UNIVERSITY OF OKLAHOMA
GRADUATE COLLEGE

MULTIMODAL NEUROERGONOMIC APPROACHES TO HUMAN BEHAVIOR AND COGNITIVE WORKLOAD IN COMPLEX HIGH-RISK SEMANTICALLY RICH ENVIRONMENTS: A CASE STUDY OF LOCAL & EN-ROUTE AIR TRAFFIC CONTROLLERS

A THESIS
SUBMITTED TO THE GRADUATE FACULTY
in partial fulfillment of the requirements for the
Degree of
MASTER OF SCIENCE

By
RICARDO PALMA FRAGA
Norman, Oklahoma 2020

MULTIMODAL NEUROERGONOMIC APPROACHES TO HUMAN BEHAVIOR AND COGNITIVE WORKLOAD IN COMPLEX HIGH-RISK SEMANTICALLY RICH ENVIRONMENTS: A CASE STUDY OF LOCAL & EN-ROUTE AIR TRAFFIC CONTROLLERS

A THESIS APPROVED FOR THE
SCHOOL OF INDUSTRIAL & SYSTEMS ENGINEERING

BY THE COMMITTEE CONSISTING OF

Dr. Ziho Kang, Chair

Dr. Edward Cokely

Dr. Randa Shehab

Acknowledgements

First and foremost, I would like to thank Dr. Ziho Kang, the chair of my thesis' committee, for acting as my mentor (and inspiration) all these years, and for introducing me to the fascinating world of eye-tracking, air traffic control, and human factors. I also want to thank Dr. Edward Cokely, for his mentorship, and whose passion for cognitive science and education have given me hefty goals to aspire and achieve. I would like to thank Dr. Randa Shehab, whose Ergonomics class sparked my interest in this field, for her inspirational comments throughout my undergraduate and graduate studies.

In addition, I would like to thank all of our amazing individuals within the FAA: Tracy Steagle, Cindy Fox, and in particular Dr. Jerry Crutchfield, whose help, passion, and involvement enabled successful completion of my thesis. In addition, I would also like to thank my colleagues at Drexel University, Pratusha Reddy, and Dr. Kurtulus Izzetoglu for their indispensable collaboration.

Finally, I would like to thank my family for their unrelentless support and companionship throughout my educational career. I would like to give my sincere thanks to all the members of the Human Factors and Simulation Laboratory, and particularly Saptarshi Mandal.

Table of Contents

List of Tables

List of Figures

Abstract

Fast-paced technology advancements have enabled us to create ecologically valid simulations of high risk, complex, and semantically rich environments in which human interaction and decision-making are the keys to increase system performance. These advances have improved our capabilities of exploring, quantifying, and measuring the underlying mechanisms that guide human behavior using sophisticated neuroergonomic devices; and in turn, improve human performance and reduce human errors. In this thesis, multimodal approaches consisted of a self-report analysis, eye-tracking analysis, and functional near-infrared spectroscopy analysis were used to investigate how veteran local & en-route air traffic controllers carry out their operational tasks. Furthermore, the correlations among the cognitive workload and physiological measures (i.e. eye movement characteristics and brain activities) were investigated. Combining the results of these experiments, we can observe that the multimodal approaches show promise on exploring the underlying mechanisms of workload and human interaction in a complex, high-risk, and semantically rich environment. This is because cognitive workload can be considered as a multidimensional construct and different devices or approaches might be more effective in sensing changes in either the task difficulty or complexity. The results can be used to find ways to better train the novices.

Keywords: air traffic control, multimodal methodology, eye-tracking, fNIRS, cognitive workload

Chapter 1: Introduction

One of the goals of Human Factors is to increase human performance by investigating how human operations interact with the interfaces and complete their tasks. Technological advancements have played a significant role in quantifying human performance and cognitive workload. The technological advancements have enabled us to realistically simulate complex environments while maintaining their ecological validity, and increase our capabilities to accurately and reliably measure both latent and salient variables.

Especially, important theories on human behavior have been discovered in the semantically rich, complex, and high-risk systems where humans play a significant role in maintaining the successful operations. Examples of such systems can be Formula 1 cockpit or the control room of a nuclear power plant. The focus of this paper is in air traffic control operations. Methods were developed to investigate the correlations among the cognitive workload and eye movement characteristics in order to better understand how veteran air traffic controllers interact with their environment to maintain the safely in the airspace.

Chapter 2 contains a comprehensive literature review on the use of the eye-tracking technology to explore human behavior and performance. In detail, methodologies and capabilities of eye-tracking research are summarized Especially, the validity of eye-tracking analysis methodologies are investigated in various semantically rich environments, such as sports, medicine, and construction. Then, the literature review is concentrated on air traffic control operations with an emphasis on how we could aid in effective training of future air traffic controllers. Lastly, the usage of other existing neuroergonomic tools are introduced.

Chapter 3 provides a multimodal research in which self-report analyses were combined with eye-tracking data obtained from veteran en-route Air Traffic Controllers in multiple ecologically valid scenarios. The goal of this experiment was to identify critical patterns that veterans apply as they carry out their tasks including: (1) visually search of the environment to detect possible conflicts among multiple aircraft; (2) apply aircraft control strategies to mitigate the identified conflicts. The general trends found from this multimodal approach might be used to train the novice air traffic controllers.

Chapter 4 provides a multimodal approach to reliably measure and quantify the cognitive workload of local air traffic controllers in a simulated tower environment. A non-intrusive functional near-infrared spectroscopy (fNIRS) neuroergonomic device was used to measure changes in oxygenation at the pre-frontal cortex of the brain. In addition, the changes in oxygenation were compared with the changes in the eye-tracking metrics. Furthermore, participants' cognitive workloads were investigated using the multiple physiological measures explained above. .

Chapter 5 provides a conclusion based on the research explained in chapters 3 and 4. Finally, future research topics are explored.

Chapter 2: Literature Review

2.1. Eye-tracking technology to Explore Human Performance

One of the key components that allows us to explore the salient and latent features that affect human performance when completing a task within an environment is the methodology behind Eye-tracking Research. This is because the process of identifying, discriminating, categorizing, and responding to most of the stimuli we face in an environment is captured by the human visual system (Williams, Davids & Williams, 1999). As such, eye-tracking allows us to identify the information of importance within an environment (Goldberg and Kotval, 1999; Ahlstrom and Friegman-Berg, 2006) which can be used to understand and map the cognitive processes behind a particular task. In particular, eye-tracking can be particularly useful in tasks that do not have "optimal" sequential solutions (e.g. to solve a problem, you must always do A, followed by B – doing B first, followed by A is not a possible alternative), but rather, the domain of solutions consists of multiple ways to complete a task in an adequate manner. Nonetheless, there exist differences between efficiency of these solutions, based on the workload they demand from the information processing system (Simon, 1979).

The objectives and goals behind applying eye-tracking to study a particular environment is both context and task dependent, yet there exists one common challenge in the literature: the functional relationship that exists between visual scanning strategies and performance in a task. This process is inherently difficult, as the tasks are affected by the dynamic complexity of the environmental information within a semantically rich domain. Simon describes these types of environments as those "domains in which successful performance calls for specific knowledge as well as general problem-solving skill" (Simon, 1979). As such, the major obstacle to understanding this fundamental relationship is observed: how can performance be adequately measured and quantified when all of these variables have to be considered. Studying these types of environments directly is often very restricted due to the hardware that is required to collect eye-tracking data, making the validity of the measures to be of low quality (Messick, 1995). In the other hand, several advances have been made within the realm of simulation technology, allowing us to re-create these semantically rich environments in a way that we can reliably measure human performance within them (Dhami, Hertwig, Hoffrage, 2004).

The recent advances in eye-tracking technology (ETT), both hardware and software, have allowed us to develop experiments that are both ecologically valid and high-validity. For example, it has improved the ability of researchers to understand how individuals behave and carry out their tasks when presented with diverse visual stimuli has improved. Elements that draw attention, or those that allow us to maintain vigilance, are capable of being identified more accurately, as it aids in observing how users interact with various interfaces (Goldberg and Kotval, 1999), their environmental information and the workload of the task (Ahlstrom and Friegman-Berg, 2006).

2.2. Eye-tracking Research Methodology

Eye-tracking research generally consists of looking at a few key metrics: fixations, saccades, and Areas of Interest (AOIs). Fixations take place when the eye remains stationary at a single place for more than 60 milliseconds, while saccades are the rapid eye movements between fixations. The combination of these two elements consist of a Scanpath: the complete eye-movements within an environment (Kang & Landry, 2014). To collect data, AOIs have to be designated, such that information can be collected in a binary manner (e.g. 0 if no fixation, 1 if fixation occurred). This process has been visualized in Figure 1 below.

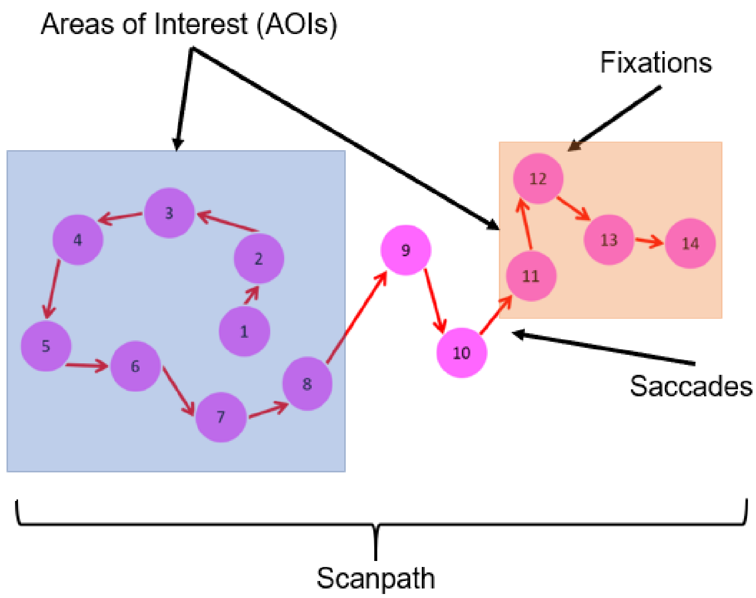


Figure 1: A Scanpath with Fixations, Saccades and AOIs highlighted.

The scanpath represents the visual scanning strategy that is applied to visually extract the relevant perceptual cues from a semantically rich environment. This process is also known as building the Situational Awareness of the environment, which is defined as: "the continuous extraction of environmental information, the integration of this information with prior knowledge to form a coherent understanding of the present situation." (Gronlund et al, 1998). As mentioned previously, the application of an eye-tracking methodology would allow us to better understand how this process is carried out, and in addition, help us develop an understanding of the functional relationship.

Most experiments, regardless of the semantically rich environment in question, tend to be the relatively similar. Due to the fact that we cannot discriminate between veterans and novices solely based on their Scanpath, we tend to work backwards: find veterans within that environment, collect and analyze their eye-movement data and identify relevant latent/salient features that affect it (e.g. Mann et al, 2007). This process works under various assumptions: (1)

veterans perform above a context dependent threshold within the domain of solutions where all of them tend to be acceptable (Figure 2); (2) performance is measured and quantified based on outcomes regarding a particular task, not directly from their visual scanning strategy. After the data is collected, it is analyzed in various ways, such as linear regression models (e.g. Currie et al, 2018) or various other data analysis tools, such as neural networks or discriminant analysis (Campana et al, 1999). The goal is to identify possible correlations between various eye-tracking metrics (e.g. average number of fixations, time spent fixating, etc.) and the performance of the veterans. With this overall framework now developed, we can begin to analyze how it is applied to various semantically rich environments, in order to find the differences and similarities between them. To be able to make direct comparisons (and due to the inherit complexity) between the eye-tracking metrics they use, this integrative review focuses on the task of identifying perceptual cues, rather than identification- comprehension-action that is generally done in most of the literature.

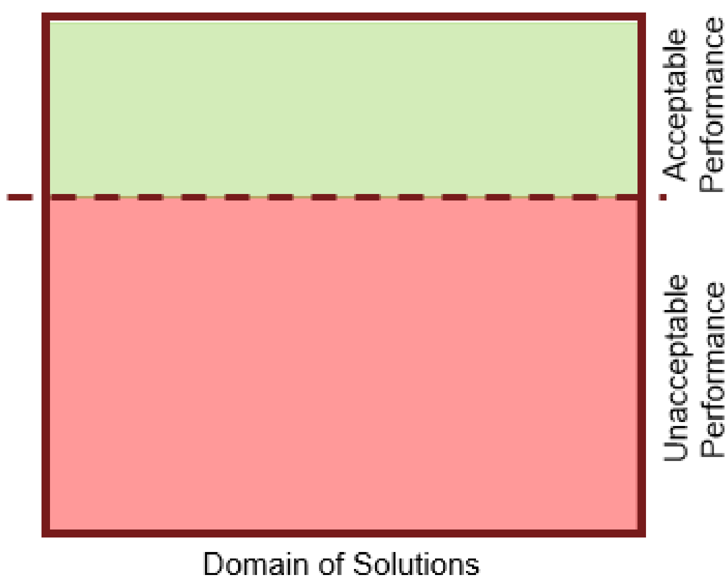


Figure 2: Visual representation for the domain of solutions within a task in a semantically rich environment.

2.3. Eye-tracking in Semantically Rich Environments

Working with Simon's definition, the number of semantically rich environments that exist are ample, from chess to Air Traffic Control. As such, this integrative review will focus only on four in which performance is measured differently: sports (e.g. ability to score), construction (e.g. hazards recognized), and medicine (e.g. identifying a lesion on an x-ray).

2.3.1. Sports

The ability of an athlete to visual scan the current state of the field is crucial to success. In addition, this cognitive process has to be tied to various motoric reaction and executing movements in order to compete at the highest level of a sport (Mann et al, 2007). In their integrative review, Hüttermann et al develop a comprehensive analysis of the eye-tracking literature in sports. They showcase one critical element within this semantically rich environment: (1) a significant number of research studies have only been carried out in static environments, or dead-ball situations (e.g. a free throw in basketball, or a penalty kick in soccer) which are generally, not very representative or ecologically valid of the fast-paced environment of a sport (Schulte-Mecklenbeck et al, 2017).

The number of studies that apply eye-tracking technology, capable of maintain a representative design are few and far in between. Nonetheless, these studies highlight key elements in terms of visual scanning strategies of veterans within this environment: high-performance athletes are better at identifying perceptual cues when compared to novices within the same sport. This is important, as it means that veteran athletes are capable of predicting the direction and force of an opponent's stroke. These features were identified through the athlete's performance in terms of response accuracy and time. An interesting finding that is mentioned in the integrative review, is the theory of Quiet Eye. Here, it is explained how adopting a goal-oriented visual scanning strategy, such as fixating on the rim prior to taking a free throw, is linked to higher performance in that task.

Based on this review, we can observe that there are two major differences between veterans and novices within this environment: their ability to pick up perceptual cues, and their implementation of the Quiet Eye Theory. These two are linked to a single performance metric: the identification of the relevant perceptual cue from the environment. From these two ideas, we can develop the following mapping to various eye-tracking measures: number of fixations prior to perceptual cue, time until fixating on perceptual cue and the time spent fixating on it.

2.3.2. Construction

Within the construction industry, the benefits of an eye-tracking analysis to hazard recognition are easily quantifiable. The number of fatal and non-fatal injuries reported in the United States during the year of 2016 were 991 and 200,000 respectively (Lingard, 2013). As such, being able to identify the cognitive processes that are utilized when visually searching for these hazards at the construction site extremely important. Similar to other semantically rich environment considered, eye-tracking can play an important role in this domain as well. In the study by Jeelani et al (2018), various participants were asked to identify the hazards shown through images of different constructions sites. During this process, their eye movements were recorded and subsequently analyzed. This semantically rich environment suffers from some of the similar problems that were highlighted in the previous one: representative design and ecological validity. Due to the limitations placed by eye-tracking technology, and this particular domain, the experiment could not be carried out at a real constructions site. As such, the various distractions that do not allow a worker to identify a hazard are not present, and the understanding as to how they impact the visual scanning strategy of an individual could not be studied. The results, nonetheless, showed that the number of perceptual cues identified between veterans and novices were a surprising 71% and 37%, respectively. The measure of accuracy they applied to identify the difference in skill levels were the accuracy and response time to identify the relevant perceptual cue.

2.3.3. Medicine

In the broad industry of medicine, there exist tasks that operate in a dynamic and complex semantically rich environment, from diverse medical procedures like surgeries to analyzing a patient's x- ray. Nurses and doctors apply visual scanning strategies to maintain the health of their patients, but nonetheless, just like in the construction industry, errors and mistakes are attached to significant consequences to all parties involved. Eye tracking research within this field works towards increasing human performance by understanding the cognitive process that take place during these visual search patterns, as explored thoroughly in the integrative review by Harezlak and Kasprowski (2018). As in the two previous semantically rich environments, the field of medicine also suffers from various difficulties in terms of representative design and ecological validity, mainly due to the limitations imposed by the eye- tracking hardware. Although various experiments have been successfully carried out, a lot of them resort to tools such as videos of operations (e.g. Khan et al, 2012) or simulations in order to create a representative design of the environment.

Due to the various applications of eye-tracking within medicine, the author focused on two particular experiments. In the first one, by Marquad et al (2011), they focused on analyzing the differences in gaze patterns of nurses during a simulated treatment of a patient. The goal was to identify their performance based on the number of identification errors throughout the scenarios.

The second experiment, by Manning et al (2006), was done in order to see the difference in visual scanning strategies between veteran and novice radiologist in a searching task on sample x-rays. Similarly, to the previous experiment, performance was measured by the ability of the participants to identify the relevant perceptual cues in the environment.

2.3.4. Conclusion

From the summaries developed for each semantically rich environment, and the boundaries set at the beginning of this project, we get the following diagram:

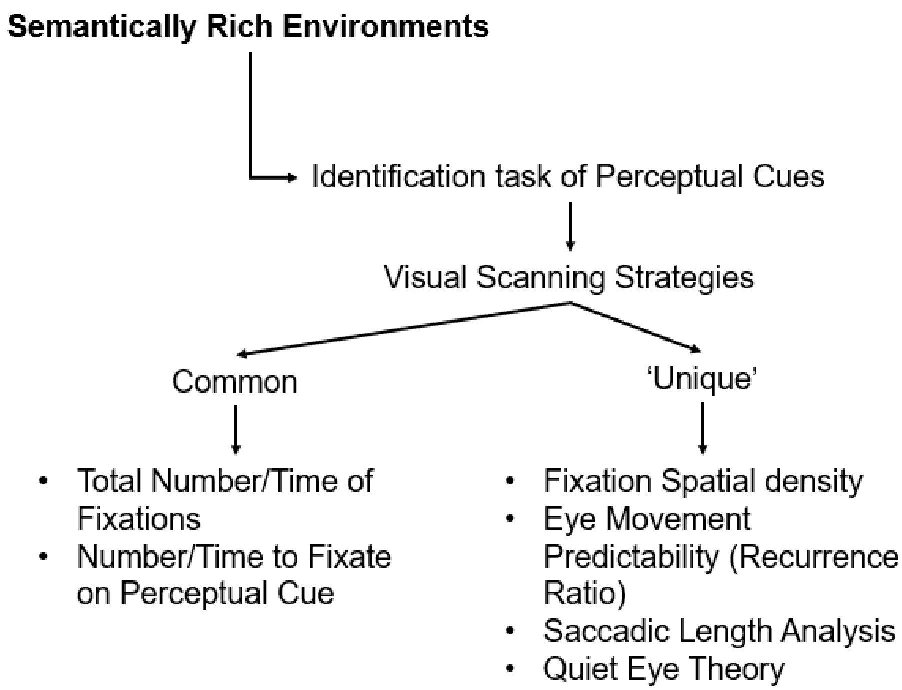


Figure 3: Integrative Review Results.

From this diagram, we can see that in a semantically rich environment, where the main task is or consists of identification of perceptual cues, there exist a number of eye-tracking measures related to performance that may be applicable to other environments. These measures are divided into two: common and unique. The former, represents various eye-metrics that appeared in all three of the semantically rich environments surveyed. In the other hand, the latter, represents measures that were in one or two of the environments considered. In addition, one can notice that there exists a temporal vs spatial element to these categories, where the common and unique categories fall under temporal and spatial, respectively. We can arrive to these two ideas and develop the following conclusions: (1) the eye-tracking metrics that correspond to performance are task dependent, and even when only the identification task was considered, the semantically rich environments utilized different metrics. This implies that we must develop a meta-analysis of the validity for the performance measures we are quantifying and mapping to performance in these environments; (2) it may be possible to combine and/or create new eye-tracking metrics that consider both: spatial and temporal elements in visual scanning strategies.

Overall, we can conclude the following from this integrative review: advances in technology, either eye-tracking or simulation, have allowed us to develop experiments that are representative of the semantically rich environments they are trying to emulate. This means that the conclusions that can be drawn between the functional relationship regarding eye-movements and performance is that, if present, there is a sufficient reason to believe that it can be measured and quantifiable.

2.4. Eye-tracking in Air Traffic Control

Within the task of Air Traffic Control, ETT allows us to analyze eye movements of ATCSs, such as moving between and within clusters of aircraft, as they develop their Situational Awareness of the sector they are currently working. This combination of eye fixations and the saccades between them are known as a “scanpath” (Kang and Landry, 2015), and they have been analyzed in the past in order to better understand how conflicts are detected (Hunter and Parush, 2009), how visual groupings are formed (Kang and Landry, 2010) and the overall visual search strategy of individuals (Aula, Marajanta and Raiha, 2005).

The difficulty in analyzing scanpaths becomes evident when data from multiple individuals is collected, as variations exist between participants under similar conditions (Chapman and Underwood, 1998; Kasarkis, et al, 2001; Underwood, 2007). These distinctions can be difficult to explain, even by participants themselves, due to the tacit nature of the knowledge. Nonetheless, advances have been made in this area by categorizing the scanpaths of veterans through various criteria, such as their geometrical shape (Kang, Bass and Lee, 2014). Most importantly, research has shown that presenting scanpaths of veterans to novices does impact the latter's visual scanning strategies in various ways, such as by reducing the rate of false alarms in an Air Traffic Control environment (Kang and Landry, 2014) and increasing their accuracy during the task (Dempere-Marco, et al, 2002). This shows that, if scanpaths of veteran Air Traffic Controllers can be identified, analyzed, and subsequently categorized, the underlying motives for such patterns can be understood, and adapted as training materials for the next generation of ATCSs.

2.5. Impact of increased air traffic demand ATCs operations.

For ATCSs to be able to carry out the task of identifying and, if needed, mitigating conflicts, they must be able to create, maintain and update their understanding of the sector as it changes over time. This process is referred to as Situational Awareness, and "the picture" within the ATC community (Niessen and Eyfeth, 2001), as has been defined as the "continuous extraction of environmental information, the integration of this information with prior knowledge to form a coherent understanding of the present situation, and the use of that coherent understanding to direct perception and anticipate future events (Gronlund, et al, 1998). Within the context of air traffic control, the environmental information is represented by the various sector-dependent characteristics (e.g. inclement weather) and the constantly changing variables: altitude, speed, and direction of aircraft, because they are all required to anticipate future events, such as possible potential conflicts (Nunes and Mogford, 2003; Durso, et al, 1998). This process is vital to the role of ATCSs, as poor Situational Awareness can lead to important sources of errors in aviation (Di Nocera, Fabrizi, Terenzi, Ferlazzo, 2006) such as poor judgement, decision making and communication (Stager, Hameluck and Jubis, 1989), which can lead to tragic consequences (O'Brien and O'Hare, 2007).

Building and maintaining Situational Awareness within the context of air traffic control is a difficult task due to the dynamic complexity of the environmental information that can affect the speed and efficiency of the ATCSs' visual search strategy as they build the "picture". Multiple variables exist that can negatively impact the development of Situational Awareness, such as the geometry of the aircraft (converging, head-on, and tailgating), identifying whether or not two aircraft are actually in conflict and the aggregated air traffic demand of the sector (Lamoureux, 1999; Remington, et al, 2000; Athènes, Averty, Puechmorel and Delahave, 2002; Boag, Neal, Loft and Halford, 2006; Neal and Kwantes, 2009). Aggregated air traffic demand is a unique variable within the "picture", as it is expected that the number of aircraft will double by 2020 (Sheridan, 2006) at a known continuous rate of 5% per year (Hollnagel, 2007). This is critical for the entire aviation industry and the National Airspace System (NAS), because if precise visual search patterns and control mitigation strategies can be identified, understood, and characterized, they can be incorporated into the training materials of candidates.

2.6. Neuroergonomic Tools to Explore Human Performance

In addition to eye-tracking technology, there exist other approaches, based on neuroergonomics, the science that investigates the relationship between human behavior and the brain (Parasuraman and Rizzo, 2007), that can be used to understand the underlying mechanisms of the brain in a variety of tasks (Parasuraman, 2003). These techniques generally consist of quantifying changes in human brain electromagnetic or hemodynamic activity, depending on the tool used, as they are both sensitive to changes in human mental workload (Parasuraman, 2011) when exposed to complex tasks. To evaluate the former, electroencephalographic (EEG) devices can be used to directly assess measures of central nervous system function (Ayaz et al, 2012). On the other hand, hemodynamic activity can be measured through a functional near infrared (fNIR) spectroscopy device to quantify changes in blood oxygenation during neural activity (Izzetoglu et al, 2004; Bunce et al, 2006).

These tools have been used in a variety of environments, tasks, and procedures to acquire reliable and accurate measures of operator workload (Prinzel et al, 2000; Bunce et al, 2006), such as Air Traffic Control (Aricò et al, 2017; Dasari et al, 2017). In particular, the emphasis has been placed in environments where the cognitive impacts of complexity and difficulty could be mitigated through the implementation of adaptive levels of automation (Aricò et al, 2016). Within these systems, the involvement of the human operator in the task at hand could vary based on their cognitive workload measures, in order to reduce the probability of operator error (Parasuraman & Wilson, 2008). In addition, it is important to highlight that these neurophysiological measures have been used alongside eye-tracking measures, such as pupil dilation (Tsang and Vidulich, 2006), in order to validate the measured cognitive workload of participants.

Chapter 3: Characterization of Air Traffic Controllers' Visual Search Patterns and Control Strategies

Author's Note: Some of the contents in Chapter 3 were published as a conference paper, titled "Characterization of Air Traffic Controllers' Visual Search Patterns and Control Strategies", in proceedings of the 2018 International Society for Engineers and Researchers (pp. 1-5), Jun. 18-19, Seoul, S. Korea. See: Palma Fraga, R., Kang, Z., & Mandal, S. (2018).

3.1. Introduction

The demand for air traffic control specialists (ATCSs) is going to increase significantly over the coming years, as the air traffic demand is expected to double by 2020 (Sheridan, 2006) at a rate of 5% every year (Hollnager, 2007). The growth in aggregated air traffic is not only creating the need to develop enhanced air traffic control systems (Metzger & Parasuraman, 2001), but also hire and train more ATCSs to meet the demands. In addition to being an occupation in which high levels of workplace stress are common (Lesiuk, 2008; Finkelman, 1994, Houksmith & Burrough, 1980; Finkelman & Kirschner, 1980), a growth in the number of aircraft they control means that there will be a significant increase in their workload (Hulburn & Jorna, 2001; Collet, Averty & Dittmar, 2009; Warm, Parasuraman & Mathews, 2008), and even more so due to difficult conditions, such as complex weather (Langan-Fox, Sankey & Canty, 2009). An increase of this magnitude in the workload of ATCSs creates a number of problems due the increased probability of human errors (Di Nocera et al, 2006), especially in regard to humans' attention, judgment, and communication (Stager, Hameluck & Jubis, 1989). For this purpose, a better understanding of the how experienced ATCSs accomplish their tasks is required. We need to better understand how veteran ATCSs keep their situational awareness through effective and efficient visual search and aircraft control strategies in order to better prepare the next generation of ATCSs who will face increased and unprecedented demands.

Situational Awareness has been defined in multiple ways, such as the “continuous extraction of environmental information, the integration of this information with prior knowledge to form a coherent understanding of the present situation, and the use of that coherent understanding to direct perception and anticipate future events” (Gronlund, Ohrt, Dougherty, Perry & Manning, 1998) and as the “perception of elements in an environment, within a volume of space and time, and comprehension of their meaning and projection of their status in the near future” (Endsley, 1995). Within Air Traffic Control, this is referred to as the “picture” (Niessen & Eyferth, 2001) and it is used to determine the existence of conflicts between aircrafts (Nunes & Mogford, 2003; Durso et al, 1998). The skill of building the “picture” is extremely important, since tragic consequences have previously occurred from a loss of situation awareness (O’Brien & O’Hare, 2007).

One way that we can better understand how Air Traffic Controllers carry out the task of building the “picture” is through the use of eye tracking data analysis, which enables the investigation of visual search strategies (Aula, Marajanta & Raiha, 2005; Rayner, 1998) and workload (Ahlstrom & Friedman-Berg, 2006). There exist multiple variables that affect the visual search strategy when building the “picture”, such as traffic load, geometry and assessing whether or not aircraft are actually in conflict with one another (Remington et al, 2000;Athenes et al, 2002; Boag et al; 2006; Lamoureux, 1999; Neal & Kwantes, 2009). Veterab ATCSs have honed this practice through years of experience, and due to the tacit nature of this knowledge, it is difficult to explain to novices.

Eye tracking technology allows us to analyze the eye movements that the veterans apply to build the “picture”. These series of eye movements, fixations and saccades is called a “scanpath” (Kang & Landry, 2015). They can provide aid in understanding the strategies used in conflict detection (Hunter & Parush, 2009) and how visual groupings are formed in such a task (Kang & Landry, 2010). There also exist clear differences in the scanpaths used by veterans and novices (Chapman & Underwood, 1998; Kasarkis et al, 2001; Underwood, 2007), and when presented as a method of instruction to the novices, their rate of false alarms was reduced (Kang & Landry, 2014) and their accuracy increased (Dempere-Marco et al, 2002).

The purpose of this research is to investigate veteran ATCSs’ visual search and mitigation strategies, with eye tracking devices, when detecting aircraft conflicts (i.e. possible collisions) using realistic scenarios in order to generate results that can be used to effective train the novices.

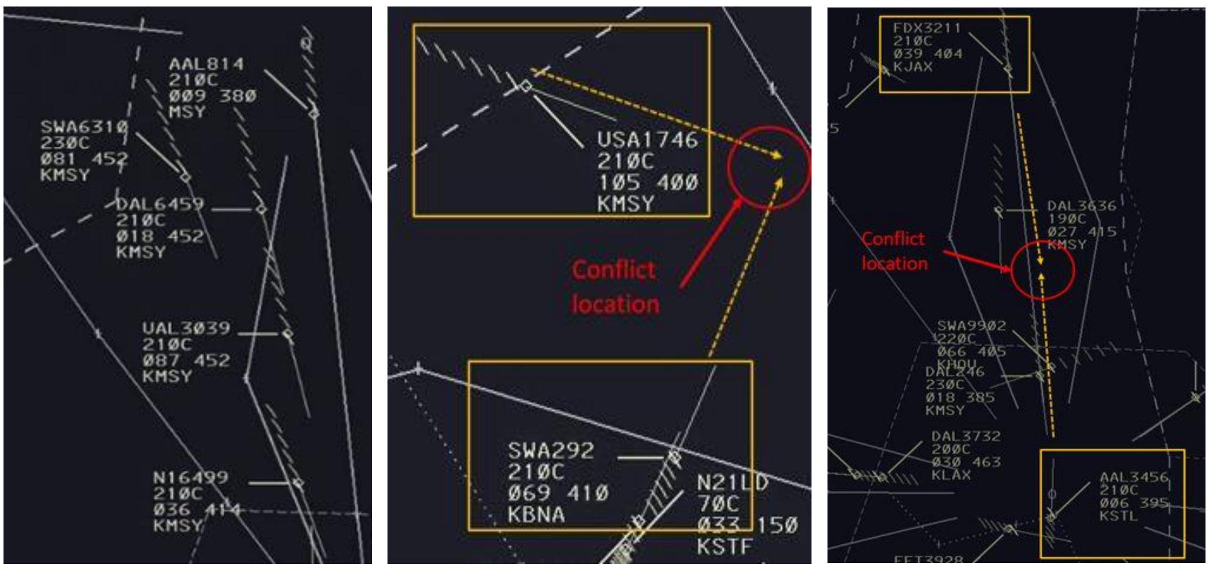
3.2. Methods

3.2.1. Data Collection

In this experiment, eleven retired veteran ATCSs participated in a within-subjects experiment with 12 representative scenarios (Table 1), in which different types of conflicts, such as converging, head-on and tailgating (Figure 4), or even none at all, were present in the simulation. Due to the high-fidelity nature of the software utilized, which was provided by the FAA, veterans were capable of giving clearances to any aircraft, not just those that were pre-determined to be in conflict. In addition, multiple requests were also given by the pseudo pilot, who managed the aircraft on the display, at different intervals during the simulation. This way, the environment the veterans have built their situational awareness all these years could be accurately represented and replicated. The scenarios were ordered randomly for each participant as to avoid any confounding effects, such as learning or fatigue.

Scenario	Type of conflict(s)
1	No conflict type I
2	No conflict type II
3	Converging conflict type I
4	Converging conflict type II
5	Head-on conflict type I
6	Head-on conflict type II
7	Converging and head-on conflicts type I
8	Converging and head-on conflicts type II
9	Tailgating conflict
10	Convergence conflict
11	Streaming type I
12	Streaming type II

Table 1: Scenarios with their respective conflict(s)



(a) **(b)** **(c)**

Figure 4: Representative examples of possible conflicts presented to participants. (a) Tailgating, (b) Converging, and (c) Head-on.

During the experiment, the eye movements of the veterans were recorded as they applied their visual search patterns and conflict mitigation strategies in all the scenarios. In addition, the voice commands instructed to the aircraft were also recorded. This was done in order to match what the participants observed, with their actions and explanations. Afterwards, an unstructured and a structured interview (Table 2) were carried-out with each participant in order to perform a self-report analysis. The former was completed with each scenario being replayed in order to comprehend the thought process of the veteran, while the latter was done individually after the experiment was completed. The goal of the final structured interview was to better understand the underlying reasoning for the participants' decision-making process, visual searching patterns and conflict mitigation strategies.

The questionnaire was designed in a structured manner to provide concise questions to the participants, as it was needed due to the tacit knowledge their answers would rely on. It was divided into four main stages, the first one, Overall Scanning Strategy, was created to better comprehend the visual search patterns applied. The second and third stages were designed to identify what information is observed, in what order, and how it aids in their decision-making process. The final stage, Control Strategies, was included in order to analyze how conflicts are mitigated and what variables do they consider prior to solving a conflict that has been detected.

Stage	#	Question name
Overall scanning strategy	1	What shape best describes your strategy overall?
	2	Why do you prefer your search method?
	3	What is your visual search strategy?
	4	How do you prioritize aircraft conflicts?
Searching for conflicts	1	What is the order in which you read information?
	2	What kind of situations cause the priorities to change?
How conflicts are detected	1	If aircraft are converging, what information do you observe and in which order?
	2	If aircraft are tailgating, what information do you observe and in which order?
Control strategies	1	In general, how do you control a conflict?
	2	Why do you prefer your chosen control strategy?

Table 2: Structured Interview Questionnaire

3.2.2. Participants

Eleven male retired ATCSs were recruited for this experiment. One FAA employee represented as a pseudo-pilot, who was responsible for executing the verbal commands and issuing requests to the participants.

3.2.3. Apparatus

A 24-inch by 24-inch monitor was used to run the I-Sim (air traffic simulation software), considered to be a high-fidelity simulation of an actual radar displayed used in Air Traffic Control, which was provided by the FAA. The eye movements were recorded using a Tobii TX300 eye tracker, in which the eye fixation threshold was 60ms and the visual angle accuracy was 0.4

degrees. A simulated radio communication channel was used for communication between he participant ATCS and the pseudo pilot with a frequency of 300 Hz.

3.2.4. Scenarios

The scenarios in this experiment were designed to challenge the veterans with a conflict detection task under different circumstances (Figure 5). None of the scenarios had weather elements present, such as wind or rain. The first two scenarios did not have any pre-determined conflicts, which meant that no control commands were necessary. The main objective of scenarios 1 and 2 was to identify the visual scanning strategies of veterans when searching for conflicts, and to recognize any differences between possible pre-emptive control measures that some ATCs could have taken, depending on how risk averse or seeking they were. This would have impacted their visual searching strategy by adding, or removing, fixations and areas of interest.

Scenarios 3 through 8 required veterans to take actions, as there existed pre-determined conflicts within the sectors that will occur in the future unless recognized and mitigated. The conflicts may have existed at the beginning of the simulation, identified over time, or through actions taken by the ATCSs. Scenarios 3 and 4 had cases of converging aircraft, followed by scenarios 5 and 6 with head-on conflicts. The next scenarios, 7 and 8, were a mix of conflicts, in which converging and head-on aircraft pairs were present. The objective of scenarios 3 through 8 was to not only better understand how the visual search strategy are carried out by veterans, but to identify conflict mitigation strategies under different environmental conditions, such as the number of aircraft or their characteristics, and distinct geometrical conflict pairs.

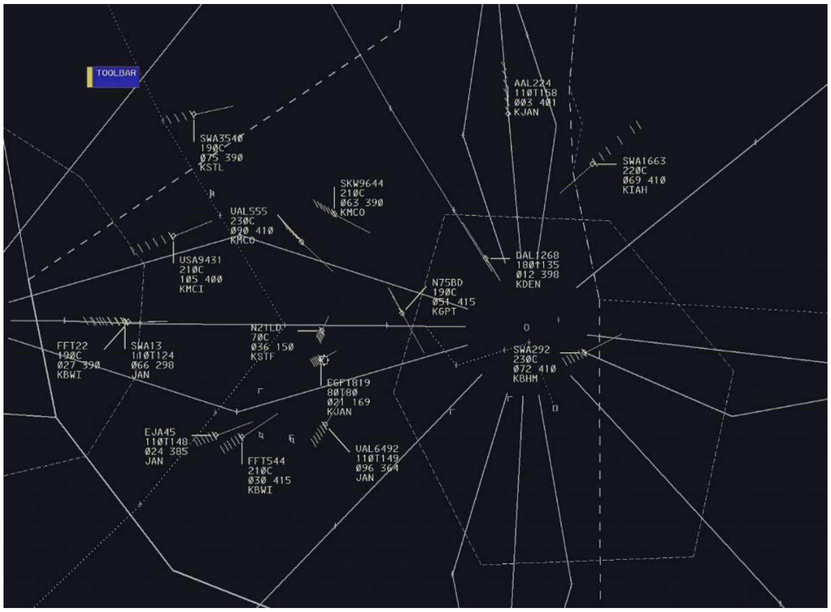
Scenarios 9 through 12 were created in order to represent special circumstances, such as a large number of tailgating or streams of aircraft flying in the same direction. In this situation, the veterans had to work with multiple large clusters of aircraft and various individual ones separate from the clusters. As such, behaviors with regards to phenomena such as tunnel vision (solely focusing on a single area) and inattentional blindness (failing to perceive a stimulus that is on plain sight) could be replicated and analyzed from the veterans' point of view.

3.2.5. Procedure

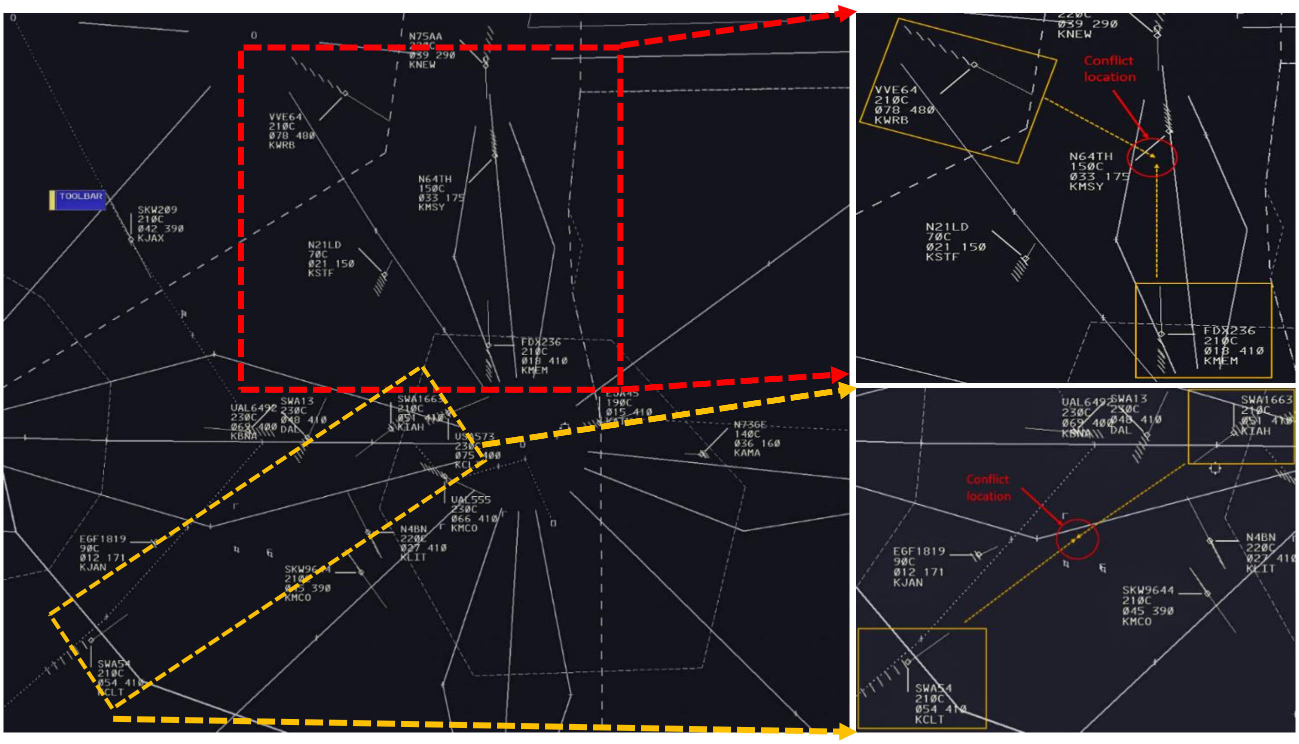
The participants were instructed to control the simulated en-route traffic scenario and resolve any potential conflict situations if identified. During the experiment, the ATCSs were not allowed to use the vector key (which allows them to see the trajectory of the aircraft for 8 mins in future). However, they could use other possible maneuvering techniques, such as changing altitude or vectoring. The ATCS provided a verbal response that they have finished once they believed all the potential conflicts have been resolved, after which the simulation was stopped. During this process, the participants eye tracking data was recorded, alongside their verbal responses, were recorded in order to be mapped together.

3.2.6. Measurements

The independent variables considered in this experiment were the 12 scenarios created. The dependent variables were the scanpaths of each participant, their responses for both the unstructured interview, and the final structured interview once the experiment was finalized. The interview was carried out alongside a replay of the recorded scenario the participant had completed. This was done in order to map their responses – the procedure they followed and their reasoning – with what they observed by matching it with their eye movement data.



(a)



(b)

Figure 5: Visual representations of scenarios. (a) Scenario 2, (b) Scenario 7 with converging and head-on conflicts highlighted in red and yellow, respectively.

3.2.7. Data analysis

The eye-tracking data collected for each participant were overlaid onto the respective scenarios in order to analyze, recognize and characterize the visual search patterns. This process is based on the participants' responses – as their eye movement data was examined after the interviews were held. This task was done by two analysts, where both had to agree on the validity and compatibility of the matching. For example, whether or not a respective scanpath used by a participant in a scenario is accurately represented by the geometrical pattern they expressed they used at that particular moment during interview (Figure 6).

The interviews were recorded and later transcribed in order to be quantitatively analyzed through a self-report analysis, in which the answers for both the structured and unstructured interviews were organized by question and arranged by frequency. This was done in order to identify similar responses from veterans to recognize trends and best practices.

3.3. Results

3.3.1. Visual search strategies of veterans

The purpose of this stage was to better comprehend the visual search patterns of veterans as they build the Situational Awareness required to identify conflicts, and if required, take necessary actions based on the environmental information they perceived. One of the most important aspects of a visual search pattern is the geometrical shape that it can be best represented by (Table 3). The table shows that most veterans prefer a continuous, rather than disjointed, search pattern in a circular shape (spiral or circular) over other geometrical patterns (linear or zigzag). There are two special cases to be considered, a quadrant search scan, where the screen is divided into two or more sections, or mixed, which can be a combination of multiple geometrical search patterns. Due to the tacit nature of the knowledge, some veterans were not capable of defining their search pattern in any known geometrical shape and explained that it was simply random. Representative examples of visual search patterns that fit the aforementioned geometrical shapes can be seen in Figure 7.

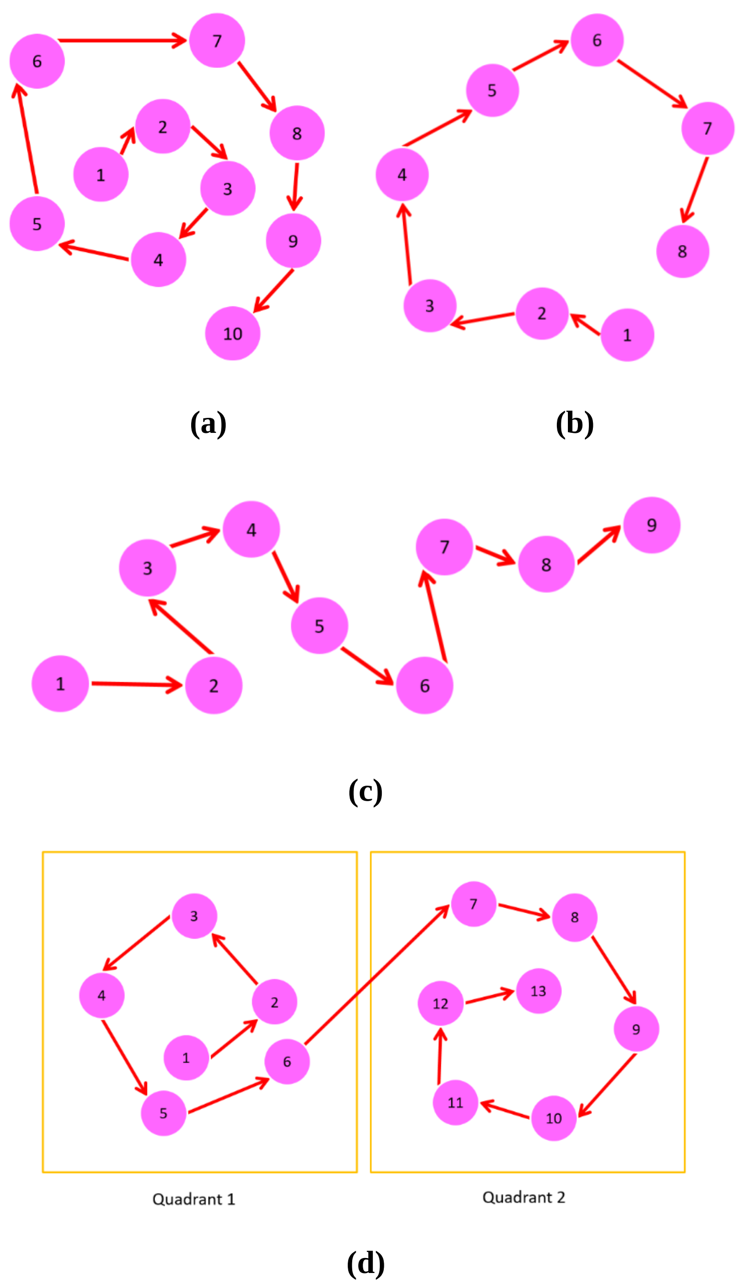


Figure 6: Representative examples of various kinds of scanpaths in a visual search task. (a) Spiral, (b) Circular, (c) Linear, and (d) Quadrant.

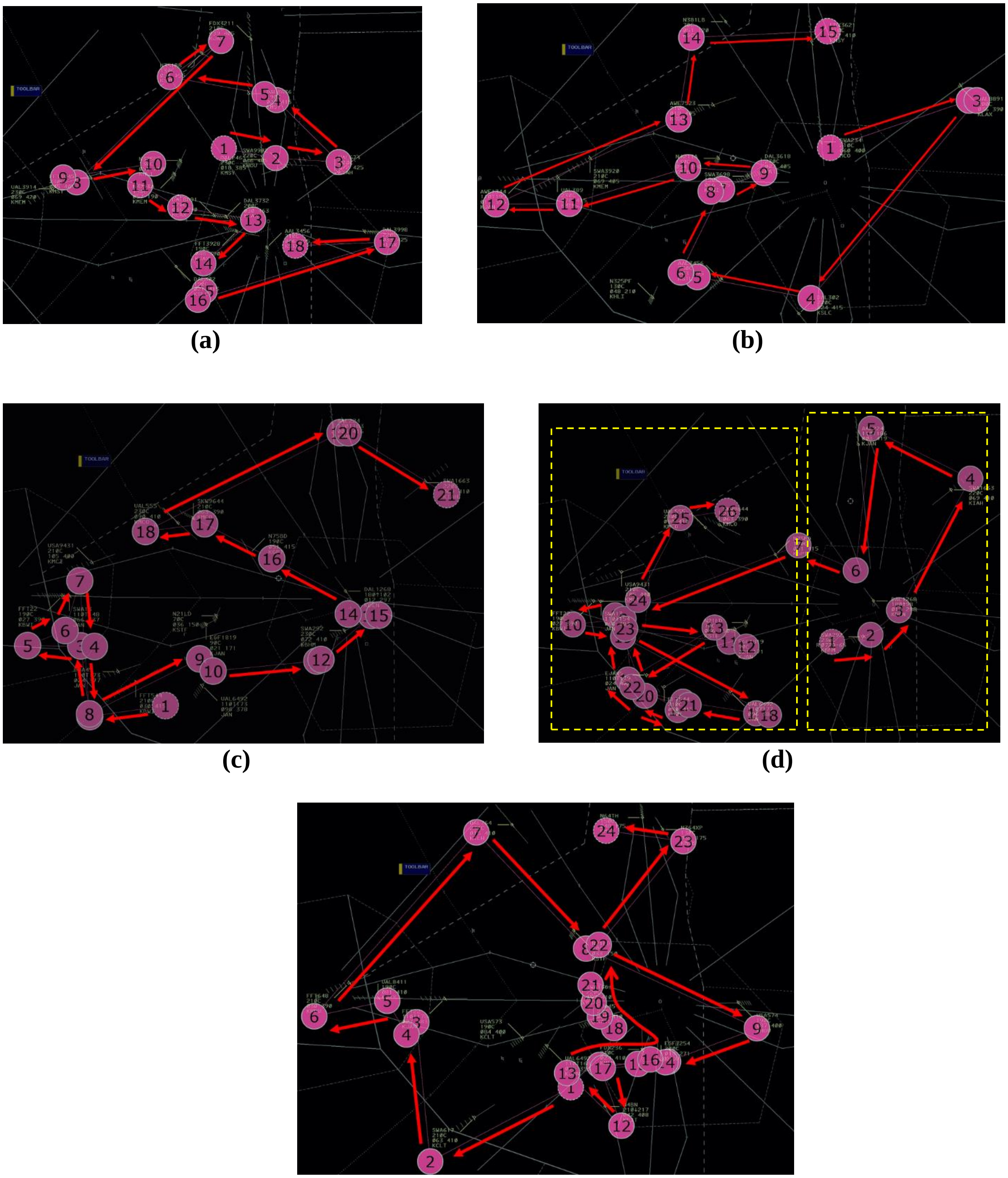


Figure 7: Representative examples of visual search patterns. (a) Spiral, (b) Circular, (c) Linear, (d) Quadrant, represented by yellow boxes, (e) Mixed visual search pattern.

Geometrical pattern	Freq.	Starting location	Freq.	Participant
Spiral	4	High density-based	3	P2, P10, P11
		Center of sector	1	P9
Circular	2	High density-based	1	P1
		Preference based on training and experience	1	P3
Linear (e.g. zigzag)	1	Low density-based	1	P5
Quadrants	1	Areas of conflict	1	P4
Mixed (circular + linear)	1	High density-based	1	P6
Random	2	Incoming sector traffic	1	P8
		High density-based	1	P7

Table 3: Visual search strategies: search pattern (i.e. geometrical shape) and starting location.

The starting location of the visual search patterns are not static and are dependent upon the environmental information of each sector, such as the number of aircraft and their location. Nonetheless, most veterans did define their preferred starting area as those in which multiple aircraft are present together (high-density based), rather than the opposite (low-density based). The motives behind each type of visual search pattern were varied, and thus, no general trend was observed from the response of the participants (Table 4) with their quotes on Table 5. In this case, a higher number of veterans explained that the major reason for their scanpath was made based on training and experience over multiple years.

Reason	Freq.	Participant
Continuous scan rather than disjointed	2	P1, P11
Focus on time-sensitive areas	2	P9, P5
Faster scan	1	P2
Focus on high-density areas	1	P10
Aid in recognizing wrong altitudes for direction of flight	1	P7
Preference based on training and experience	4	P3, P4, P8, P6

Table 4: Reason behind preferred visual search pattern.

Reason	Quotes from ATCs on reasons for visual search strategy
Continuous scan rather than disjointed	**(P1: circular)** I guess it has more continuity rather than jumping all over? You have more of a flow to what you are looking at. **(P11: spiral)** Well, I think if you stay with a sweep around the sector you won't miss or are at least less likely to miss anybody.
Focus on time-sensitive Areas	**(P9: spiral)** If the group is in the center of the sector it is more time critical to make control judgements **(P5: linear)** The whole thing is a big puzzle with smaller puzzles. If you eliminate the smaller conflicts first, it gives you more time to look at what else is going on.
Faster Scan	**(P2: spiral)** The main reason is because that is the fastest way to get to everybody
Focus on high-density Areas	**(P10: spiral)** I usually try to focus on the most complex area.
Aid in recognizing wrong altitudes for direction of flight	**(P7: random)** If they are at the wrong altitude for the direction of their flight, I need to fix that
Preference based on training and experience	**(P3: circular)** It's just me and there is no particular reason. It's just how I do it. **(P4: quadrants)** I just found that it works better for me that way. **(P6: mixed)** I guess I was trained that way and I think it's a more thorough way to do it. **(P8: random)** That really comes with experience. To tell somebody that there is a certain order is something that I'm not sure would be very helpful.

Table 5: ATC quotations on reason behind preferred visual search pattern.

The effect of conflict type was also studied, as it could be possible that certain conflicts may be more important (Table 6). Here, the veterans identified that a converging conflict has priority over any other type of conflict, but after that, the hierarchy is not well defined. Similarly, the motives behind the hierarchy of conflicts were distinct, yet the general trend moves towards safety, often due to time being a critical factor, rather than providing service to an aircraft.

Order of conflict observed			Freq.	Reason	Freq.	Participant
Converging	Head-on	Tailgating	4	Immediate Safety Issue	3	P1, P2, P3
				Loss of Separation	1	P4
	Tailgating	Parallel/Diverging	2	Time Sensitive	2	P7, P9
	Diverging	Tailgating/Parallel	1	Symbol Recognition	1	P5
	Parallel	Tailgating/Diverging/Head-on	1	Immediate Safety Issue	1	P6
Head-on	Converging	Parallel	1	Immediate Safety Issue	1	P10
Order based on training and experience			2	Immediate Safety Issue	1	P8
				Time Sensitive	1	P11

Table 6: Priority order of aircraft conflict types and associated reason.

3.3.2. Conflict detection process

After understanding how the visual search pattern is carried out, the next step is to understand what happens after a fixation occurs on an aircraft. The information that represents the status of an aircraft is found within the data tag (Figure 8) and it provides all the variables needed to know if a conflict will actually occur, such as altitude, direction, velocity and final destination. Due to the three-dimensional nature of ATC, not all variables have to match for there to be a conflict. For example, two aircraft could be going at the same direction and speed, but if their altitudes do not match, no conflict is present.

Figure 8: An airplane and its data tag.

Due to the necessity to understand the situation as fast and accurately as possible, in order to build Situational Awareness, veterans may read the data tag differently (Table 7). Here, there is a very strong trend in regard to a hierarchy of information observed, in which altitude is first, followed by direction and lastly the speed of an aircraft. Only a single participant had a different hierarchy, but even then, altitude was still considered the highest priority, as it is a key variable in identifying the existence of conflicts. This means that, whenever a fixation occurred in an aircraft, the participant was reading the data tag to build their awareness of the situation applying such a hierarchy.

Order of information observed			**Freq.**	**Participant**
Altitude	Direction	Speed	10	P1, P6, P7 P10, P11, P8, P5, P9, P4, P3
Altitude	Speed	Direction	1	P2

Table 7: Order of information observed.

This process becomes more complicated when scenario characteristics are present in the sector, which may exacerbate conflicts, which include elements such as inclement weather, not following letters of agreement or even pilot requests, as they may change the order in which the information is observed (Table 8). For example, an airplane heading parallel to the direction of the wind, would naturally speed up. Therefore, a generally slower aircraft may suddenly go at a speed higher than usual, increasing the likelihood of a conflict. As seen in table 8, the majority of veterans agree that situation characteristics do not affect the order of information observed. Yet, their chosen conflict control action, taken based on such characteristics, may actually affect the order in which they observe the information.

Order of information	**Freq.**	**Reason**	**Freq.**	**Participant**
Situation characteristics does not affect order of information observed	7	Control action affects order of information	5	P2, P6, P7, P8, P9
		Used to predict aircraft behavior	1	P4
		Preference based on training and experience	1	P3
Situation characteristics affects order of information observed	4	Sector characteristics	3	P5, P10, P11
		Conflict urgency	1	P1

Table 8: Effect of scenario characteristics in order the information is read in.

Similarly, to visual searching patterns, the type of conflict, such as a converging, head-on or tailgating pair of aircraft, could play a role on how the data tag is read. This could mean that certain information provided by the tag may be discarded, or simply omitted, when attempting to fix a conflict. In the case of tailgating aircraft (Table 9), the hierarchy of information presents the same trends shown in table 7, with altitude being the top priority. Nonetheless, there is a clear difference with the second most important information observed, as the speed of aircraft becomes more important than its direction. This is mainly due to the fact that tailgating aircraft are, by definition, behind each other.

Order of information observed			Freq.	Participant
Altitude	Speed	*	5	P3, P5, P8, P10, P11
Altitude	Speed	Direction	1	P4
Altitude	Speed	Destination	1	P9
Altitude	Direction	Speed	1	P1
Speed	Altitude	Turns	1	P2
Altitude	Direction	*	1	P7
Speed	*	*	1	P6

Table 9: Order of information observed for tailgating aircraft.
* denotes that information was not mentioned by the participant(s).

The case of the hierarchy of information observed with pairs of converging aircraft (Table 10) was similar to the cases explained previously. As such, most veterans agree that altitude remains the highest priority when reading the data tag once a fixation occurs, followed by the direction of the aircraft and finally, the speed.

Order of information observed			Freq.	Participant
Altitude	Direction	Speed	6	P4, P6, P7, P8, P10, P11
Altitude	Relative Position	Speed	1	P5
Altitude	Speed	*	1	P1
Altitude	Destination	*	1	P9
Altitude	*	*	1	P2
Preference based on training and experience			1	P3

Table 10: Order of information observed for converging aircraft.
* denotes that information was not mentioned by the participant(s).

3.2.3. Conflict mitigation strategies

The goal of this stage was to further analyze the decision-making process once the visual search has been completed, the "picture" has been built, and conflicts were identified. There are multiple ways to solve a conflict, even when ignoring the impact of any sector characteristics, due to the fact that they are working in a three-dimensional environment. Altitudes can be changed to separate aircraft vertically, an aircraft may be slowed down or accelerated, and an aircraft may be vectored off course to avoid another (Table 11). Similarly, to the lack of general trends in a hierarchy of conflicts observed in table 6, there is not a particular consensus regarding the best way to solve a conflict. A high number of veterans set the hierarchy to start with changing altitudes as a preferred method to mitigate conflicts, while another group mention that modifying vectors is the best way. Nonetheless, it is important to highlight two trends: one, no veteran mentioned that changing speed was their preferred way to stop a conflict. Second, some participants explained that they cannot provide a hierarchy of conflict control strategies, due to the fact that whatever strategy they choose at that particular moment, heavily depends on the impact sector characteristics may have on the possible solutions.

Order of conflict control strategies			**Freq.**	**Participant**
Altitude change	Vector change	Speed change	4	P3, P5, P8, P10
Vector change	Altitude change	*	4	P4, P6, P7 P9
Altitude change	Vector change	*	1	1
Sector Characteristics			2	P2, P11

Table 11: Preferred conflict mitigation strategies.
* denotes that information was not mentioned by the participant(s).

These preferred conflict mitigation strategies depend on multiple factors (Table 12). Here we can see that there are three major emerging trends. One is that the hierarchy is dependent upon the nature of the conflict itself such as the geometrical shape of the conflict or the angle of approach between aircraft. The second is that it could also be dependent on sector characteristics such as inclement weather or even the number of aircraft present in a sector. The last trend is that changing vectors may provide a shortcut and entirely avoid a conflict. Detailed quotations from the veterans can be seen in Table 13.

Reason	**Freq.**	**Participant**
Conflict characteristics	3	P3, P6, P11
Customer service (short-cut)	3	P4, P5, P7
Sector characteristics	3	P1, P2, P9
Practicality of altitude change	1	P10
Preference base on training and experience	1	P8

Table 12: Motives for preferred conflict mitigation strategies.

Reason	Quotes from ATCs regarding reasons for conflict control strategies
Conflict characteristics	**(P3)** It's hard to give a specific example because you are reacting in many instances **(P6)** Normally the speed comes into play as you are looking for the traffic and you decide: "am I going to turn the slower aircraft behind the faster aircraft" **(P11)** I can't really say that. Sometimes routes are so much easier, you can just tweak a route or turn them behind him.
Customer service (short-cut)	**(P4)** If I could give an aircraft a more direct route, he is happy. I am not changing his altitude, so everyone is happy in that situation. **(P5)** Airplanes like to stay on course and anytime you turn them, it takes longer and costs more money as burn more gas. **(P7)** If it's shorter for the aircraft, that's more efficient, as it saves them time and fuel, which equates to efficiency. It's more efficient for me.
Sector characteristics	**(P1)** Altitude or vectoring would be best. Where I have worked, these guys are coming cross country. If you start slowing down, they are going to get irritated **(P2)** If you are working on a low altitude near an airport, using speeds and having them slow down, it works really well but not so much in different airports **(P9)** Sometimes you have to do just choose one due to the scenario. There may be a scenario where you have to vector because there are no altitudes
Practicality of altitude change	**(P10)** I think altitude is the easiest because the pilot just has to go up or down. In vectors, the pilots may start wondering, why are they on this vector?
Preference based on training and experience	**(P8)** To me, it's just the easiest way to work airplanes.

Table 13: ATC quotations on motives for preferred conflict mitigation strategies.

3.4. General trends

The present experiment was able to identify a number of trends that veterans tend to apply every time they are carrying out the task of conflict control. As mentioned previously, building their Situational Awareness is a process that goes beyond visual search patterns, due to the fact that projecting how the impact of their actions would affect the current state of the sector is a key component of their task.

From interviews, and the analysis of the recorded eye-tracking data, the majority of veterans prefer to apply a circular or spiral visual search pattern over other possible geometrical shapes. This allows them to fixate on every aircraft faster in a single swoop, rather than in a disjointed manner all over the sector. Nonetheless, this process is also very dependent upon the starting location the veteran chooses, as it is not static due to the fact the environmental information between sectors is dynamic, such as weather or the geometrical shape of possible conflicts.

Generally, the hierarchy of conflicts defined by veterans tends to begin with a converging geometry as the highest priority, followed by head-on and lastly tailgating aircraft. This was mainly due safety, as even though the veterans would prefer to provide increased customer service to the aircraft and airlines, the underlying motive behind their actions in this experiment seems to always be in regard to safety.

The process of recognizing conflicts is dependent upon the information provided by each airplane's data tag, and its effect on the environmental information. Once a fixation occurs, veterans strongly tend to observe the altitude of an aircraft, followed by its vector and then the speed. Based on this experiment, this hierarchy of observed information was agreed upon by every participant, except one – nonetheless, it shows that the most element in the data tag is always altitude. This is due to the three-dimensional nature of the environment the ATCSs work with. Although the environmental information of a sector may affect the starting location of a visual search pattern, the hierarchy of information observed is not affected by the sector characteristics. Rather, the priority is dependent upon the control action the veteran has determined would be the most valuable one, based on any possible conflicts they may have identified through the Situational Awareness they have built of the sector. Changing the altitude of an aircraft is significantly more complicated than, for example, changing speeds. An altitude change means that an aircraft would have to go through a number of transitory altitudes, in which multiple aircraft may already be in. Without proper knowledge of the altitudes of all the aircraft in the sector, an altitude change is more dangerous.

Regarding the preferred action to control conflicts, there is no consensus in the hierarchy of conflict control. Most veterans set changing altitude as their highest priority, followed by the changing. Conflict characteristics, such as the angle of convergence between aircraft and the geometry of the conflict, play a role alongside sector characteristics such as inclement weather, or the presence of multiple aircraft in various clusters. This environmental information can impact the action taken by a veteran. For example, in the latter, vectoring an aircraft into the direction of the wind would slow it down, allowing another faster airplane to pass the slower aircraft without the need to change altitude. Meanwhile, in the former, a small angle of convergence can make a vector change take significantly longer than just descending, or ascending, an aircraft.

In summary, veterans prefer to have a continuous visual search pattern, which may be best represented by a spiral or circular geometrical shape. When reading the data tag of an aircraft, they generally observe the altitude, followed by the direction and lastly speed. This is done in order to build the Situational Awareness of the sector they are working with, which allows them to project into the future and recognize conflicts. Once a possible conflict has been identified, the order in which the hierarchy of information observed is read in may change based upon what the veteran thinks would be the easiest solution to the conflict. Regarding conflict mitigation strategies, the veterans do not have a preference but tend to choose either changing altitudes, or vectors, over changing speed.

3.5. Limitations and future research

This research project had a number of limitations and avenues for future research projects. For example, the participants were all retired veterans of a previous generation, whose experience with technology had to be built over training and experience. The future air traffic controllers, consisting of millenials and generation z, who will have to deal with the increased air traffic demand, have grown their entire life surrounded by technology, and thus, possess an inherent ability and understanding that previous generations may not have had. Thus, it is possible that an experiment with the newer generation of ATCSs may provide an insightful view of how they build their own situational awareness with this deeper understanding of technology.

This project may have benefitted from a higher number of participants, in order to create a deeper self-report analysis. Due to the high variability of responses, we can assume that there exist multiple answers that we, as researchers, may not be able to come up with due to a lack of experience as air traffic controllers. Some of these may include different types of visual search patterns or different motives to a preferred conflict mitigation strategy.

A major future project would be to identify significant differences between the distinct visual search patterns characterized in this paper, improve them based on various criteria such as time and false alarms, teach them to novices in Air Traffic Control, and then compare their performance with that of veterans.

3.6. Conclusion

Recorded eye-tracking data were utilized, alongside the verbal explanations for the different visual search patterns and conflict mitigation strategies. General trends utilized by veteran ATCSs, when carrying out the task of conflict detection were identified. The methods they apply, as they build their situational awareness, identify conflicts, and stop them from occurring were identified. The hierarchies of information observed and conflict mitigation strategies, alongside preferred visual search patterns amongst veterans, including the underlying reasons as to why they are used, were discovered by the research team.

3.7. Acknowledgements

This research is supported by the Federal Aviation Administration Center of Excellence (Project No. A17-0160). The FAA has sponsored this project through the Center of Excellence for Technical Training and Human Performance. However, the agency neither endorses nor rejects the findings of this research. This information is provided in the interest of invoking technical community comment on the results and conclusions of the research. We worked closely with the FAA Civil Aerospace Medical Institute at Mike Monroney Aeronautical Center (AAM-520).

Chapter 4: Multimodal Analysis using Neuroimaging and Eye Movements to Assess Cognitive Workload

Author's Note: The majority of the contents in Chapter 4 were published as a conference paper, titled "Multimodal analysis using neuroimaging and eye movements to assess cognitive workload", in Proceedings of the 22nd International Conference on Human-Computer Interaction, Jul. 19-24, Copenhagen, Denmark. See: Palma Fraga, R., Reddy, Y. P., Kang, Z., Izzetoglu, K. (2020). In addition, the data collection and analysis of fNIRS data was done by Drexel Universities' Pratusha Reddy & Dr. Kurtulus Izzetoglu.

4.1. Introduction

4.1.1. Background

Air Traffic Control (ATC) specialists are crucial individuals within the system of civil and military aviation in which they maintain safety, while at the same time, expediting the flow of aircraft. In addition to the already established high levels of workplace-related stress (Lesiuk, 2008; Finkelman, 1994), the environment where they carry out their tasks is expected to become significantly more complex in the near future. This is partially due to numerous long-term projections, estimating that by 2040, future air traffic volume may reach 60 million aircraft annually (FAA, 2018) and potentially serve around 10 billion airline passengers (ICAO, 2017). In response to the predicted increase in traffic volume, and the continuous efforts by the industry to improve efficiency and safety in air transportation, new technologies and procedures are likely to be adopted in the ATC environment as well. Change to the roles and responsibilities of controllers poses a potential to increase complexity as well. This implies that ATC will impose more mental demands on human operators (Vogt et al, 2006; Ahlstrom & Friedman-Berg, 2006) and subject them to a higher risk of human error (Di Nocera et al, 2006).

Due to the potential increase in complexity and workload, there has been a need to conduct a closer, inspection of the interaction between controllers and computers using methods such as NASA-Task Load Index (NASA-TLX) (Hart & Staveland, 1988), the Instantaneous Self-Assessment techniques (Kirwan et al, 1997) and the Air Traffic Workload Input Technique (Stein, 1985). The aforementioned methods obtain workload ratings post-task, or during interruptions to the task itself. On the other hand, there exists alternative neural and physiological methods that provide direct and real-time evaluation of workload during the task. Additionally, the monitoring of physiological variables has the potential to not only enable a real-time evaluation of an operator's mental state, but to also aid in the informed design of safe and effective adaptive automation systems and future training methods.

Over the last decade, functional Near Infrared Spectroscopy (fNIRS) and eye-tracking are among many devices developed to measure various aspects of cognitive functioning (Bhavsar, 2017). fNIRS is an emerging, noninvasive, affordable, and portable neuroimaging modality that exploits the optical properties of biological tissues and hemoglobin chromophores in assessing changes in brain activity. It does so by deploying wavelengths between 700nm to 900nm, where the chromophores of oxygenated and deoxygenated hemoglobin (HbO and HbR, respectively) are found to be the main absorbers (Jobsis, 1977; Cope, 1988). The changes in HbO and HbR are directly associated with changes in brain activity (Villringer et al., 1997; Izzetoglu et al., 2004), therefore making it an attractive modality to study cognitive performance within the environment of ATC. Alternatively, eye-tracking has established itself as another key methodology that can be used to measure the efficient and timely acquisition of visual information (Bruder & Hasse, 2019). The metrics derived from eye movements, such as fixation durations & counts, saccadic lengths, gaze event durations, and pupil dilation have been used to quantify the effort, information processing capabilities, attention, and decision-making ability of participants in numerous complex environments (Otero et al, 2011; Causse et al, 2019; Rudi et al, 2019).

4.1.2. Multidimensionality of cognitive workload

It is a widely accepted fact that one of the major human performance factors in ATC is cognitive workload (Edwards, 2013, Ball et al, 2007), which is affected by the airspace design and traffic requirements (Durso and Manning, 2008). The number of aircraft under control has been studied numerous times and shown to be associated with mental workload measures (Ahlstrom and Friedman-Berg, 2006). However, it is not the only environmental variable capable of increasing cognitive workload and may not necessarily represent the only quantifiable measure of complexity for the ATC task (Kirwan et al, 2001; Athenes et al. 2002). There exist additional operator-dependent factors, such as strategy, that play an important role in quantifying cognitive workload. How a controller chooses to prioritize tasks, or compensatory strategies used to respond to workload fluctuation, all influence an ATC specialist’s cognitive workload (Koros et al, 2003).

This multidimensional aspect of cognitive workload can be defined by the interaction between three categories: (1) drivers (activity-based estimators), which evaluate the prescription and quantity of information to process; (2) mediators (operator-based estimators) where changes of strategy are quantified; (3) indicators (activity or operator-based estimators) which include performance metrics, the participant’s subjective experience with the task, and physiological measures (Kostenko et al, 2016). In the ATC setting, the drivers are represented by the task difficulty (e.g. number of aircraft or weather conditions), mediators portray how controllers strategize with the complexity of the task (e.g. number of commands), and indicators can be characterized by quantifiable performance measures (e.g. neural and psychophysiological measures).

fNIRS and eye-tracking have been widely used to measure controller workload (e.g. Truschzinski et al., 2018; Marchitto et al, 2016; Harrison et al, 2014; Tsai et al, 2012). Results from the fNIRS studies indicated that oxygenation measures were sensitive to task load changes and correlated highly with behavioral performance measures (Izzetoglu et al., 2004 & 2019; Ayaz et al, 2012; Reddy et al., 2018). Similarly, eye-tracking metrics such as pupil dilation, fixation number & duration, and saccades have shown to be closely related with cognitive workload (Van der Wel and Van Steenbergen, 2018; Eckstein et al, 2017).

4.1.3. Multimodal approach

The utility & reliability of a multimodal approach is imperative to achieve well-rounded and concise results, as one singular approach may fail to capture all the critical elements of cognitive workload. For example, researchers have shown that application of a multimodal methodology has helped in understanding how levels of automation impact cognitive workload (Evans & Fendley, 2017), assess multiple factors that influence situational awareness (Friedrich et al, 2018), and identify differences between veterans and novices (İşbilir et al, 2019). We expand on this topic by investigating the correlations that exist between the variables captured from a multimodal approach, consisting of fNIRS and eye-tracking, alongside behavioral measures. These have been defined by considering cognitive workload as a multidimensional construct and quantified in terms of the number of aircraft (task difficulty) & number of commands (task complexity).

4.2. Methodology

4.2.1. Participants

Three retired veteran ATC specialists, each with over 20 years of experience as tower controllers, participated in the experiment by switching and performing three different roles: (1) observer; (2) operating the Local control position; (3) operating the Ground control position. The participants were recruited with the help of the Federal Aviation Administration's Civil Aerospace Medical Institute (CAMI) in Oklahoma City, Oklahoma.

4.2.2. Simulator and sensors used

The simulated air and ground traffic experienced at an airport was displayed via eight 55" HD (1080p) screens. Three separate monitors were used for the Airport Surface Detection Equipment (ASDE), Bright Radar Indicator Tower Equipment (BRITE) and Status Information Area (Figure 9). The simulator software – MaxSim, developed by Adacel Systems Inc – was used to generate the scenario. Participants instructed the aircraft through verbal commands called clearances, which were processed by voice recognition software embedded in the simulator. The recognition software responded accordingly with a simulated voice reply and aircraft behavior consistent with the command. The simulator had a pseudo-pilot, whose role was to monitor the interpretations of the voice commands given by controllers and take the necessary steps to correct any misrecognized or unrecognized utterances.

For data collection purposes, a pair of Tobii Pro Glasses II (100 Hz) was used to capture the eye movements, while fNIRS 1000 (2 Hz, 10 detectors, 4 LEDs that operated at 730nm, 850nm, and ambient wavelengths) created by fNIRS Systems, LLC was used to measure raw light intensity data from the prefrontal cortex (PFC).

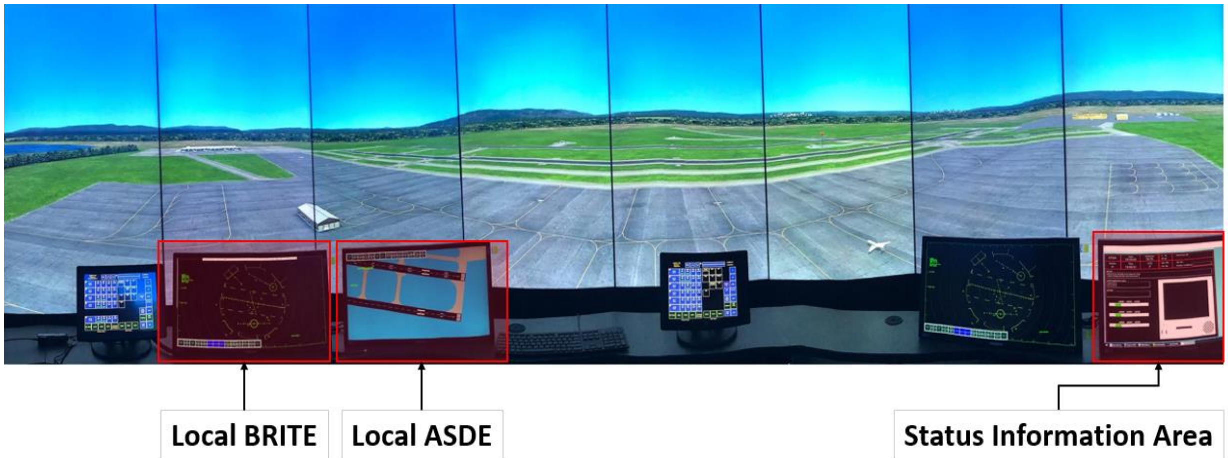


Figure 9: The simulation environment was presented to the participants during the experiment. The additional displays that can be used by the local controller are highlighted in red.

4.2.3. Scenario & Task

All participants underwent a single 32-minute scenario developed by the Federal Aviation Administration (FAA). The scenario consisted of 33 aircraft, with 19 arrivals and 14 departures. Within this scenario, participants carried out duties (issue landing and takeoff instructions to pilots, monitor and direct the movement of aircraft on the ground and in the air) of a Local controller. No pre-determined conflicts were incorporated into the scenario and the weather conditions (clear blue skies) remained consistent throughout the scenario. This procedure was chosen in order to elicit normative behavior of controlling traffic in a tower control environment from the participants.

4.2.4. Data Analysis

The eye-tracking data was subjected to the automated I-VT filter (Komogortsev & Karpov, 2013) within Tobii to extract the fixations and saccades. Here, fixations were defined at a threshold of at least 60ms, while everything else was considered a saccade. The data obtained from these eye-movements included: left and right pupil dilation, number of fixations, gaze event duration, and the sequential saccadic distance. To factor individual differences in natural pupil dilation, those estimates were normalized by calculating a grand pupil dilation mean across all participants, which was then used to divide each pupil dilation data point (Engelhardt et al, 2010).

On the other hand, the fNIRS signal was first corrected for light leaks by simply subtracting ambient signal from the other two signals (730nm, and 850nm). The signals were then low pass filtered to remove instrument noise, physiological noise, and motion artifacts (M. Izzetoglu et al., 2005; Reddy et al., 2018). Then, modified Beer- Lambert Law was used to calculate the oxygenated (HbO) and deoxygenated-hemoglobin (HbR) changes at each channel (Villringer & Chance, 1997). Using HbR and HbO measures, oxygenation (Oxygenation = HbO - HbR) and total hemoglobin (Hbtot = HbO + HbR) were derived. Lastly, samples that were three standard deviations above the expected values were classified as outliers and removed (0.34%) from further analysis.

Task difficulty and complexity factors were quantified using number of aircraft and number of clearances measures. The former was identified by summing the number of aircraft managed by the Local controller as seen on the ASDE and BRITE radars. This approach is different than simply considering the total number of aircraft present in the scenario at any particular instant, as not all of them fall under the direct command of the Local controller (e.g. aircraft taxiing from their respective gate to the runways are managed by the Ground controller, but they also appear on the radar displays). The latter was determined by counting clearances given per time bin associated with each aircraft condition. Additional information regarding the type of clearance given, and the aircraft that the clearance was given to were also noted.

The times at which the task difficulty changed were used to bin eye-tracking and fNIRS measures. The resulting average measures of the extracted features were assessed for correlation between each other. One-way ANOVA or non-parametric Kruskal–Wallis test was used to determine the difference between neural and psychophysiological metrics across the task load measures. Post-hoc Tukey or Mann Whitney tests were used to determine where the differences come from. An alpha level of 0.05 was used as the significance criteria for tests conducted.

Data from the first three minutes were removed from the analysis, because participants used this time to become familiar with the task environment. In addition, the data segment associated with 12-aircraft was removed as only one participant controlled 12 aircraft simultaneously.

4.3. Results

4.3.1. Correlation between task, behavioral and physiological measures

A correlation analysis was conducted to investigate the relationship between task factors, neural and psychophysiological metrics and the results of this analysis are seen in (Figure 10). Average oxygenation measures from both left (Channels: 3, 4, 5, 6) and right (Channels: 11, 12, 13, 14) prefrontal cortex (PFC) negatively correlated with normalized left & right pupil dilation ($r = -0.47$ & -0.48, with both having a p-val < 0.01), and gaze duration ($r = -0.38$, p-val < 0.01), but not with other eye-tracking measures. Both left and right hemisphere cortical oxygenation increased as the number of aircraft increased. However, no trend was observed between the oxygenation changes and the number of clearances. Alternatively, the number of fixations increased with the number of clearances, but not with the number of aircraft. Besides the fixation number, no other eye-tracking metrics revealed any correlation with the task factors.

4.3.2. Interaction between task and behavioral measures

The correlation between the number of aircraft and number of clearances was low ($r = 0.36$), when it was expected to be high. To understand the relationship between these two behavioral metrics better, we investigated the temporal profiles of each variable. The results of this analysis are shown in Figure 11, where the number of aircraft managed by the controller changed quadratically with time, while the number of clearances changed cubically. By overlaying the two plots or plotting number of aircraft vs number of clearances, three phases of task demand can be identified. The first phase (3 to 9 minutes) shows an increase in both number of clearances and number of aircraft. The second phase (9 to 21 minutes) is defined by an increase in the number of aircraft, but no change in number of clearances. Finally, the last phase (21 to 33 minutes) shows an increase in the number of clearances, while the number of aircraft does not change. These three phases were used to define three levels of task demand - Low, Medium, and High. A one-way ANOVA was used to assess whether the number of clearances differed across groups or not. Results indicated significant difference across task demand groups ($F_{2,64} = 3.329$, $p = 0.04$), with significant difference between low and high group ($p = 0.01$).

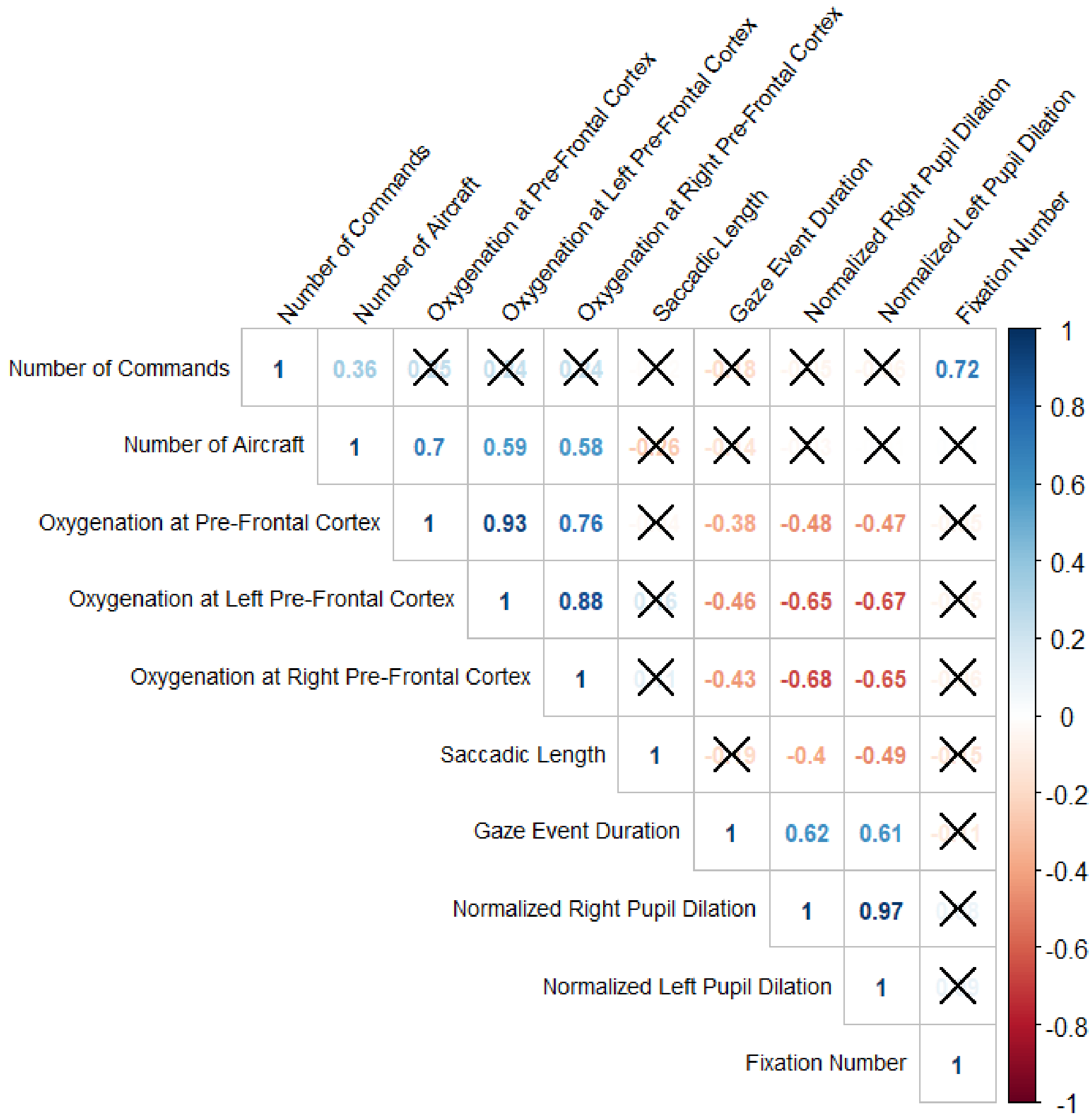


Figure 10: Correlation plot (R package: corrplot, Wei & Simko, 2017) between the behavioral measures, fNIRS and eye-tracking metrics. The X's represent non-significant correlations (p-val > 0.01).

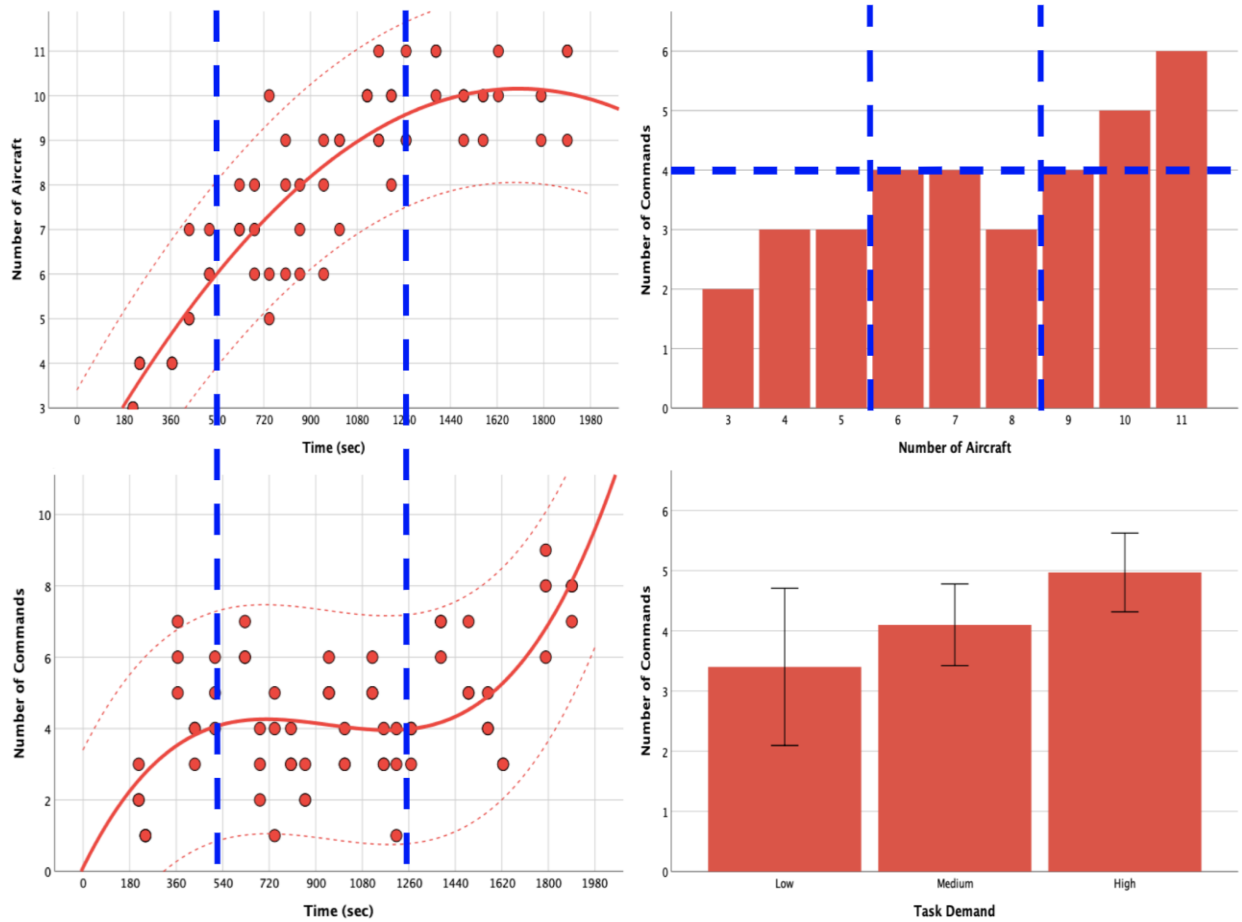


Figure 11: Change in number of aircraft across time (left, top) and change in number of clearances across time (left, bottom). Median changes in the number of clearances relative to the number of aircraft (left, top). Average number of clearances per low, medium, and high load conditions. Blue lines represent the partitions defined by multidimensional workload model. Error bars are +/- 2 standard error of the mean (SEM).

4.3.3. Effects of changes in drivers and mediators on neural and psychophysiological measures

After checking for normality and parametric assumptions, oxygenation measures from the prefrontal cortex (PFC) were submitted to a one-way ANOVA to assess the effect of the number of aircraft. Significant effects of the aircraft number on oxygenation measures were reported ($F_{8,56} = 6.95$, $p < 0.01$). Post hoc comparisons using Tukey test indicated no differences between aircraft conditions ranging from 3 to 8 and 9 to 11. However, conditions 9 to 11 revealed significantly higher oxygenation measures in comparison to 3 and 4, while condition 11 was higher than 6 and

7. This result suggested three levels of task load, which were defined as low (3 to 5 aircrafts), medium (6 to 8 aircrafts), and high (9 to 11 aircrafts). After establishing these three groups, we conducted a one-way ANOVA again to assess the effect of multidimensional task load on oxygenation from left and right PFC. The results as shown in Figure 12 indicate significant effect of workload on the left ($F_{2,62} = 10.61$, $p = 0.00$) and on the right ($F_{2,62} = 10.29$, $p = 0.00$) PFC, with differences across all load levels per region of interest.

A non-parametric Kruskal–Wallis test was used to examine the effect of workload on eye-tracking measures (fixation number, gaze event duration and saccadic length). This was performed since the measures failed to meet normality and parametric assumptions. Only saccadic length showed significant decrease ($\chi_2 = 9.3$, $p = 0.01$, $df = 2$) across task load. A post-hoc Mann Whitney test suggested that these decreases were significant between high and other load groups but not low and medium groups. An increasing trend was seen in fixation number and decreasing trend was observed in the gaze event duration (Figure 13).

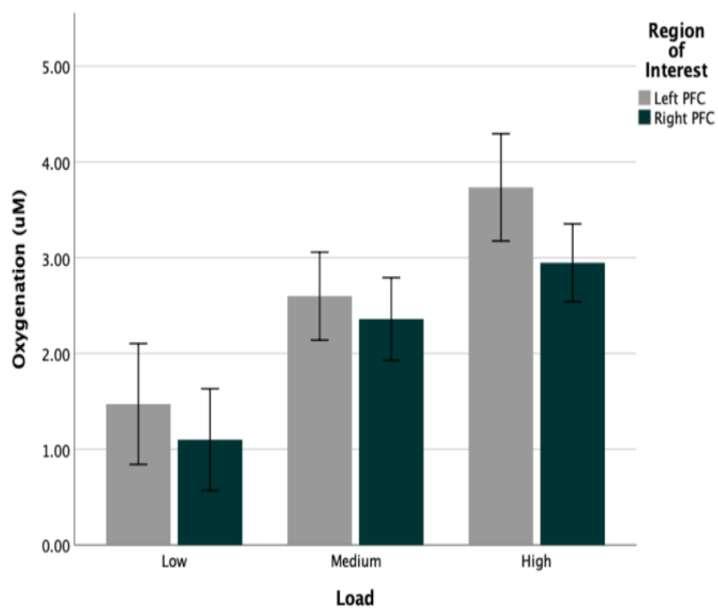


Figure 12: Average oxygenation measures from right and left PFC across low, medium and high multidimensional load conditions. Error bars are +/- 2 standard error of the mean (SEM).

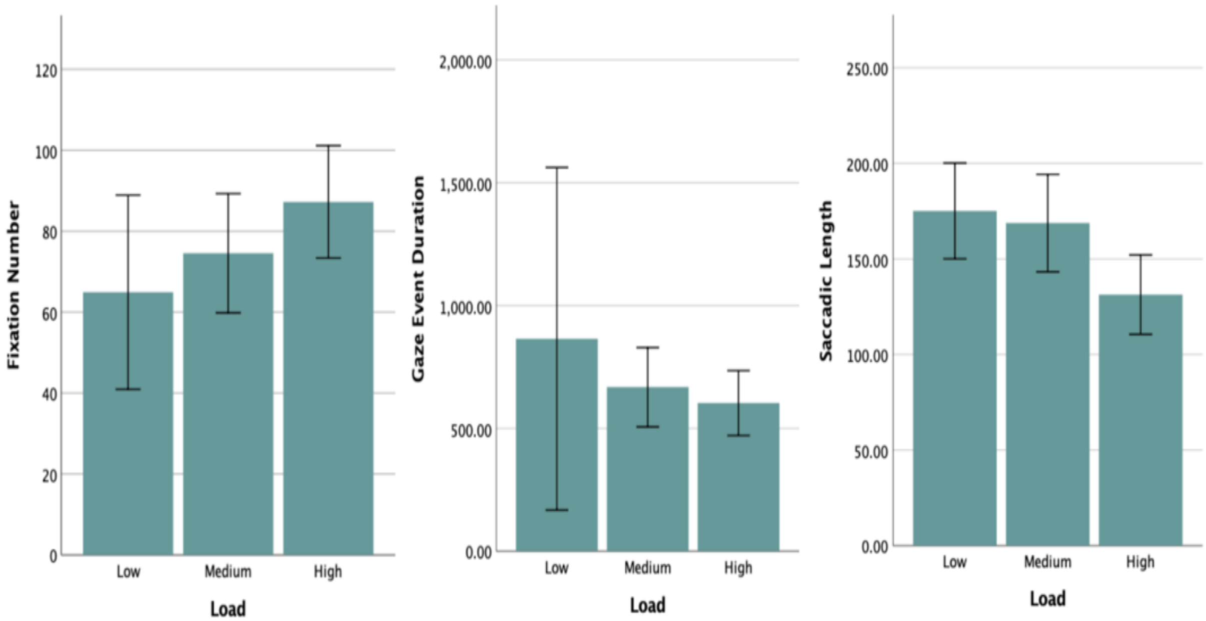


Figure 13: Average eye-tracking measures (fixation number, gaze event duration and saccadic length) across low, medium, and high multidimensional load conditions. Error bars are +/- 2 SEM.

4.3.4. Changes in neural and physiological measures relative to each other

The correlation coefficient of r = 0.97 between normalized left and right pupil dilation points towards a common causal factor (Hansen et al, 2018). As such, normalized right and left pupil dilation were averaged and will be referred to as the normalized pupil dilation. The oxygenation changes from the PFC were plotted against the normalized pupil dilation across participants (Figure 14) to observe the different workload capabilities elicited through the scenario. Here, participant 1 and 2 had a relatively decreasing trend, but the former maintained the lowest levels of oxygenation and the highest normalized pupil dilation, while in the latter, the opposite trend was observed. On the other hand, participant 3 had an increasing trend but maintained similar characteristics as participant 2. Additionally, the normalized pupil dilation levels appear to remain relatively constant without strong changes.

4.4. Discussion

Aircraft density and air traffic complexity have long been known to be two major factors involved in workload imposed on ATC specialists (Kirwan et al, 2001; Athenes et al. 2002). This study aimed to quantify these factors' influence on workload using a multimodal approach.

The results indicate that (i) eye-tracking and fNIRS measures are sensitive to increases in either task complexity or difficulty, but not to both; (ii) task difficulty and complexity are non-linearly related; (iii) increase in task demands, quantified by the interaction between task difficulty and complexity, cause an increase in both fNIRS and eye-tracking measures (iv) implementation of behavioral and physiological measures could serve as metric for assessing expertise even within a skilled ATC officer group.

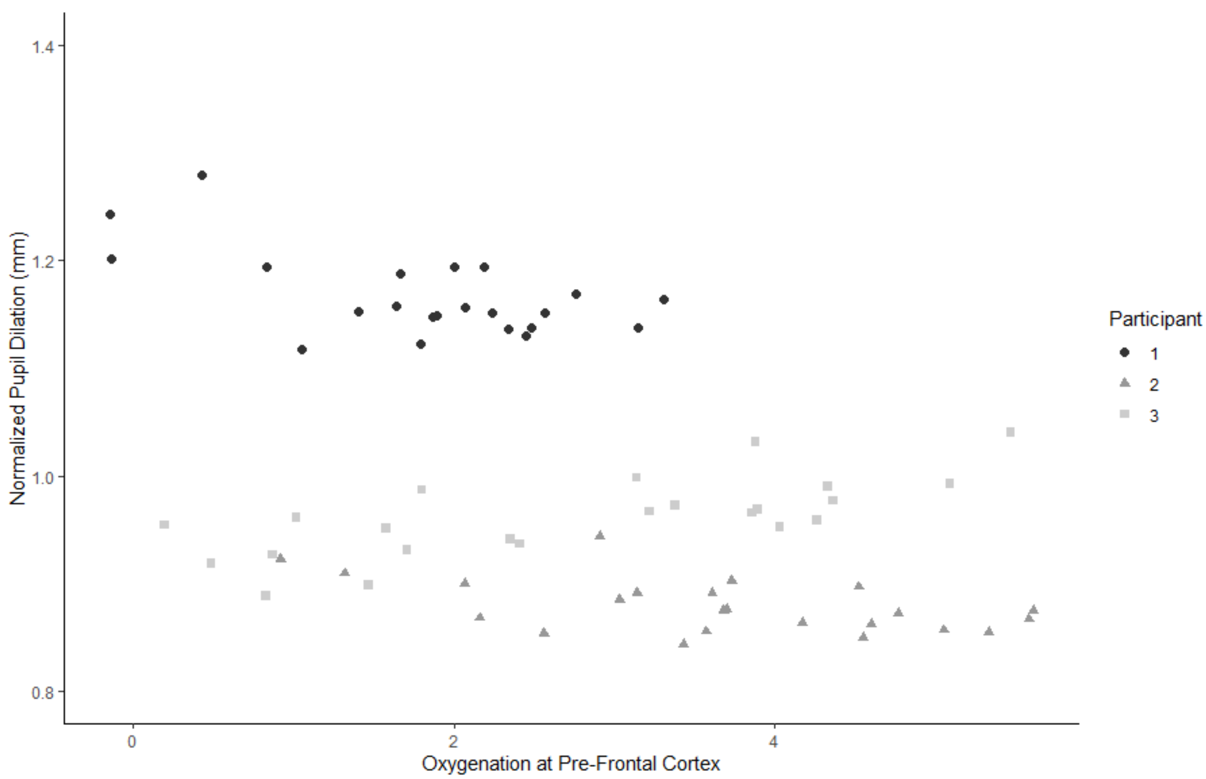


Figure 14: Scatterplot (R package: ggplot2, Wickham, 2016) of normalized pupil dilation plotted versus oxygenation levels at the PFC.

These preliminary results revealed that different modalities could be sensitive to different task demand factors. In this study, eye-tracking measures (fixation number) increased linearly with task complexity (number of clearances), while fNIRS measures (oxygenation) across both hemispheres increased linearly with task difficulty (number of aircraft). These findings agree with unimodal fNIRS or eye-tracking studies that assessed workload of controllers (Ayaz et al, 2012; Izzetoglu & Richards, 2019; Reddy et al., 2018; Van der Wel and Van Steenbergen, 2018; Tsai et al, 2012). Specifically, fNIRS studies assessing cognitive workload in ATC specialists, used the number of aircraft as an indicator of task demand and reported that as the demand increased, so did blood oxygenation response in both right and left hemisphere. Eye-tracking studies have shown that there exists a positive correlation between fixation number and workload (Ha et al, 2006). In this case, similar results were found as there exists a strong correlation between the number of fixations and the number of clearances ($r = 0.72$). The finding as to why fNIRS measures correlated with number of aircrafts, but not with number of clearances, and vice-versa for eye-tracking measures, calls for further study with more participants. Considering the limited sample size, this is still a critical finding and should be investigated further.

Task complexity, assumed in this study by the number of clearances, did not behave linearly with the task difficulty as defined by number of aircraft. At the beginning of the scenario there was an increase in aircraft, which was correlated with an increase in clearances. This positive correlation may be due to the fact that an experienced ATC officer can divide his/her attention to execute multiple simultaneous tasks. Following this trend, there was no change in clearances even though the number of aircraft increased, which may be attributed to the fact that attendance to multiple tasks is limited in number by the human perceptual system (Wickens & Hollands, 2003). In order to prevent an overload, an experienced controller would take note of information

(Wickens, 2002), building their situational awareness, but only attending to a set number of aircraft at a given moment. The latter portion of this trend line shows that the number of aircraft remained relatively constant on the final stages of the scenario, while the number of clearances increased. One possible explanation for this behavior may be that – once that many aircraft are present on both the ground and the airspace – the controller may need to micromanage aircraft in more detail to maintain the required separation between aircraft to ensure safety.

The final finding of this study is related to expertise. We discovered that the first participant was much more tolerable to increases in task demand in comparison to the other two participants. Specifically, he had no significant differences in behavioral performance when compared with other participants but showed significantly lower increase in fNIRS and higher increase in pupil dilation. This result could be in line with the hypothesis that expertise tends to be associated lower brain activity relative to novices, particularly in PFC areas (Milton et al., 2004). The reasoning as to why this happens can be explained by the fact that PFC is involved in higher order functioning, so once one becomes a veteran, their PFC frees up space for novel incoming stimuli or task demand. However, this evaluation of expertise is dependent on the expectation that behavioral performance positively correlated with task demand, as decrease in performance and brain activity would be indicative of a participant not engaging with the task or being overloaded (Izzetoglu et al., 2004).

This study was conducted under a number of limitations. This is the first time, as far as the authors are aware, that the number of clearances was used as a measure of complexity that was compared with physiological measures, therefore the use of this metric as a task demand factor needs to be investigated and thus validated with a larger sample size. The resulting distinction between participant 1 from the other two in an expertise context may be insignificant if the experiment were to be repeated more than once. It is important to highlight that the statistical analysis presented in this paper should be approached with caution as it was based on three participants – this exploratory phrase will be further expanded on later experiments. In this same train of thought, the results presented in this study need to be supported by enrolling a larger sample population, considering the fact that other multi-modal studies have reported contradictory results. For instance, Hogervorst et al, 2014 investigated which physiological variables provide the most accurate information on cognitive workload and found non-significant trends of improved accuracy for a combination of electroencephalogram (EEG) and eye-tracking over EEG alone, 91% and 86% accuracy, respectively. Another study identified that neurophysiological measures are more sensitive, in comparison to eye-tracking metrics, when discriminating the impact that task difficulty and complexity have in a high-fidelity driving scenario (Di Flumeri et al, 2018). In contrast, Bernhardt et al, 2019, reached the conclusion that EEG workload and pupil dilation are not correlated to each other, a finding also shared by Matthews et al, 2015, and hypothesized that it may be that pupil dilation may not be a direct measure of cognitive workload. Regardless, fNIRS and eye-tracking are portable, safe, affordable and unobtrusive systems, which can present a potential multi-sensor approach to a multidimensional evaluation of an ATC specialist's cognitive workload.

4.5. Acknowledgment

The authors would like to gratefully thank and recognize Dr. Jerry Crutchfield (FAA – Civil Aeronautical Medical Institute in Oklahoma City, Oklahoma) for his continuous support through the experimental protocol, data acquisition phases of the experiment; and for facilitating access to the simulation environment. In addition, his comments throughout the writing process of this paper were invaluable.

Chapter 5: Conclusion and Future Works

In this thesis, the state-of-the-art eye tracking and fNIRS devices were used together to investigate human interaction behaviors using high fidelity air traffic simulations. The multimodal approaches presented in this thesis showed promise on reliably quantifying the underlying mechanisms that dictate human performance and decision making.

In Chapter 2, a combined multimodal approach that consisted of self-report analyses and eye-tracking analyses enabled the investigation of how veteran en-route air traffic controllers carry out their tasks and build their situational awareness. In detail, the air traffic controllers participated in multiple scenarios that contained different types of possible conflicts. Eye-tracking data were recorded and mapped with the self-report analyses results. The mappings were conducted based on the multiple steps of visual search, detect, and mitigation of the possible aircraft conflicts.

The results showed generalizable strategies that veteran air traffic controllers apply under similar conditions. In detail, they applied a “continuous” visual search pattern applied to extract information from the environment to create and update their situational awareness. In addition, the controllers process information by first observing the altitude of an aircraft followed by its direction and speed. The information observation sequence can change based on the aircraft conflict conditions , meaning that certain mitigation strategies might require different information observation sequences. When choosing between conflict mitigation strategies, the controllers preferred to change an aircraft’ altitude or vector, rather than their speed. Lastly, the scenarios containing different factors such as weather, number of aircraft present, and type of conflict, can be useful to develop the mitigation guidelines when training novice air traffic controllers.

Chapter 3 consisted of another multimodal approach that combined an eye-tracker with a functional near infra-red spectroscopy (fNIRS), a neuroergonomic device that measures blood oxygenation to identify brain activity. This combination would allow us to quantify the cognitive workload of veteran local traffic controllers in a simulated tower environment. The goal of this experiment was two-fold: Investigate the correlations among the measurements obtained from the two devices and Explore the air traffic controllers' cognitive workloads when they are carrying out their tasks.

One of the key takeaways of this research was the multidimensional definition used to analyze the cognitive workload. By dividing workload into task complexity and difficulty, several interesting preliminary results were found. First, the devices used were sensitive to different behavioral measures. In detail, the number of eye fixations correlated with the number of commands, whereas the oxygenation at the pre-frontal cortex was related to the number of aircraft. It is unknown why certain measures are better correlated with certain cognitive measures and additional experiments are required. It is noted that oxygenation levels were inversely correlated with multiple eye-tracking measures such as gaze event duration and the pupil dilation. Second, the difficulty and complexity of a task appear to be non-linearly correlated. One possible reason might be that the controllers continuously builds "the whole picture" by attending to a set number of aircraft at a given moment. Lastly, the multimodal approach was able to identify the differences in the cognitive workloads among the veterans.

Combining the results of these two experiments, we can observe that the multimodal approaches show promise on exploring the underlying mechanisms of workload and human interaction in a complex, high-risk, and semantically rich environment. This is because cognitive workload can be considered as a multidimensional construct and different devices or approaches might be more effective in sensing changes in either the task difficulty or complexity. The results can be used to find ways to better train the novices.

5.1. Future research

One of the major avenues in which this work can be expanded upon is the development of multimodal approaches to measure multidimensional cognitive workload. Some future research questions are: (1) What are the system conditions that enables the multimodal approach to accurately capture the multidimensional cognitive workload, and (2) under which conditions a unimodal methodology might be sufficient. One possible method to accurately measure cognitive workload might be the use of NASA-TLX and investigate the correlation between the NASA-TLS outputs with the outputs obtained from the neuroergonomic devices.

Another future research topic is to compare the outcomes between the veterans and novices. This way, we could measure and quantify the differences that exist between the two groups. After, a set of novice controllers could undergo some additional training using the materials developed from the research provided in Chapter 2 to investigate whether any learning has taken place.

Another topic might be exploring other air traffic control environments which were not considered. For example, in the case of the en-route control environment, air sectors can be created using varied complexity, such as having major airports near them. For tower control, more intricate and complex airports can be introduced.

Lastly, more participants should be recruited to increase the validity of the research results and revise the results as needed.

References

Ahlstrom, & Friedman-Berg. (2006). Using eye movement activity as a correlate of cognitive workload. International Journal of Industrial Ergonomics, 36(7), 623-636.

Alemdag, E., Cagiltay, K. (2018). A systematic review of eye tracking research on multimedia learning. Computers and Education 125, 413–428.

Athènes, S., Averty, P., Puechmorel, S., & Delahaye, D.C. (2002). Complexity and controller workload: Trying to bridge the gap. In Proceedings of the 2002 International Conference on Human-Computer Interaction in Aeronautics, 56-60

Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Bonelli, S., Golfetti, A., et al. (2016). Adaptive automation triggered by EEG-Based mental workload index: a passive brain-computer interface application in realistic air traffic control environment. Front. Hum. Neurosci. 10. https://doi.org/10.3389/fnhum.2016.00539.

Aricò, P., Borghini, G., Di Flumeri, G., Bonelli, S., Golfetti, A., Graziani, I., et al. (2017). Human factors and neurophysiological metrics in air traffic control: a critical review. IEEE Rev. Biomed. Eng. 10, 250–263. https://doi.org/10.1109/rbme.2017.2694142

Aula, A., Majaranta, P., & Raiha, K. (2005). Eye-tracking reveals the personal styles for search result evaluation. Lecture Notes in Computer Science, 3585, 1058–1061.

Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., Onaral, B. (2012). Optical brain monitoring for operator training and mental workload assessment. NeuroImage 59(1), 36–47.

Ball M, Barnhart C, Nemhauser G, Odoni A. (2007). Air transportation: irregular operations and control. In: Barnhart C, Laporte G (eds) Handbooks in operations research and management science, vol 14. Elsevier, Amsterdam, pp 1–67.

Bernhardt, K. A., Poltavski, D., Petros, T., Ferraro, F. R., Jorgenson, T., Carlson, C., Iseminger, C. (2019). The effects of dynamic workload and experience on commercially available EEG cognitive state metrics in a high-fidelity air traffic control environment. Applied Ergonomics 77, 83–91.

Bhavsar, P., Srinivasan, B., Srinivasan, R. (2017). Quantifying situation awareness of control room operators using eye-gaze behavior. Computers and Chemical Engineering 106, 191–201.

Boag, C., Neal, A., Loft, S., & Halford, G.S. (2006). An analysis of relational complexity in an air traffic control conflict detection task. Ergonomics, 49, 1508-1526.

Bruder, C., & Hasse, C. (2019). Differences between experts and novices in the monitoring of automated systems. International Journal of Industrial Ergonomics 72, 1–11.

Bunce, S.C., Izzetoglu, M., Izzetoglu, K., Onaral, B., Pourrezaei, K. (2006). Functional near-infrared spectroscopy: an emerging neuroimaging modality. IEEE Engineering in Medicine and Biology Magazine 25, 54–62

Campana, A., Duci, A., Gambini, O., & Scarone, S. (1999). An artificial neural network that uses eye- tracking performance to identify patients with schizophrenia. Schizophrenia Bulletin, *25*(4), 789– 799. https://doi.org/10.1093/oxfordjournals.schbul.a033419

Causse, M., Lancelot, F., Maillant, J., Behrend, J., Cousy, M., Schneider, N. (2019). Encoding decisions and expertise in the operator’s eyes: Using eye-tracking as input for system adaptation. International Journal of Human Computer Studies 125, 55–65.

Chapman, P. R., Underwood, G. (Ed.) (1998). Eye guidance in reading and scene perception. Oxford, England: Elsevier.

Currie, J., Bond, R. R., McCullagh, P., Black, P., Finlay, D. D., & Peace, A. (2018). Eye Tracking the Visual Attention of Nurses Interpreting Simulated Vital Signs Scenarios: Mining Metrics to Discriminate between Performance Level. IEEE Transactions on Human-Machine Systems, 48(2), 113–124.

Dasari, D., Shou, G., Ding, L. (2017). ICA-derived EEG correlates to mental fatigue, effort, and workload in a realistically simulated air traffic control task. Front. Neurosci. 11. https://doi.org/10.3389/fnins.2017.00297.

Delpy, D. T., Cope, M., van der Zee, P. (1988). Estimation of optical path length through tissue from direct time of flight measurement. Physic in Medicine and Biology 33(12), 1433–1442.

Dempere-Marco, L., Hu, X., MacDonald, S., Ellis, S., Hansell, D., & Yang, G. (2002). The use of visual search for knowledge gathering in image decision support. IEEE Transactions on Medical Imaging, 21(7), 741–754.

Dhami, M. K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin*, *130*(6), 959–988. https://doi.org/10.1037/0033- 2909.130.6.959

Di Flumeri, G., De Crescenzio, F., Berberian, B., Ohneiser, O., Kramer, J., Aricò, P., Piastra, S. (2019). Brain–Computer Interface-Based Adaptive Automation to Prevent Out-Of-The-Loop Phenomenon in Air Traffic Controllers Dealing With Highly Automated Systems. Frontiers in Human Neuroscience 13.

Di Nocera F, Fabrizi R, Terenzi M, Ferlazzo F. (2006). Procedural errors in air traffic control: effects of traffic density, expertise, and automation. Aviation Space Environ Med 77, 693-643.

Durso FT, Manning CA. (2008). Air traffic control. Rev. Hum. Factors. Ergon. 4, 195–244.

Durso, F. T., Hackworth, C., Truitt, T. R., Crutchfield, J., Nikolic, D. & Manning, C. A. (1998). Situation awareness as a predictor of performance in en route air traffic controllers. Air Traffic Control Quarterly

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

Edwards T. (2013). Human performance in air traffic control. Dissertation, University of Nottingham.

Engelhardt, P. E., Ferreira, F., Patsenko, E. G. (2010). Rapid Communication Pupillometry reveals processing load during spoken language comprehension. The Quarterly Journal of Experimental Psychology 63(4), 639–645.

Evans, D. C., & Fendley, M. (2017). A multi-measure approach for connecting cognitive workload and automation. International Journal of Human Computer Studies 97, 182–189.

Federal Aviation Administration. (2018). Forecasts of IFR aircraft handled by FAA air route traffic control centers FY 2017-2040.

Finkelman, JM. (1994). A large database study of the factors associated with work-induced fatigue. Human Factors 36, 232-43.

Friedrich, M., Biermann, M., Gontar, P., Biella, M., Bengler, K. (2018). The influence of task load on situation awareness and control strategy in the ATC tower environment. Cognition, Technology and Work, 20(2), 205–217.

Goldberg, J., & Kotval, X. (1999). Computer interface evaluation using eye movements: Methods and constructs. International Journal of Industrial Ergonomics, 24, 631–645.

Gronlund, S. D., Ohrt, D. D., Dougherty, M. R. P., Perry, J. L., Manning, C. A. (1998). Role of memory in air traffic control. Journal of Experimental Psychology: Applied, 4, 263–280.

H. Wickham. (2016). R package "ggplot2": Elegant Graphics for Data Analysis. Springer-Verlag New York.

Ha, C.H., Kim, J.H., Lee, S.J., Seong, P.H. (2006). Investigation on relationship between information flow rate and mental workload of accident diagnosis tasks in NPPs. Nucl. Sci. IEEE Trans. 53(3), 1450-1459.

Hansen, J. P., Hardenberg, D., Biermann, F., Bækgaard, P. (2018). A gaze interactive assembly instruction with pupillometric recording. Behavior Research Methods 50(4), 1723–1733.

Harezlak, K., & Kasprowski, P. (2018). Application of eye tracking in medicine: A survey, research issues and challenges. *Computerized Medical Imaging and Graphics*, *65*, 176–190. https://doi.org/10.1016/j.compmedimag.2017.04.006

Harrison, J., Izzetoğlu, K., Ayaz, H., Willems, B., Hah, S., Ahlstrom, U., Onaral, B. (2014). Cognitive workload and learning assessment during the implementation of a next-generation air traffic control technology using functional near-infrared spectroscopy. IEEE Transactions on Human-Machine Systems 44(4), 429–440.

Hart, S.G., Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In Peter A. Hancock and Najmedin Meshkati (ed.), Advances in Psychology, Human Mental Workload. North-Holland, pp. 139–183.

Herbert, S. (1979). Information Processing Models of Cognition. *Annual Review Psychology*, *43*(6), 323. https://doi.org/10.1097/00006199-199411000-00001

Herrmann, M. J., Walter, A., Schreppel, T., Ehlis, A.-C., Pauli, P., Lesch, K.-P., Fallgatter, A. J. (2007). D4 receptor gene variation modulates activation of prefrontal cortex during working memory. European Journal of Neuroscience 26(10), 2713–2718.

Hilburn, B., Jorna, P. G. A. M. (2001). Workload and air traffic control. Human Factors in Transportation. Stress, Workload, and Fatigue, 384–394.

Hilburn, B., Jorna, P. G., Byrne, E. A., Parasuraman, R. (1997). The effect of adaptive air traffic control (ATC) decision aiding on controller mental workload. Human-Automation Interaction: Research and Practice, eds M. Mouloua and J. Koonce (Mahwah, NJ: Erlbaum), 84–91.

Hogervorst, M. A., Brouwer, A. M., van Erp, J. B. F. (2014). Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload. Frontiers in Neuroscience 8, 1–15.

Hollnagel, E. (2007). Flight decks and free flight: Where are the system boundaries? Applied Ergonomics, 38, 409-416.

Hunter, A., Parush, A. (2009). Using eye movements to uncover conflict detection strategies. In Proceedings of the Human Factors and Ergonomics Society 53rd annual meeting 1729–1733

Hüttermann, S., Noël, B., & Memmert, D. (2019). Eye tracking in high-performance sports: Evaluation of its application in expert athletes. *International Journal of Computer Science in Sport*, *17*(2), 182–203. https://doi.org/10.2478/ijcss-2018-0011

ICAO. (2005). Doc 9328: Manual of Runway Visual Range Observing and Reporting Practices, International Civil Aviation Organization, 3rd edn.

ICAO. The World of Air Transport in 2017, https://www.icao.int/annual-report-2017/Pages/the-world-of-air-transport-in-2017.aspx, last accessed 2020/02/09.

İşbilir, E., Çakır, M. P., Acartürk, C., Tekerek, A. Ş. (2019). Towards a Multimodal Model of Cognitive Workload Through Synchronous Optical Brain Imaging and Eye Tracking Measures. Frontiers in Human Neuroscience 13, 1–13.

Izzetoglu, K., Bunce, S., Onaral, B., Pourrezaei, K., Chance, B. (2004). Functional Optical Brain Imaging Using Near-Infrared During Cognitive Tasks. International Journal of Human-Computer Interaction 17(2), 211–227.

Izzetoglu, K., Richards, D. (2019). Human Performance Assessment: Evaluation of Wearable Sensors for Monitoring Brain Activity. In M. Vidulich & P. Tsang (Eds.), Improving Aviation Performance through Applying Engineering Psychology: Advances in Aviation Psychology (1st ed., pp. 163–180). Boca Raton, FL: CRC Press.

Izzetoglu, M., Izzetoglu, K., Bunce, S., Ayaz, H., Devaraj, A., & Onaral, B. (2005). Functional Near-Infrared Neuroimaging. IEEE Transactions on Neural Systems And Rehabilitation Engineering 13(2), 153–159.

Jeelani, I., K. Han, and A. Albert. (2018). "Automating and scaling person- alized safety training using eye- tracking data." Autom. Constr. 93 (1): 63–77. https://doi.org/10.1016/j.autcon.2018.05.006.

Jobsis, F. F. (1977). Noninvasive, Infrared Monitoring of Cerebral and Myocardial Oxygen Sufficiency and Circulatory Parameters. Science 198(4323), 1264–1267.

Kasarskis, P., Stehwien, J., Hickox, J., Aretz, A., Wickens, C. (2001). Comparison of expert and novice scan behaviors during VFR flight. In Proceedings of the 11th International Symposium on Aviation Psychology (n.p.). Columbus: Ohio State University.

Khan, R.S.A., Tien, G., Atkins, M.S., Zheng, B., Panton, O.N.M., Meneghetti, A.T. (2012). Analysis of eye gaze: do novice surgeons look at the same location as expert surgeons during a laparoscopic operation? Surg. Endosc. 26 (12), 3536–3540, http://dx.doi.org/10.1007/s00464-012-2400-7.

Kirwan, B. (1997). Validation of human reliability assessment of techniques: Part 1 - Validation issues. Safety Science 27(1), 25–41.

Kirwan, B., Scaife, R., Kennedy, R. (2017). Investigating complexity factors in UK Air Traffic Management. Engineering Psychology and Cognitive Ergonomics, 189–195.

Komogortsev, O. V., Karpov, A. (2013). Automated classification and scoring of smooth pursuit eye movements in the presence of fixations and saccades. Behav. Res. Methods 45, 203–215.

Koros, A., Rocco, P. S., Panjwani, G., Ingurgio, V., D'Arcy, J. F. (2003). Complexity in Air TrafficControl towers: A field study. Part 1: Complexity Factors.

Kostenko, A., Rauffet, P., Chauvin, C., Coppin, G. (2016). A dynamic closed-looped and multidimensional model for Mental Workload evaluation. IFAC - Papers On Line 49(19), 549–554.

Lamoureux, T., (1999). The influence of aircraft proximity data on the subjective mental workload of controllers on the air traffic control task. Ergonomics, 42, 1482–1491.

Langan-Fox, J., Sankey, M.J., Canty, J.M. (2009). Human factors measurement for future air traffic control systems. Hum. Factors J. Hum. Factors Ergon. Soc. 51(5), 595-637.

Lesiuk, T. (2008). The effect of preferred music listening on stress levels of air traffic controllers. Arts in Psychotherapy 35, 1-10.

Lingard, H. (2013). “Occupational health and safety in the construction industry.” Constr. Manage. Econ. 31 (6): 505–514. https://doi.org/10 .1080/01446193.2013.816435.

Mann, D. T. Y., Williams, A. M., Ward, P., & Janelle, C. M. (2007). Perceptual-cognitive expertise in sport: A meta-analysis. Journal of Sport & Exercise Psychology, 29, 457- 478.

Manning, D., Ethell, S., Donovan, T., Crawford, T. (2006). How do radiologists do it? The influence of experience and training on searching for chest nodules. Radiography 12 (2), 134–142, http://dx.doi.org/10.1016/j.radi.2005.02.003

Marchitto, M., Benedetto, S., Baccino, T., Cañas, J. J. (2016). Air traffic control: Ocular metrics reflect cognitive complexity. International Journal of Industrial Ergonomics 54, 120–130.

Marquard, J.L., Henneman, P.L., He, Z., Jo, J., Fisher, D.L., Henneman, E.A., (2011). Nurses’ behaviors and visual scanning patterns may reduce patient identification errors. J. Exp. Psychol. Appl. 17 (3), 247

Martijn Jansma, J., Ramsey, N. F., Coppola, R., Kahn, R. S. (2000). Specific versus Nonspecific Brain Activity in a Parametric N-Back Task.

Matsuo, K., Glahn, D. C., Peluso, M. A. M., Hatch, J. P., Monkul, E. S., Najt, P., Sanches, M., Zamarripa, F., Li, J., Lancaster, J. L., Fox, P. T., Gao, J. H., Soares, J. C. (2007). Prefrontal hyperactivation during working memory task in untreated individuals with major depressive disorder. Molecular Psychiatry 12(2), 158–166.

Matthews, G., Reinerman-Jones, L.E., Barber, D.J., Abich, J. (2015). The psychometrics of mental workload: multiple measures are sensitive but divergent. Hum. Factors 57(1), 125–143.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons’ responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

Milton, J. G., Small, S. S., Solodkin, A. (2004). On the Road to Automatic: Dynamic Aspects in the Development of Expertise. Journal of Clinical Neurophysiology 21(3), 134–143.

Neal and P. J. Kwantes. (2009). An evidence accumulation model for conflict detection performance in a simulated air traffic control task. Human Factors, vol. 51, no. 2, 164–180.

Niessen, C., Eyferth, K. (2001). A model of the air traffic controller's picture. Safety Science 37(2), 187-202.

Nunes, A., Mogford, R. H. (2003). Identifying controller strategies that support the ‘picture’. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 71-75

O’Brien, K.S., & O’Hare, D. (2007). Situational awareness ability and cognitive skills training in a complex real-world task. Ergonomics, 50, 1064-1091.

Otero, S.C., Weekes, B.S., Hutton, S.B. (2011). Pupil size changes during recognition memory. Psychophysiology 48(10), 1346-1353.

Palma Fraga, R. , Kang, Z., and Mandal, S. (2018). Characterization of air traffic controllers’ visual search patterns and control strategies. In Proceedings of the 2018 International Society for Engineers and Researchers (pp.1-5), Jun. 18-19, Seoul, S. Korea

Palma Fraga, R., Reddy, Y. P., Kang, Z., Izzetoglu, K. (2020). Multimodal analysis using neuroimaging and eye movements to assess cognitive workload. In Proceedings of the 22nd International Conference on Human-Computer Interaction, Jul. 19-24, Copenhagen, Denmark.

Parasuraman, R. (2003). Neuroergonomics: research and practice. Theoretical Issues in Ergonomics Science 4, 5–20.

Parasuraman, R. (2011). Neuroergonomics: brain, cognition, and performance at work. Current Directions in Psychological Science. doi:10.1177/0963721411409176.

Parasuraman, R., and Rizzo, M. (2007). Neuroergonomics: The Brain at Work, 1 Edn. New York, NY: Oxford University Press.

Parasuraman, R. and Wilson, G. (2008). Putting the brain to work: neuroergonomics past, present, and future. Human Factors 50, 468.

Prinzel, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., and Pope, A. T. (2000). A closed-loop system for examining psychophysiological measures for adaptive task allocation. Int. J. Aviat. Psychol. 10, 393–410. doi: 10.1207/ S15327108IJAP1004_6

Reddy, P., Richards, D., Izzetoglu, K. (2018). Cognitive Performance Assessment of UAS Sensor Operators via Neurophysiological Measures. Frontiers in Human Neuroscience 12.

Remington, R. W., Johnston, J. C., Ruthruff, E., Gold, M., Romera, M. (2000). Visual search in complex displays: Factors affecting conflict detection by air traffic controllers. Human Factors: The Journal of the Human Factors and Ergonomics Society, 42(3), 349–366.

Rudi, D., Kiefer, P., & Raubal, M. (2019). The instructor assistant system (iASSYST) utilizing eye tracking for commercial aviation training purposes. Ergonomics 63(1), 61–78.

Schulte-Mecklenbeck, M., Fiedler, S., Renkewitz, F., & Orquin, J. L. (2017). Reporting standards in eye- tracking research. In M. Schulte-Mecklenbeck, A. Kühberger, & J. Johnson (Eds.), A handbook of process tracing methods. New York: Routledge

Sheridan, T.B. (2006). Next Generation air transportation systems: Human-automation interaction and organizational risks. Paper presented at the 2nd Symposium on Resilience Engineering, Cannes, France.

Stager, P., Hameluck, D., & Jubis, R. (1989). Underlying factors in air traffic control incidents. In Proceedings of the Human Factors Society 33rd Annual Meeting, 43-46

Stein, E.S. (1985). Air traffic controller workload: An examination of workload probe. Atlantic City International Airport: Federal Aviation Administration Technical Center. (DOT/FAA/CT-TN84/24).

Truschzinski, M., Betella, A., Brunnett, G., Verschure, P.F.M.J. (2018). Emotional and cognitive influences in air traffic controller tasks: an investigation using a virtual environment. Appl. Ergon. 69, 1–9.

Tsai, M.-J., Hou, H.-T., Lai, M.-L., Liu, W.-Y., Yang, F.-Y. (2012). Visual attention for solving multiple-choice science problem: an eye-tracking analysis. Comput. Educ. 58(1), 375-385.

Tsang, P.S., Vidulich, M.A. (2006). Mental workload and situation awareness. In: Gavriel, S. (Ed.), Handbook of Human Factors and Ergonomics. John Wiley & Sons, Inc, Hoboken, NJ, pp. 243–268.

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

van der Wel, P., van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. Psychonomic Bulletin and Review 25(6), 2005–2015.

van Schaik, F. J., Roessingh, J. J. M., Lindqvist, G., Fält, K. (2010). Assessment of Visual Cues by Tower Controllers , with Implications for a Remote Tower Control Centre. IFAC Proceedings Volumes 43(13), 123–128.

Veltman, D. J., Rombouts, S. A. R. B., Dolan, R. J. (2003). Maintenance versus manipulation in verbal working memory revisited: an fMRI study. NeuroImage 18(2), 247–256.

Villringer, A., Chance, B. (1997). Noninvasive optical spectroscopy and imaging of human brain function. Trends in Neurosciences 20(10), 435–442.

Vogt, J., Hagemann, T., Kastner, M. (2006). The impact of workload on heart rate and blood pressure in en-route and tower air traffic control. J. Psychophysiology 20, 297–314.

Wei, T., Simko, V. (2017). R package "corrplot": Visualization of a Correlation Matrix (Version 0.84).

Wickens, C. D., Hollands, J. G. (2000). Engineering Psychology and Human Performance, Prentice Hal*l* (N. Roberts & B. Webber (eds.); Third). Prentice-Hall.

Wickens, Christopher D. (2002). Situation Awareness and Workload in Aviation. Current Directions in Psychological Science 11(4), 128–133.

Williams, A. M., Davids, K., & Williams, J. G. (1999). Visual perception and action in sport. London: E & F.N Spon.

Z. Kang and S. J. Landry (2014). Using scanpaths as a learning method for a conflict detection task of multiple target tracking. Human Factors, vol. 56, no. 6, 1150–1162, 2014.

Z. Kang and S. J. Landry. (2010). Capturing and analyzing visual groupings of multiple moving targets in an aircraft conflict detection task using eye movements. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 54, no. 23, 1906–1910.

Z. Kang and S. J. Landry. (2015). An Eye Movement Analysis Algorithm for a Multielement Target Tracking Task: Maximum Transition-Based Agglomerative Hierarchical Clustering," in IEEE Transactions on Human-Machine Systems, vol. 45, no. 1, 13-24.

Z. Kang, E. J. Bass, and D. W. Lee. (2014). Air traffic controllers' visual scanning, aircraft selection, and comparison strategies in support of conflict detection. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 58, no. 1, 77–81.