light simulator software (FSX) Human Factor Lab at the University of Oklahoma.

3.2.3. Tasks and procedures

Before the Participants signed the consent form, they were first given a brief tutorial on the experiment and on each scenario; then they were seated in front the computer screen to calibrate their eye movements. Each participant was instructed to make eye movement toward the target locations that appear on the screen. Participants were asked to fly the airplane for 30 minutes from Glasgow in Scotland to Leeds in England using the FSX flight simulator. Fig. 3.1 shows the simulator screen with the areas of interests (AOIs) designated for the experiment. Fig. 3.2 shows how the engine failure signals would show up during flight.

Figure 3.2. Cockpit display (B-52) with associated important AOIs: (1) engine oil pressure; (2) attitude indicator; (3) enhanced visual screen; (4) engines indicators; (5) altimeter; (6) airspeed indicator.

Figure 3.3. Engine failure lights (enhanced screen shot of AOI (4) in Figure 3. 1)

3.2.4. Scenarios

Two scenarios were developed using the flight simulator application. The participants were instructed to fly the aircraft for 30 minutes using instrument flight rules (IFR). The scenario begins with taxing phase, moving the aircraft on the ground by its own power from the terminal to the starting line of the defined runway. Following the taxi motion, pilot began preparing for the takeoff phase: The first step, run up the engines at high power to check for engine-related problems. The second step is to release the brakes and set the flaps downward and then accelerating the engine rapidly until the necessary speed “maximum thrust” for take-off is achieved. The next phase is start climbing to the defined altitude (ten thousand feet). Once the aircraft complete climbing phase, the pilot will start cruising phase follow the defined line towards the destination as planned until approach the destination airport. Then the descent phase will start. The pilot should start reducing engine power to the low power and losing the altitude gradually. Landing is the last phase of the flight mission, where the aircraft power is slowing down and starts descending to the runway. The first scenario was to fly the aircraft under normal weather without any engine failure. The second scenario was to fly the aircraft under the severe thunderstorm condition with two engine set failures (out of four engines set in total) occurring approximately five minutes after takeoff. The administrations of the scenarios were counter balanced. In addition, in order to prevent confounding effects, we requested the pilots to first perform one scenario, and then come back after taking a long break to perform the second scenario.

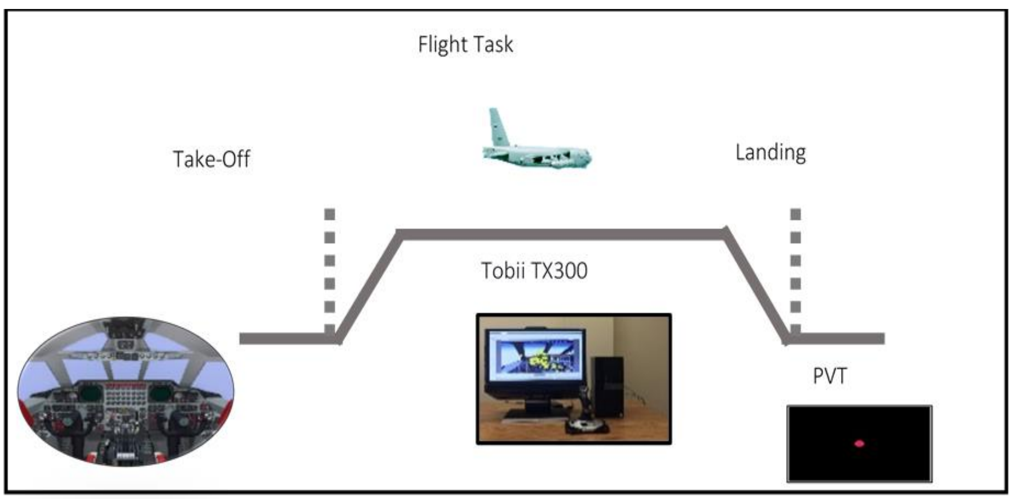


Figure 3.4. Experimental design and virtual scenario. The diagram illustrates an example of the tasks performed by pilots. Each pilot underwent two scenarios, pilots’ real time eye movement data were collected using Tobii TX300.

3.2.5. Variables

In order to control the variability and potential of differences on experimental conditions, we applied within-subject design. The independent variable was the severity of the incidents with two levels: normal condition vs. extreme condition. The dependent variables were the total numbers of the eye fixations on the AOIs (see Figure 3.2) and the total durations of eye fixations. In addition, the visual scan path (i.e. order of eye fixations and saccades) were sampled. Each eye-fixation was identified using a built-in velocity threshold identification (or I-VT) algorithm within the Tobii Studio software connected to the Tobii TX300 eye tracking hardware.

3.3. Results

Figure 3.5 (a) shows the expert pilots total fixation duration on AOIs during extreme scenario. For each AOI, the following metrics were calculated: number of fixations and mean fixation. Those features were calculated for the normal flight condition scenario (30m) as well as for adverse flight scenario (30m). The statistical analysis was performed to investigate the research question "Is there any differences in eye behavior between participants with high and low flight experience. Figure. 3.5 (b) shows the mean eye fixation numbers and durations of the novice and experienced pilots for both scenarios. The mean eye fixation numbers were higher for the experienced pilots, whereas the mean eye fixation durations were lower for the experienced pilot. In details, the Mann-Whitney-Wilcoxon tests showed that there were significant differences on fixation durations ($p < 0.001$) and fixation numbers ($p < 0.001$) between the novices and the experts for both normal and extreme conditions.

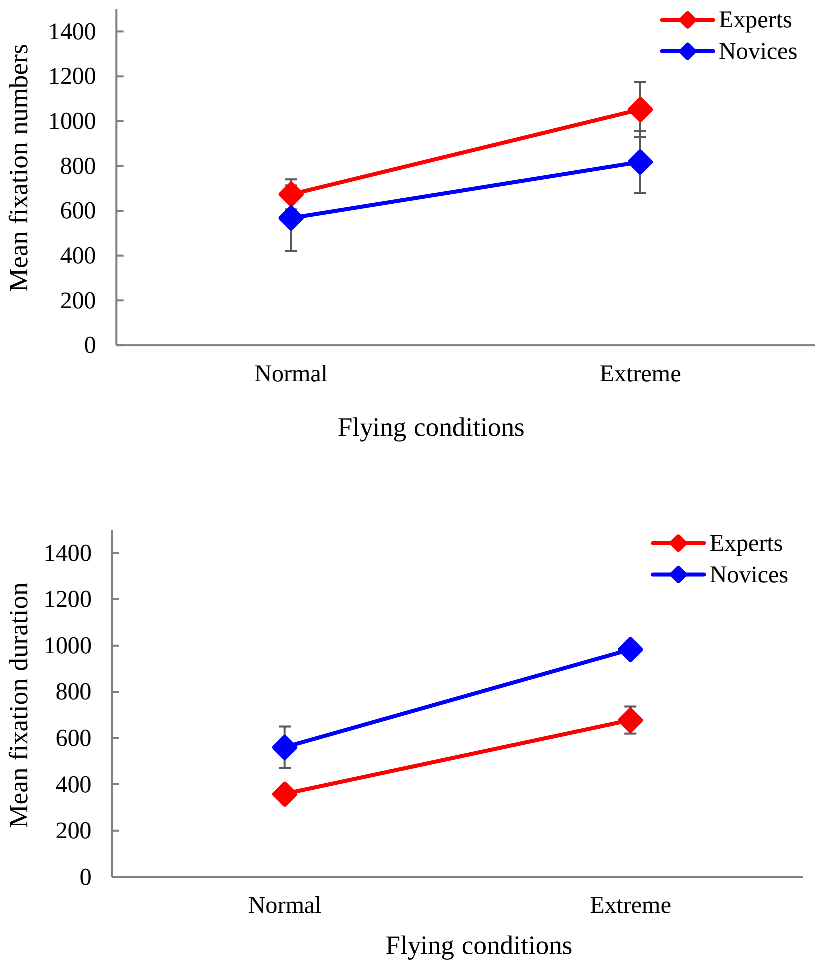


Figure 3.5. Mean eye fixation numbers and durations based on expertise and flight conditions.

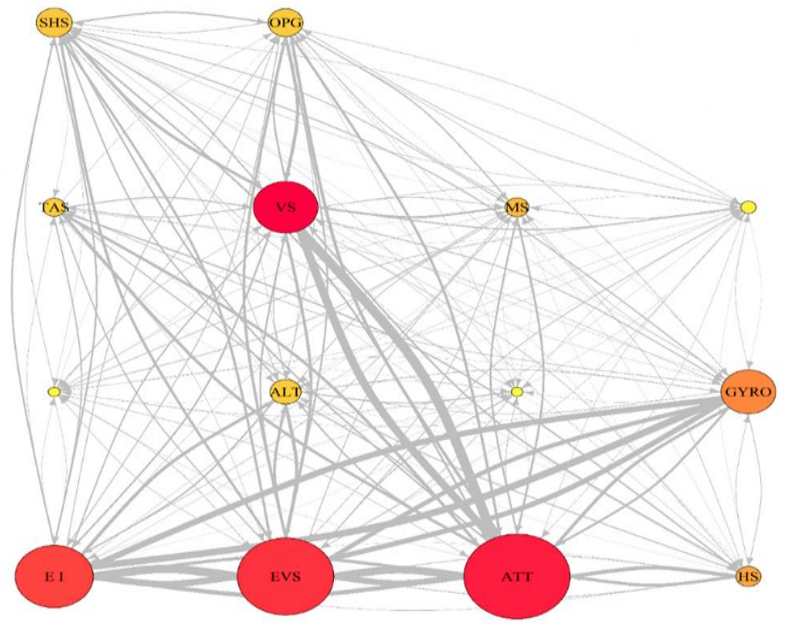
3.3.1. Directed weighted network analysis

Figure 3.6 shows four DWN that were developed using a transition matrix approach to visualize expert and novice pilots' visual transition characteristics during normal and extreme flight conditions. We defined 14 AOIs on the cockpit display, as shown in Figure 3.1. The next step of developing the transition matrix (Holmqvist et al., 2011). The columns and the rows of the transition matrix represent the AOIs in the cockpit display and the cells in the matrix indicate the number of transitions from the row AOI to the column AOI. In the AOI transition matrices, we consider AOIs as nodes and transitions as edges, then we visualize the transition matrix using a directed weighted network (see Figure 3.6).

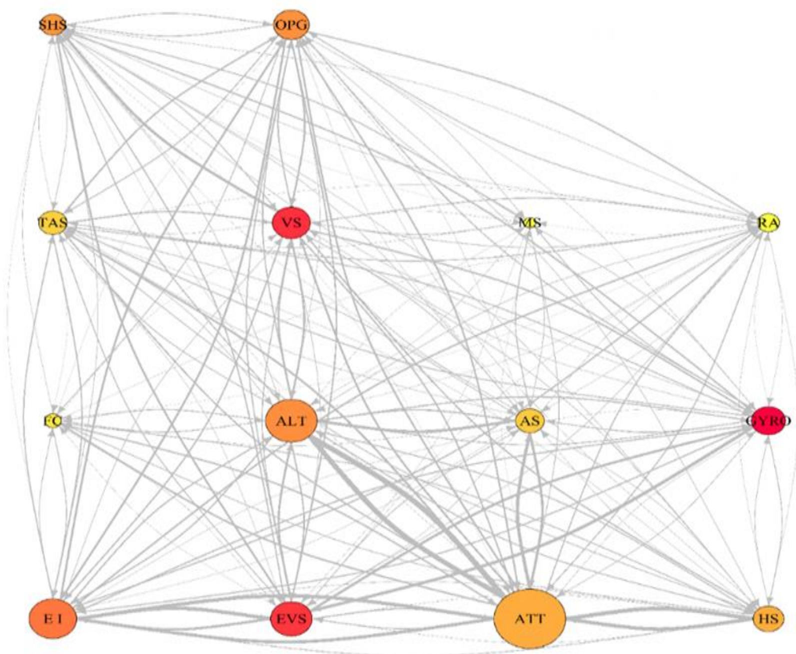
Each edge represents the beginning and ending of the eye fixation transition between two AOIs $i \text{ and } j, t_{ij} > 0$. The width of the edge t_{ij} is proportional to the number of transitions between two AOIs; the wider the edge the higher the transition frequency. The size of the node is proportional to the weight of the edges. The color of the nodes is proportional to the average eye fixation duration in each AOI. Red color represents longer fixation duration, while yellow represents shorter fixation duration. These networks clearly demonstrate the importance of how pilots move their eyes among the areas of interest (AOIs) on the cockpit display.

First, the experts' search strategy was different under adverse from normal flight conditions. Noting that, expert pilot had longer fixation duration and more fixation transition under adverse flight conditions than normal conditions. ATT, EVS, EI and VS AOIs in the experts were the most viewed under adverse flight conditions, this could indicate that pilots at higher expertise levels saw more relevant information contained in specific AOIs. Therefore, attempting to clarify seek out further information to help them acquire more information to safely land the aircraft. This may also suggest AOIs that are fixated on more frequently are valued more by the expert pilots.

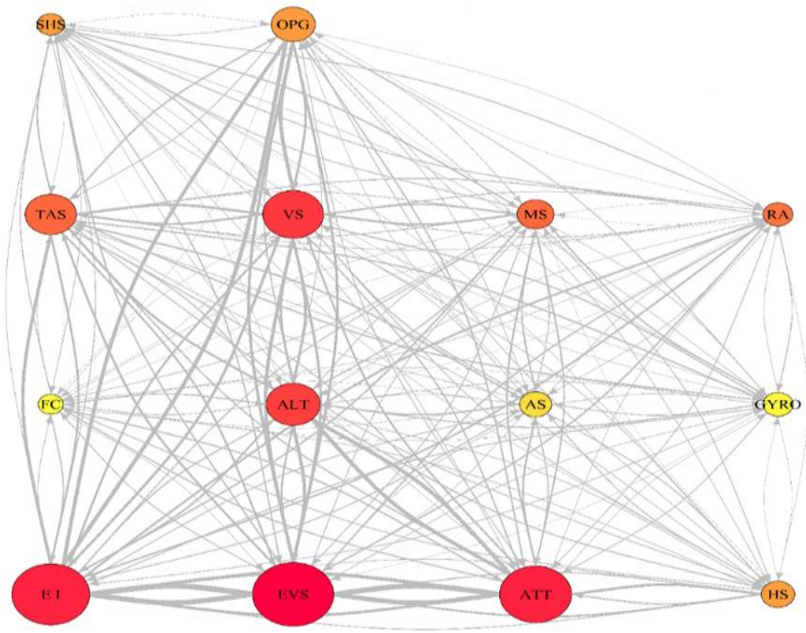
Second, novice pilots exhibit less transition frequency among AOIs and spent more time viewing AOIs under both flight conditions than experts, this could indicate that pilots at lower expertise levels were unable to interpret the information the AOIs contained. It was assumed that longer fixations implied a more difficult task (Jacob & Karn, 2003). Additionally, more fixation transitions were assumed to imply the importance of information (Jacob & Karn, 2003). The results confirm that pilots who have more transitions among AOIs and shorter fixation duration exhibit more efficient search strategy.



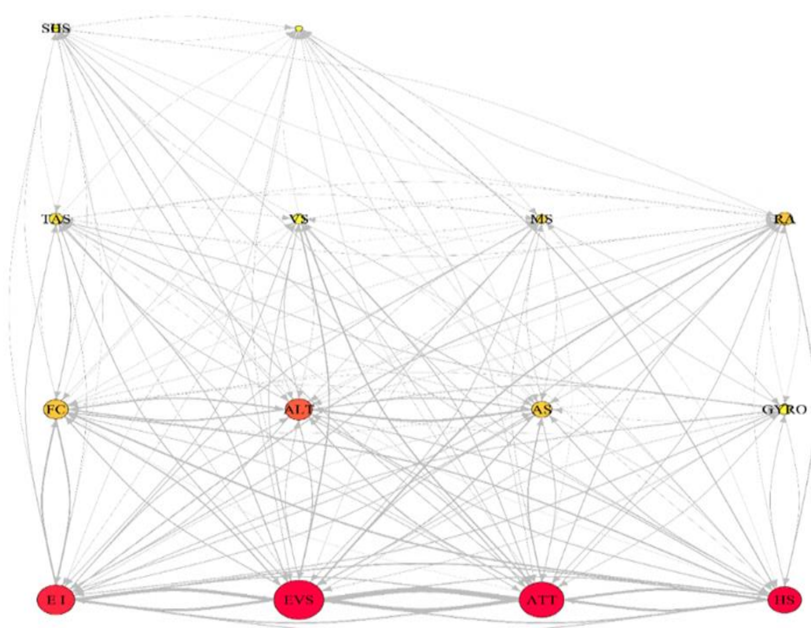
a. Expert: Extreme condition



b. Expert: Normal condition



c. Novice: Extreme condition



d. Novice: Normal condition

Figure 3.6. Sample visual transition characteristics of a novice and an experienced pilot under normal and extreme conditions.

3.3.2. Centrality measures

To understand relational states between network nodes, three measures of centrality: degree, closeness and betweenness have been used. Centrality indegree measures the number of edges from other network nodes that are directed towards a node (the number of nodes that the focal node is connected to). It is a local measure since it does not consider the rest of the network; closeness measures the shortest path from the focal node to all other nodes; betweenness measures the number of times a node acts as a bridge along the shortest path between other nodes (Newman, 2001; Brandes, 2001). Both last two measures are global measures since they consider the whole of the network. Thus, when evaluating each node, the higher the values for all three-centrality metrics, the more important the node is in the overall network (Opsahl et al., 2009). Freeman (1978) formalized these three different measures of node centrality:

I. Indegree

$$C_{D-in}^{w\alpha}\,(i) = \; k_i^{(in)} \; \times \left(\frac{s_i^{in}}{k_i^{in}}\right)^{\alpha}$$

$Where\ \alpha = 1,\ k_i^{(in)} = \ C_D^{in}(i) = \sum_j^N x_{ij}\,, i$ represent the focal node and j represent all other nodes, N is the number of all network nodes, represent the transition matrix, and x_{ij} represent each cell in the matrix weighted 1 if node i is connected, and 0 if not connected.

II. Closeness

$$C_c^{w\alpha}(i) = \left[\sum_j^N d^{w\alpha}(i,j)\right]^{-1}$$

$Where\ \alpha = 1,\ d^{w\alpha}(i,j) \ = \min\left[\frac{1}{w_{ih^{\alpha}}} + \cdots \frac{1}{hj^{\alpha}}\right]$, h represents intermediary nodes on paths between node i and node j

III. Betweenness

$$C_B^{w\alpha}(i) = \frac{g_{jk}^{w\alpha}(i)}{g_{jk}^{w\alpha}}$$

where $g_{jk}^{w\alpha}$ represents the shortest path between two nodes and $g_{jk}^{w\alpha}(i)$ is the number of nodes that flow through node i.

Figure 3.7 shows the plot of indegree centrality measures of experts and novices for the flight mission under extreme conditions. Figure 3.7 shows that ATT, EVS, EI, VS and GYRO are the most important nodes for the expert. The plot of novices indegree measure shows ATT, EVS, EI, ALT and VS are the most important nodes. However, the novice exhibited less indegree level than experts (see Table 3.2). Figure 3.8 shows the plot of indegree centrality under normal flight conditions, the expert pilots had exhibited similar trend to the that under adverse flight conditions. By comparing indegree centrality plots with DWN plots shown in figure 3.6, we noted that ATT, EVS, EI, VS are the most important areas of interests and hence it is having greater impact over rest of the other nodes.

Figure 3.9 shows the plot of closeness centrality measures of experts and novices for the flight mission under adverse conditions. Figure 3.9 shows that ATT, EVS, EI, GYRO and VS have the highest closeness centrality nodes for the expert (see Table 3.2). The plot of novices' closeness centrality measure shows ATT, EVS, EI, GYRO and ALT have the highest closeness centrality nodes. Figure 3.10 shows the plot of closeness centrality under normal flight conditions; the expert pilots had exhibited almost similar trend to the adverse flight conditions. By comparing closeness centrality plots with DWN plots shown in figure 3.6, we noted that ATT, EVS, EI, VS and GYRO are the closest nodes to other network nodes. It takes the fewest distance for any nodes to reach them. Hence it is having greater impact over rest of the other nodes.

Figure 3.11 shows the plot of betweenness centrality measures of experts and novices for the flight mission under adverse conditions. Figure 3.11 shows that ATT and EVS have the highest betweenness centrality nodes for the expert (see Table 3.2). The plot of novices' betweenness centrality measure shows ATT, EVS and EI have the highest betweenness centrality nodes (see Table 3.1). Figure 3.12 shows the plot of betweenness centrality under normal flight conditions; the expert pilots had exhibited almost similar trend to the adverse flight conditions. By comparing betweenness centrality plots with DWN plots shown in figure 3.6, we noted that ATT and EVS are often important controller of information.

Table 3.1. Ccentrality measures of experts and novices under normal flight conditions

Nodes	**Extreme flight conditions**					
	Indegree centrality		Closeness centrality		Betweenness centrality	
	Expert	Novice	Expert	Expert	Novice	Expert
EI	225	214	2.20	225	214	2.20
EVS	278	223	2.39	278	223	2.39
ATT	306	198	2.35	306	198	2.35
HS	72	92	1.17	72	92	1.17
FC	34	69	0.42	34	69	0.42
ALT	89	148	1.14	89	148	1.14
AS	33	88	0.54	33	88	0.54
GYRO	158	83	1.82	158	83	1.82
TAS	65	141	1.08	65	141	1.08
VS	186	167	1.97	186	167	1.97
MS	68	102	0.93	68	102	0.93
RA	44	82	0.67	44	82	0.67
SHS	104	75	0.98	104	75	0.98
OPG	100	121	1.09	100	121	1.09

Table 3.2. Ccentrality measures of experts and novices under extreme flight conditions

Nodes	Normal flight conditions					
	Indegree centrality		Closeness centrality		Betweenness centrality	
	Expert	Novice	Expert	Novice	Expert	Novice
EI	133	103	1.78	1.77	0.45	0.51
EVS	116	137	1.35	2.13	0.26	0.88
ATT	200	123	2.09	1.90	0.95	0.56
HS	86	92	1.41	1.47	0.00	0.24
FC	46	67	0.85	1.22	0.00	0.06
ALT	144	74	1.68	1.09	0.17	0.06
AS	80	55	1.23	1.13	0.05	0.01
GYRO	95	35	1.12	0.81	0.05	0.00
TAS	79	38	1.15	0.83	0.00	0.00
VS	106	33	1.20	0.91	0.15	0.00
MS	34	30	0.55	0.72	0.00	0.00
RA	61	44	0.83	0.78	0.00	0.12
SHS	70	25	0.80	0.54	0.08	0.00
OPG	99	19	1.14	0.75	0.03	0.00

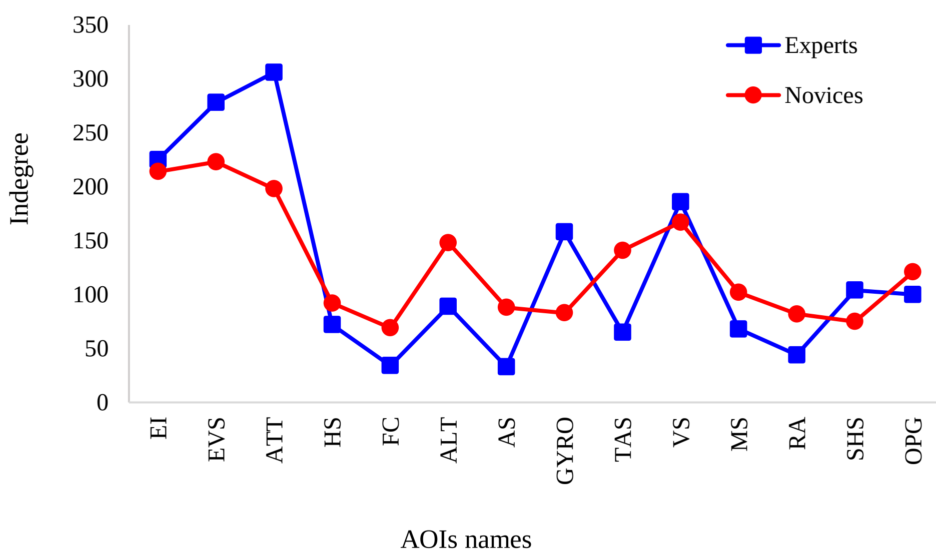


Figure 3.7. Extreme conditions indegree measures.

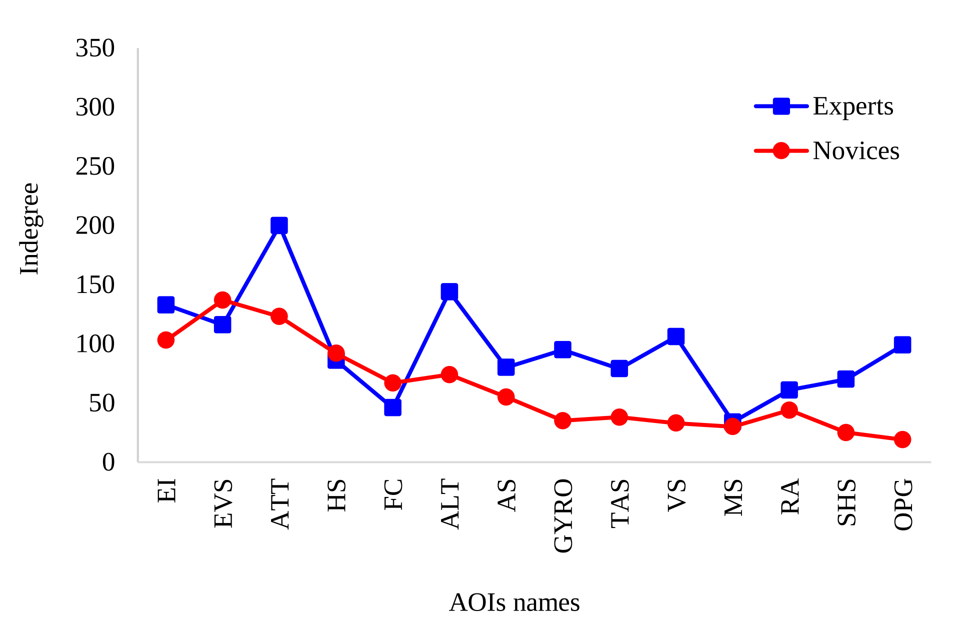


Figure 3.8. Normal conditions indegree measures.

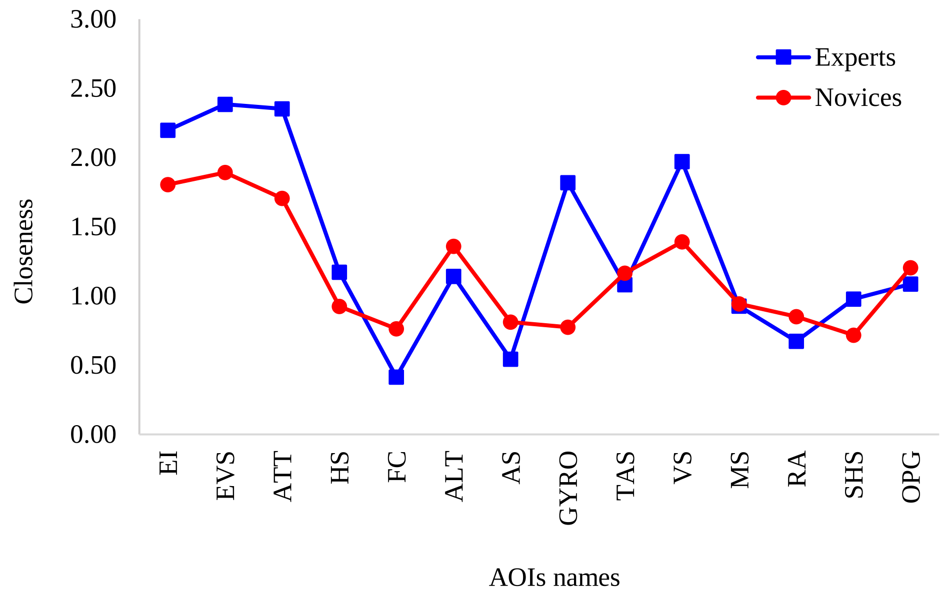


Figure 3.9. Extreme conditions closeness measures.

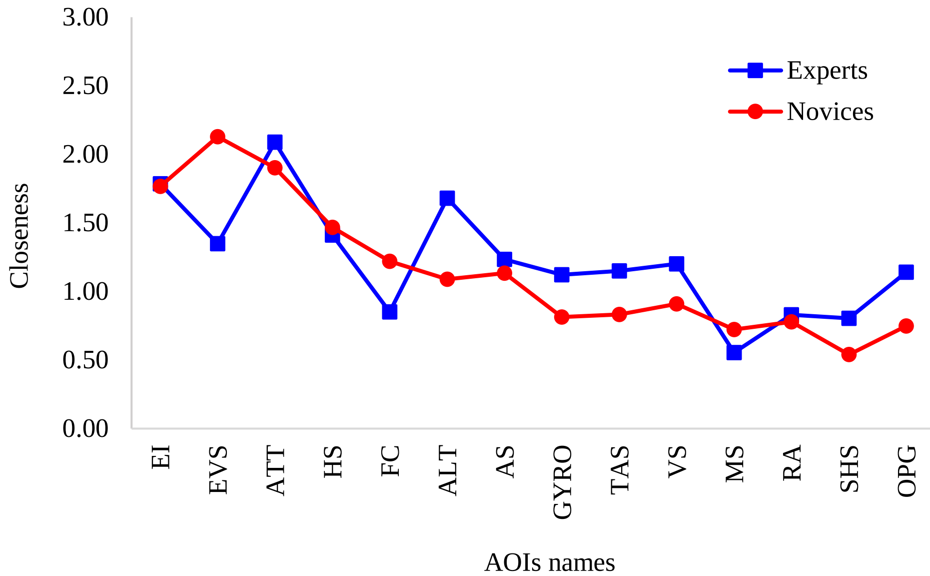


Figure 3.10. Normal conditions closeness measures.

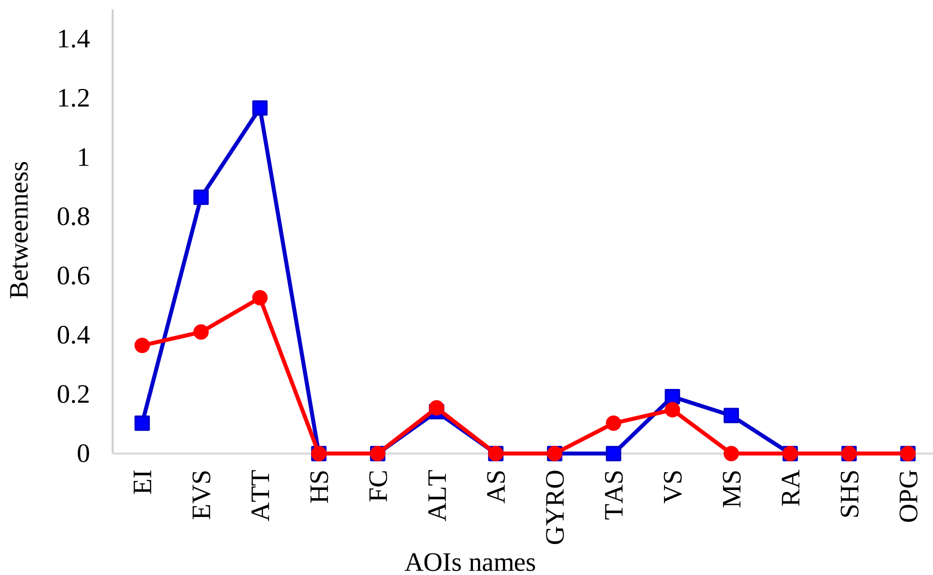


Figure 3.11. Extreme conditions betweenness measures.

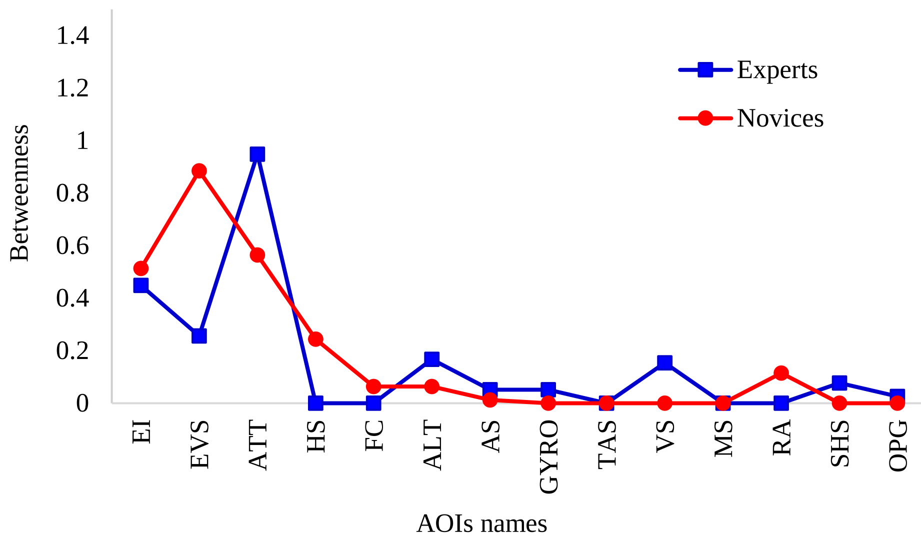


Figure 3.12. Normal conditions betweenness measures.

3.4. Discussion

As stated in the introduction, most of the existing metrics and measurement techniques have its own limitations. Therefore, it is required to apply new metrics that can handle the limitations of the available techniques. With the application of the directed weighted network, we were able to visually represent and observe the characteristics of pilots' visual behaviors on each AOI, as well as the amount of visual transitions among the selected AOIs under different flight situation. In detail, we could identify which AOIs were visited more often and how the time were spent on each AOI.

The directed weighted network approach can provide deep insight on the expert pilots underlying strategy as well as their cognitive process. The representation of the expert's pilot DWN suggest that there is an interactive influence between expertise and the visual scanning techniques of pilots. We were able to observe that the experienced pilots performed more rapidly and extract the required information from the important AOIs than novices' pilot under normal and extreme flight conditions (see Figure. 3.5). The expert's pilot had exerted shorter durations and checks the information with more frequency (see Figure 3.5). The increase of the visual scanning complexity seems to increase with expertise and accords with the finding of Underwood (2007). Expert pilots show high sensitivity to emerged flight situation (i.e. under extreme weather condition and engine failures) (see Figure 3.5). According to Jarodzka, Scheiter, Gerjets, & Van Gog. (2010), shorter fixation duration and wide spread of transitions certainly allow the expert to have more time available to scan other AOIs and correct any errors that might have appeared on the cockpit display. A possible interpretation is that expert pilots applied an appropriate scanning strategy and could adjust their scanning strategy to attend more to relevant information according to the flying phase.

With respect to the centrality-based metrics; indegree, closeness and betweenness the results show that experts' pilot did not make random decisions on transition among AOIs and they tend to more revisit some of the AOIs based on the situation. For example, they tend to revisit EI, EVS, ATT, GARO and VS more than other AOIs during extreme flight condition. Comparing expert and novice pilot in terms of their centrality values, it was found that expert pilots tend to connect multiple AOIs in treating complex problems, whereas novices' pilots are more likely to consider information in isolation AOI. Based on the above analysis, it can be inferred that pilot expertise and deep flight knowledge are key elements in determining the right course of action to handle any emergency problem during flight missions. Sullivan, Yang, Day & Kennedy (2011) found that flight operation influenced by the way pilots distribute their attention. Therefore, to improve their visual scanning pattern of novice pilots, they should train on how to distribute and control their scanning techniques. Janelle & Hatfield (2008) found that if persons receive visual feedback their reactions, their awareness increases and being motivated to improve reactions and control their behavior.

Chapter 4: Exploring the impact of fatigue on the pilots' when interacting with the cockpit interfaces.

Author's Note: The contents in Section 4.1 was published as a conference paper, titled "Exploring the relationship between pilot's and fatigue level when interacting with cockpit interfaces". In Proceedings of the Institute of Industrial and Systems Engineers (IISE) Annual Conference (See Naeeri and Kang (2018)). The contents in Section 4.2 was published as a conference paper, titled "Exploring the effect of fatigue on pilot during single and multi-takeoffs and landings flight missions". In Proceedings of the 7th Annual World Conference of the Society for Industrial and Systems Engineering (See Naeeri, Kang and Mandal (2018)).

4.1. Introduction

Fatigue has been identified as a major reason for accidents and losses with the implications that a fatigued person is most likely to produce unsafe actions in a range of settings (e.g. aviation, rail, maritime and other transport operations) (Drury et al., 2012). Fatigue takes different forms, including mental and physical fatigue depending on the nature of its causes. According to Gislason, Bogdane and Vasiļevska-Nesbita (2017) fatigue is mainly affected by time awake, time of the day, and work-related tasks. Therefore, insufficient recovery from fatigue leads to increasing fatigue. Fatigue in the pilot may build up slowly over several working days or weeks, including memory, concentration, judgment, decision-making, reaction time (RT), attention, and eye fixation (Jackson & Earl, 2006). Fatigue is one of the major factors in aviation accidents, though the contribution of fatigue to accidents is often underestimated in official reporting (Williamson et al., 2011). Jackson & Earl (2006) found that fatigue was a cause of many aviation accidents.

According to Petrilli, Roach, Dawson, and Lamond (2006), fatigue in flight operations has diverse and complex causes, such as insufficient or irregular sleep time, irregular work, rest cycles, and crossing time zones. The National Aeronautics and Space Administration (NASA's.) Aviation Safety Reporting System has shown that 21% of the reported aviation incidents were fatigue related (Jackson Earl, 2006). Several investigations have identified fatigue-related factors as major causes of aviation accidents. Fatigue was involved in at least 8% of aviation mishaps (Caldwell & Caldwell, 2016). The costs associated with fatigue in aviation are tremendous, both in terms of money (aviation accidents can exceed $500 million) and the loss of life. Considering the direct relation between pilot fatigue and accidents, the U.S. Federal Aviation Administration (FAA) has introduced and enforced new rules about flight crew duty and rest requirements and regulations (Federal Aviation Administration, 2011). Fatigue is still one of the primary concerning factors, which negatively affects the pilot's ability to concentrate and impairs his capacity to analyze complex processes (Avers and Johnson, 2011). Pilots' fatigue, which has been recognized by National Transportation Safety Board (NTSB) as one of the foremost concerns contributing to impairments for over four decades (Federal Aviation Administration, 2011).

The most important contributing factors that produce significant levels of fatigue and are specific to the pilot occupation were night flights, jet lag, early morning wakeups, time zone transitions (these desynchronize), multiple flight tasks, and excessive working hours without adequate rest. These factors can lead to high levels of fatigue (Powell et al., 2007; Powell et al., 2008). Something that characterizes fatigue, especially in aviation, is the increasing number of errors that may endanger the flight safety (Cabon et al., 2012). Therefore, it is important to understand the influence of fatigue on the pilot while interacting with cockpit interface design (Thomas et al., 2015).

4.2. Study I: Investigation of the relationship between pilots' and fatigue level when interacting with cockpit interfaces

4.2.1. Purpose of the study

In the present study, we have examined the effect of fatigue on novices and expert pilots' when interacting with cockpit interfaces under extreme conditions.

4.2.2. Method

4.2.2.1. Participants

Six pilots with different expertise levels (i.e. three novices and three experienced) were recruited from the industrial and system engineering department, as well as from the Department of Aviation at the University of Oklahoma. Those who had completed the introductory flight program with flight experiences that were less than one year were defined as novice pilots, and those who had rated, and had more than one-year experience been defined as expert pilots. The age of participants ranged from 29 to 32 years with a mean of 30.5 years and S.D. 2.12 hours. All had normal vision.

4.2.2.2. Apparatus

Experimental measurement systems included Microsoft flight simulator software (FSX) with Logitech extreme 3D Pro Joystick and eye movement measurement system. The prototype of the simulator was a Boeing B-52H model. The eye movement measurement system was a remote eye tracking system Tobii TX300 with a sampling frequency of 60 Hz. The eye tracker average systematic gazes' error (the average deviation between actual gaze position and observed gaze position) was less than 0.5 degree 5 (Horsley, Eliot, Knight & Reilly, 2013). Tobii TX300 system could measure eyes numerical data, fixation position, and the sequence to scan instrument of subjects. Many eye-movement data could be obtained, such as the fixation numbers, duration, pupil diameters, and so on (Holmqvist et al., 2011). Psychomotor Vigilance task (PVT) PVT was administered to measure pilot fatigue. The collected data were analyzed using SAS software version 9.3 and R programming.

4.2.2.3. Tasks and procedures

Before the participants signed the consent form, they were first given a brief tutorial on the experiment; and then they were seated in front the computer screen to calibrate their eye movements. Each participant was instructed to make an eye movement toward the target locations that appear on the screen. With respect to this study eye tracking area of interests (AOIs), are defined as engines' indicators, enhanced visual screen, attitude indicator, altimeter, and airspeed indicator (see Figure 4.1). Participants were asked to fly the airplane B-52 for two hours from Oklahoma Regl airport to Florida Opa Locka airport in the USA using the FSX flight simulator. Between each two flight phases, fatigue testing occurred once every 25 minutes. Each participant did fatigue testing individually.

Figure 4.1. AOI of B-52 Cockpit display: (1) Engines indicators; (2) Enhanced visual screen (3) Attitude indicator; (4) Altimeter; (5) Airspeed indicator.

4.2.2.4. Scenario

The participants were instructed to fly the aircraft for two hours using instrument flight rules (IFR). The phases of the scenario are as follows: 1. Taxing: moving the aircraft on the ground by its own power from the terminal to the starting line of the defined runway. 2. Preparing for the takeoff: The first step, run up the engines at high power to check for engine problems. The second step is to release the brakes and set the flaps downward and then accelerating the engine rapidly until the necessary speed "maximum thrust" for take-off is achieved. The next step starts climbing to the defined altitude (ten thousand feet). Once the aircraft completes the climbing phase, the pilot will start the cruising phase, following the defined line towards the destination as planned, until approaching the destination airport. Then the descent phase will start. The pilot should start reducing engine power and lose altitude gradually. Landing, where the aircraft is slowing down and landed on the runway. The scenario was to fly the aircraft under the extreme condition with two engine set failures (out of four engines set in total) occurring approximately at 10 minutes before landing.

4.2.2.5. Variables

The independent variables were the flight phases with five levels. The dependent variables were the total numbers of the eye fixations on the AOIs. In addition, we extracted the following variables from each PVT test as a measure of the overall level of pilot fatigue: mean reaction time (RT), mean of slowest 10% of RT, mean of the fastest 10% of RT, mean number of false alarms, defined as the number of reaction times $\leq$ 100 ms, and the number of lapses in attention, defined as the number of reaction times $\geq$500 ms. Visual transition characteristics were calculated and used to analyze the visual scan path sampled (i.e. Order of eye fixations and saccades). Each eye-fixation was identified using a built-in velocity threshold identification (or I-VT) algorithm within the Tobii Studio software connected to the Tobii TX300 eye tracking hardware.

4.2.3. Results

4.2.3.1. Results of Psychomotor Vigilance task (PVT)

Figure 4.2 shows the mean reaction time of the novice and expert pilots during five flight phases. The mean reaction times were shorter for the expert pilots (See Figure 4.2). The number of lapses were also lower for the expert pilot (See Figure 4.3). More specifically, the Mann-Whitney-Wilcoxon tests showed that there were significant differences on reaction time ($p < 0.001$) and number of lapses ($p < 0.001$) between the novices and the experts. Figure 4.2and 4.3 also shows there were remarkable increases in reaction time and the number of lapses for both novice and expert pilot’s as the flight went on to the final phase.

Figure 4.4 shows the mean of slowest 10% RT, the mean of the fastest 10% of RT, the mean number of false alarms. The mean of the slowest 10% of RT were shorter for the experienced pilots (See Figure 4.4). The mean of the fastest 10% of RT were shorter for the experienced pilots (See Figure 4.5). The number of false alarms were also lower for the experienced pilots (See Figure 4.6). More specifically, the Mann-Whitney-Wilcoxon tests showed that there were significant differences on slowest 10% of RT ($p < 0.031$), fastest 10% of RT ($p < 0.015$), and the number of false alarms ($p < 0.030$) between the novices and the experts. Figure 4.4, 4.5 and 4.6 also shows there were significant increases in the slowest 10% of RT, the fastest 10% of RT, and the number of false alarms, for both novice and expert pilots over the course of the flight

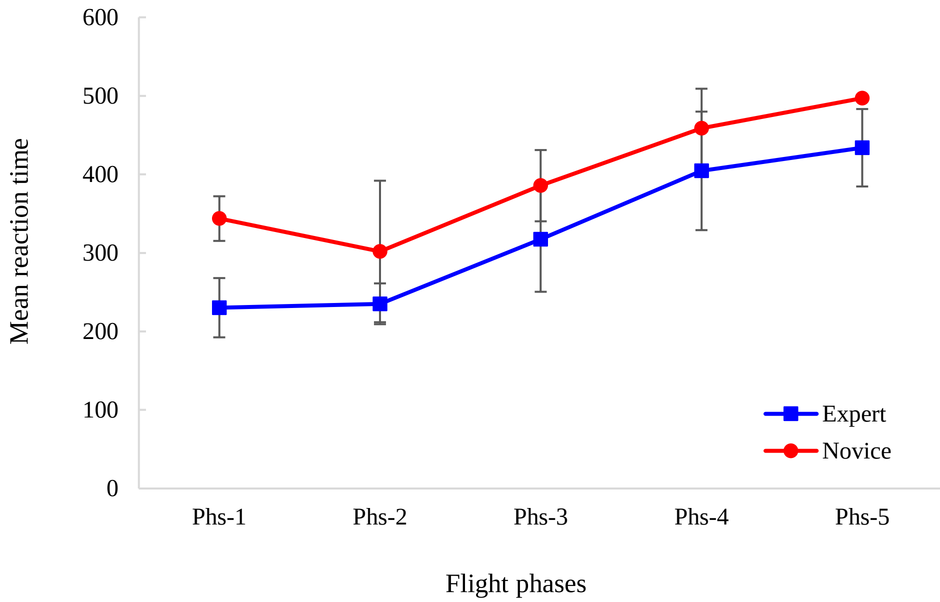


Figure 4.2. Mean reaction time based on expertise and flight phases.

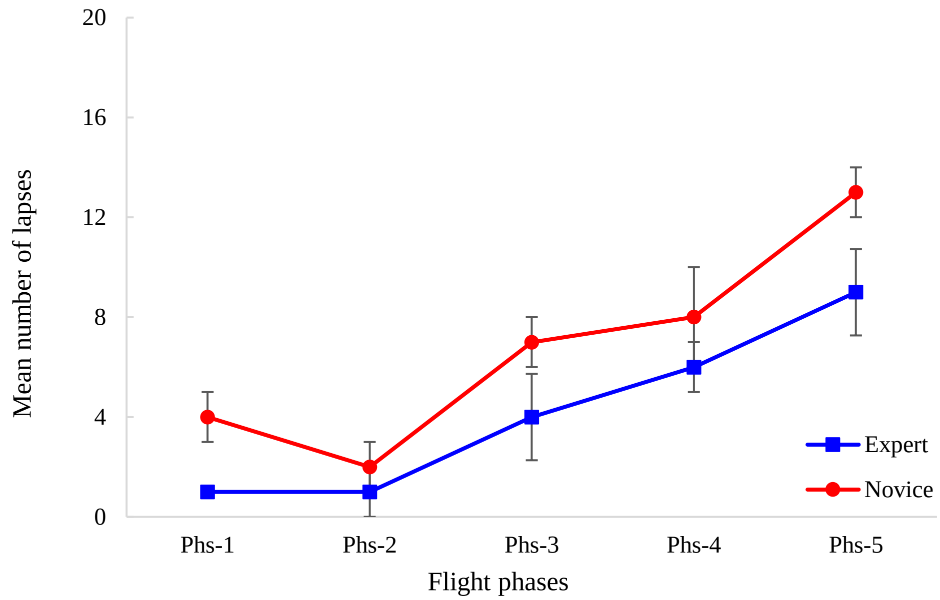


Figure 4.3. Mean number of lapses based on expertise and flight phases.

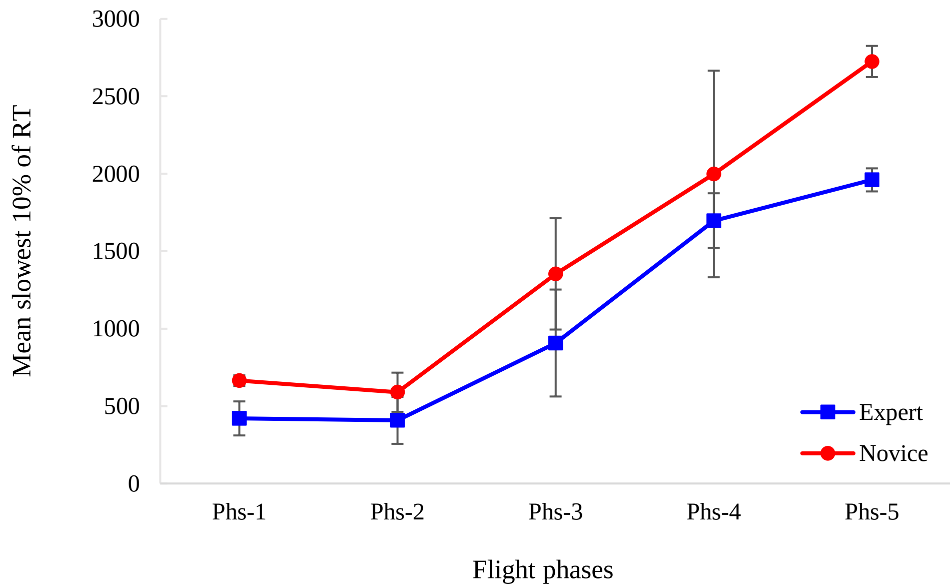


Figure 4.4. Mean slowest 10% of RT based on expertise and flight phases.

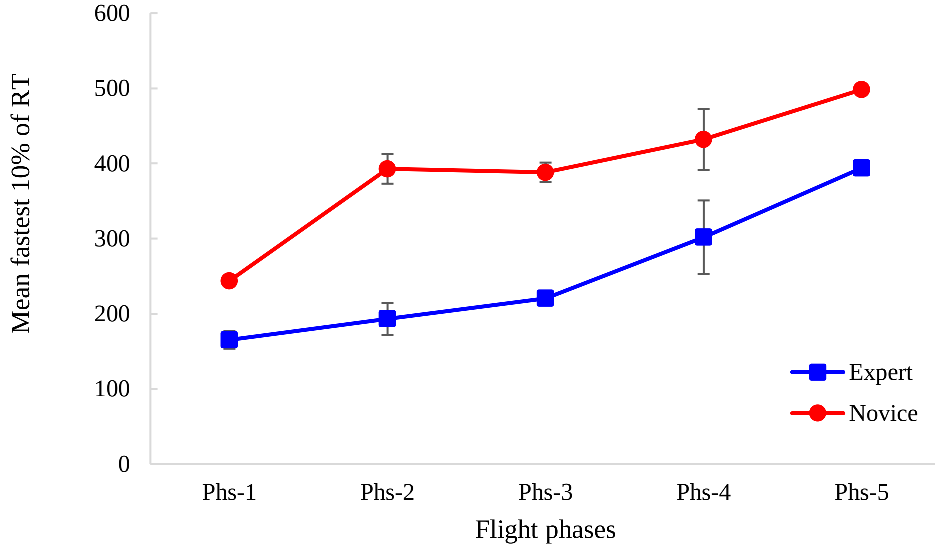


Figure 4.5. Mean fastest 10% of RT based on expertise and flight phases.

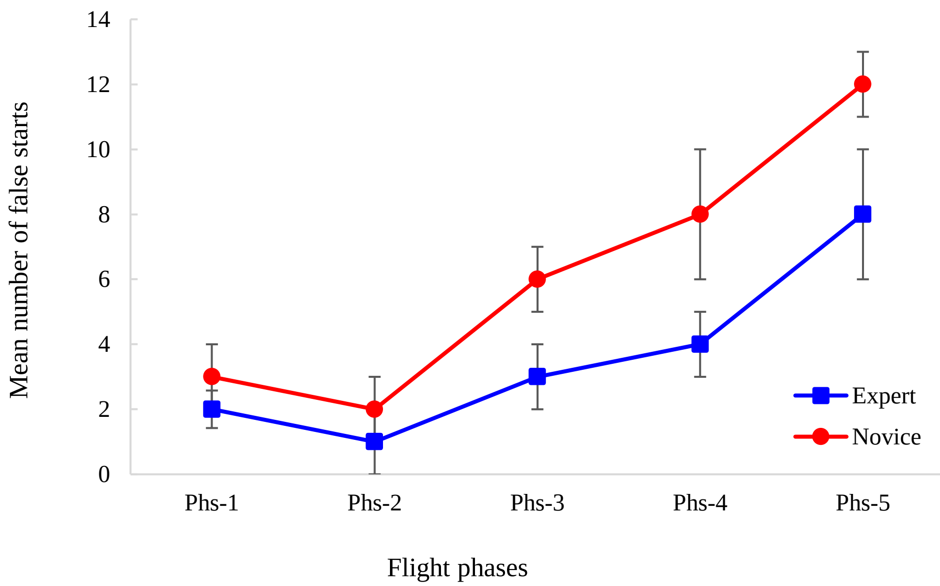


Figure 4.6. Mean number of false starts based on expertise and flight phases

4.2.3.2. Results of eye fixation

Figure 4.7 shows that the mean eye fixation numbers were higher for the experienced pilots than for the novices, whereas the mean eye fixation durations were lower for the experienced pilot than for novices (See Figure 4.8). More specifically, the Mann-Whitney-Wilcoxon tests showed that there were significant differences on fixation durations ($p < 0.0001$) and fixation numbers ($p < 0.0001$) between the novices and the experts. Figure 4.7 shows there were remarkable increases in mean eye fixation numbers for the experts over the course of flight. But there were slight increases in mean eye fixation numbers for novices. Figure 4.8 shows there were insignificant increases in mean eye fixation duration for the experts but there was significant increase in mean fixation duration for novices.

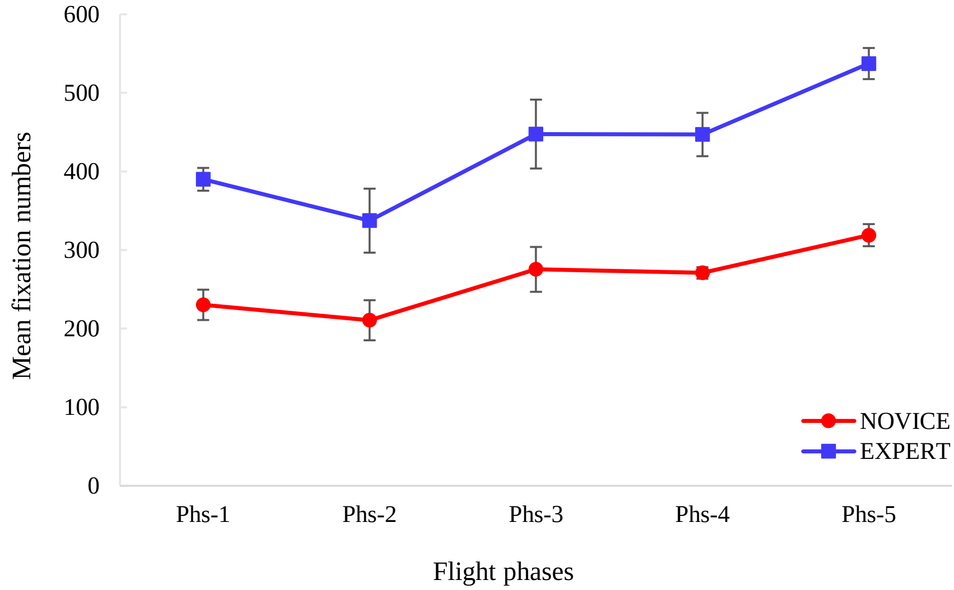


Figure 4.7. Mean eye fixation numbers based on expertise and flight phases

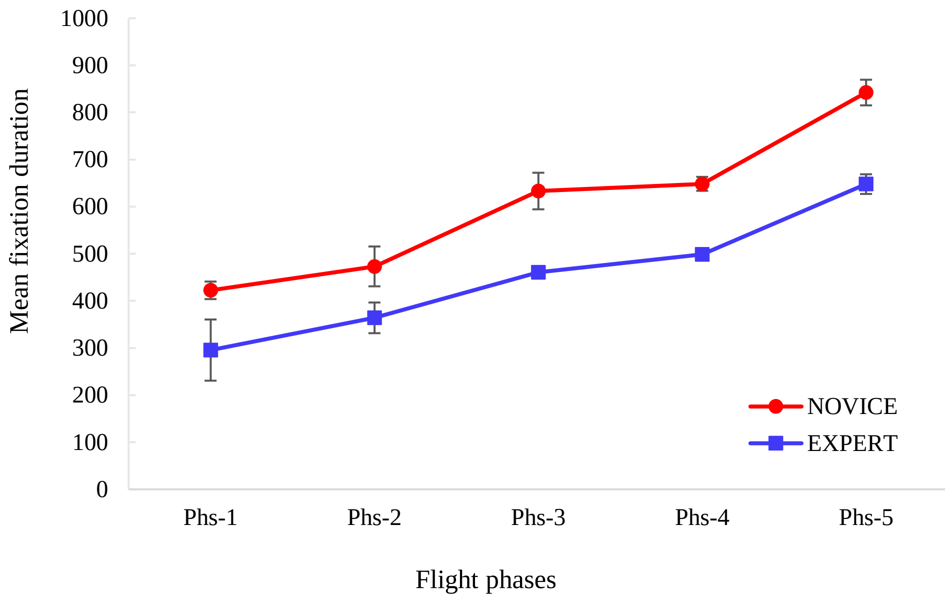


Figure 4.8. Mean eye durations based on expertise and flight phases

4.2.3.3. Results of eye transition metrics

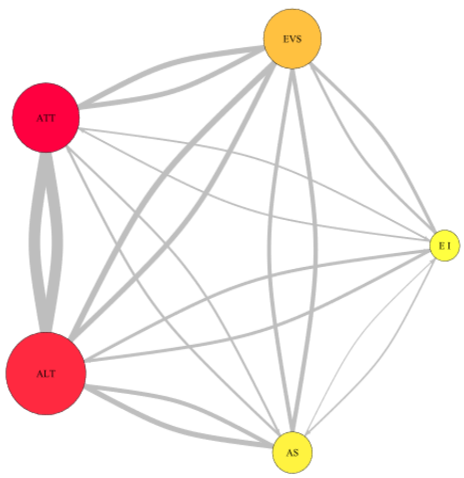
The eye transition sequences of the pilot during five flight phases were analyzed using the transition matrix approach. For better visualization of eye transition sequences, DWN were developed. Figure 4.9 shows 10 directed weighed networks that were developed using a transition matrix approach to visualize three novices and three expert pilots' visual transition characteristics during the five flight phases. Each network represents pilot eye transition sequences during each one of the five flight phases (each phase lasts for about 25 minutes). Phase 1 involved take off phase, phase 2, 3 and 4 involved climbing, cruising and approach phases respectively and phase 5 involved approach and landing phase. These networks clearly demonstrate how a pilot transits his/her eyes fixation among the areas of interest (AOIs) on the cockpit display during flight mission (see Figure 4.1).

The network was divided into five areas: Engine's indicators (EI), Enhanced visual screen (EVS), Attitude indicator (ATT), Altimeter indicator (ALT), and Airspeed indicator (AS) (see

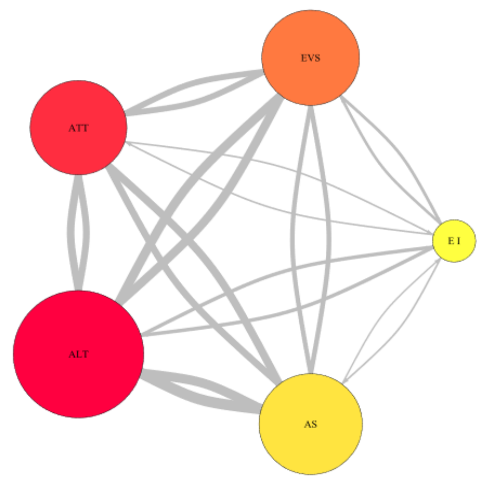
Figure 4.9). Each one of these AOIs represented on the DWN by one node (Yellow circle). The size of the node is proportional to the average fixation time spent on each AOI. These nodes are connected by edges (Black arrows). The width of the edges connecting the nodes is proportional to the number of fixation transitions between each two AOIs. They represent the average weights of the eye fixation transition among the above mentioned AOIs. The analysis of directed weighed networks was based on two measures. A first measure was the percent time spent on the AOIs based on the total fixation duration. A second measure was the transition frequencies between AOI's. The dominant saccades were calculated for each flight phase by transition matrices.

As Figure 4.9 shows, differences were found between expert and novice pilots visual search patterns during five flight phases. The novice DWN during the first flight phase (See Figure 4.9 (a)) shows unbalanced patterns of visual search among AOI's and they also spent more fixation time on ATT and ALT. In contrast, the expert pilot DWN shows a clearly defined visual search pattern. As Figure 4.9 (b) shows, the expert pilots' transit their fixation among the AOI's more frequently than the novice pilot did and they also spent less fixation time on an AOI's. The dominant transition of expert pilots was between ATT, ALT, EVS and AS; this would be because they are the most important indicators during the takeoff phase which is involved in this phase.

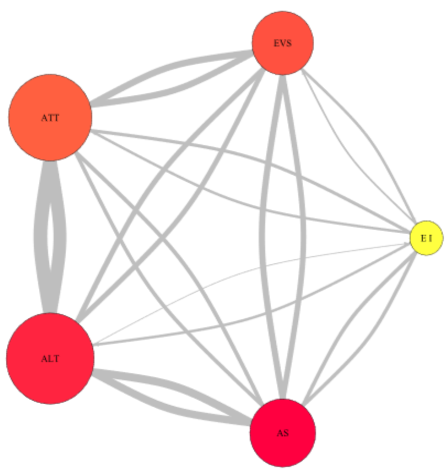
The novice DWN's of the second, third and fourth flight phases, which was the climbing, cruising and approach phases (See Figure 4.9 (c, e and h)), show that novice pilots exhibit a weak scanning pattern. The transition frequencies between some AOI's, such as EI and AS or AS and EVS were very low. In contrast, Expert DWN's (See Figure 4.9 (d, f and h)), showed a clearly defined scanning pattern among all AOI's and less fixation duration on them. Expert DWN's show that pilots had allocated more attention to the attitude indicator, altitude and enhanced visual screen than other AOI's; which is very important for controlling when the aircraft reach one of these flight phases. The novice DWN of the fifth phase (See Figure 4.9 (i)) shows that novice pilots still exhibited a weakened pattern of visual scanning search among AOI's. The dominant fixation transitions of novices were between the ATT, ALT. Expert DWN of the fifth phase (See Figure 4.9 (k)) shows that they had remarkable differences in their scanning pattern. They showed more balanced visual scanning searches among AOI's than a novice. They showed a wider spread of transition.



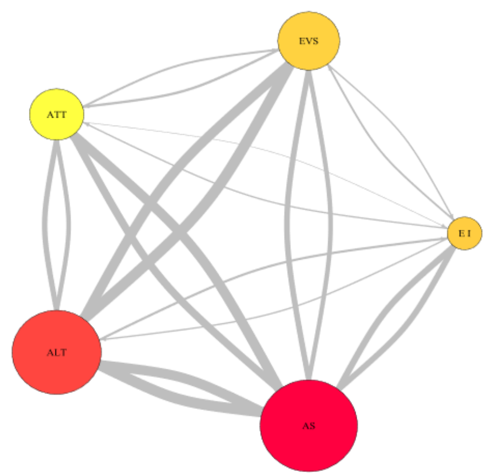
(a) Novice: DWN of Phase (1)



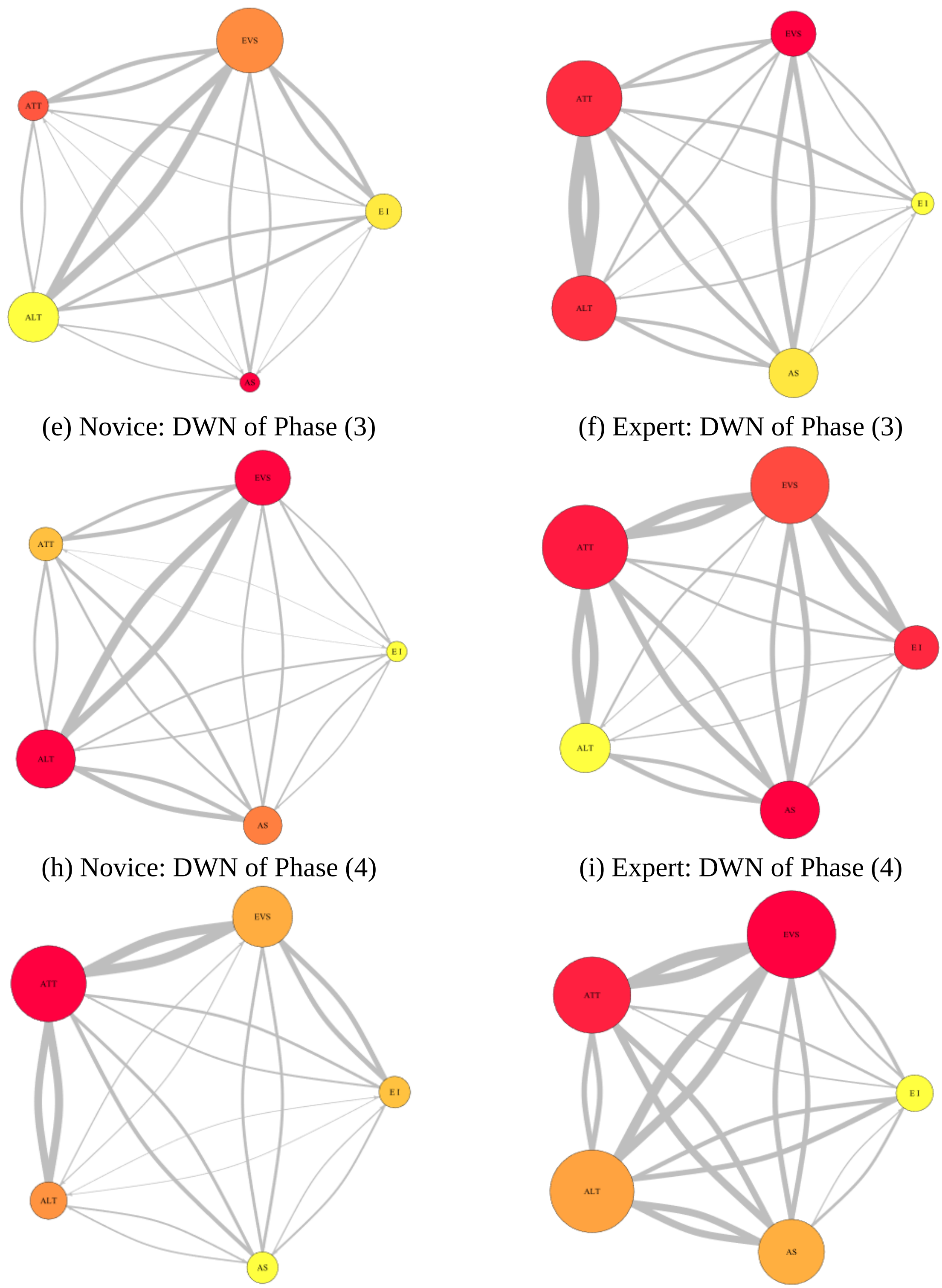
(b) Expert: DWN of Phase (1)



(c) Novice: DWN of Phase (2)



(d) Expert: DWN of Phase (2)



(e) Novice: DWN of Phase (3)
(f) Expert: DWN of Phase (3)
(h) Novice: DWN of Phase (4)
(i) Expert: DWN of Phase (4)
(j) Novice: DWN of Phase (5)
(k) Expert: DWN of Phase (5)

Figure 4.9. Sample visual transition characteristics of a novice and an experienced pilot during five flight phases.

4.2.4. Discussion

The findings of this study show that experts had fewer lapses and less reaction time during all the defined flight phases (see Figure 4.2 and 4.3). The results also show that experts had less slow 10% RT, more fastest 10% RT and a smaller number of false alarms. Even though the results demonstrate that experts performed their flight tasks in a more efficient way than novices, there were remarkable increases in reaction time, slowest 10% of RT, fastest 10% of RT, the number of lapses, the number of false alarms, fixation number, and fixation duration, for both novice and expert pilot's as the flight went on to the final phase. This accords with the findings of Lee et al., (2010), which demonstrated that increasing in these measures were strongly associated with the increasing level of fatigue (Lee, Bardwell, Ancoli-Israel, & Dimsdale, 2010).

The analysis of DWN's (see Figure 4.9), shows that the scanning strategy of experts was different from novice pilots. Expert pilots demonstrated more frequent fixations with shorter duration on AOI's than novice pilots did during all five flight phases. Novices paid most of their attention to the attitude and altimeter indicators during all flight phases, while experts not only focused on the attitude indicator and altimeter, but also regularly scanned the EI, EVS and the AS indicators in fast and flexible eye movement. Based on the DWN analysis, expert pilots demonstrated a stronger and more defined scan pattern than novice pilots. Even though the analysis demonstrates that experts had better scanning pattern results than novices, there were notable increase in fixation time and less fixation transition among AOI's for both novice and expert pilot's as the flight went on to the final phase.

These findings are in accordance with earlier findings (Kramer et al., 1994), showing that the difference between expert and novice pilots scanning strategy was due to two main reasons. The first was that the expert had more familiarity with tasks, therefore the cognitive process tended to be programed and automated. The second reason was peripheral vision. Experts developed better peripheral vision than novices because of the flight experience (Horsley et al., 2013; Powell et al., 2007; Powell et al., 2008).

Based on the above analysis, it can be inferred that the analysis of DWN's concurs with PVT results. DWN's results (see Figure 4.9) showed that experts and novices started with strong visual scanning pattern and became weak as the flight mission became more demanding and pilots became more fatigued over the course of the flight. This also had been demonstrated by the eye fixation results (see Figure 4.7 and 4.8) which showed notable increases in fixation duration and slight increase in fixation number for both novices and experts as the flight went on to the final phase. In accordance with these results, PVT results (see Figure 4.3 & 4.6) also showed that both experts and novices started responding faster with a smaller number of lapses and less false alarms; however, their response became slower and the number of lapses and false alarms had increased as the flight went on to the final flight phase.

4.3. Study II: Exploring the effect of fatigue on the pilots' during single and multi-takeoffs and landings flight missions

4.3.1. Purpose of the study

In this study we have examined the effect of fatigue on expert pilots under different conditions: single-takeoff and landing versus multi-takeoff and landing flight missions.

4.3.2. Methods

4.3.2.1. Participants

Four expert pilots (age range 28-36 years, mean 30.75 years, and S.D. 3.10 years), recruited from the Department of Aviation at the University of Oklahoma, participated for the experiment data collection. All of them were instrument flight rated and had more than three years of experience with normal vision.

4.3.2.2. Scenarios

Two scenarios were developed using the flight simulator application. Both flight scenarios were conducted in accordance with FAA instrument flight rules (IFR). The participants were instructed to fly B-52 simulated aircraft like a real flight for four hours. The administrations of the scenarios were counterbalanced. To prevent confounding effects, we requested the pilots to first perform one scenario, and then come back another day to perform the second scenario at the same time. Although the flight scenarios contained emergencies in the last phase, the pilots were instructed to be ready to respond to any unexpected event that may emerge during the flight time. All participants were instructed to fill out a paper-and-pencil Samn-Perelli (SP) Subjective Fatigue Check card and the Karolinska Sleepiness Scale (KSS) subjective sleepiness questionnaire at the scheduled time. They also practiced a 5-minute Psychomotor Vigilance task (PVT) and the eye tracking system collected their fixation data. The yellow circles indicate the times the fatigue tests were administered to the pilots during the multi-takeoff and landing versus the single-takeoff and landing flight missions.

Scenario 1: This scenario was a multi-takeoff and landing flight in which participants flew four flight missions, each mission involved takeoff and landing (see Figure 4.10 (a)).

Scenario 2: This scenario was a single-takeoff and landing flight mission in which participants flew from Ft Collins-Loveland Mun, CO to East Texas Rgnl, TX, and this scenario involved one takeoff and one landing (see Figure 4.10 (b)).

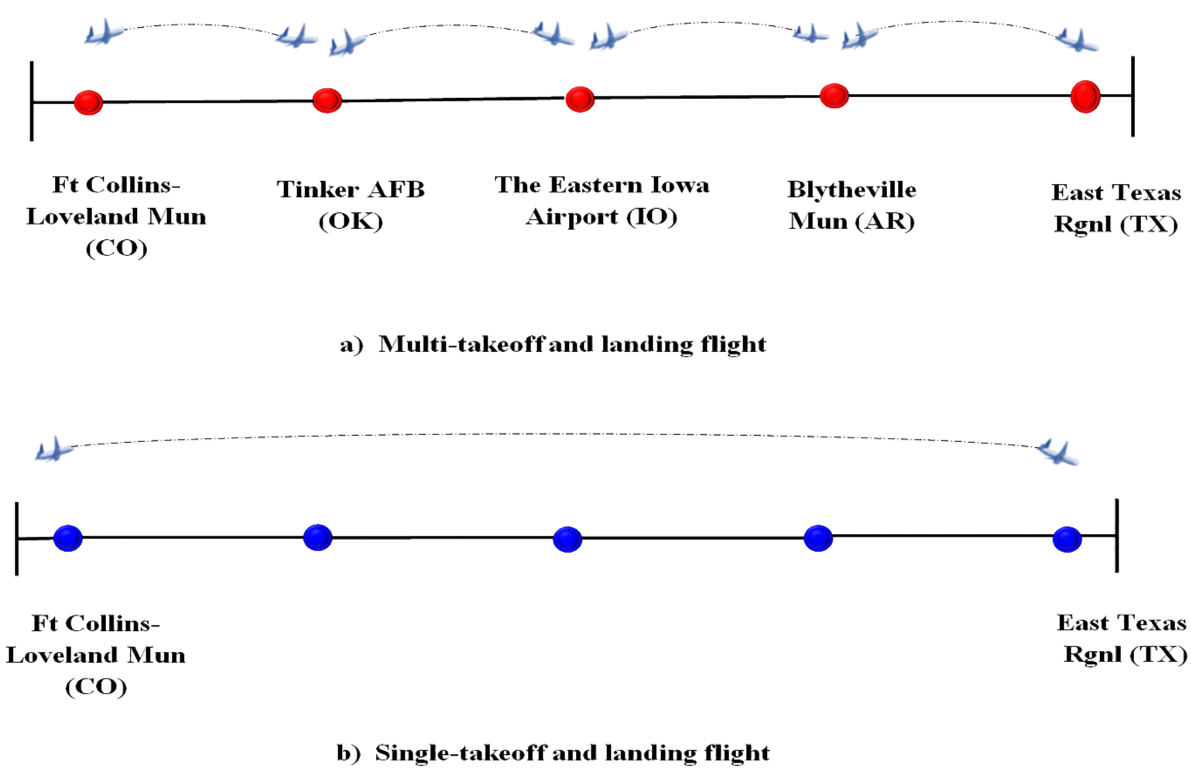


Figure 4.10. Experiment runs timeline. The starting and end of the 4-hour flight missions for both flight scenarios. Blue line (multi- flight mission) and red line (single flight).

4.3.2.3. Tasks

Before the participants signed the consent form, they were first given a brief tutorial on the experiment for each scenario, and then they were seated in front of the computer screen to calibrate their eye movements. Each participant was instructed to make an eye movement toward the target locations that appeared on the screen. Since the four pilots participated in both flight scenarios, there were eight flight missions. Participants were asked to fly the airplane B-52 based on the scheduled scenario using the FSX flight simulator. Figure 4.10 depicts the timeline of the flight scenarios, as actually flown in the simulator. Each pilot started preparing as if he was preparing a real flight operation. He settled into the cockpit interface and reviewed his flight plan. He then began flight with taxing to the runway. Following the taxing, the pilot began preparing for the takeoff: the first step, run up the engines. The second step was to release the flaps downward and then power up the aircraft until the take-off was achieved. The next step was climbing to the defined cruising altitude (twenty thousand feet) and following the defined en-route towards the destination as planned. At the end of each flight, the pilots conducted a routine descent and landed the airplane. In the multi-takeoff and landing flight, the pilot utilized the time between flight takeoff and landing to complete PVT, SP, and KSS tests immediately after landing. In the single-takeoff and landing flight, the pilot completes the tests while on autopilot. These tests were planned to occur at the same time as in the multi-takeoff and landing flight mission. Figure 4.11 depicts the areas of interests (AOIs) designated for this experiment.

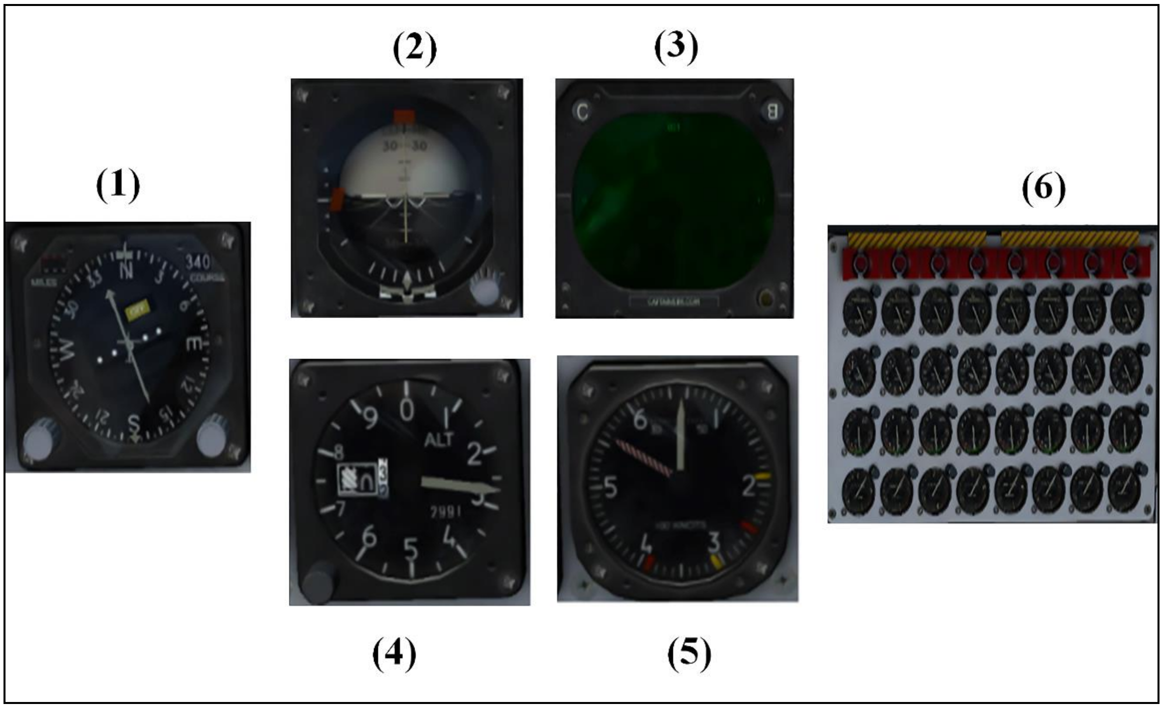


Figure 4.11. AOI's of B-52 Cockpit display: (1) Horizontal situation indicator; (2) Attitude-director indicator (3) Enhanced visual screen (4) Engines indicators (5) Indicated airspeed indicator (6) Altimeter.

4.3.2.4. Experiment apparatus

Hardware: Computer screen (24 inch) to display the simulator, keyboard to control and adjust the cockpit display, Logitech extreme 3D Pro Joystick to fly and control the airplane and Tobii TX300 with a sampling frequency of 300 Hz and visual angle accuracy of 10 was used to collect the eye movement data of the participants. **Software:** Microsoft flight simulator software (FSX) (a full-flight simulator of the Boeing B-52H jet airplane) was used to simulate the cockpit display. Tobii studio was used to analyze the raw eye movement data obtained from the eye tracker. Psychomotor Vigilance task (PVT) was used to measure the reaction time. Fatigue tests also comprised the SP scale, and the KSS. The collected data were analyzed using excel, SAS software version 9.3 and R software.

4.3.2.5. Variables

I. **Independent variables**

The independent variables were the flight types with two levels (single verses multi-takeoff and landing).

Dependent variables

The dependent variables were the mean reaction time (RT), the number of lapses in attention and mean number of false starts, slowest 10% of reaction time, fatigue rating and sleeping rating. In addition, we extracted the following variables from the eye tracking system as a measure of the overall level of pilot visual transition sequences to analyze the visual scan path sampled (i.e. Order of eye fixations and saccades). Each eye-fixation was identified using a built-in velocity threshold identification (or I-VT) algorithm within the Tobii Studio software connected to the Tobii TX300 eye tracking hardware.

4.3.3. Results

4.3.3.1. Psychomotor Vigilance task (PVT) results

During each experimental run, the participants were instructed to conduct the following tests: the 5-minute PVT test, the 7-point Samn-Perelli scale (SP), and the 9-point Karolinska Sleepiness Scale (KSS). The results of PVT tests, both mean RTs and the number of memory lapses found to increase as fatigue increases (See Figure 4.12 & 4.13). More specifically, the Mann-Whitney-Wilcoxon tests showed that there were significant differences in RT ($p < 0.037$) and the number of lapses ($p < 0.047$) between single and multi-takeoff and landing (higher for the multi-takeoff and landing), with the single takeoff and landing having, on average, 103.81 milliseconds faster mean RT and 3.26 fewer lapses. The difference in mean RT and number of memory lapses between the multi-takeoff and landing and single-takeoff and landing appeared to remarkably increase at the end of the experimental duration, i.e. when the flight was about to reach to its destination. Figure 4.14 and 4.15 represents the mean number of false starts versus the rating time and the mean of slowest 10% RT versus the rating time respectively. The mean number of false starts was lower for the single takeoff and landing flight mission (see Figure 4,16). The mean of the slowest 10% of RT was also smaller for the single takeoff and landing than multi-takeoff and landing flight mission (see Figure 4.15). The Mann-Whitney-Wilcoxon tests showed that there were significant differences in the number of false starts ($p < 0.039$) and slowest 10% of RT ($p < 0.032$) and between the single and multi-flight takeoffs and landings. Figure 4.14 and 4.15 also shows there were significant increases in the number of false starts and the slowest 10% of RT for both single and multi-takeoff and landing as the flight went on to the destination.

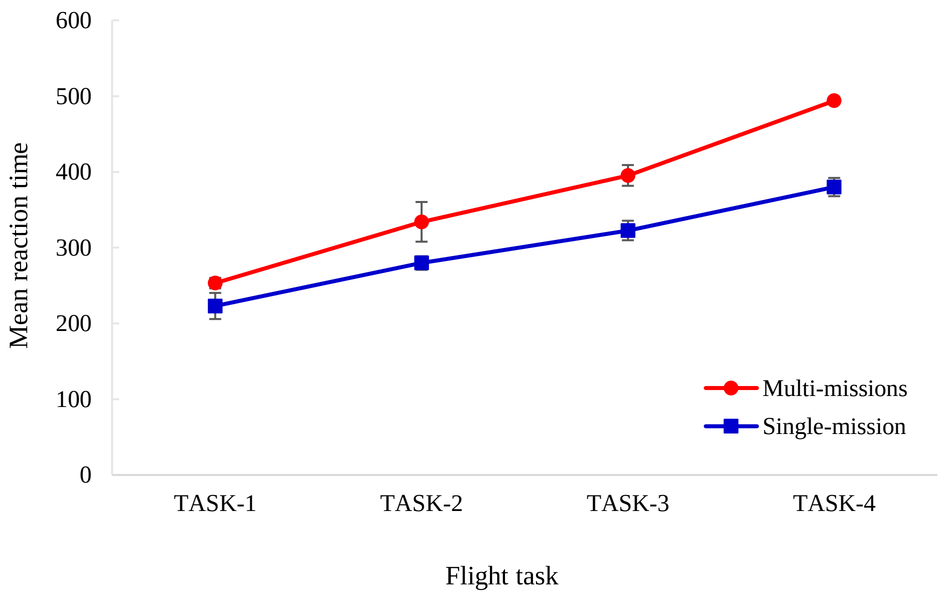


Figure 4.12. Mean reaction time based on flight scenarios

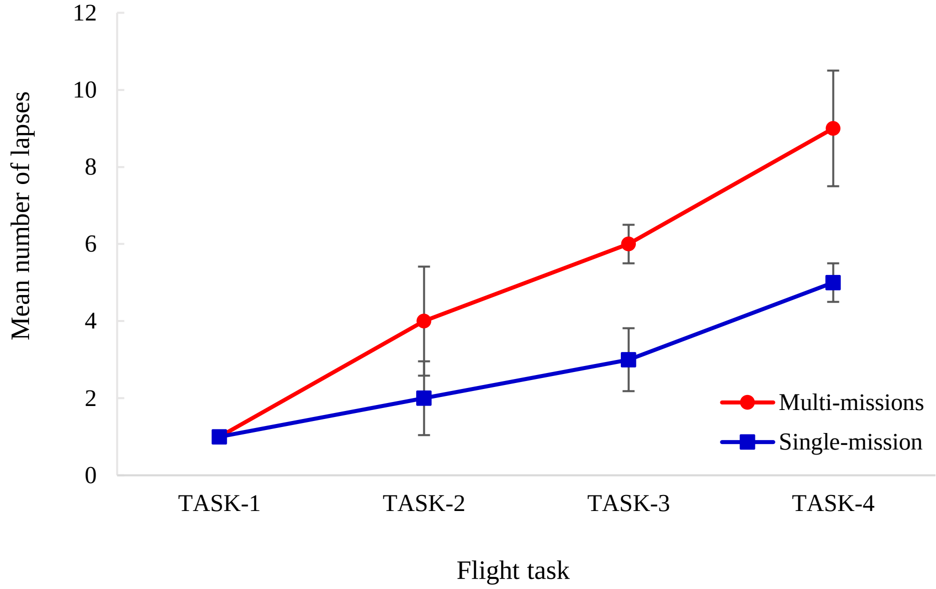


Figure 4.13. Mean number of lapses based on flight scenarios

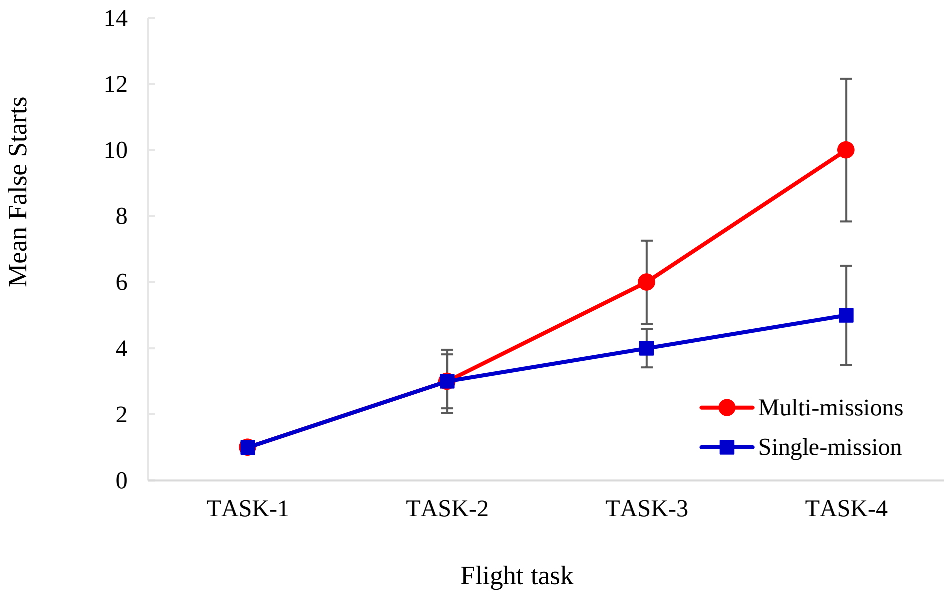


Figure 4.14. Mean Number of false starts based on flight scenarios

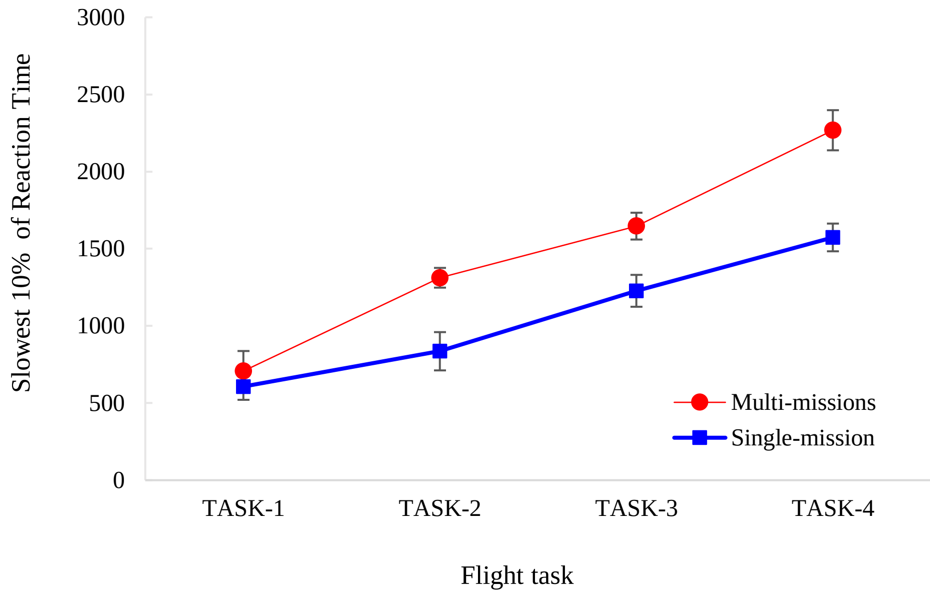


Figure 4.15. Mean Slowest 10% of Reaction Time based on flight scenarios

4.3.3.2. Self-reported fatigue level results

Results of the Samn-Perelli scale and KSS indicate greater subjective fatigue in the multi-takeoff and landing compared to the single-takeoff and landing flight missions (See Figure 4.16 & 4.17). The Mann-Whitney-Wilcoxon tests showed that there were significant differences on KSS scale ($p < 0.039$). The results from SP ratings also showed a statistically significant difference between single and multi-takeoff and landing. The Mann-Whitney-Wilcoxon tests showed that there were significant differences on Samn-Perelli scale ($p < 0.031$). SP and KSS scales show change over time of the flight mission.

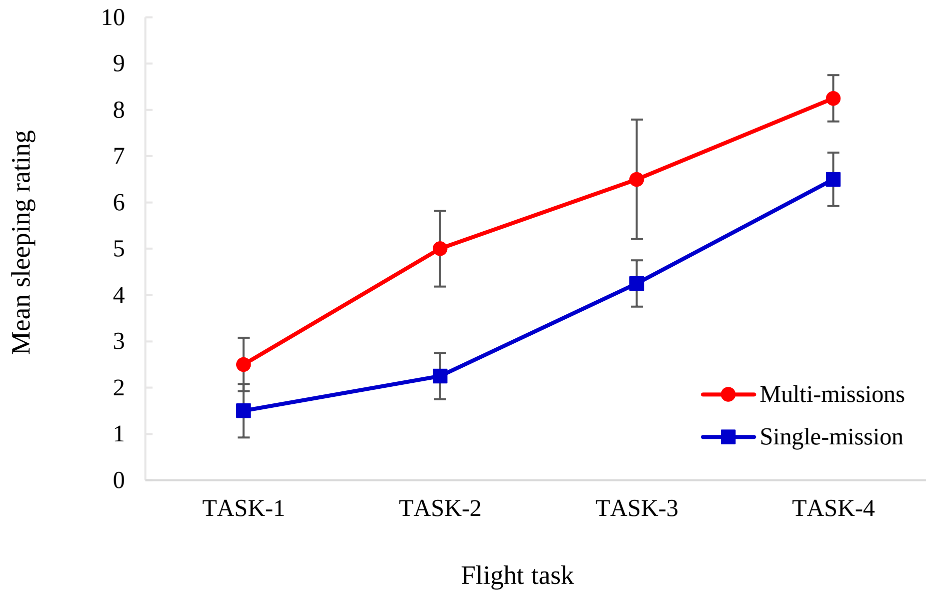


Figure 4.16. Mean sleeping rating based on the flight scenario.

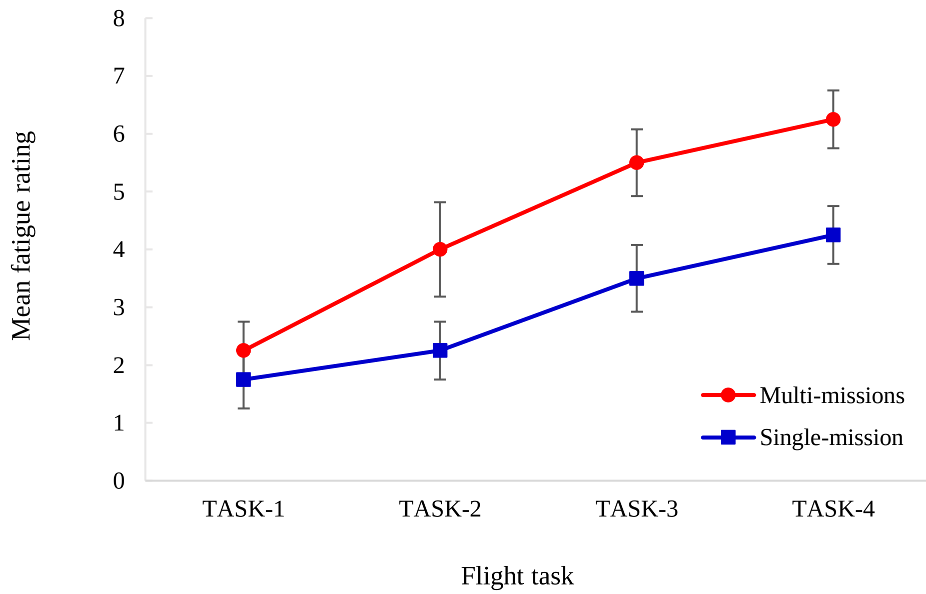


Figure 4.17. Mean fatigue rating based on the flight scenario.

4.3.3.3. Results of eye movement data

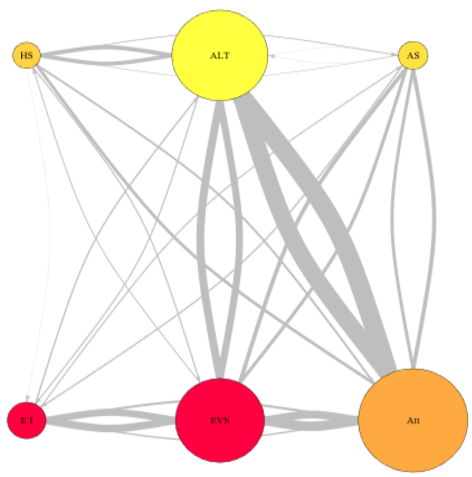
Figure 4.18 shows eight DWNs which represent the average of the eye fixation transitions data of the four pilots for each of the flight missions. The cockpit display is divided into six areas of interests (AOIs): (1) Horizontal situation indicator; (2) Attitude-director indicator (3) Enhanced visual screen (4) Engines indicators (5) Indicated airspeed indicator (6) Altimeter (see Figure 4.2). Each network represents pilot transition sequences during each one of the four flight missions (each lasting for approx. 60 minutes). These network visualizations demonstrate the pilots' eye fixation transition among the various AOIs on the cockpit display. Since one of the major goals of this study is to understand the effect of fatigue on pilot fixation strategy during the single and the multi-takeoff and landing flight missions, the analysis of DWNs was done based on two measures (i) total number of eye fixation transitions between AOIs, (ii) total eye fixation time duration spent on AOIs. As Figure 4.18 shows, there were differences between pilot visual search patterns during the single and the multi-takeoff and landing.

The multi-takeoff and landing DWN of the first flight mission (See Figure 4.18 (a)) shows balanced patterns of visual search among AOIs and they spent more fixation time on AOIs. Likewise, the single takeoff and landing DWN of the first flight mission (See Figure 4.18 (b)) shows a balanced defined visual search pattern.

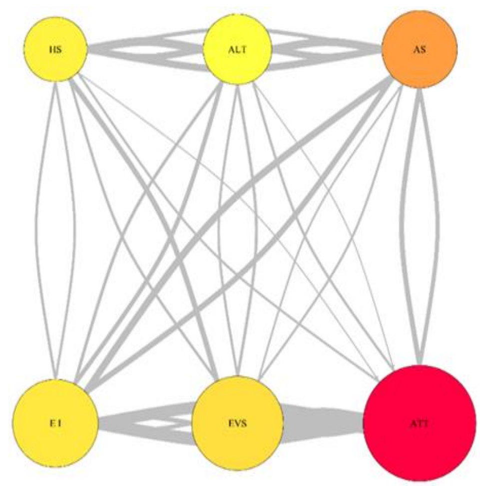
The second DWN of the multi- takeoff and landing shows that the pilots transit their fixation among the AOIs less frequently and they also spent more time on AOIs than they did during the single-takeoff and landing flight ((See Figure 4.18 (c) & (d)). The dominant transition was between ATT, AS and ALT; this would be because these indicators are the most important indicators during the multi-takeoff and landing phases.

The DWN of the third flight mission of the multi-takeoffs and landings phases (See Figure 4.18 (e)), shows that pilots exhibit a weak scanning pattern. The transition frequencies were quite low among AOIs. In contrast, the single-takeoff and landing flight mission DWN (See Figure 4.18 (f)) shows a clearly defined scanning pattern among all AOIs and a shorter fixation duration.

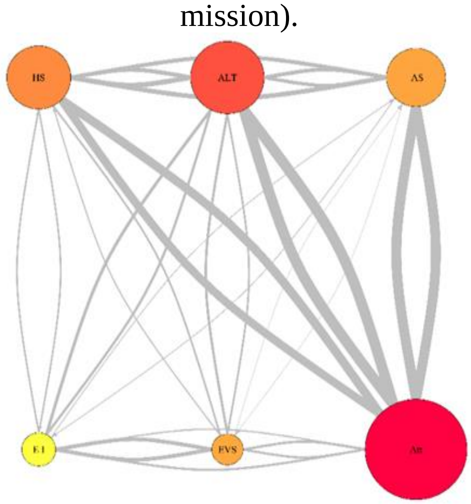
The fourth DWN of the multi-takeoff and landing (See Figure 4.18 (g)) shows that pilots also had exhibited a weakened pattern of visual scanning search among AOIs, the dominant fixation transitions were between the ATT and ALT. The DWN of the fourth flight mission of the single-takeoff and landing (See Figure 4.18 (h)) shows that they have remarkable differences in their scanning pattern. Pilots showed more balanced visual scanning searches among AOIs during the single takeoff and landing than they did during the multi-takeoff and landing. They showed a wider spread of transition and shorter fixation duration on AOIs.



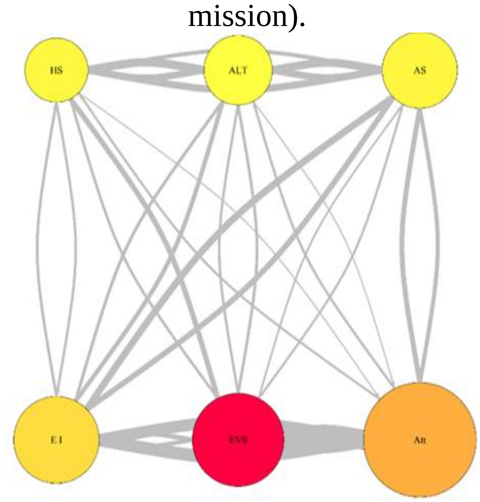
(a) DWN of Multi-takeoff and landing (First mission).



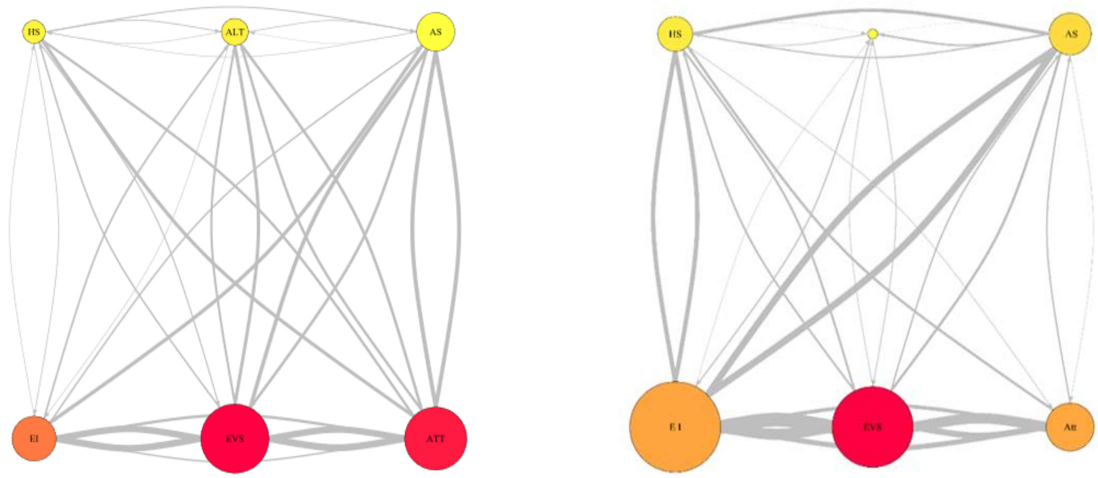
(b) DWN of single takeoff and landing (First mission).



(c) DWN of Multi-takeoff and landing (Second mission).

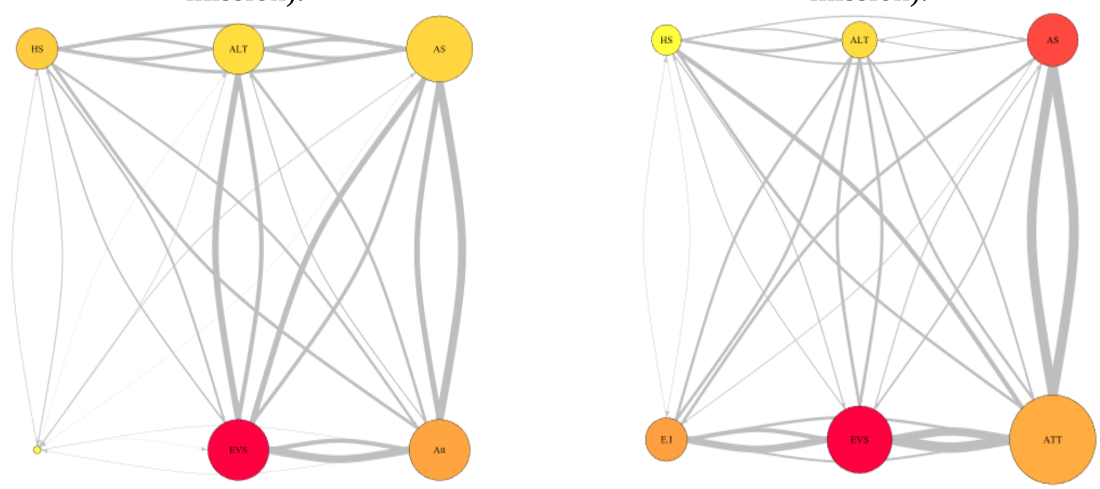


(d) DWN of single takeoff and landing (Second mission).



(e) DWN of Multi-takeoff and landing (Third mission).

(f) DWN of single takeoff and landing (Third mission).



(g) DWN of Multi-takeoff and landing (Fourth mission).

(h) DWN of single takeoff and landing (Fourth mission).

Figure 4.18. Sample visual transition characteristics of the single and the multi-takeoff and landing of an expert pilot on important AOIs

4.3.4. Discussion

In this study, two different scenarios were developed, and more data collection techniques were used to explore the influence of multi-takeoff and landing verses single takeoff and landing on pilots' fatigue level when interacting with cockpit interfaces during prolonged flight missions. Each experimental scenario in this study involved the same conditions except the number of takeoffs and landings phases. These conditions included a fixed configuration of a medium- fidelity flight simulator of the airplane; randomized and counterbalanced participants; all participants were certified to fly; the flights started and ended at the same times; fatigue tests were administrated at the same times. Participants were informed about the fatigue testing procedures before they started the flight scenario to ensure the best effort during testing. Fatigue tests were administered while on autopilot (after take-off) or parked on a taxiway (after landing). The primary measures of fatigue in this study were the RT, mean number of false starts, slowest 10% of RT and number of lapses on the PVT. The secondary measures of fatigue were the Samn–Perelli (SP) fatigue check card (Samn and Perelli, 1982) and Karolinska Sleepiness Scale (Akerstedt, Anund, Axelsson, & Kecklund, 2014). In addition to the visual transition characteristics.

Subjective results (Fig. 4.16 & 4.17) have been shown to follow similar trends to objective results (Fig. 4.13 & 4.16) throughout the flight mission; they show a higher buildup of fatigue in the multi-takeoff and landing flight mission than in the single-takeoff and landing flight mission. Figure 4.13 & 3.16 compare PVTs results during the single-takeoff and landing and multi-takeoff and landing flight missions. As Figures 4.12 and 4.13 show, the number of PVT RTs and lapses scored in the multi-takeoff and landing were higher than in the single-takeoff and landing flight mission. Figure 4.15 also shows that single-takeoff and landing had slowest 10 % RT and a smaller number of false starts. Even though the results demonstrate that pilots performed their tasks in a more efficient way during single-takeoff and landing than multi-takeoff and landing, there were notable increases in RT, the number of lapses and the number of false starts for both single-takeoff and landing and multi-takeoff and landing as the flight went on to the destination. The increased number of take-offs and landings for multi-takeoff and landing could explain this. When a pilot is conducting multi take-offs and landings, he is suffering from a high fatigue level, and there is insufficient capacity for him to process received information quickly. If the fatigue level is too high, pilot might take more time to extract the required information. On the other hand, when a pilot is conducting the same task with a single take-off and landing, he might feel less fatigued and take less time to collect the required information.

Based on the analysis of DWNs (Figure 4.18), pilots' fixation transition and duration are increased as the flight goes to the destination. There is a significant difference in fixation duration between AOIs with different takeoff and landing. The analysis of DWNs (Figure 4.18) shows that the scanning strategy of pilots during the single-takeoff and landing was different from multi-takeoff and landing. Pilots demonstrated more frequent fixations and shorter duration on AOIs during single-takeoff and landing than multi-takeoff and landing. Throughout, single-takeoff and landing pilots demonstrated a stronger and more defined scan pattern than they did in multi-takeoff and landing.

Chapter 5: Analysis of fatigue in novice and expert pilots for prolonged multi-flight missions: multi-modal analysis approach using vigilance test and eye tracking

Author's Note: The contents in Section 5.1 was published as a conference paper, titled "Analyzing pilots' fatigue for prolonged flight missions: Multi-modal analysis approach using vigilance test and eye tracking" in the Proceedings of the Human Factors and Ergonomics Society Annual Meeting (See Naeeri, Mandal and Kang (2019)). The contents in Section 5.2 is currently being written into a journal paper.

5.1. Introduction

Although airplanes have become more advanced and reliable, however, human factors issues, especially pilot error has still been one of the main causes of the major aviation crashes (Avers & Johnson, 2011). The onset of pilot fatigue during aircraft operation is a grave safety concern, as it impacts their ability to stay alert and attentive to the demands required for controlling the flight safely. Various past studies found that high levels of fatigue severely impact pilots' ability to interpret complex information, detect safety threats, and solve inflight emerging problems by providing timely responses (Dismukes et al., 2015; Hartzler, 2014; Lee & Kim, 2018). Furthermore, from the flight safety standpoint, it becomes imminent to evaluate the impact of fatigue for some scenarios, such as commercial flights having a multiple layover itinerary (multiple take-off and landing) with the same pilot crew flying them. Therefore, for the present study, we analyze the impact of fatigue on pilot for multiphase flights. To serve the above-mentioned objective, one pre-requisite is to explore various fatigue measures (both objective and subjective) for their efficacy for the present task situation concerned (i.e. long duration pilot operation). The various fatigue measures can be broadly categorized into three groups (a) Subjective measures, (b) Bio-mathematical model measures, and (c) Objective measures (Gander et al., 2013).

As the name suggests, the various subjective measures, e.g. Samn-Perelli Fatigue rating, Karolinska Sleepiness Scale, mostly consist of self-assessment scores (Gander et al., 2013; Van Drongelen, Van Der Beek, Hlobil, Smid & Boot, 2013; Gander et al., 2014). Although these measures help evaluate how fatigued the person feels, it fails to show the adverse effect of fatigue. According to Goker (2018), the results of subjective measures usually are affected by external and internal factors (Goker, 2018).

On the other hand, bio-mathematical models (e.g. Boeing Alertness Model, and Fatigue Risk Index) mostly deal with predicting the fatigue levels associated with various tasks (Powell et al., 2014). Although these models are helpful in comparing work schedules to develop regulations for duty time limitations, bio-mathematical models cannot definitively answer the question whether the work schedule is acceptable and safe (Williamson, Lombardi, Folkard, Stutts, Courtney & Connor, 2011; Branford, Lowe, Hayward, Pacific & Cabon, 2014). Unlike these, the objective measures consist of physiological (Marandi et al., 2018). The various objective measures, e.g. Electroencephalogram (EEG), Electrooculogram (EOG), psychomotor vigilance tasks (PVT) and eye movement measures, are highly sensitive to changes in the fatigue levels, which allows for an unbiased (in absence of any self-reported data) perspective on the amount of fatigue arising due to different tasks. However, using these measures individually might fail to represent the overall aspect of pilot fatigue.

As an example, EEG technique requires considerable preprocessing because the noise caused by pilots' head and eye movements (Cao et al., 2014). Similarly, EOG data are contaminated by undesirable signals that makes the analysis very complex (Heide et al., 1999).

The PVT measures provides some direct behavioral response measures; however, they require the ongoing tasks to be disturbed before the fatigue data can be collected (Arsintescu et al., 2017; McMahon & Newman, 2015).

Eye tracking technology measures various eye movement characteristics and has shown to be a powerful tool for capturing human visual behavior reliably and conveniently without being intrusive (Eckstein et al., 2017; Mandal et al., 2016; Naeeri & Kang, 2018; Naeeri et al., 2018, Mandal & Kang, 2018). Recent study found a strong correlation between eye movement characteristics (e.g., eye fixation numbers, durations, pupil size and blink rate) and individual cognitive state (Glaholt, 2014). However, basic metrics such as eye fixation number and duration data, provides only the localized aspect of the visual attention deployment. Additionally, blink rate measure has been found to be somewhat noisy and may not be able to monitor fatigue accurately (Thomas et al., 2015). In That Way, other eye tracking metrics needs to be explored that can tackle the above limitations. In this context, gaze visual entropy would be a useful as it provides a single quantitative value to represent the overall attention distribution (Harris, et al., 1986; Shiferaw, Downey et al., 2019).

Gaze entropy-based measures have been applied to examine the observe the pilot visual attention distribution over various simulated flight phases. (Di Nocera, Camilli & Terenzi, 2007). However, the impacts of fatigue level on pilots ‘gaze entropy are yet to explored, especially for long haul flight missions associated with multiple landing and take-off flight missions. Additionally, the impacts of fatigue on pilot’s with different expertise levels has also not been studied. Tackling these issues considering the limitations of common fatigue measurement. we propose a multi-modal approach that is non-intrusive, easier to apply, and analyze. The proposed fatigue measurement multi-modal combines both PVT method and eye tracking method to identify physiological indications of fatigue and then identify a relationship, if any, between fatigue impacts and pilot's eye movement characteristics.

Identify the relationship between pilot fatigue levels and the eye movement characteristics could help us estimate fatigue levels in near real time thus develop the appropriate action to reduce the fatigue levels to the accepted limits. To validate our approach, we conducted a high-fidelity simulated experiment involving a long duration flight operation which consists of multiple takeoff and landing phases to explore how fatigue level affects pilot behavior. Both participated groups of pilots (i.e. experts and novices) performed four standardized (flying under instrument flight rules) prolonged flight missions.

5.2. Study I: Analyzing pilots' fatigue for prolonged flight missions: Multi-modal analysis approach using vigilance test and eye tracking

5.2.1. Purpose of the study

The purpose of this research is to conduct a multi-modal analysis by combining the PVT fatigue measures and eye movement characteristics (i.e. entropy, eye fixation numbers, and eye fixation durations) to investigate whether those measures might show mutual supporting or possibly conflicting evidence in measuring pilots' fatigue for a prolonged flight mission composed of multiple tasks.

5.2.2. Methods

5.2.2.1. Participants

Six pilots from the Department of Aviation at the University of Oklahoma were recruited. All of them had been rated and on average had 18 months of experience with normal vision. The age parameters are: Range = 21-29 yrs., Mean = yrs., SD = 2.65 years.

1.6.1.1. Apparatus

Software specifications: Microsoft light simulator software (FSX) for generating airplane (Boeing B-52 model) flight scenario. Psychomotor Vigilance task (PVT) was administered to measure pilot fatigue. The collected data were analyzed using SAS software version 9.3 and R. Hardware specifications: 21-inch monitor for displaying simulated flight scenarios. Logitech extreme 3D Pro Joystick to control the simulated aircraft. To collect EM data, eye tracking system Tobii TX300 with a sampling frequency of 300 Hz was used. The visual angle accuracy of the eye tracker is 0.5 degree.

5.2.2.2. Scenario

The experimental scenario lasted approximately four hours and involved four flight tasks each lasting one hour (each task involved take-offs and landings). The participants performed the flight tasks under instrument flight rules (IFR). The last flight task involved responding to an extreme event (i.e., extreme weather conditions, engine failure).

5.2.2.3. Tasks and procedures

The complete task involved flying a plane for four hours, which was further divided into four sub tasks, each of one-hour duration. Task -1: Take off from Ft Collins-Loveland, Mun (CO) airport and arrive at Tinker AFB (OK); task -2: Take off from Tinker AFB (OK) and land at Eastern Iowa Airport (IO); task -3: Take off from Eastern Iowa Airport (IO) and arrive at Blytheville Mun (AR); task -4: Take off from Blytheville Mun (AR) and arrive at East Texas Rgnl (TX) airport. The flight departure time for the task - 1 was 08:15, and flight arrival time for the task - 4 was 13:00. The duty period began at 07:45 am allowing for 30 min of preparation time and ended at 13:15 pm (total of 5 hrs. and 30 mins on duty). The pilots were instructed to be ready to respond to any emergencies. All flights were conducted in accordance with FAA instrument flight rules (IFR).

5.2.2.4. Data analysis

We divided the experimental duration into four equal parts of one hour each for the analysis, each corresponding to one task. The eye tracking data were analyzed using MATLAB and R software. The PVT data were analyzed using the SAS statistics program (version 9.3). No data were excluded from the analyses. Kruskal-Wallis test was undertaken to find the effect of the task sequences on the pilots' fatigue level. Post hoc analysis was also done to find pairwise differences between PVT and eye movement attributes for various tasks.

5.2.2.5. Entropy calculation

For the entropy calculations we only considered 14 different AOIs as shown in Figure 5.1. AOI transition frequency matrix was calculated from the AOI fixation sequence as per the design principles described in Mandal et al. (2016) and Krejtz, Szmidt, Duchowski, and Krejtz. (2014). The $\boldsymbol{i^{th}}$ row and $\boldsymbol{j^{th}}$ column element of the transition frequency matrix $\boldsymbol{p_{ij}}$ can be interpreted as the probability of eye fixation transition from $\boldsymbol{i^{th}}$ AOI to the $\boldsymbol{j^{th}}$ AOI (Holmqvist et al., 2011). The complexity associated with this transition frequency matrix can be evaluated by the two different entropy measures namely, stationary visual entropy and transition visual entropy (Krejtz et al., 2014). Stationary visual entropy represents the randomness associated with the cumulative distribution of visual attention among AOIs, while transition visual entropy indicates level of uncertainty associated with switching of eye fixations among various AOIs (Acartürk & Habel, 2012). Stationary visual entropy is calculated as follows.

Stationary visual entropy $(H_s) = -\sum_{i \in S} \pi_i \; log \, \pi_i$ (1)

Here, π is the stationary distribution associated with AOI transition frequency matrix, and i is the index of AOIs. Thus, πi shows the overall share of the visual attention devoted to the i^{th} AOI. Higher π_i means more fixations on the i^{th}AOI. S represents the set of AOIs. High value of Hs indicates that the pilot's visual attention tends to be distributed more evenly among all AOIs, otherwise, a low value means that the fixations are concentrated on specific AOIs (Krejtz et al., 2014). Transition visual entropy is calculated as follows.

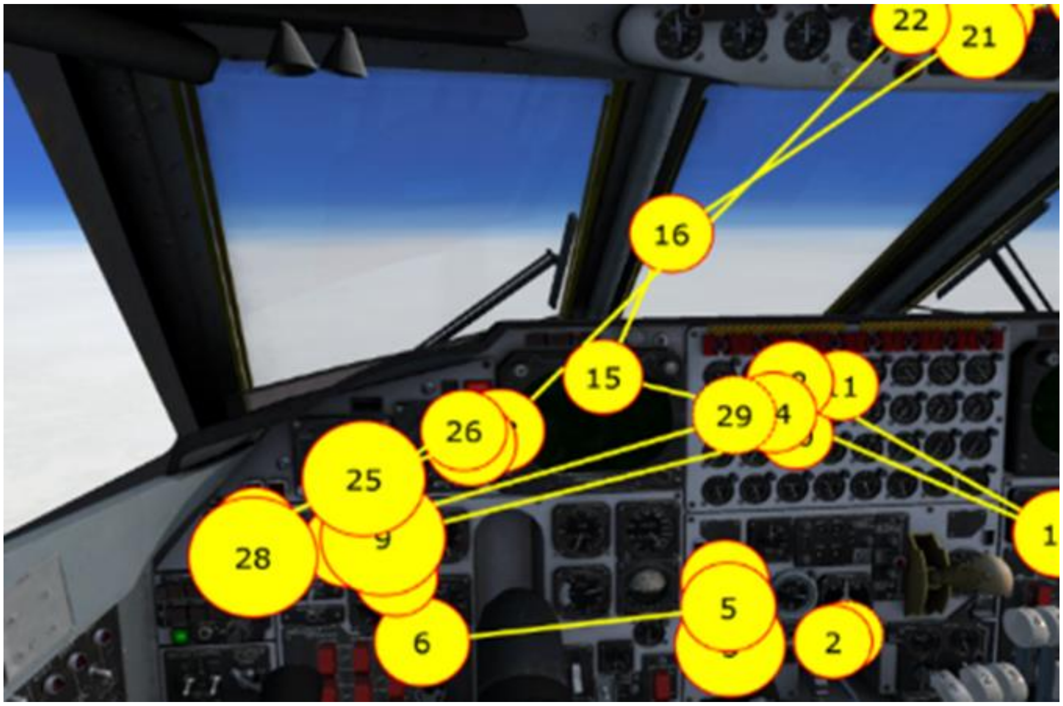
Transition visual entropy $(H_t) = \sum_{i \in S} \pi_i \sum_{i \in S} p_{ij} \, log \, (p_{ij})$ (2)

Here, p_{ij} is the probability of transitioning from i^{th} AOI to j^{th} AOI, where $i \neq j$. H_t represents the randomness associated with switching of visual attention among AOIs. Higher $H_t$values suggest more random transitions among AOIs and vice-versa.

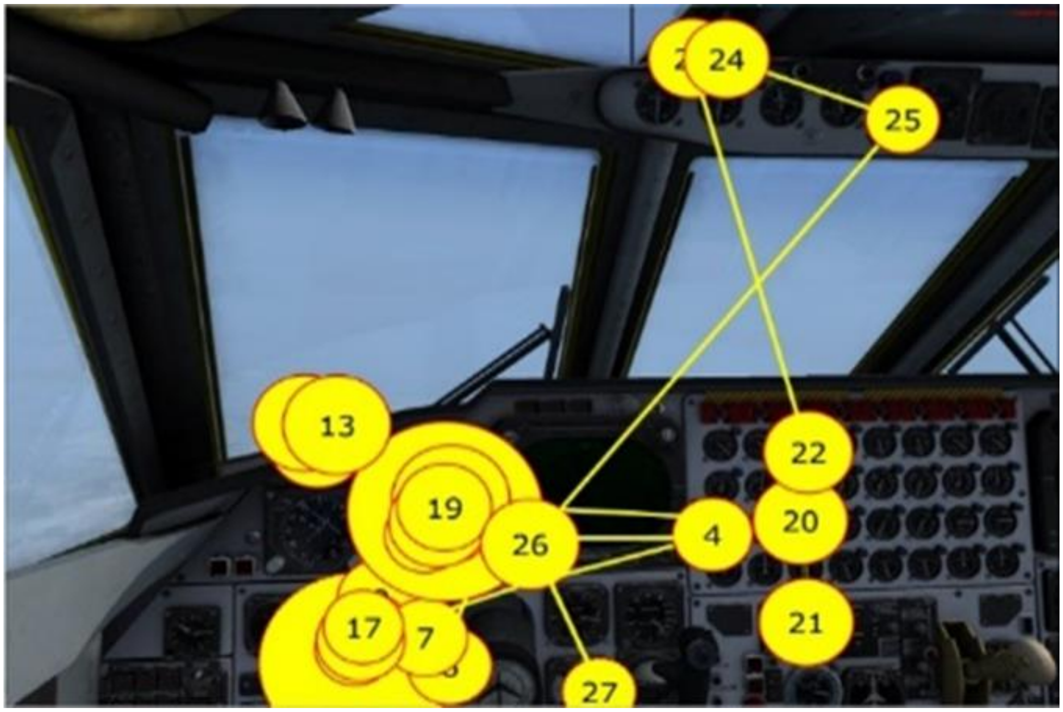
5.2.3. Results

5.2.3.1. Visual scan path

Figure 5.2 shows examples of two different visual scan path plots for the same participant for 10 seconds duration just prior to the finishing flight tasks, task -1 and task -4. The Figure shows change in the visual scanning strategy of the same pilot for the same landing phases (as the time in very near the completion of the two tasks) with progress in the flight mission. The visual scan path for task -4 shows more instance of longer fixation duration as compared to task -1.



(a) 10 seconds prior to finishing Task 1.
(b)



(b) 10 seconds prior to finishing Task 4.

Figure 5.1. Examples of visual scan path plots for one pilot, prior to just finishing (10 seconds before) the two tasks, task -1 and task- 4. The numbers show the eye fixation index, and the size of the circle is proportional to the fixation duration. The yellow lines represent the saccades.

5.2.3.2. Eye fixation number and duration

Figure 5.3 shows the change in mean eye fixation number for various tasks, where we can see it is steadily decreasing over the course of flight. A Kruskal-Wallis test showed that there exists a significant difference in average number of fixations for various tasks ($H(3) = 14.60$, $p < .0022$). The post-hoc tests demonstrated that the average fixation number at task -4 was significantly different compared to the other three tasks. Unlike this, the mean eye fixation duration showed an increasing trend over the course of flight (Figure 5.4). The mean duration was significantly affected by flight tasks sequence ($H(3) = 15.56$, $p < .0014$), as it was significantly higher for task -4 compared to other three tasks. The post-hoc tests showed that there was a significant difference between task -4 and task -1 and task -2 also. There was no significant difference between tasks 1 and 2, and between tasks 2 and 3. All the other comparisons were shown to be significantly different.

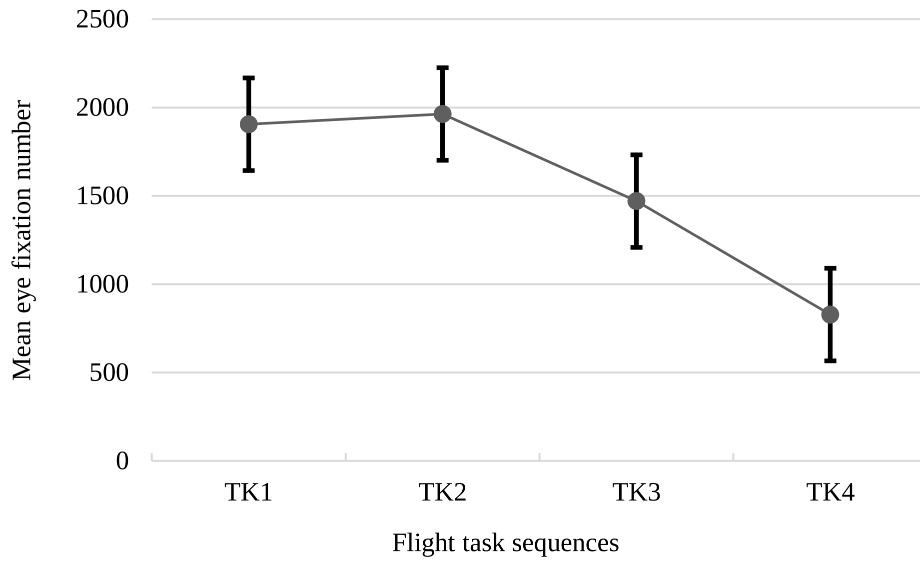


Figure 5.2. Plot of mean eye fixation number for various flight tasks. TK_i shows the i^{th} task, where, $i \in \{1, 2, 3, 4\}$.

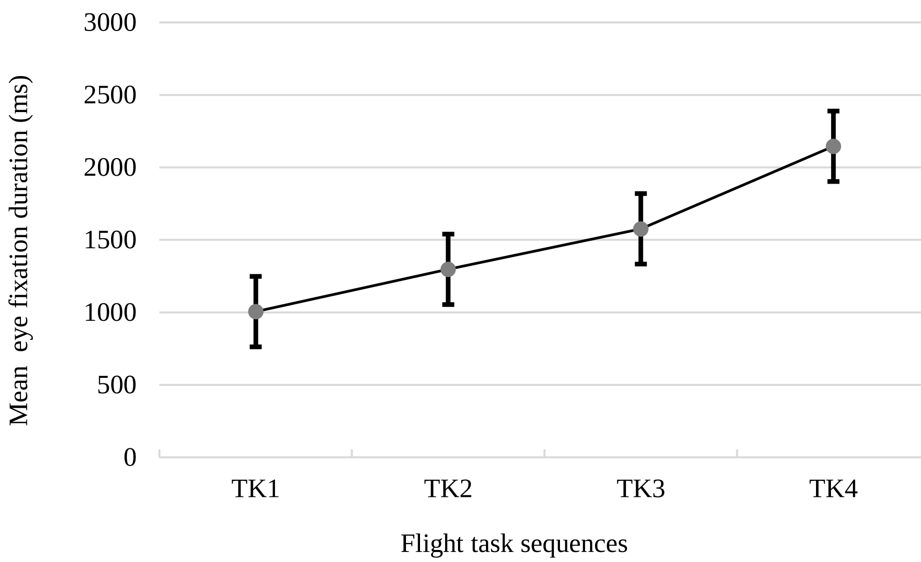


Figure 5.3. Plot of mean eye fixation duration for various flight tasks. TK_i shows the i^{th} task, where, $i \in \{1, 2, 3, 4\}$.

5.2.3.3. PVT results

The evolution of the NL and RT for six participating pilots is shown in Figure 5.5. Both mean RT and NL have been found to increase steadily over the course of flight. On average pilots became approx. 50.60 milliseconds slower per task. A Kruskal-Wallis test showed that there exists significant difference in the mean RT value across various tasks ($H(3) = 21.59$, $p < .0001$). The mean RT was shorter for task -1 and longest at task -4 (Figures 5.5- 5.10). The post-hoc tests demonstrated that the RTs at task -4 were different from RTs of task -1 and task -2, but not for TASK-3. All the other comparisons showed significant difference. Similarly, the NL were also significantly affected by flight tasks sequence ($H(3) = 19.24$, $p < .0002$). The NL was significantly higher for task-4 as compared to. The post-hoc tests showed that there was a significant difference between task 4 and all other three tasks, while there was no difference between task -1 and task -2. All the other comparisons were shown to be significantly different.

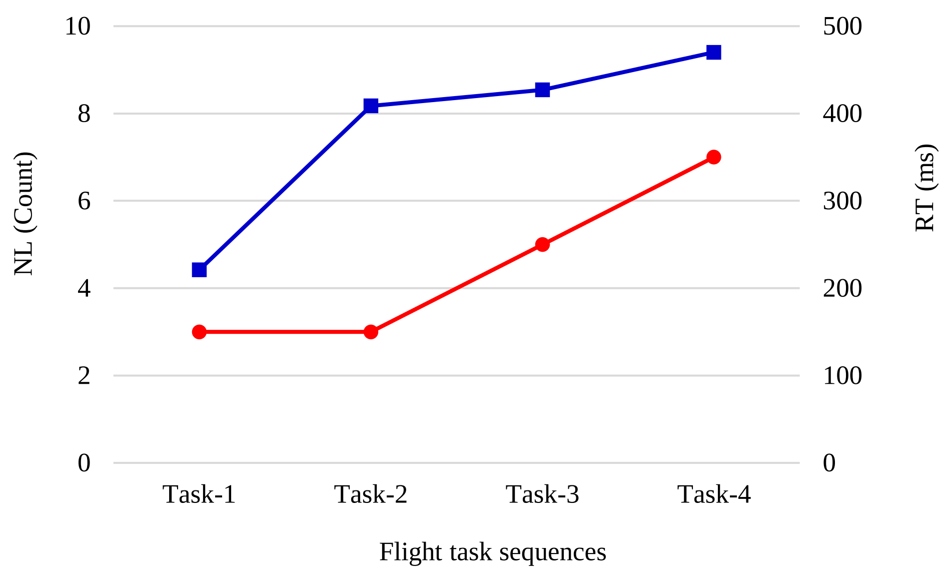


Figure 5.4. Plot of the NL and RT for participant 1 for four different tasks.

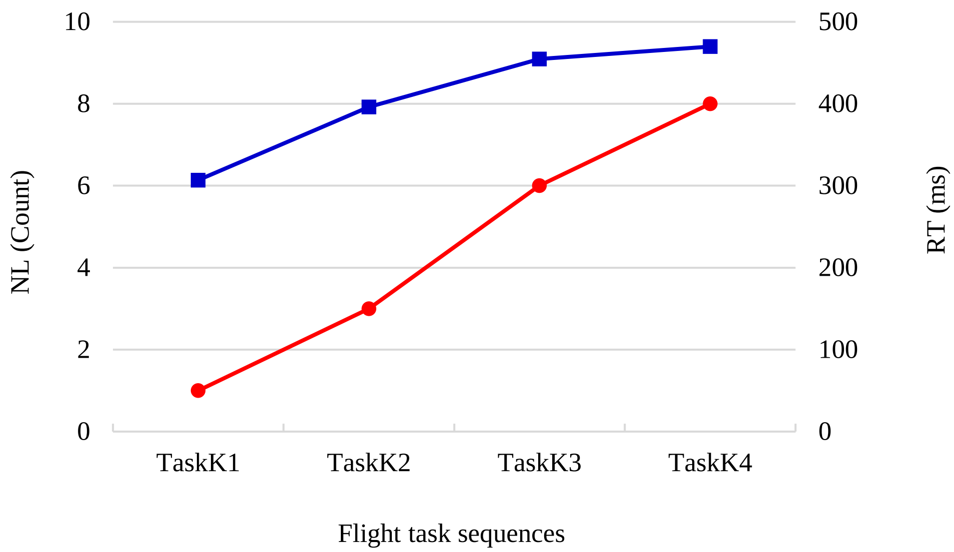


Figure 5.5. Plot of the NL and RT for participant 2 for four different tasks.

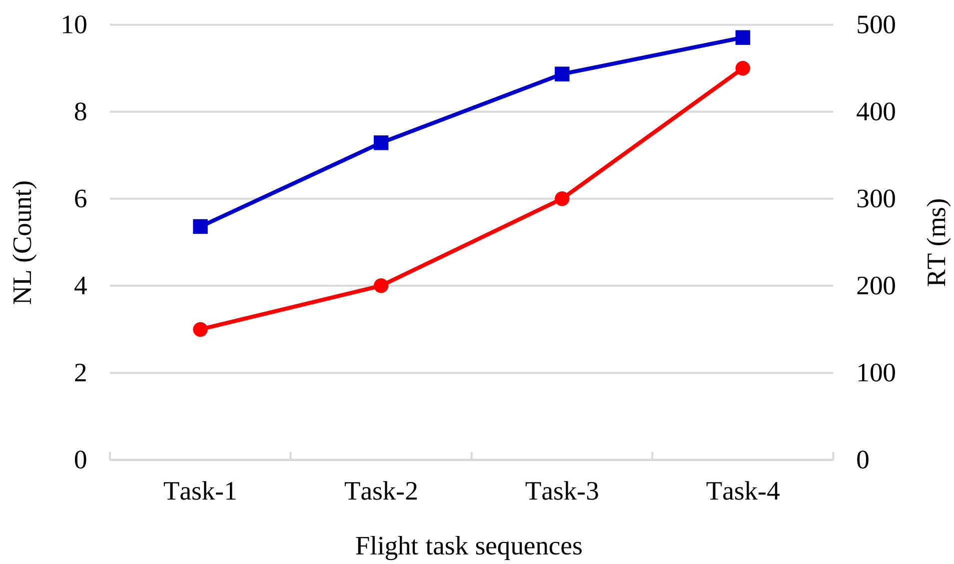


Figure 5.6. Plot of the NL and RT for participant 3 for four different tasks.

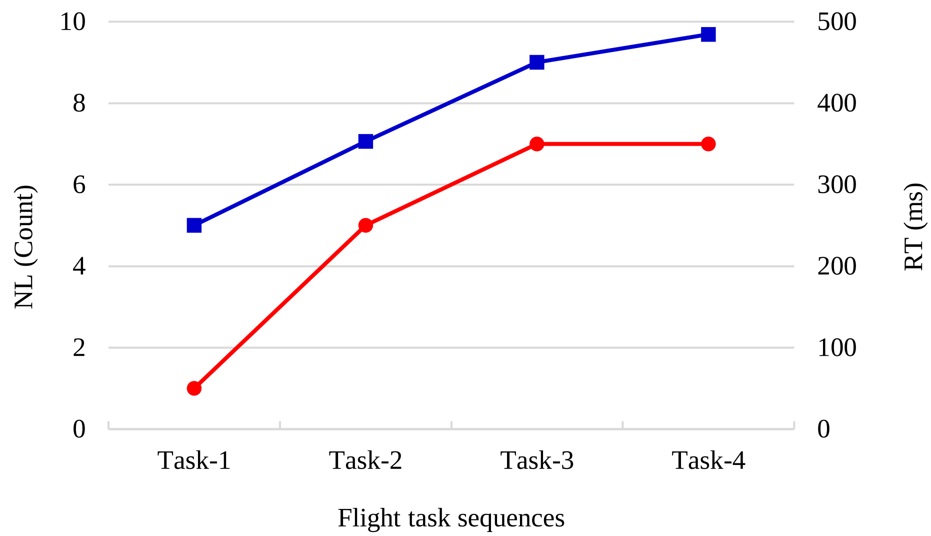


Figure 5.7. Plot of the NL and RT for participant 4 for four different tasks.

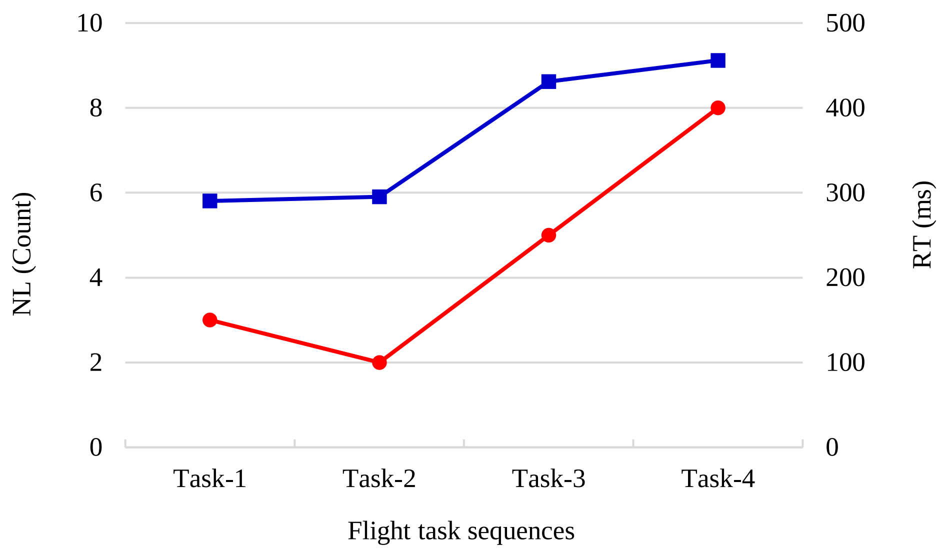


Figure 5.8. Plot of the NL and RT for participant 5 for four different tasks.

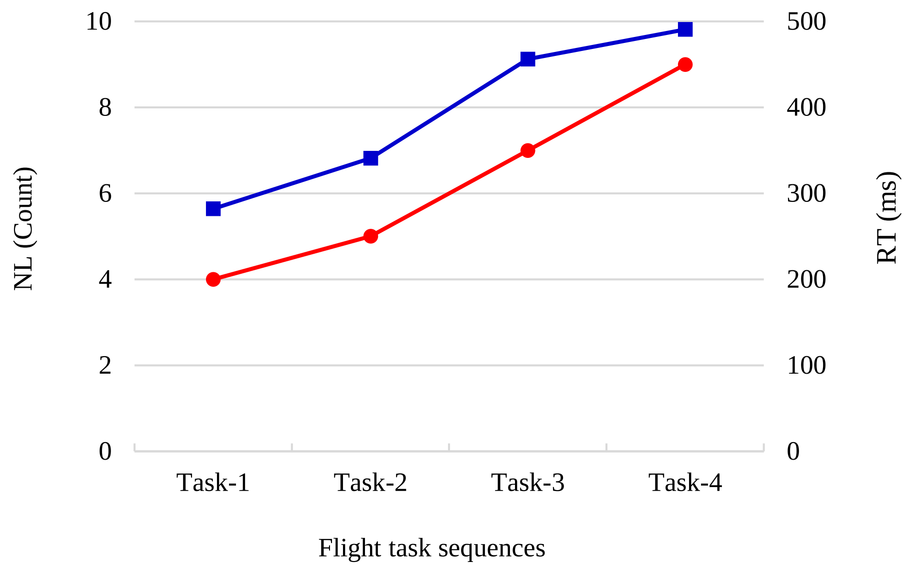


Figure 5.9. Plot of the NL and RT for participant 6 for four different tasks.

5.2.3.4. Entropy results

Figures 5.11- 5.16 show a steady increase in values of $\boldsymbol{H_s}$ as more tasks are being performed. Kruskal-Wallis test also showed that there was a statistically significant difference in $\boldsymbol{H_s}$ and $\boldsymbol{H_t}$. The $\boldsymbol{H_s}$was significantly affected by flight tasks sequence (H (3) = 15.57, p <.0023). The post-hoc tests showed that there was a difference task-4 and all other three tasks, while there was no difference between task -2 and task -3. All the other comparisons were shown to be significantly different. Figure 4 also shows that $\boldsymbol{H_t}$ was significantly higher for task -4 compared to all other tasks. The $\boldsymbol{H_t}$ was also significantly affected by the number of the task (H (3) = 15.90, p <.0012). The post-hoc tests showed that there was a difference between task -4 and tasks task - 1 and task -2, while there was no difference between task -4 and task -3 as well as between task - 1 and task -2. All the other were shown to be significantly different.

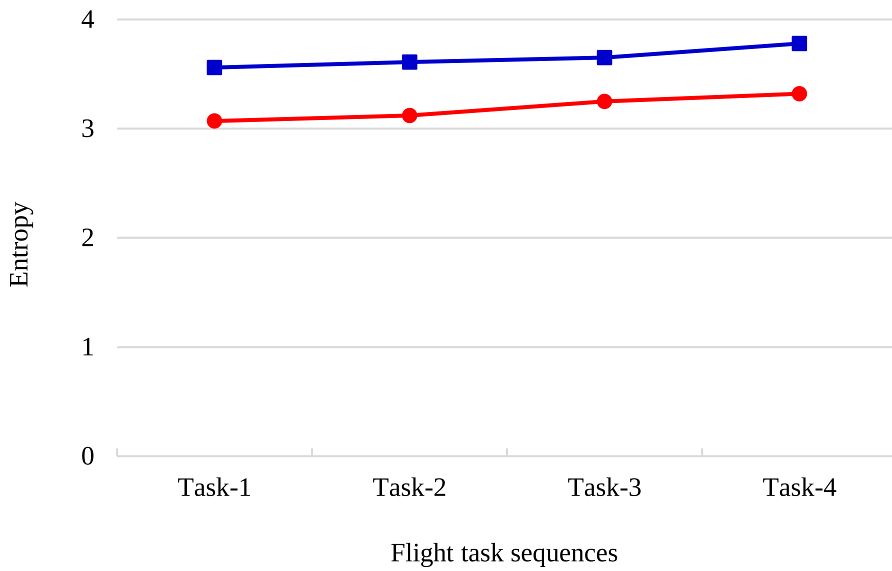


Figure 5.10. Participant 1 mean entropy values (both transition and stationary state).

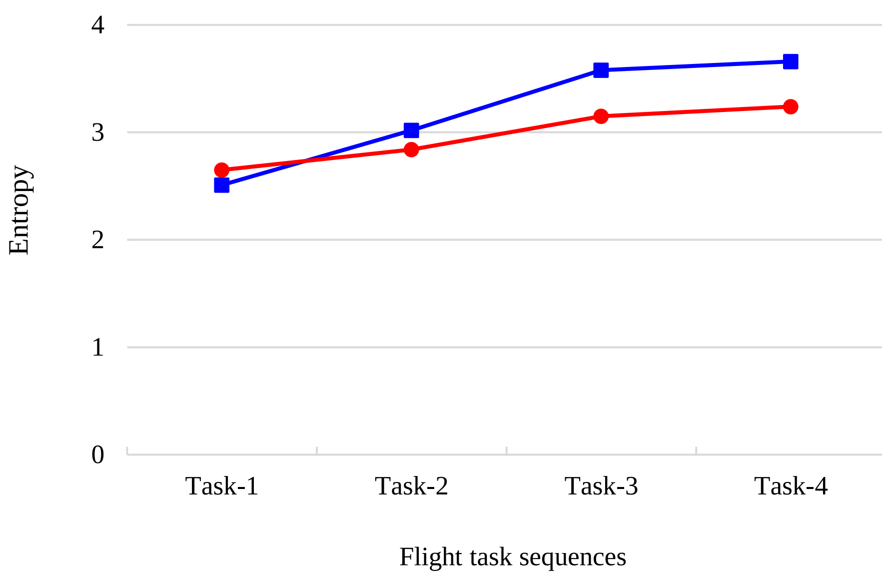


Figure 5.11. Participant 2 mean entropy values (both transition and stationary state).

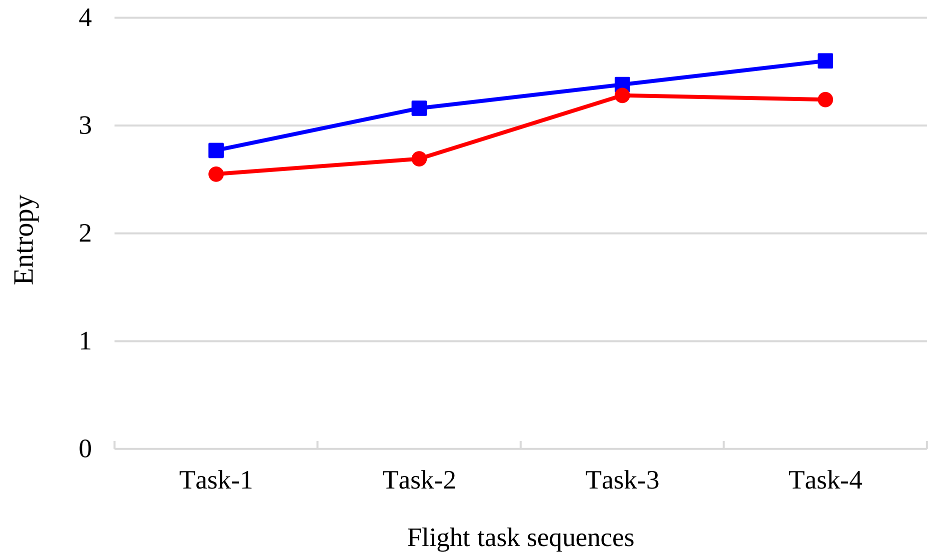


Figure 5.12. Participant 3 mean entropy values (both transition and stationary state).

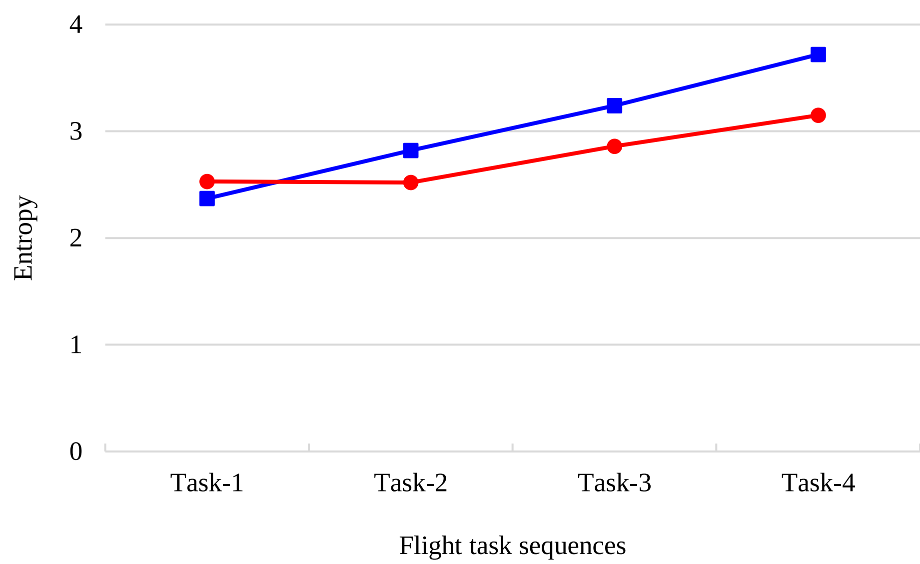


Figure 5.13. Participant 4 eye movement data mean entropy values (both transition and stationary state).

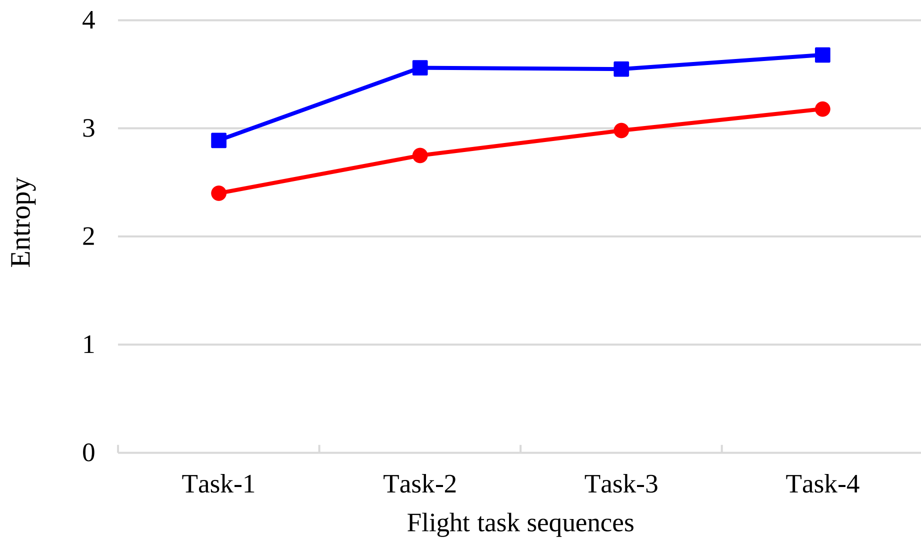


Figure 5.14. Participant 5 mean entropy values (both transition and stationary state).

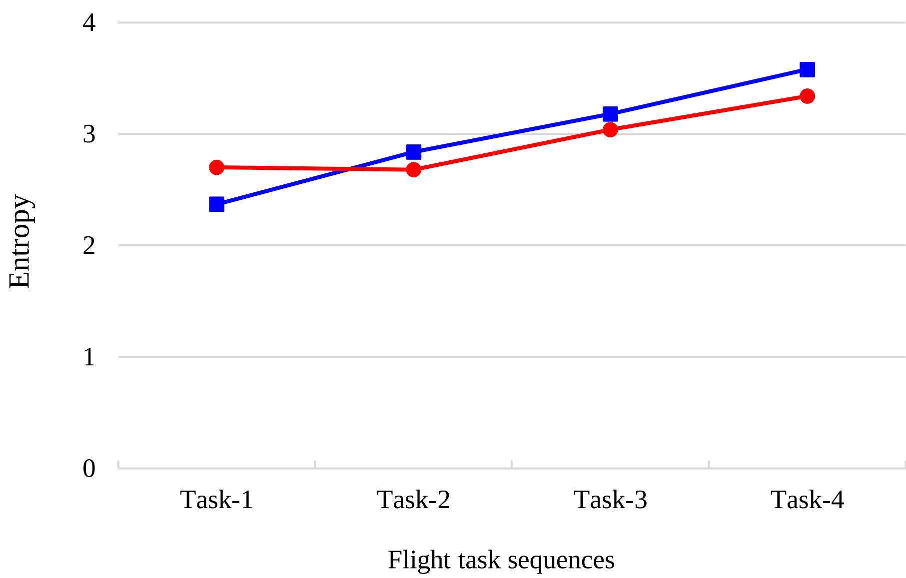


Figure 5.15. Participant 6 mean entropy values (both transition and stationary state).

5.2.4. Discussion

This study examined the impact of fatigue on the pilots for long duration flights. The four major findings which occurred with increase in the flight duration are as follows: (1) decrease in the mean number of eye fixations, (2) increase in the mean eye fixation duration, (3) increased mean RT and NL values, showing that fatigue impacted pilots', and (4) pilots' eye movements became more random, as shown by increase in the both the entropy measures H_s and H_t.

The various trends observed in the three-eye movement characteristics for instance number of eye fixations, durations and entropy demonstrates tendencies of deterioration of pilots' concentration. These results are in accordance with the previous research findings that suggest uncertainty in search pattern (more random eye fixation behavior as shown by increase in entropy values) is related to more difficulty in extracting the required information (increased eye fixation duration) (Harris et al., 1986).

In addition, the increase in RT values from task -1 to task -4 suggests that as the flight task sequences increased, it might have impaired pilots. Moreover, a higher NL value for task -4 as compared to earlier tasks demonstrates that as fatigue induces with progress of time, it negatively impacts pilots' capacity to control aircraft especially during emergencies and prolonged flights (Ziv & Ziv, 2017).

5.3. Study II: Comparative analysis of fatigue in novice and expert pilots for prolonged multi-flight missions: Multi-modal analysis approach using vigilance test and eye tracking

5.3.1. Purpose of the study

The purpose of this study is investigating effects of pilots' fatigue on gaze entropy, especially for long haul flight missions associated with multiple landing and take-off phases. Furthermore, how fatigue impacts pilots with different expertise levels has also been explored. Some other questions that include whether the increase in fatigue level follows a linear trend or not and if it follows the same trend among both novices and experts have also investigated.

5.3.2. Methods

Ten expert and novice pilots participated in a four-hour simulated multi-phased (multi-takeoffs & landings) flight mission. The methodology applied consists of a multi-modal analysis combining eye tracking measures (i.e. eye fixation numbers and durations, gaze entropy, pupil size) to measure visual scanning behavior and PVT (i.e., reaction time, false starts and number of lapses) indirectly measure pilot fatigue level after each flight task. An experiment was conducted to explore expert–novice differences in visual scanning behavior under the effect of the fatigue.

5.3.2.1. Participants

Twenty pilots (10 novices and 10 experts) were recruited from the Department of Aviation at the University of Oklahoma (see details in Table 5.1). The study was approved by the University of Oklahoma Institutional Review Board (IRB). All participants have signed the informed consent. The data collected from the individual pilots have been kept confidential and were not shared with any parties.

Table 5.1. Descriptive statistics of participants age and experience.

Subject	Age (years)			Experience (months)		
	Mean	SD	Range	Mean	SD	Range
Novice	24.8	2.66	29-21	18	2.4	22-14
Expert	30.9	2.97	36-28	42	4.5	51-37

5.3.2.2. Apparatus

Software specifications: Microsoft light simulator software (FSX) was used for generating the airplane (Boeing B-52 model) flight scenario simulations. The PVT was performed on the Psychology Experiment Building Language (PEBL) software version 0.13 (Mueller & Piper, 2014). Tobi Studio was used to extract eye fixation data from the raw eye tracking collected during the experiment. The collected data were analyzed using SPSS software version 25, SAS software version 9.3, MATLAB and R software.

Hardware specifications: A 21-inch monitor was used for displaying the simulated flight scenarios. Logitech extreme 3D Pro Joystick was applied to control the simulated aircraft in a given scenario. A keyboard was used to provide the responses required for the PVT task. To collect EM data, eye-tracking system Tobii TX300 with a sampling frequency of 300 Hz was used. The visual angle accuracy of the eye tracker is 0.5 degree.

5.3.2.3. Fatigue detection model

The applied fatigue detection model involves two types of measures (see Figure 5.16), that would be useful for inferring mental fatigue: The eye-tracking measures sets which have been used for characterizing eye movements in other domains to infer brain diseases (Hopstaken et al., 2015).

The PVT measures were used in previous studies on mental fatigue during cognitive tasks (Roma et al., 2010; Wilson et al., 2007).

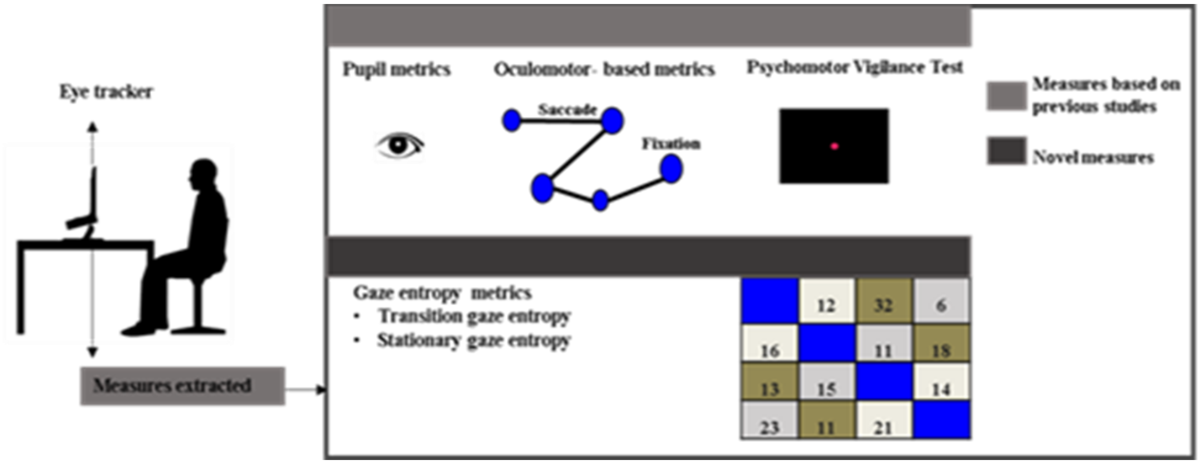


Figure 5.16. An overview of fatigue-detection model

5.3.2.4. Tasks and procedures

The full-flight simulator experiment took place at the Department of Industrial and System Engineering at the University of Oklahoma. Before each experiment, the researcher explained the study procedures to each participating pilot. The eye tracking system was set up and we performed the calibration procedures. The calibration procedure was repeated between flight tasks so as to ensure accuracy of the eye tracking data. Right after each flight task, pilots underwent the PVT which lasts for about 5 minutes. The PVT testing was administered 4 times at the end of each flight task. The response variables extracted from the PVT task were Mean reaction times, Total PVT false starts (i.e. defined as the number of reaction times less than 150 milliseconds) and Total number of lapses (i.e. number of reaction times greater than 500 milliseconds). Figure 5.18 shows the overall sequential organization of various tasks. Since we used simulator software, the runway configuration was similar among all airports and no other aircraft were existed on the runway. All participants were asked to not share the contents of the experiment with their colleagues.

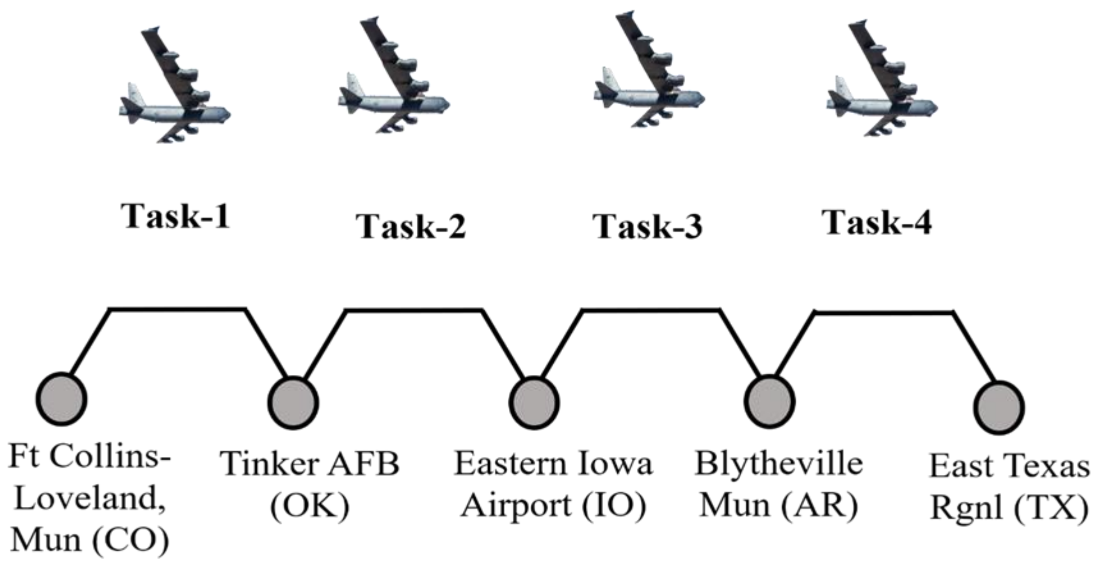


Figure 5.17. Timeline of the various tasks involved in the experimental process, there are four tasks each lasting approximately for 60 minutes.

5.3.2.5. Simulator flights

Pilots flew the high-fidelity fixed-base full-flight simulator Boeing B-52 airplane. The complete scenario involved flying a plane for four hours, which was further divided into four sub tasks of one hour each (see Figure 5.18). Each sub task involved takeoff and landing phases. All flights were conducted in accordance with FAA instrument flight rules (IFR). Figure 5.19 shows all the 14 AOIs of B-52 flight deck considered for the analysis. The various AIOs and their associated code names are as follows: Engines oil pressure (EOP); Engines indicators (EI); Enhanced visual screen (EVS); Attitude indicator (ATT); horizontal situation indicator (HS); flight command indicator (FC); Altimeter (ALT); Airspeed indicator (AS); True airspeed indicator (TS); Heading indicator (HI); Vertical velocity indicator (VV); Radar altimeter (RA); Mach indicator (MI); Standby horizon indicator (SHS).

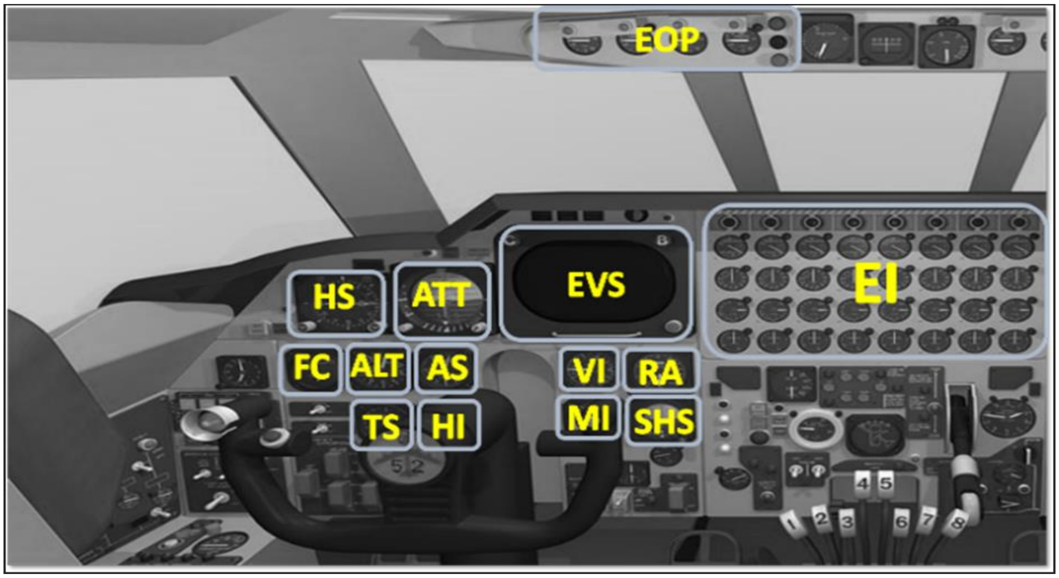


Figure 5.18. The various AOIs considered for the analysis. The yellow characters represent the code name used for various AOIs.

5.3.2.6. Measurements techniques

5.3.2.6.1. PVT

PVT was held for approximately five-minutes after each task was finished (Lim and Dinges, 2008). The PVT testing is highly sensitive to fatigue level (Balkin et al., 2004). It is a simple reaction time task with stimuli occurring randomly at intervals of 2–10 seconds. The PVT testing was administered 4 times at the end of each flight task. The response variables extracted from the PVT task were Mean reaction times, Total number of lapses (i.e. number of reaction times greater than 500 milliseconds) (Lee et al., 2010).), and Total PVT false starts (i.e. defined as the number of reaction times less than 150 milliseconds) (Basner et al., 2011).

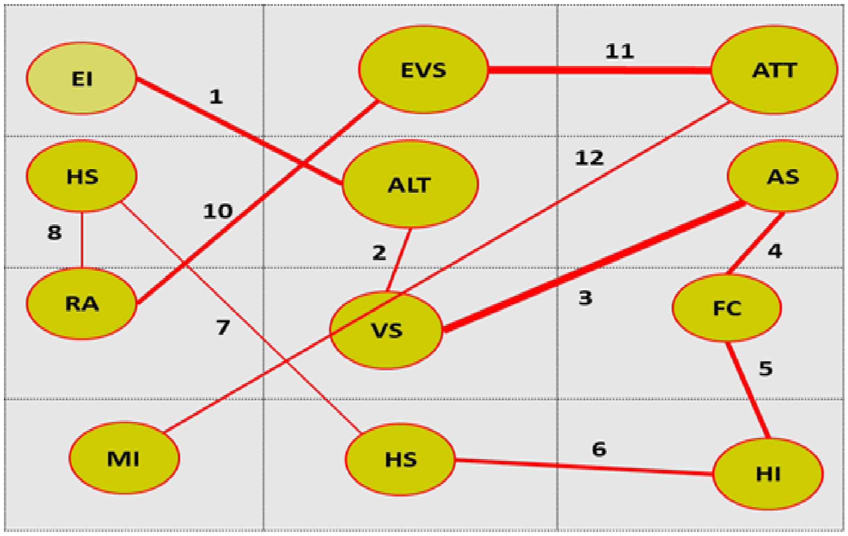
5.3.2.6.2. Eye movement

The response variables extracted from the eye tracking data were, (i) Number of eye fixations (i.e. the total count of eye fixations falling inside an AOI boundary), (ii) Eye fixation duration (i.e. the time for which the gaze remains approximately fixated at a certain location), (iii) Pupil size, and (iv) Gaze entropy (both transition and stationary visual entropy). The variable gaze entropy evaluates the amount of randomness associated with the visual scanning strategy of the pilots (details mentioned in the following section).

5.3.2.6.3. Gaze entropy metrics

Figure 5.19 shows 14 AOIs considered for gaze entropy calculations. A pre-requisite for entropy calculation is the transition probability values among AOI pairs. Transition between these AOIs was evaluated by the eye fixations transition frequency matrix as per the design principles described in Mandal et al. (2016) and Krejtz et al., (2014). In a transition frequency matrix, each cell shows the number of transitions from the AOI placed on a row to an AOI placed on a column. Figure 5.20 shows a sample gaze map of twelve AOIs and their transition matrix. In the gaze map shown in Figure 5.20, the arrows represent the transitions between the AOIs. The width of an arrow reflects its weight in the transition matrix (Mandal et al., 2016).

The transition frequency matrix was then transformed into a transition probability matrix. $\boldsymbol{P} = [p_{ij}]$, where p_{ij} can be interpreted as the probability of eye fixation transition from i^{th} AOI towards the j^{th} AOI (Holmqvist et al., 2011).



To / From	E I	EVS	ATT	HS	FC	ALT	AS	HI	VS	MI	RA	sum (from)
E I	0	0.31	0.19	0.08	0	0.06	0.08	0.05	0.07	0.04	0.12	1
EVS	0.15	0	0.4	0.09	0.02	0.08	0.05	0.02	0.07	0.03	0.09	1
ATT	0.08	0.38	0	0.27	0.02	0.06	0.08	0.06	0.05	0	0	1
HS	0.09	0.23	0.38	0	0.08	0.09	0.06	0	0.03	0	0.04	1
FC	0.20	0.20	0.20	0.20	0	0.20	0	0	0	0	0	1
ALT	0.06	0.18	0.12	0.18	0.06	0	0.23	0.08	0.09	0	0	1
AS	0.05	0.2	0.15	0.05	0.09	0.15	0	0.08	0.1	0.08	0.05	1
HI	0.11	0.22	0.11	0.18	0	0.15	0.11	0	0.12	0	0	1
TS	0.1	0.3	0.1	0	0	0	0.2	0.1	0	0.1	0.1	1
VI	0	0.14	0.17	0	0	0.16	0.18	0.16	0.19	0	0	1
MI	0	0.31	0.19	0.08	0	0.06	0.08	0.05	0.07	0.04	0.12	1
RA	0.15	0	0.4	0.09	0.02	0.08	0.05	0.02	0.07	0.03	0.09	1
SUM (to)	1.55	3.19	2.46	1.4	0.27	1.05	1.37	0.65	0.91	0.6	0.55	

Figure 5.19. A sample gaze map and the probability transition matric for the illustrated gaze map.

Subsequently, the pilots' scan behavior was quantified using two different entropy measures, namely, transition and stationary visual entropy (Krejtz et al., 2014; Acartürk & Habel, 2012).

The equation for calculating these two entropy values are shown below:

Transition visual entropy:

$$H_t = -\sum_{i \in S} \pi_i \sum_{i \in S} p_{ij} \log(p_{ij}) \quad , i \neq j \qquad \text{(i)}$$

Stationary visual entropy:

$$\mathrm{H_S} = -\sum_{i \in S} \pi_i \log(\pi_i) \qquad \text{(ii)}$$

Where, $P_{ij} = \frac{n_{ij}}{\sum_{j \in S}(n_{ij})}$, $\pi = \pi P,\ i, j \in S$, and π is the stationary distribution associated with AOIs transition probability matrix, i and j is index of AOIs, S is the set of AOIs. π_i shows the overall percentage share of the visual attention devoted to the i^{th} AOI, and thus higher π_i means more fixations on the i^{th} AOI (Krejtz et al., 2014). n_{ij} is the number of transitions from i^{th} AOI to j^{th} AOI. Thus, Ht can be interpreted as the randomness associated with switching of visual attention among various AOIs. Hence, higher Ht means more complex and more random search patterns. On the other hand, H_S can be interpreted as the randomness associated with distribution of visual attention among various AOIs. Higher H_s suggests a more uniform dispersion of visual attention among various AOIs, whereas a lower H_s value is suggestive of more selective visual attention distribution behavior.

5.3.2.7. Data analysis

The analysis includes the following dependent variables: fixations number, fixation duration, pupil size, gaze entropy, reaction time, and as well as the other extracted from the reaction time. The following two independent variables were involved in the analysis: flight tasks sequence as the within subject factor (Four levels: Task 1, Task 2, Task 3 and Task 4) and the pilot expertise as the between subject factor (Two levels: Expert and Novice). The mixed design between-within subjects' analysis of variance (ANOVA) was conducted to assess the impact of expertise levels on pilot fatigue level across four flight tasks (see Figure 5.21), while the descriptive statistics were adopted to show the average and the standard errors of the measures of the expert and novice pilots across the four flight tasks. The mathematical model that describes the relationship between the response and treatment for the two-Factor ANOVA with Repeated Measures on one Factor is given by:

$$Y_{ijk} = \mu + \rho_{i(j)} + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \varepsilon_{ijk} \qquad \text{(iii)}$$

Where Y_{ijk} represents the observation taken at k^{th} task from the i^{th} subject in group j, μ is the overall mean response, $\rho_{i(j)}$ is effect of subject nested in group, α_j is the main effect of expertise, β_k is the main effect of task , $(\alpha\beta)_{jk}$ is the interaction effect of expertise and task and ε_{ijk} is the error effect

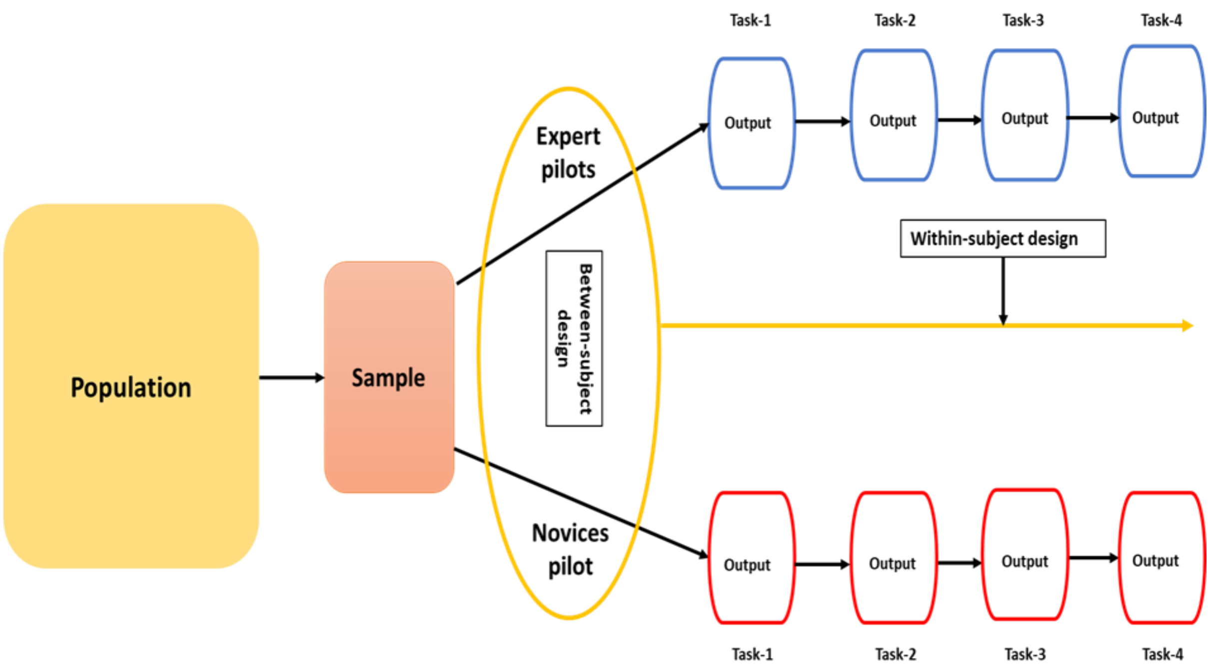


Figure 5.20. Comparison between novices and experts pilot groups across time or within subject design

We also conducted a mixed model correlation analysis to identify the relationship between the eye movement and PVT measures of fatigue. We consider the linear mixed effect model given by:

$$Y_i = X_i\beta + Z_i\gamma_i + \varepsilon_i \qquad \text{(iv)}$$

Where Y_i is the vector of observed outcomes, X_i is the known matrix of fixed , β is the unknown vector of fixed effects parameters and Z_i is known matrix of random effect, γ_i is a vector of unknown random effect parameter and ε_i is the unknown random error. Both γ_i and ε_i are normally distributed.

5.3.2.8. Gaze entropy

For ease of understanding, an example of the eye fixation transition probability matrix for a pilot has been shown in Table 5.2. Note that, for the present study, we have used collapsed eye fixation sequences, i.e. all self-transitions are considered same, and thus, have been ignored for calculation. Consequently, the diagonal elements in Table 5.2 are zero. Using the information in Table 5.2, the stationary distribution obtained is (using equation (iii)) π = (0.099, 0.22, 0.2, 0.11, 0.02, 0.07, 0.07, 0.03, 0.04, 0.04, 0.03, 0.04, 0.01, 0.02)^T. Subsequently, the two entropy values are H_t = 2.81, and H_s = 3.34 (using equation (i) and (ii)).

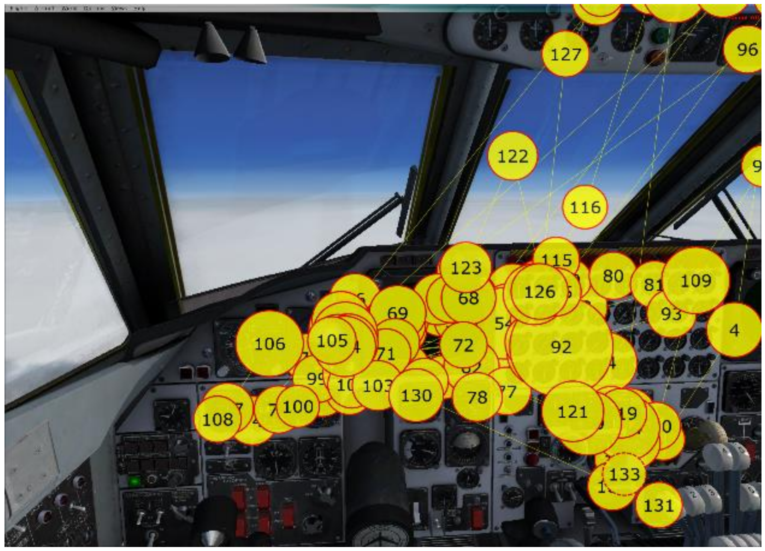
Table 5.2. AOI transition matrix developed from the AOI fixation sequence for one task.

		E I	EVS	ATT	HS	FC	ALT	AS	HI	TS	VI	MI	RA	SHS	EOP
	E I	0	0.31	0.19	0.08	0	0.04	0.04	0.03	0.04	0.07	0.04	0.12	0	0.04
	EVS	0.15	0	0.4	0.1	0.02	0.05	0.05	0.02	0.04	0.02	0.02	0.07	0.02	0.04
	ATT	0.08	0.38	0	0.27	0.02	0.06	0.08	0.02	0.02	0.03	0	0	0.02	0.02
	HS	0.09	0.22	0.38	0	0.04	0.09	0.04	0	0.04	0.03	0	0.04	0	0.03
	FC	0.2	0.2	0.2	0.2	0	0.2	0	0	0	0	0	0	0	0
	ALT	0.06	0.18	0.12	0.18	0.06	0	0.23	0.06	0.05	0.06	0	0	0	0
	AS	0.05	0.2	0.15	0.05	0.05	0.15	0	0.05	0.1	0.1	0.05	0.05	0	0
***P*=**	**HI**	0.11	0.22	0.11	0.11	0	0.11	0.11	0	0.11	0.12	0	0	0	0
	TS	0.1	0.2	0.1	0.1	0	0.1	0.1	0.1	0	0.1	0.1	0	0	0
	VI	0.1	0.3	0.1	0	0	0	0.2	0.1	0	0	0.1	0.1	0	0
	MI	0	0.13	0.13	0	0	0.13	0.13	0.12	0.12	0.12	0	0	0.12	0
	RA	0.25	0.25	0.08	0.09	0	0	0.09	0	0.08	0.08	0	0	0	0.08
	SHS	0.25	0.25	0.25	0	0	0	0	0	0	0	0.25	0	0	0
	EOP	0.11	0.33	0.11	0.11	0	0	0.11	0	0	0	0	0.11	0.12	0

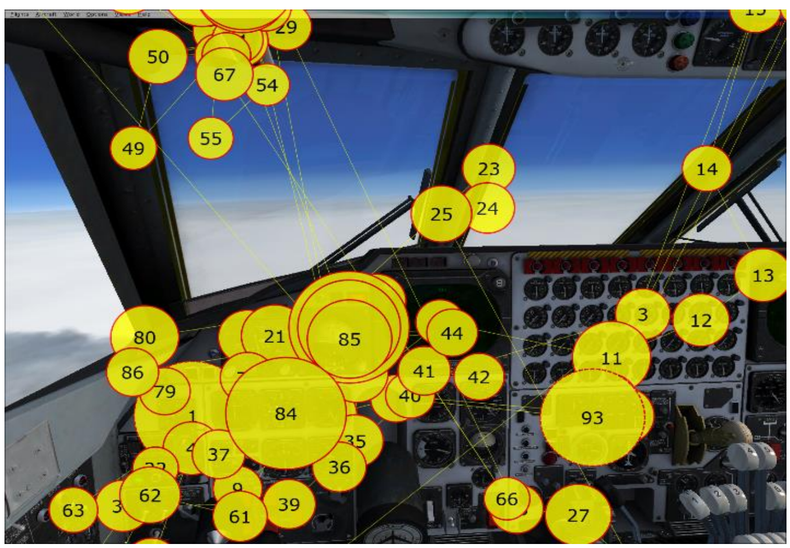
5.3.3. Results

5.3.3.1. Eye movement pattern

Figure 5.22 shows examples of different visual scan paths, of 40 seconds duration, observed just prior to the finish of the task 1 and task 4. The Figure shows that with the progress in the flight mission (as we move from task 1 to task 4), the visual scan path consists of more instances of longer eye fixation durations.

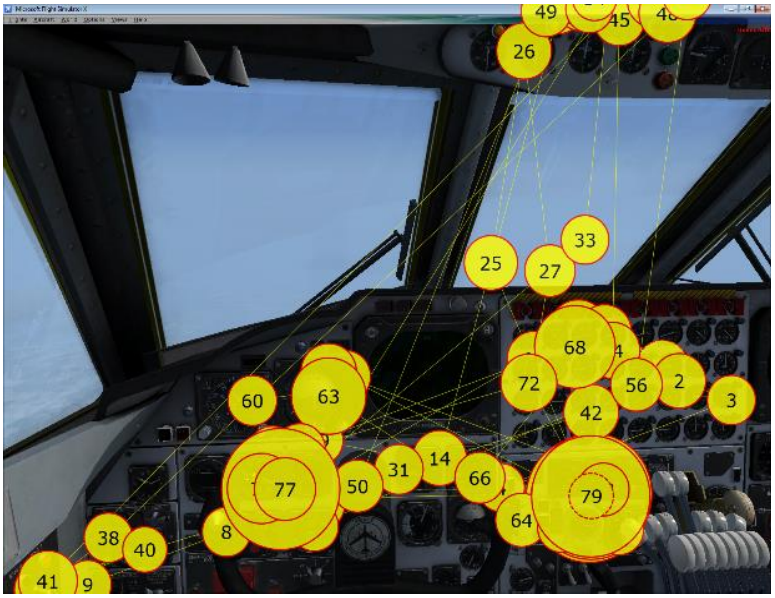


(a) Expert pilot.

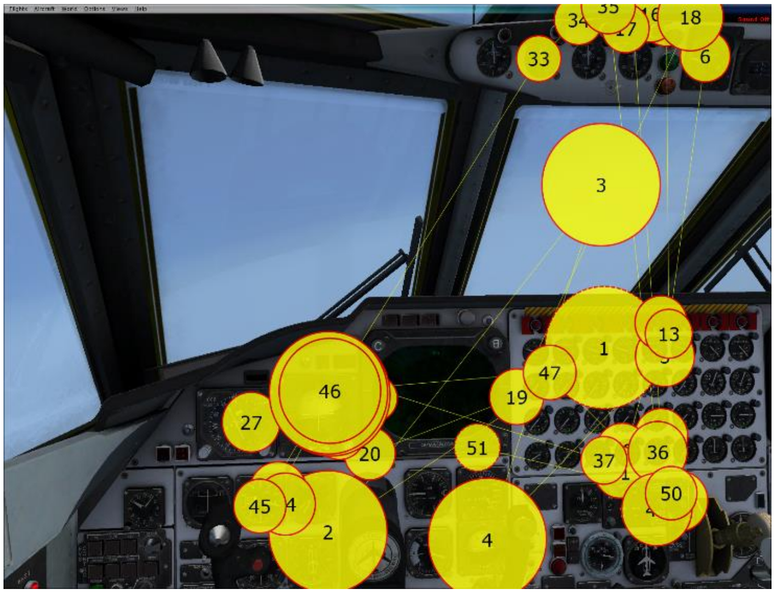


(b) Novice pilot.

(1) 40 seconds prior to finishing Task 1



(a) Expert pilot



(b) Novice pilot

(2) 40 seconds prior to finishing Task 4

Figure 5.21. Example visual scan path plots for one expert and one novice pilot for the two tasks, Task 1 and Task 4. The numbers represent the eye fixation index, and the size of the circle is proportional to the eye fixation duration. The red lines represent the saccades.

5.3.3.2. Eye fixation number

The main effect of expertise (F (1, 18) = 72.41, p =.001, η_P^2 = 0.80) was significant, the expert pilots' number of fixations was greater than novice pilots. The main effect of flight tasks sequence (F (3, 54) = 157.12, p = 0.001, η_P^2 = 0. 89) was also significant, both expert and novice pilots were influenced by flight task sequences, their mean number of fixations showed a decreasing trend over the course of flight (see Figure 5.23). The results also showed that there was a remarkable interaction effect ((F (3, 54) = 5.25, p = 0.003, η_P^2 = 0.22). The post-hoc tests within subject showed that pilots' number of fixations task 1 was not different from the number of fixations at task 2 (p = 1.0). All the other comparisons showed significant differences (p = 0.001). The post-hoc between subject comparison showed significant differences between expert and novice pilots' number of fixations (p = 0.001).

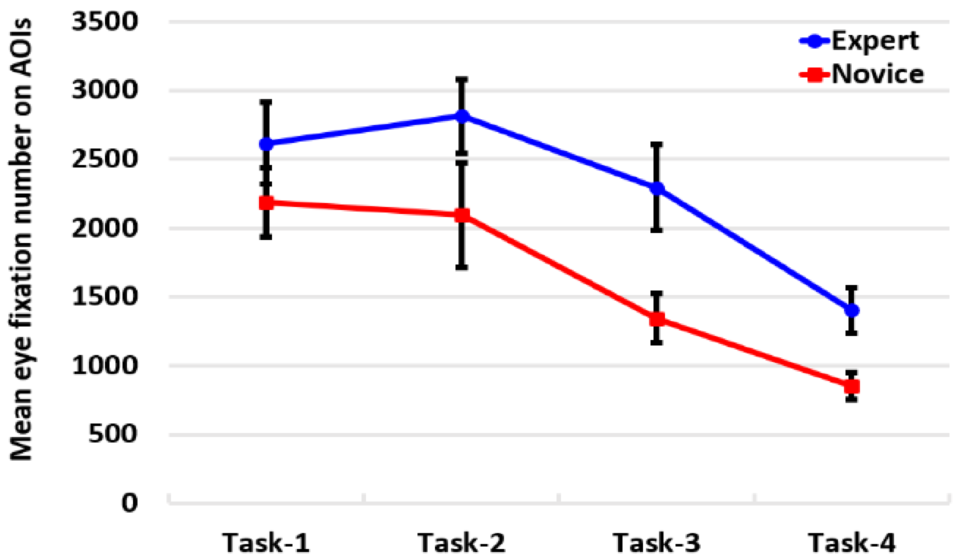


Figure 5.22. Mean eye fixation number for all tasks for both expert and novice pilots.

5.3.3.3. Eye fixation duration

Mauchly's test indicated that the assumption of sphericity had been violated ($X^2(5) = 11.81, p = 0.038$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (ε =0.78). Main effect of flight tasks sequence (F (2.34, 54) = 168.76, p = 0.001, η_P^2 = 0.90) was significant, both expert and novice pilots were influenced by flight task sequences, their mean fixation duration showed an increasing trend over the course of flight (see Figure 5.24). The results also revealed significant main effect of expertise levels (F (1, 18) = 459.9, p = 0.001, η_P^2 = 0. 96) on the fixation duration, indicating a significant difference between expert and novice pilots. The average fixations duration of the expert pilots was shorter than that of the novice pilots. These main effects were qualified by an interaction between expertise and tasks sequences (F (2.34 54) = 7.71, p = 0.001, η_P^2 = 0.30). The post-hoc tests within subject showed significant differences of pilots' fixation duration among all tasks. The post-hoc between subject comparison showed significant differences between expert and novice pilots fixation duration (p = 0.001).

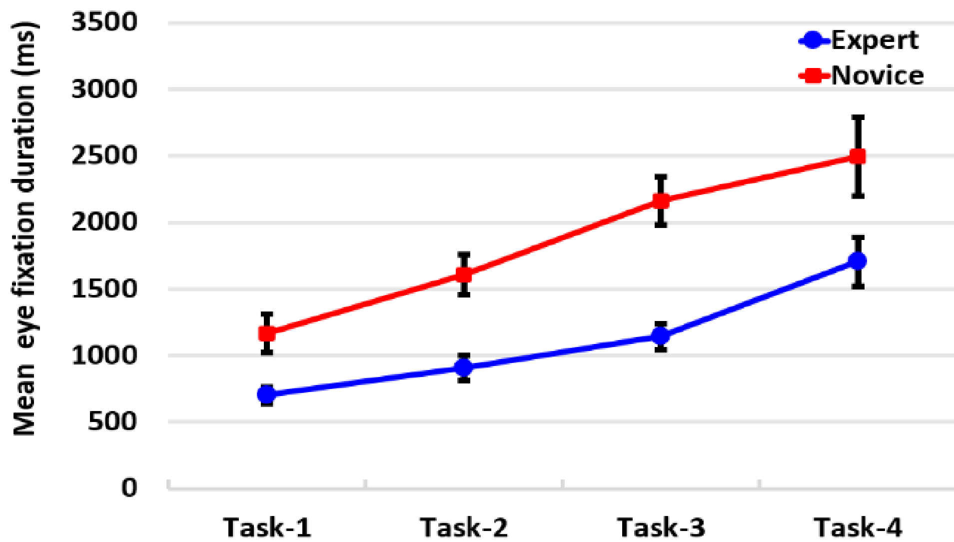


Figure 5.23. Mean eye fixation duration for all tasks for both expert and novice pilots

5.3.3.4. Pupil size

The main effect of expertise ($F\ (1, 18) = 101.89$, $p = .001$, $\eta_P^2 = 0.85$) was significant, the expert pilots' pupil diameter was greater than novice pilots. The main effect of flight tasks sequence ($F\ (3, 54) = 408.79$, $p = 0.001$, $\eta_P^2 = 0.\ 96$) was also significant, both expert and novice pilots were influenced by flight task sequences, their mean pupil diameter showed a decreasing trend over the course of flight (see Figure 5.25). The results also showed that there was an interaction effect ($F\ (3, 54) = 21.24$, $p = 0.001$, $\eta_P^2 = 0.54$). The post-hoc tests within subject showed that pilots' pupil diameter during task 1 was not different from the pupil diameter at task 2 ($p = 0.99$). All the other comparisons showed significant differences ($p = 0.001$). The post-hoc between subject comparison showed significant differences between expert and novice pilots' pupil diameter ($p = 0.001$).

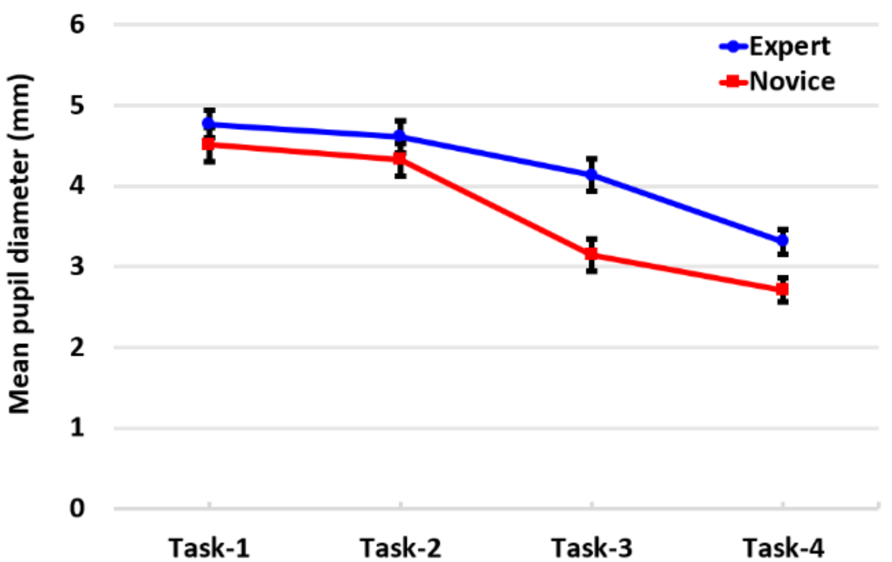


Figure 5.24. Mean eye pupil for all tasks for both expert and novice pilots.

5.3.3.5. Transition visual entropy

The results revealed a main effect of expertise level (F (1, 18) = 210.88, p =.001, $\eta_P^2 = 0.92$) on the transition entropy, the expert pilots' mean transition entropy was lower than novice pilots. The results also revealed a significant effect of flight tasks sequences (F (3, 54) = 200.75, p = 0.001, $\eta_P^2 = 0.92$). Both expert and novice pilots' groups were influenced by flight task sequences, their mean transition entropy showed an increasing trend over the course of flight (see Figure 5.26). The analysis also showed the interaction effect to be significant, (F (3, 54) = 9.15, p = 0.001, $\eta_P^2 = 0.34$). The post-hoc tests within subject showed that pilots' transition entropy was different across all flight tasks (p = 0.001). The post-hoc between subject comparison showed significant differences between expert and novice pilots' transition entropy (p = 0.001).

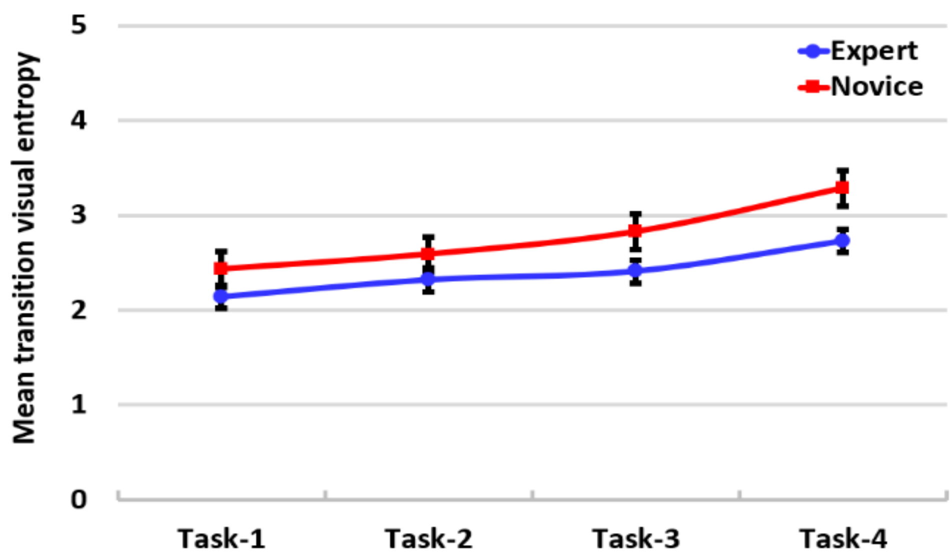


Figure 5.25. Mean transition visual Entropy for all tasks for both expert and novice pilots

5.3.3.6. Stationary visual entropy

The results revealed a main effect of expertise level (F (1, 18) = 119.11, p = 0.001, $\eta_P^2 = 0.87$) on the stationary entropy, indicating a significant difference between expert and novice pilots.

Both expert and novice pilots were influenced by flight tasks sequences, their mean stationary entropy showed an increasing trend over the course of flight (see Figure 5.27). The expert pilots' stationary entropy was lower than novice pilots across all flight tasks. The main effect of flight tasks ($F(3, 54) = 75.99$, $p = 0.001$, $\eta_P^2 = 0.81$), indicating that the effect for flight tasks sequence were significant. The analysis also showed the interaction effect to be remarkable, as ($F(3, 54) = 3.85$, $p = 0.014$, $\eta_P^2 = 0.18$). The post-hoc tests within subject showed that pilots' stationary entropy task 1 was not different from the stationary entropy at task 2 ($p = 0.18$). All the other comparisons showed significant differences. The post-hoc between subject comparison showed significant differences between expert and novice pilots' stationary entropy ($p = 0.001$).

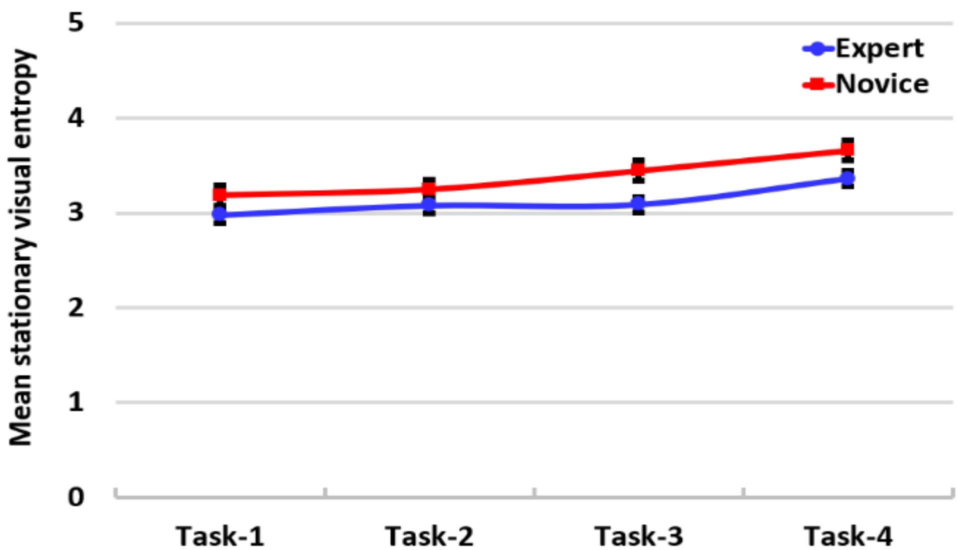


Figure 5.26. Mean stationary visual Entropy for all tasks for both expert and novice pilots.

5.3.3.7. PVT reaction time

The main effect of expertise ($F(1, 18) = 82.45$, $p = 0.001$, $\eta_P^2 = 0.82$) was significant. The results also revealed a significant effect of flight tasks sequence ($F(3, 54) = 177.02$, $p = 0.001$, $\eta_P^2 = 0.91$), reflecting higher reaction time in the final flight task than the first flight task (see

Figure5.28). Both expert and novice pilots were influenced by flight task sequences, their mean reaction time showed an increasing trend over the course of flight (see Figure 5.28). The interaction of flight tasks sequence by the pilot expertise was also statistically significant ($F (3, 54) = 5.26$, $p = 0.003$, $\eta_p^2 = 0.23$). The post-hoc tests showed that pilots' reaction times were different across all flight tasks ($p = 0.001$). The post-hoc between subject comparison showed significant differences between expert and novice pilots reaction time ($p = 0.001$).

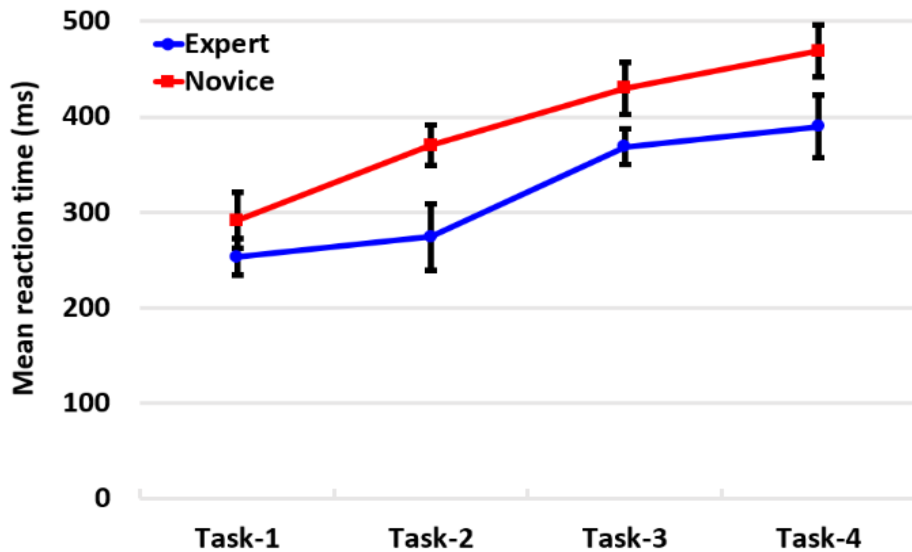


Figure 5.27. Mean reaction time for all tasks for both expert and novice pilots.

5.3.3.8. PVT number of lapses

The main effect of expertise ($F (1, 18) = 104.72$, $p = 0.001$, $\eta_p^2 = 0.85$) on number of lapses was significant. The results also revealed a significant effect of flight tasks sequence ($F (3, 54) = 35.11$, $p = 0.001$, $\eta_p^2 = 0.66$) reflecting number of lapses in the final flight task than the first flight task (see Figure 5.29). Both expert and novice pilots were influenced by flight tasks, where an additional flight task lead to an increase in the mean number of lapses. The interaction of flight tasks by the pilot expertise was also statistically significant ($F (3, 54) = 4.67$, $p = 0.006$, $\eta_p^2 = 0.21$).

The post-hoc tests within subject showed that pilots' number of lapses task 1 was not different from the number of lapses at task 2 ($p = 0.066$). All the other comparisons showed significant differences ($p = 0.001$). The post-hoc between subject comparison showed significant differences between expert and novice pilots' number of lapses ($p = 0.001$).

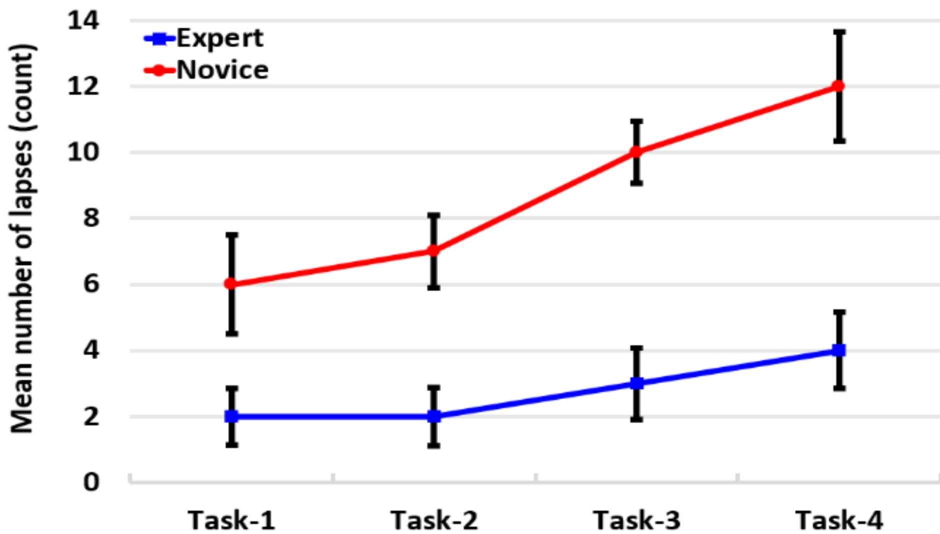


Figure 5.28. Mean number of Lapses for all tasks for both expert and novice pilots.

5.3.3.9. PVT false starts

The main effect of expertise ($F(1, 18) = 90.72$, $p = 0.001$, $\eta_P^2 = 0.83$) on the false starts was significant. The results also revealed a significant effect of flight tasks sequence ($F(3, 54) = 39.87$, $p = 0.001$, $\eta_P^2 = 0.69$), reflecting more false starts in the final flight task than the first flight task (see Figure 5.30). Both expert and novice pilots' groups were influenced by flight tasks sequence, their mean false starts showed an increasing trend over the course of flight. The interaction of flight tasks sequence by the pilot expertise was statistically significant ($F(3, 54) = 6.99$, $p = 0.001$, $\eta_P^2 = 0.28$). The post-hoc tests within subject showed that pilots false starts task 2 was not different from the pilots' false starts at task 3 ($p = 0.21$). All the other comparisons showed significant differences (p = 0.001). The post-hoc between subject comparison showed significant differences between expert and novice pilots' false starts (p = 0.001.

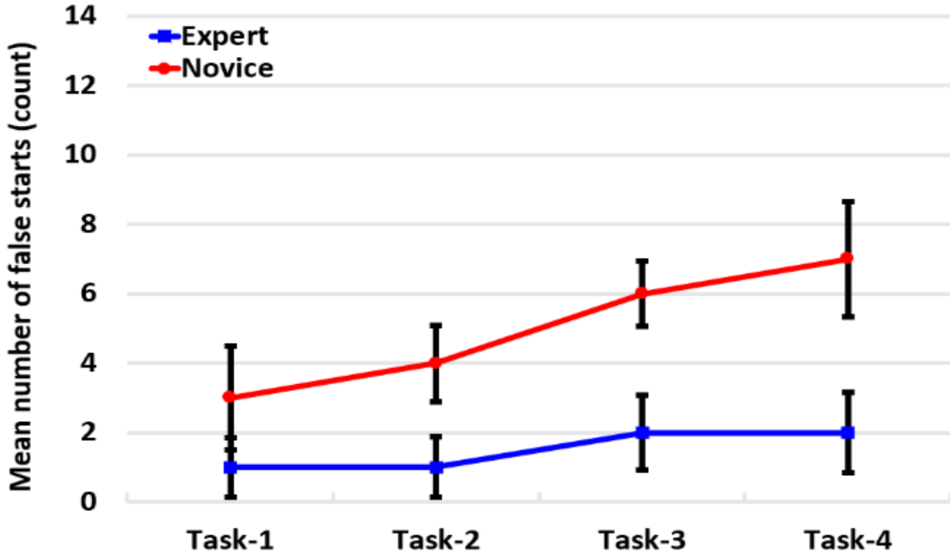


Figure 5.29. Mean number of False Alarm for all tasks for both expert and novice pilots.

5.3.3.10. Correlation results

In line with our predictions, we found all measures to change consistently as task index increased. These findings were obtained using mixed design ANOVA. However, ANOVA does not provide direct insight into the relationships between the different measures used in the present study. Correlation is an appropriate measure to quantify the association between two variables. However, widely used techniques for correlation, e.g. Pearson correlation, assume independence of error between observations and does not consider the different number of replicate measurements belonging to the same subject (Aarts, Verhage, Veenvliet, Dolan, & Van Der Sluis, 2014). Various approaches have been suggested for measuring correlations between variables in the presence of replications, e.g. partial correlation coefficient (Bland & Altman, 1995). However, in the case of many subjects, this measure suffers loss in power caused by the increased number of parameters that are to be estimated. To address this issue, Hamlett, Ryan, & Wolfinger. (2004) suggested mixed model approach to calculate the between-subject correlation and within-subject correlation. Presently, we have applied this approach to measure the correlation structure between our response variables while accounting for repeated measures (task sequences) into the analysis.

Figures 5.31-5.33 shows the correlation between various eye movements and PVT measures for both novice and expert pilots. The correlation plots show that the all three PVT measures have similar trends of association (positive/negative) with the eye movement variables for both expert and novice groups. For example, all three PVT measures have strong positive correlation with both the entropy measures and eye fixation duration whereas, they are negatively associated eye fixation number and pupil size for both novice and expert pilots (see Table 5.3 & 5.4). This suggests that, as fatigue increases both gaze entropy and eye fixation duration increase, on the other hand, eye fixation number and pupil size decrease with higher fatigue levels.

Table 5.3. Correlation between expert pilot's eye movement measures and PVTs measures.

Variables	Reaction time	Number of lapses	Number of false starts
Transition visual entropy	0.61	0.63	0.76
Stationary visual entropy	0.64	0.49	0.57
Fixation number	-0.69	-0.61	-0.49
Fixation duration	0.76	0.68	0.62
Pupil size	-0.81	-0.56	-0.59

Table 5.4. Correlation between novice pilot's eye movement measures and PVTs measures.

Variables	Reaction time	Number of lapses	Number of false starts
Transition visual entropy	0.74	0.69	0.61
Stationary visual entropy	0.71	0.64	0.57
Fixation number	-0.84	-0.76	-0.76
Fixation duration	0.88	0.78	0.74
Pupil size	-0.86	-0.78	-0.70

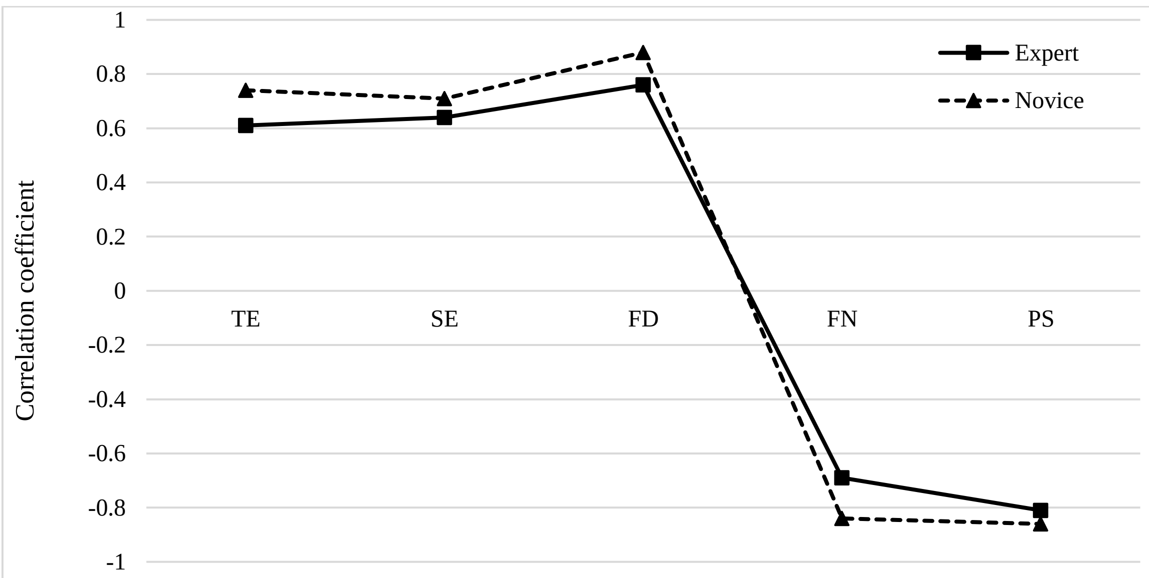


Figure 5.30. Plot of correlation between Reaction time and TE: Transition visual Entropy, SE: Stationary visual Entropy, FD: Fixation Duration, FN: Fixation Number, PS: Pupil Size for expert and novice pilots.

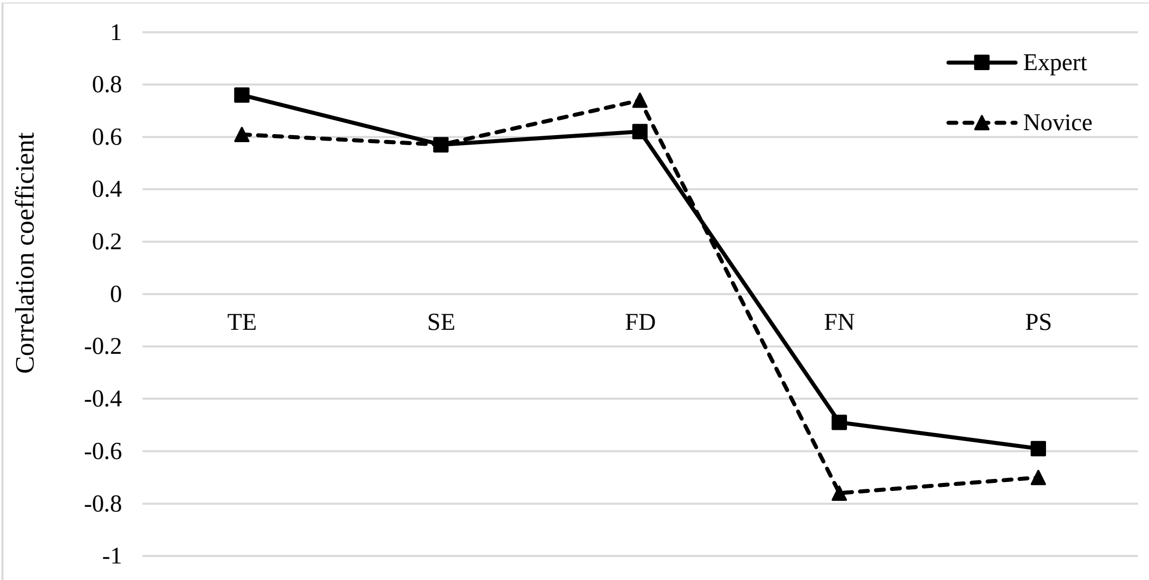


Figure 5.31. Correlation between false starts and TE: Transition visual Entropy, SE: Stationary visual Entropy, FD: Fixation Duration, FN: Fixation Number, PS: Pupil Size for expert and novice pilots.

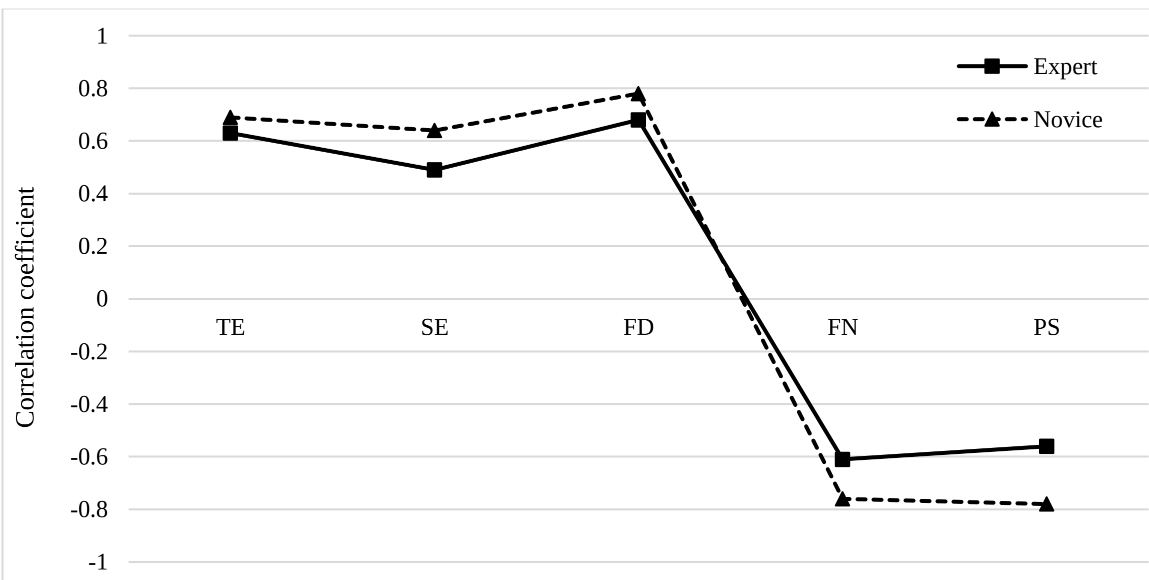


Figure 5.32. Correlation between Number of lapses and TE: Transition visual Entropy, SE: Stationary visual Entropy, FD: Fixation Duration, FN: Fixation Number, PS: Pupil Size for expert and novice pilots.

5.3.4. Discussion

In this study, we proposed a multi-modal framework, combing PVT and eye tracking measures, for fatigue measurement in the case of long duration flights with multiple takeoff and landing phases. In addition, a comparative analysis of the effect of fatigue on novice and expert pilots were performed. Two new eye tracking measures, namely transition visual entropy and stationary visual entropy, have been introduced for the fatigue measurement purpose. The study findings, which was fatigue increases as the number of takeoff and landing operations increase, agree with previous fatigue studies (Samel, Wegmann & Vejvoda, 1997; Rosekind et al., 1994; Rosekind et al., 1995; Caldwell et al., 2009; Gander et al., 1998).

The obtained results suggest that our model is able to capture changes in fatigue levels and pilot's behavioral aspect (i.e. visual information scanning) that gets severely affected by increased fatigue levels. Furthermore, a correlation analysis between the PVT measures (reaction time, false start, number of lapses) and eye tracking measures (eye fixation number and duration, pupil size, H_t, H_s) shows that strong association exists between both the measures; thus, suggesting that the intrusive PVT approach can be substituted by the non-intrusive eye tracking approach for fatigue measurement.

In detail, the trends observed for the eye tracking measures suggest deterioration in concentration ability of pilots' (i.e. decreased eye fixation number and pupil size, and increased eye fixation duration) as the flight mission progressed. These results are in accordance with the previous research findings, which suggests an uncertainty in the visual search pattern, indicative of the difficulty in extracting the relevant information (Harris et al., 1986; Diez et al., 2001).

Moreover, the comparative analysis showed that, unlike novices, the expert pilots performed better and on average made a higher number of eye fixations and shorter eye fixation durations on AOIs throughout the flight mission. It might be possible that, compared to novice pilots, experts may have better ability of identifying task relevant AOIs, thus more efficient visual search strategy and more information processing time for a given task (Kirby et al., 2014; Ho, Su, Li, Yu & Braithwaite 2016; Xiong, Wang, Zhou, Liu & Zhang, 2016; Ziv, 2016). In the case of pupil size, it became progressively smaller (for both expert and novices) as pilot fatigue increased over the course of the flight mission. However, as expected, the experts had larger pupil size compared to novices, signaling that expert pilots might have experienced a lower visual information processing load than novices.

In the case of the newly introduced eye tracking measures, results showed that both the gaze entropy measures H_s and H_t increased with higher fatigue levels. One possible reason might be that, with higher fatigued levels, pilots' visual search strategy became more random in nature (both in terms of overall attention distribution and switching behavior among AOIs) resulting in higher entropy values. Interestingly, our findings are in accordance with various results observed in non-aviation settings, e.g. several healthcare studies have found the same trend for gaze entropy with increased levels of operator's fatigue (Diaz-Piedra, Sanchez-Carrion, Rieiro & Di Stasi, 2017). Although, it is to be noted that expert pilots have significantly lower gaze entropy (H_s and H_t) compared to novice cases, indicative of a more systematic visual search strategy.

Results of the PVT measures, i.e. increasing mean reaction time, number of lapses and false starts over the course of the flight mission, suggests that fatigue level became higher with time and additional flight tasks (Ziv, 2016). This resulted in the increasing of fatigue level increase for both pilot groups. However, the fatigue level of novices. Was significantly higher than experts

The correlation analysis showed that eye fixation duration and entropy measures are strongly positively correlated with PVT metrics, unlike pupil size and eye fixation number which were strongly negatively associated with PVT response variables. In summary, the present multi-modal study demonstrates the validity of the eye tracking measures in detecting pilot fatigue, and its utility in identifying the particular behavioral aspect that gets affected due to high fatigue levels.

Chapter 6: Research summary, recommendations, limitations, and future research

6.1. Research summary

The main aim of this research was to investigate the relationship between the pilot the level of fatigue during a prolonged flight mission. This research presents four hypotheses that evaluate the relationship between the pilot and the level of fatigue. The first hypothesis addresses the question: How does flight conditions (extreme vs normal) impact the pilot when the pilot is observing the cockpit interface? The second hypothesis addresses the question: How does expertise level impact the fatigued pilot? The third hypothesis addresses the question: How does a prolonged flight mission (based on sequential multi-flight tasks) impact the pilot? The fourth hypothesis addresses the question: Can we use eye movement characteristics (i.e. eye fixation durations, eye fixation numbers & gaze entropy) to identify fatigue level?

Five studies included in this research used quantitative research methods:

Study I: Investigation of the pilots' eye movement characteristics under normal and extreme flight conditions.

Study II: Investigation of the relationship between pilots' and fatigue level when interacting with cockpit interfaces.

Study III: Investigation of the effect of fatigue during single and multi-takeoffs and landings.

Studies IV and V: Comparison of the effect of fatigue between novice pilots and expert pilots for a prolonged multi-flight mission using the proposed multi-modal analysis approach.

The findings from Study I (related to Hypothesis 1) suggest an increase in the fatigue level leads to increased eye fixation durations and decreased eye fixation numbers. Additionally, the extreme conditions during the flight was also found to be a significant factor affecting the pilots' eye-scan behavior. In other words, eye movement networks (i.e. directed weighted networks (DWNs)) of the expert pilots and novice pilots were substantially different in response to the emergent flight conditions (i.e. severe weather accompanied by engine failures).

Respect to the correlation between pilot eye-scan behavior, eye-tracking data was processed as a DWN to characterize pilots' eye fixation transitions among AOIs. The DWN enabled us to quantify the pilots' eye fixation transitions. In addition, the eye movement analysis revealed that the pilot's visual attention was in much higher demand during emergency. The finding of this study validates the hypothesis that extreme (i.e., severe whether & engine failures) flight conditions impact the pilot eye-scan behavior.

The findings from Study II (related to Hypothesis 2) is that expert pilots performed better than novices during all the flight phases. In detail, expert pilots had fewer number of lapses, less reaction times, slowest 10% RTs, fastest 10% RTs, and less false alarms. Although, remarkable increases in reaction times, slowest 10% of RTs, fastest 10% of RTs, number of lapses, number of false alarms, eye fixation numbers, and eye fixation durations were discovered for both the novice pilots and the expert pilots during the final phase of the flight. These findings are consistent with the findings of existing research, in which pilots who did not get enough sleep over a long period or pilots who were involved in ongoing physical or mental activities with insufficient rest (Hartzler, 2014; Lee & Kim, 2018).

In addition, the pilots who had higher PVT scores were also more efficient in their visual scanning behavior evidenced by the DWNs. In other words, they showed more eye fixations transition among the important AOIs and shorter eye fixation duration on AOIs.

The findings from Study III (related to Hypothesis 3) is that the pilots performed better during the single flight mission compared to the multi-takeoffs and landing flight mission. Subjective fatigue ratings were investigated along with the PVT. The pilots showed higher total fatigue ratings, reaction times, and number of lapses during the multi-takeoffs and landings flight mission. The result is aligned with the finding that workload and fatigue can increase during takeoffs and landings (Vejvoda, et al., 2014). The pilots showed more structured visual scanning behaviors during the takeoffs and landings evidenced by the DWNs.

The findings from Study IV (related to Hypothesis 3) is that pilot fatigue level will increase with increasing in the flight duration. The results showed decrease of the eye fixation numbers, increase of the eye fixation durations, increase of RT and NL values, and increase of the randomness of the pilots' eye movements as evidenced by the increase of both of the transition and the stationary visual entropy values. The eye movement characteristics seem to demonstrate the deterioration of the pilots' concentration.

These findings are somewhat consistent with the findings from a previous research (Harris et al., 1986), which asserted that the increase of uncertainty in visual search patterns are related to the increase of difficulty in extracting the required information. In addition, the increase in reaction times and the number of lapses show that the increase of the fatigue level negatively impacts the pilots' capability to effectively control the aircraft.

The proposed multi-model analysis framework was developed and applied in Study V (related to hypothesis 4). The results showed that the changes in the fatigue levels were highly correlated with the pilots' eye movement characteristics. In addition, the novice pilots had less eye fixation numbers and longer eye fixation durations on the important AOIs throughout the flight mission suggesting that less amount of information might have been processed by the novices.

In addition, the pupil size became smaller (for both expert and novices) as pilot fatigue level increased over the course of the flight mission and the visual entropy values increased with higher fatigue levels. Furthermore, the novices had smaller pupil sizes compared to experts, signaling that novice pilots might have experienced a higher visual information processing load than experts. Furthermore, applying the multi modal measurement framework enabled us to detect the differences between the novice and expert pilot fatigue level. The results have shown that expert pilots experienced lower fatigue level under all flight conditions compared to the novice pilots. Therefore, the effective and practical models of experts can help in training novices.

6.2. Research strengths

The studies included in this dissertation combined data from different flights characteristics (routes & ranges). Each of the experiments was conducted with different groups of pilots, which increased the validity of this research. In addition, multiple experiments throughout this research resulted in developing and evaluating the proposed multi-modal analysis framework.

6.3. Research limitations

There are some limitations in this research. First our estimates may be conservative and underestimate the individual and work factors that influence the development of fatigue in pilots, as (1) all experiments were conducted during the day and thus changes in fatigue levels and the effects of night shift was not considered in this research; (3) participants were relatively young, (35.5 years old) compared to the U.S. pilots, which is ageing with an average projected age of 44.7 years as of 2015; (4) as many aviation research on pilots there are limited participants since it is difficult to recruit many pilots for prolonged flight missions experiments ; (5) the visual scan paths of experts were not analyzed in great detail so as to find what strategy aspects leads to their less exhaustive a more efficient search strategy. The study was also limited by the moderate-fidelity simulator and conducting experiments in higher-fidelity flight simulator might further improve the validity. One more limitation is the need for more high effect size measures to our developed modal in order to identify real change in a pilot fatigue level. Response speed which is a reciprocal of reaction time is a more statistically sensitive measure to detect fatigue levels compared to reaction time.

There is a need for looking in the most prevalent transition patterns of the Directed Weighted Network (DWN) visualizations of the expert pilots at different time intervals. These patterns of expert pilots' transitions among AOIs can help to develop several optimal and efficient visual scanning strategies to train novices and student pilots.

There is also a need to investigate and analyze the consistency of directed weighted network outputs. There are various powerful methods (i.e., match graphs, graph embedding and kernels). Match graphs compares the sequence of visual scan. Graph embedding converts the graph data into a low dimensional space, and graph kernels compares graphs based on the adjacency matrices eigenvectors.

There have been many studies recently concerning task complexity and pilot fatigue level. Diaz-Piedra et al (2019) studied the impact of flight phase complexity on the pilot gaze entropy. They found that as flight phase complexity increased (more stimulus) the pilot's gaze entropy decreased. This result suggested that pilots showed a systematic visual scanning and less randomness as he received more stimulus during the high complex flight phases. Moreover, Schieber and Gilland (2008) had also observed that visual scanning decreases as the numbers of stimulus increased. These results are in contradiction with the results of this study. The results of my study showed an increasing trend in the randomness of visual scanning pattern with the increased levels of pilot fatigue over the course of the flight even during the flight tasks which involved high complex flight phases where the number of stimulus increases. Therefore, analyzing the influence of complex cognitive operations on pilot fatigue and workload simultaneously need to be investigated more in the future research.

6.4. Recommendations

The analysis of the data indicates that the impact of fatigue is often unpredictable, and each pilot may experience symptoms of fatigue while performing her/his tasks. Pilots are negatively impacted in many ways by fatigue. These negative impacts may and do lead to pilot errors and accidents. Given the findings of this research, our recommendations, are as follows: All pilots should be cultured regarding the causes of fatigue, the impacts of fatigue, the dangers of fatigue, and the fatigue countermeasures. In addition, consider integrating fatigue detection technologies, such as real-time eye tracking analysis, as part of a safety management approach.

Aviation regulator bodies should define additional procedures in conjunction with other fatigue countermeasures, as required, to manage both unexpected fatigue and to reduce the risk of Increasing fatigue level experienced during flight operations. For example, they would use a controlled rest at times when pilot does not interfere with required operational duties and during low workload phases of flight (e.g., during cruise phase). This controlled rest must be limited for only one pilot at a time on his/her seat in the flight deck and the other pilot shall be fully alert, maintain situational awareness and has the full capacity to control the aircraft. All safety procedures should be applied to prevent any unintentional interaction with the flight control system. All the required information and procedures for the pilot-controlled rest on the flight deck should be clearly established. Like, when the pilot will take rest, when the rest and wake-up should be arranged. In the meantime, a crew member should observe the pilot throughout controlled rest. These procedures should also be included in both the flight safety manual and the fatigue countermeasures training program.

6.5. Future research

Future research involves examining the effect of pilot fatigue levels across both day and night flight shifts. Additionally, this research can be expanded to explore the relationships among fatigue, workload, visual scanning behaviors, and brain activities over a longer flight period.

Furthermore, regarding the visual scanning behaviors, detailed time-ordered visual can paths can be compared between the expert pilots and the novice pilots. It would be useful to investigate in which order the cockpit information were interrogated and how the order is affected by fatigue. In addition, the relationships between fatigue and eye movement characteristics can be investigated based on different task types (e.g. cruising vs. approach), different cockpit interfaces (e.g. different flight instruments), and flight procedures (e.g. instrument flight rule (IFR) vs. visual flight rule (VRF). There is a need for developing mathematical models to use the multi-measures outputs. This can help to understand and explore the meanings of functional relationships among these measures outputs and how they work out together. It may also help to identify the effects of changes in the measure's outputs of the modal.

Chapter 7: References

Aarts, E., Verhage, M., Veenvliet, J. V., Dolan, C. V., & Van Der Sluis, S. (2014). A solution to dependency: using multilevel analysis to accommodate nested data. Nature neuroscience, 17(4), 491.

Acartürk, C., & Habel, C. (2012). Eye tracking in multi-modal comprehension of graphs. In Proceedings of the workshop on technology-enhanced diagrams research (Vol. 887, pp. 11-25). CEUR Canterbury, UK.

Ahsberg, E. (1998). Perceived fatigue related to work.

Akerstedt, T., Anund, A., Axelsson, J., & Kecklund, G. (2014). Subjective sleepiness is a sensitive indicator of insufficient sleep and impaired waking function. Journal of sleep research, 23(3), 242-254.

Akerstedt, T., Folkard, S., & Portin, C. (2004). Predictions from the three-process model of alertness. Aviation, space, and environmental medicine, 75(3), A75-A83.

Al-Rahayfeh, A., & Faezipour, M. (2013). Eye Tracking and Head Movement Detection: A State-of-Art Survey. IEEE Journal of Translational Engineering in Health and Medicine, 1, 2100212–2100212.

Arsintescu, L., Kato, K. H., Cravalho, P. F., Feick, N. H., Stone, L. S., & Flynn-Evans, E. E. (2017). Validation of a touchscreen psychomotor vigilance task. Accident Analysis & Prevention.

Aviation safety reporting system. (2019). Human factors as contributor's factors of aircraft accidents. Retrieved from https://akama.arc.nasa.gov/ASRSDBOnline/QueryWizard_Filter.aspx

Avers, K., & Johnson, W. B. (2011). A review of Federal Aviation Administration fatigue research: Transitioning scientific results to the aviation industry. Aviation Psychology and Applied Human Factors, 1(2), 87.

Balkin, T. J., Bliese, P. D., Belenky, G., Sing, H., Thorne, D. R., Thomas, M., ... & Wesensten, N. J. (2004). Comparative utility of instruments for monitoring sleepiness-related performance decrements in the operational environment. Journal of sleep research, 13(3), 219-227.

Basner, M., Mollicone, D., & Dinges, D. F. (2011). Validity and sensitivity of a brief Psychomotor Vigilance task (PVT-B) to total and partial sleep deprivation. Acta astronautica, 69(11-12), 949-959.

Beurskens, A. J., Bültmann, U., Kant, I., Vercoulen, J. H., Bleijenberg, G., & Swaen, G. M. (2000). Fatigue among working people: validity of a questionnaire measure. Occupational and environmental medicine, 57(5), 353-357.

Bland, J. M., & Altman, D. G. (1995). Calculating correlation coefficients with repeated observations: Part 2—Correlation between subjects. Bmj, 310(6980), 633.

Blascheck, T., Kurzhals, K., Raschke, M., Burch, M., Weiskopf, D., & Ertl, T. (2014, June). State-of-the-Art of Visualization for Eye Tracking Data. In EuroVis (STARs).

Boksem, M. A., Meijman, T. F., & Lorist, M. M. (2005). Effects of mental fatigue on attention: an ERP study. Cognitive brain research, 25(1), 107-116.

Boksem, M. A., & Tops, M. (2008). Mental fatigue: costs and benefits. Brain research reviews, 59(1), 125-139.

Borbely, A. A. (1982). A two-process model of sleep regulation. Hum neurobiol, 1(3), 195-204.

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. Neuroscience & Biobehavioral Reviews, 44, 58-75.

Bourgeois-Bougrine, S., Carbon, P., Gounelle, C., Mollard, R., & Coblentz, A. (2003). Perceived fatigue for short-and long-haul flights: a survey of 739 airline pilots. Aviation, space, and environmental medicine, 74(10), 1072-1077.

Brandes, U. (2001). A faster algorithm for betweenness centrality. Journal of mathematical sociology, 25(2), 163-177.

Branford, K., Lowe, A., Hayward, B., Pacific, A., & Cabon, P. (2014). Biomathematical Fatigue Models. Guidance Document Civil Aviation Safety Authority (CASA) Australia.

Bris, R., & Soares, C. G. (2009). Reliability, Risk, and Safety, Three Volume Set: Theory and Applications. CRC Press.

Brown, L., & Whitehurst, G. (2012). Effects of Commuting on Crewmember Fatigue

Bruseberg, A. (2009). Cost-Benefit Analysis for Human Factors Integration: A Practical Guide. HFI DTC/WP, 2(2), 3.

Cabon, P., Deharvengt, S., Grau, J.Y., Maille, N., Berechet, I. and Mollard, R., 2012. Research and guidelines for implementing Fatigue Risk Management Systems for the French regional airlines. Accident Analysis & Prevention, 45, pp.41-44

Caldwell, J. A. (2005). Fatigue in aviation. Travel medicine and infectious disease, 3(2), 85-96.

Caldwell, J. A., Mallis, M. M., Caldwell, J. L., Paul, M. A., Miller, J. C., & Neri, D. F. (2009). Fatigue countermeasures in aviation. Aviation, space, and environmental medicine, 80(1), 29-59.

Caldwell, J. A. (2012). Crew schedules, sleep deprivation, and aviation performance. Current Directions in Psychological Science, 21(2), 85-89.

Caldwell, J. A., & Caldwell, J. L. (2016). Fatigue in aviation: A guide to staying awake at the stick. Routledge.

Cao, T., Wan, F., Wong, C. M., da Cruz, J. N., & Hu, Y. (2014). Objective evaluation of fatigue by EEG spectral analysis in steady-state visual evoked potential-based brain-computer interfaces. Biomedical engineering online, 13(1), 28.

Cheung, B., Vartanian, O., Hofer, K., & Bouak, F. (2010). General recommendations on fatigue risk management for the Canadian Forces (No. DRDC-TR-2010-056). Defense research and development Toronto (Canada).

Civil Aviation Safety Authority of Australia. (2014). Biomathematical Fatigue Models Guidance Document.

Dekker, S. W. (2002). The field guide to Human Error Investigations, Bedford, UK.

Deng, L. Y., Hsu, C. L., Lin, T. C., Tuan, J. S., & Chang, S. M. (2010). EOG-based Human–Computer Interface system development. Expert Systems with Applications, 37(4), 3337-3343.

Detwiler, C., Hackworth, C., Holcomb, K., Boquet, A., Pfleiderer, E., Wiegmann, D., & Shappell, S. (2006). Beneath the tip of the iceberg: A human factors analysis of general aviation accidents in Alaska versus the rest of the United States.

Dhillon, B. S. (2009). Human reliability, error, and human factors in engineering maintenance: with reference to aviation and power generation. CRC Press.

Di Nocera, F., Camilli, M., & Terenzi, M. (2007). A random glance at the flight deck: Pilots' scanning strategies and the real-time assessment of mental workload. Journal of Cognitive Engineering and Decision Making, 1(3), 271-285.

Diaz-Piedra, C., Rieiro, H., Cherino, A., Fuentes, L. J., Catena, A., & Di Stasi, L. L. (2019). The effects of flight complexity on gaze entropy: An experimental study with fighter pilots. Applied ergonomics, 77, 92-99.

Diaz-Piedra, C., Sanchez-Carrion, J. M., Rieiro, H., & Di Stasi, L. L. (2017). Gaze-based technology as a tool for surgical skills assessment and training in urology. Urology, 107, 26-30.

Diez, M., Boehm-Davis, D. A., Holt, R. W., Pinney, M. E., Hansberger, J. T., & Schoppek, W. (2001). Tracking pilot interactions with flight management systems through eye movements. In Proceedings of the 11th International Symposium on Aviation Psychology (Vol. 6).

Dillard, M. B., Orhan, U., & Letsu-Dake, E. (2016). Nonintrusive pilot fatigue monitoring. In Processing of AIAA/IEEE 35th Digital Avionics Systems Conference Vol. 2016–December (pp. 1–6).

Dinges, D. F., Mallis, M. M., Maislin, G., & Powell, J. W. (1998). Evaluation of techniques for ocular measurement as an index of fatigue and as the basis for alertness management (No. DOT-HS-808-762). United States. National Highway Traffic Safety Administration.

Dismukes, R. K. (2017). Human error in aviation. Routledge.

Dismukes, R. K., Goldsmith, T. E., & Kochan, J. A. (2015). Effects of acute stress on aircrew performance: literature review and analysis of operational aspects. NASA Technical Memorandum, 218930.

Donaldson, M. S., Corrigan, J. M., & Kohn, L. T. (Eds.). (2000). To err is human: building a safer health system (Vol. 6). National Academies Press.

Drury, D. A., Ferguson, S. A., & Thomas, M. J. (2012). Restricted sleep and negative affective states in commercial pilots during short haul operations. Accident Analysis & Prevention, 45, 80-84.

Duchowski, A. T. (2007). Eye tracking methodology. Theory and practice, 328(614), 2-3.

Dul, J., Bruder, R., Buckle, P., Carayon, P., Falzon, P., Marras, W. S., ... & van der Doelen, B. (2012). A strategy for human factors/ergonomics: developing the discipline and profession. Ergonomics, 55(4), 377-395.

Eckstein, M. K., Guerra-Carrillo, B., Singley, A. T. M., & Bunge, S. A. (2017). Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? Developmental cognitive neuroscience, 25, 69-91.

Edwards, E. (1972). Man, and machine: Systems for safety. Proceedings of British Airline Pilots Association Technical Symposium (pp. 21–36). London: British Airline Pilots Association.

Embrey, D. (2005). Understanding human behavior and error. Human Reliability Associates, 1(2005), 1-10.

European Cockpit Association. (2012). Pilot fatigue barometer. Brussels: Rue du Commerce.

Farnsworth, B. (2018). What is EEG (Electroencephalography) and How Does it Work?

Federal Aviation Administration (2011). Final rule: flight crew member duty and rest requirements. 14 CFR, Parts 117, 119, and 121.

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. Journal of vision, 8(2), 6-6.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. Social networks, 1(3), 215-239.

Gander, P. H., Gregory, K. B., Graeber, R. C., Connell, L. J., Miller, D. L., & Rosekind, M. R. (1998). Flight crew fatigue II: short-haul fixed-wing air transport operations. Aviation Space and Environmental Medicine, 69, 8-15.

Gander, P. H., Gregory, K. B., Graeber, R. C., Connell, L. J., Miller, D. L., & Rosekind, M. R. (1998). Flight crew fatigue V: Long-haul air transport operations. Aviation, space, and environmental medicine.

Gander, P. H., Mangie, J., van den Berg, M. J., Smith, A. A. T., Mulrine, H. M., & Signal, T. L. (2014). Crew fatigue safety performance indicators for fatigue risk management systems. Aviation, space, and environmental medicine, 85(2), 139-147.

Gander, P. H., Signal, T. L., van den Berg, M. J., Mulrine, H. M., Jay, S. M., & Jim Mangie, C. (2013). In-flight sleep, pilot fatigue and P sychomotor V igilance T ask performance on ultra-long range versus long range flights. Journal of sleep research, 22(6), 697-706.

Gao, X., Cao, H., Ming, D., Qi, H., Wang, X., Wang, X., ... & Zhou, P. (2014). Analysis of EEG activity in response to binaural beats with different frequencies. International Journal of Psychophysiology, 94(3), 399-406.

Gislason, S. H., Bogdane, R., & Vasiļevska-Nesbita, I. (2017). Fatigue Monitoring Tool for Airline Operators (FMT). Transport and Aerospace Engineering, 5(1), 67-74.

Glaholt, M. G. (2014). Eye tracking in the cockpit: a review of the relationships between eye movements and the aviators cognitive state. Defense Research and Development Toronto (No. DRDC-RDDC-2014-R153) (Canada).

Goel, N., Basner, M., Rao, H., & Dinges, D. F. (2013). Circadian rhythms, sleep deprivation, and human performance. In Progress in molecular biology and translational science (Vol. 119, pp. 155-190). Academic Press.

Goker, Z. (2018). Fatigue in The Aviation: An Overview of The Measurements and Countermeasures. Journal of Aviation, 2(2), 185–194.

Golz, M., Sommer, D., Trutschel, U., Sirois, B., & Edwards, D. (2010). Evaluation of fatigue monitoring technologies. Somnologie-Schlafforschung und Schlafmedizin, 14(3), 187-199.

Gong, L., Zhang, S., Tang, P., & Lu, Y. (2014). An integrated graphic–taxonomic–associative approach to analyze human factors in aviation accidents. Chinese Journal of Aeronautics, 27(2), 226-240.

Gregory, K. B., Winn, W., Johnson, K., & Rosekind, M. R. (2010). Pilot fatigue survey: exploring fatigue factors in air medical operations. Air medical journal, 29(6), 309-319.

Greubel, J., & Nachreiner, F. (2013). The validity of the risk index for comparing the accident risk associated with different work schedules. Accident Analysis & Prevention, 50, 1090-1095.

Guimarans, D., Arias, P., Tomasella, M., & Wu, C. L. (2019). A Review of Sustainability in Aviation: A Multidimensional Perspective. In Sustainable Transportation and Smart Logistics (pp. 91-121). Elsevier.

Gundel, A., Marsalek, K., & ten Thoren, C. (2007). A critical review of existing mathematical models for alertness. Somnologie-Schlafforschung und Schlafmedizin, 11(3), 148-156.

Hamlett, A., Ryan, L., & Wolfinger, R. (2004). On the use of PROC MIXED to estimate correlation in the presence of repeated measures. Proc Statistics and Data Analysis, 198-29.

Hansen, D. W., & Ji, Q. (2009). In the eye of the beholder: A survey of models for eyes and gaze. IEEE transactions on pattern analysis and machine intelligence, 32(3), 478-500.

Harris, R. L., Glover, B. L., & Spady, A. A. (1986). Analytic techniques of pilot scanning behavior and their application (Technical Paper No. 2525). Hampton, VA: NASA Langley Research Center.

Hartzler, B. M. (2014). Fatigue on the flight deck: the consequences of sleep loss and the benefits of napping. Accident Analysis & Prevention, 62, 309-318.

Heide, W., Koenig, E., Trillenberg, P., Kömpf, D., & Zee, D. S. (1999). Electrooculography: technical standards and applications. Electroencephalogr Clin Neurophysiol Suppl, 52, 223-240.

Helmreich, R. L. (2000). On error management: lessons from aviation. Bmj, 320(7237), 781-785.

Ho, H. F., Su, H. S., Li, W. C., Yu, C. S., & Braithwaite, G. (2016, July). Pilots' latency of first fixation and dwell among regions of interest on the flight deck. In International Conference on Engineering Psychology and Cognitive Ergonomics (pp. 389-396).

Hollnagel, E. (2018). Safety-I and safety-II: the past and future of safety management. CRC press.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). Eye tracking: A comprehensive guide to methods and measures. OUP Oxford.

Hopstaken, J. F., Van Der Linden, D., Bakker, A. B., & Kompier, M. A. (2015). A multifaceted investigation of the link between mental fatigue and task disengagement. Psychophysiology, 52(3), 305-315.

Horsley, M., Eliot, M., Knight, B. A., & Reilly, R. (Eds.). (2013). Current trends in eye tracking research. Springer Science & Business Media.

Hudson, P. (2001). Safety culture-theory and practice. Leiden univ (Netherlands) center for safety science.

International Civil Aviation Organization. (2004). Manual on the regulation of international air transport (Vol. 9626). International Civil Aviation Organization.

International Civil Aviation Organization. (2010). Annex 6–Operation of aircraft. Part I–International Commercial Air Transport–Airplanes.

International Civil Aviation Organization. (2013). Safety management manual (SMM). International Civil Aviation Organization.

Jacob, R. J., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In The mind's eye (pp. 573-605). North-Holland.

Janelle, C. M., & Hatfield, B. D. (2008). Visual attention and brain processes that underlie expert performance: Implications for sport and military psychology. Military Psychology, 20(S1), S39.

Jarodzka, H., Scheiter, K., Gerjets, P., & Van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. Learning and Instruction, 20(2), 146-154.

Jing, M. J., Lin, W. Q., Wang, Q., Wang, J. J., Tang, J., Jiang, E. S., ... & Wang, P. X. (2016). Reliability and construct validity of two versions of chalder fatigue scale among the general population in mainland China. International journal of environmental research and public health, 13(1), 147.

Kang, Z., & Landry, S. J. (2014). Using scanpaths as a learning method for a conflict detection task of multiple target tracking. Human factors, 56(6), 1150-1162.

Kirby, C. E., Kennedy, Q., & Yang, J. H. (2014). Helicopter pilot scan techniques during low-altitude high-speed flight. Aviation, space, and environmental medicine, 85(7), 740-744.

Kramer, A., Tham, M., Konrad, C., Wickens, C., Lintern, G., Marsh, R., ... & Merwin, D. (1994). Instrument scan and pilot expertise. Proceedings of the Human Factors and Ergonomics Society, 1, 36-40.

Krejtz, K., Szmidt, T., Duchowski, A. T., & Krejtz, I. (2014, March). Entropy-based statistical analysis of eye movement stationary. In Proceedings of the Symposium on Eye Tracking Research and Applications (pp. 159-166). ACM.

Krejtz, K., Duchowski, A., Szmidt, T., Krejtz, I., González Perilli, F., Pires, A., ... & Villalobos, N. (2015). Gaze transition entropy. ACM Transactions on Applied Perception (TAP), 13(1), 1-20.

Krey, N. (2007). The Nall Report 2007: Accident Trends and Factors for 2006. AOPA Air Safety Foundation.

Lacagnina, M. (2007). Defusing the Ramp: Progress Report on FSF efforts to stem the toll of ground accidents. Flight Safety Foundation, AeroSafety World May.

Lamond, N., Jay, S. M., Dorrian, J., Ferguson, S. A., Roach, G. D., & Dawson, D. (2008). The sensitivity of a palm-based psychomotor vigilance task to severe sleep loss. Behavior research methods, 40(1), 347-352.

Lee, I.-S., Bardwell, W. A., Ancoli-Israel, S., & Dimsdale, J. E. (2010). Number of Lapses during the Psychomotor Vigilance Task as an Objective Measure of Fatigue. Journal of Clinical Sleep Medicine: JCSM: Official Publication of the American Academy of Sleep Medicine, 6(2), 163–168.

Lee, S., & Kim, J. K. (2018). Factors contributing to the risk of airline pilot fatigue. Journal of air transport management, 67, 197-207.

Li, G., Baker, S. P., Grabowski, J. G., & Rebok, G. W. (2001). Factors associated with pilot error in aviation crashes. Aviation, space, and environmental medicine, 72(1), 52-58.

Li, W. C., Lin, J. J., Braithwaite, G., & Greaves, M. (2016). The development of eye tracking in aviation (ETA) technique to investigate pilot's cognitive processes of attention and decision-making.

Lim, J., & Dinges, D. F. (2008). Sleep deprivation and vigilant attention. Annals of the New York Academy of Sciences, 1129(1), 305-322.

Lorist, M. M., Klein, M., Nieuwenhuis, S., De Jong, R., Mulder, G., & Meijman, T. F. (2000). Mental fatigue and task control: planning and preparation. Psychophysiology, 37(5), 614-625.

Lupu, R. G., & Ungureanu, F. (2013). A survey of eye tracking methods and applications. Buletinul Institutului Politehnic din Iasi, Automatic Control and Computer Science Section, 3, 72-86.

Mallis, M. M., Mejdal, S., Nguyen, T. T., & Dinges, D. F. (2004). Summary of the key features of seven biomathematical models of human fatigue and performance. Aviation, space, and environmental medicine, 75(3), A4-A14.

Mandal, S., & Kang, Z. (2018). Using Eye Movement Data Visualization to Enhance Training of Air Traffic Controllers: A Dynamic Network Approach. Journal of Eye Movement Research, 11(4).

Mandal, S., Kang, Z., & Millan, A. (2016, September). Data Visualization of Complex Eye Movements Using Directed Weighted Networks: A Case Study on a Multi-Element Target Tracking Task. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 60, No. 1, pp. 106-110). Sage CA: Los Angeles, CA: SAGE Publications.

Marandi, R. Z., Madeleine, P., Omland, Ø., Vuillerme, N., & Samani, A. (2018). Eye movement characteristics reflected fatigue development in both young and elderly individuals. Scientific reports, 8(1), 13148.

Marcus, J. H., & Rosekind, M. R. (2017). Fatigue in transportation: NTSB investigations and safety recommendations. Injury prevention, 23(4), 232-238.

Maurino, D. E., Reason, J., Johnston, N., & Lee, R. B. (2017). Beyond aviation human factors: Safety in high technology systems. Routledge.

McMahon, T. W., & Newman, D. G. (2015). A methodology to determine the psychomotor performance of helicopter pilots during flight maneuvers. Aerospace medicine and human performance, 86(7), 641-646.

Mellor, D. (2012). Use of Bio-mathematical Models to Assess Rosters and Pairings: A predictive alertness model designed specifically for the aviation industry. In FRMS Forum. FRMSc Ltd.

Miranda, A. T. (2018). Understanding human error in naval aviation mishaps. Human factors, 60(6), 763-777.

Mueller, S. T., & Piper, B. J. (2014). The psychology experiment building language (PEBL) and PEBL test battery. Journal of neuroscience methods, 222, 250-259.

Muller, J. A., Wendt, D., Kollmeier, B., & Brand, T. (2016). Comparing eye tracking with electrooculography for measuring individual sentence comprehension duration. PloS one, 11(10), e0164627.

Naeeri, S. M., Kang, Z. (2018). Exploring the relationship between pilot's performance and fatigue when interacting with cockpit interfaces. In Proceedings of the 2018 Institute of Industrial and Systems Engineers (IISE) Annual Conference, May 19-22, Orlando, FL.

Naeeri, S., and Kang, Z. (2017). Analysis of Pilot's Visual Scanning Characteristics under Normal and Extreme Flight Conditions. In Proceedings of the 6th Annual World Conference of the Society for Industrial and Systems Engineering, Herndon, VA. 227-233.

Naeeri, S., Kang, Z., & Mandal, S. (2018). Exploring the effect of fatigue on pilot performance during single and multi-takeoffs and landings flight missions. In Proceedings of the 7th Annual World Conference of the Society for Industrial and Systems Engineering, 7(1) 174-181

Naeeri, S., Mandal, S., & Kang, Z. (2019, November). Analyzing pilots' fatigue for prolonged flight missions: Multi-modal analysis approach using vigilance test and eye tracking. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 63, No. 1, pp. 111-115). Sage CA: Los Angeles, CA: SAGE Publications.

Neuberger, G. B. (2003). Measures of fatigue: the fatigue questionnaire, fatigue severity scale, multidimensional assessment of fatigue scale, and short form-36 vitality (energy/fatigue) subscale of the short form health survey. Arthritis Care & Research: Official Journal of the American College of Rheumatology, 49(S5), S175-S183.

Newman, M. E. (2001). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. Physical review E, 64(1), 016132.

Opsahl, T., & Panzarasa, P. (2009). Clustering in weighted networks. Social networks, 31(2), 155-163.

Oster Jr, C. V., Strong, J. S., & Zorn, C. K. (2013). Analyzing aviation safety: problems, challenges, opportunities. Research in transportation economics, 43(1), 148-164.

Peibl, S., Wickens, C. D., & Baruah, R. (2018). Eye-tracking measures in aviation: a selective literature review. The International Journal of Aerospace Psychology, 28(3-4), 98-112.

Peng, H., & Bouak, F. (2015). Development of bio-mathematical models for human performance under fatigue. DRDC Scientific Report, DRDC-RDDC-2015-R280.

Perrow, C. (2011). Normal accidents: Living with high risk technologies-Updated edition. Princeton university press.

Petrilli, R. M., Roach, G. D., Dawson, D., & Lamond, N. (2006). The sleep, subjective fatigue, and sustained attention of commercial airline pilots during an international pattern. Chronobiology international, 23(6), 1357-1362.

Phillips, R. O. (2014). An assessment of studies of human fatigue in land and sea transport. TØI Report, (1354/2014).

Pierre, J., & Peters, B. G. (2009). From a club to a bureaucracy: JAA, EASA, and European aviation regulation. Journal of European Public Policy, 16(3), 337-355.

Poole, A. & Ball, L.J. (2005). Eye tracking in HCI and usability research. Encyclopedia of Human Computer Interaction, C. Ghaoui, Ed. Idea Group Reference. 1. 211-219. 10.4018/978-1-59140-562-7.ch034.

Powell, D., Spencer, M. B., Holland, D., Broadbent, E., & Petrie, K. J. (2007). Pilot fatigue in short-haul operations: Effects of number of sectors, duty length, and time of day. Aviation, Space, and Environmental Medicine, 78(7), 698-701.

Powell, D., Spencer, M. B., Holland, D., & Petrie, K. J. (2008). Fatigue in two-pilot operations: implications for flight and duty time limitations. Aviation, space, and environmental medicine, 79(11), 1047-1050.

Powell, D., Spencer, M. B., & Petrie, K. J. (2014). Comparison of in-flight measures with predictions of a bio-mathematical fatigue model. Aviation, space, and environmental medicine, 85(12), 1177-1184.

Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. IEEE transactions on systems, man, and cybernetics, (3), 257-266.

Rasmussen, J. (1990). Human error and the problem of causality in analysis of accidents. Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 327(1241), 449-462.

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

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. Psychological bulletin, 124(3), 372.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. The quarterly journal of experimental psychology, 62(8), 1457-1506.

Reder, L. M., Weber, K., Shang, J., & Vanyukov, P. M. (2003). The adaptive character of the attentional system: statistical sensitivity in a target localization task. Journal of Experimental Psychology: Human Perception and Performance, 29(3), 631.

Reis, C., Mestre, C., Canhão, H., Gradwell, D., & Paiva, T. (2016). Sleep and fatigue differences in the two most common types of commercial flight operations. Aerospace medicine and human performance, 87(9), 811-815.

Richardson, D. C., & Spivey, M. J. (2004). Eye tracking: Characteristics and methods. Encyclopedia of biomaterials and biomedical engineering, 3, 1028-1042.

Robinson, K. W. (2011). Improving Human Factors Training: Perceptions of Retired Aircraft Maintenance Technicians. In 16th International Symposium on Aviation Psychology (p. 327).

Roma, P. G., Mallis, M. M., Hursh, S. R., Mead, A. M., & Nesthus, T. E. (2010). Flight Attendant Fatigue Recommendation 2: Flight Attendant Work/Rest Patterns, Alertness, and Performance Assessment (No. DOT/FAA/AM-10/22). INSTITUTE FOR BEHAVIOR RESOURCES BALTIMORE MD.

Rosekind, M. R., Graeber, R. C., Dinges, D. F., Connell, L. J., Rountree, M. S., Spinweber, C. L., & Gillen, K. A. (1994). Crew factors in flight operations IX: effects of planned cockpit rest on crew performance and alertness in long-haul operations.

Rosekind, M. R., Smith, R. M., Miller, D. L., Co, E. L., Gregory, K. B., Webbon, L. L., ... & Lebacqz, J. V. (1995). Alertness management: strategic naps. In operational settings. Journal of Sleep Research, 4, 62-66.

Russ, A. L., Fairbanks, R. J., Karsh, B. T., Militello, L. G., Saleem, J. J., & Wears, R. L. (2013). The science of human factors: separating fact from fiction. BMJ Qual Saf, 22(10), 802-808.

Sadeghniiat-Haghighi, K., & Yazdi, Z. (2015). Fatigue management in the workplace. Industrial psychiatry journal, 24(1), 12.

Salas, E., Maurino, D., & Curtis, M. (2010). Human factors in aviation: an overview. In Human factors in aviation (pp. 3-19). Academic Press.

Salazar, G. J. (2007). Fatigue in aviation, series: Medical facts for pilots. OK-07-193.

Salmon, P. M., Cornelissen, M., & Trotter, M. J. (2012). Systems-based accident analysis methods: A comparison of Accimap, HFACS, and STAMP. Safety science, 50(4), 1158-1170.

Samel, A., Wegmann, H. M., & Vejvoda, M. (1997). Aircrew fatigue in long-haul operations. Accident Analysis & Prevention, 29(4), 439-452.

Samn, S. W., & Perelli, L. P. (1982). Estimating aircrew fatigue: a technique with application to airlift operations (No. SAM-TR-82-21). School of Aerospace Medicine Brooks Afb tx.

Schieber, F., & Gilland, J. (2008, September). Visual entropy metric reveals differences in drivers' eye gaze complexity across variations in age and subsidiary task load. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 52, No. 23, pp. 1883-1887). Sage CA: Los Angeles, CA: SAGE Publications.

Sexton, J. B., Thomas, E. J., & Helmreich, R. L. (2000). Error, stress, and teamwork in medicine and aviation: cross sectional surveys. Bmj, 320(7237), 745-749.

Shappell, S. A., & Wiegmann, D. A. (2000). The human factors analysis and classification system--HFACS.

Shappell, S. A., & Wiegmann, D. A. (2004). HFACS analysis of military and civilian aviation accidents: A North American comparison. In Proceedings of the Annual Meeting of the International Society of Air Safety Investigators. Australia: Gold Coast.

Shappell, S., Detwiler, C., Holcomb, K., Hackworth, C., Boquet, A., & Wiegmann, D. (2006). Human error and commercial aviation accidents: A comprehensive, fine-grained analysis using HFACS.

Shappell, S., Detwiler, C., Holcomb, K., Hackworth, C., Boquet, A., & Wiegmann, D. A. (2017). Human error and commercial aviation accidents: an analysis using the human factors analysis and classification system. In Human Error in Aviation (pp. 73-88). Routledge.

Shiferaw, B. A., Downey, L. A., Westlake, J., Stevens, B., Rajaratnam, S. M., Berlowitz, D. J., ... & Howard, M. E. (2018). Stationary gaze entropy predicts lane departure events in sleep-deprived drivers. Scientific reports, 8(1), 1-10.

Shiferaw, B., Downey, L., & Crewther, D. (2019). A review of gaze entropy as a measure of visual scanning efficiency. Neuroscience & Biobehavioral Reviews, 96, 353-366.

Smets, E. M. A., Garssen, B., Cull, A., & De Haes, J. C. J. M. (1996). Application of the multidimensional fatigue inventory (MFI-20) in cancer patients receiving radiotherapy. British journal of cancer, 73(2), 241.

Stokes, A. F., & Kite, K. (2017). Flight stress: Stress, fatigue and performance in aviation. Routledge.

Stolzer, A. J. (2017). Safety management systems in aviation. Routledge.

Sullivan, J., Yang, J. H., Day, M., & Kennedy, Q. (2011). Training simulation for helicopter navigation by characterizing visual scan patterns. Aviation, space, and environmental medicine, 82(9), 871-878.

Thomas, L. C., Gast, C., Grube, R., & Craig, K. (2015). Fatigue detection in commercial flight operations: Results using physiological measures. Procedia Manufacturing, 3, 2357-2364.

Underwood, G. (2007). Visual attention and the transition from novice to advanced driver. Ergonomics, 50(8), 1235-1249.

Valko, P. O., Bassetti, C. L., Bloch, K. E., Held, U., & Baumann, C. R. (2008). Validation of the fatigue severity scale in a Swiss cohort. Sleep, 31(11), 1601-1607.

Van Drongelen, A., Van Der Beek, A. J., Hlobil, H., Smid, T., & Boot, C. R. L. (2013). Development and evaluation of an intervention aiming to reduce fatigue in airline pilots: Design of a randomised controlled trial. BMC Public Health, 13(1).

Vejvoda, M., Elmenhorst, E. M., Pennig, S., Plath, G., Maass, H., Tritschler, K., ... & Aeschbach, D. (2014). Significance of time awake for predicting pilots' fatigue on short-haul flights: implications for flight duty time regulations. Journal of sleep research, 23(5), 564-567.

Velazquez, J. (2018). The Presence of Behavioral Traps in U.S. Airline Accidents: A Qualitative Analysis. Safety, 4(1), 2.

Wenner, C. (2011). What is Human Factors and Why is it Important? [Power points slides]. Retrieved from https://www.osti.gov/servlets/purl/1110674

Webster, K., Cella, D., & Yost, K. (2003). The F unctional A ssessment of C hronic I llness T herapy (FACIT) Measurement System: properties, applications, and interpretation. Health and quality of life outcomes, 1(1), 79.

Wewers, M. E., Rachfal, C., & Ahijevych, K. (1990). A psychometric evaluation of a visual analogue scale of craving for cigarettes. Western journal of nursing research, 12(5), 672-681.

Wickens, C. D., Gordon, S. E., & Liu, Y. (1998). An introduction to human factors engineering.

Wiegmann, D. A., & Shappell, S. A. (2017). A human error approach to aviation accident analysis: The human factors analysis and classification system. Routledge.

Williamson, A., Lombardi, D. A., Folkard, S., Stutts, J., Courtney, T. K., & Connor, J. L. (2011). The link between fatigue and safety. Accident Analysis & Prevention, 43(2), 498-515.

Wilson, G. F., Caldwell, J. A., & Russell, C. A. (2007). Performance and psychophysiological measures of fatigue effects on aviation related tasks of varying difficulty. The international journal of aviation psychology, 17(2), 219-247.

Wolfe, J. M., Alvarez, G. A., & Horowitz, T. S. (2000). Attention is fast but volition is slow. Nature, 406(6797), 691-691.

Woods, D. D., & Cook, R. I. (2003). Mistaking error. Patient safety handbook, 95-108.

Xiong, W., Wang, Y., Zhou, Q., Liu, Z., & Zhang, X. (2016, July). The research of eye movement behavior of expert and novice in flight simulation of landing. In International Conference on Engineering Psychology and Cognitive Ergonomics (pp. 485-493).

Zhang, Y. F., Gao, X. Y., Zhu, J. Y., Zheng, W. L., & Lu, B. L. (2015). A novel approach to driving fatigue detection using forehead EOG. International IEEE/EMBS Conference on Neural Engineering, NER, 2015–July 707–710. https://doi.org/10.1109/NER.2015.7146721

Ziv, G. (2016). Gaze behavior and visual attention: A review of eye tracking studies in aviation. The International Journal of Aviation Psychology, 26(3-4), 75-104. Abeyratne, R. (2012). Strategic issues in air transport: legal, economic and technical aspects. Springer Science & Business Media.

Ziv, G., & Ziv, G. (2017). Gaze Behavior and Visual Attention: A Review of Eye Tracking Studies in Aviation Gaze Behavior and Visual Attention: A Review of Eye Tracking Studies in Aviation. The International Journal of Aviation Psychology, 26(3–4), 75–104.

Chapter 8: Appendices

Appendix A: Samn-Perelli Ten-point fatigue Scale

Please read them carefully and choose the column that best corresponds to the statement describing your level of Fatigue.

STATEMENT	BETTER THAN	SAME AS	WORSE THAN
1) Very Lively			
2) Extremely Tired			
3) Quite fresh			
4) Slightly Pooped			
5) Extremely Peppy			
6) Somewhat Fresh			
7) Petered Out			
8) Very Refreshed			
9) Fairly Well Pooped			
10) Ready to Drop			

Appendix B: Karolinska Sleeping Scale

Please read them carefully and CIRCLE the number that best corresponds to the statement describing your level of sleepiness

1.	extremely alert	
2.		
3.	alert	
4.		
5.	Neither alert nor sleepy	
6.		
7.	Sleepy but no difficulty remaining awake	
8.		
9.	extremely sleepy fighting sleep	