EFFECTS OF WORKING MEMORY, ATTENTION, AND EXPERTISE ON PILOTS' SITUATION AWARENESS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS INSTITUTE
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

SERKAN ÇAK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
PHILOSOPHY OF DOCTORATE
IN
THE DEPARTMENT OF COGNITIVE SCIENCES

JUNE 2011

Approval of Graduate School of Informatics

____

Prof.Dr. Nazife BAYKAL
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Doctor of Philosophy.

____

Prof.Dr. Deniz ZEYREK
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Doctor of Philosophy.

_____________	______________________________
Dr. Bilge SAY	Assist.Prof.Dr. Mine MISIRLISOY
Co-Supervisor	Supervisor

Examining Committee Members

Assoc.Prof.Dr.Kürşat ÇAĞILTAY	(METU, CEIT)	_______________________
Assist.Prof.Dr.Mine MISIRLISOY	(METU, PSY)	_______________________
Dr. Bilge SAY	(METU, SEM)	_______________________
Assist.Prof.Dr. Didem GÖKÇAY	(METU, COGS)	_______________________
Dr. Oliver WRIGHT	(BAHÇEŞEHİR, PSY)	__________________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: SERKAN ÇAK

Signature :

ABSTRACT

EFFECTS OF WORKING MEMORY, ATTENTION, AND EXPERTISE ON PILOTS' SITUATION AWARENESS

Çak, Serkan
Ph.D., Department of Cognitive Sciences
Supervisor : Assist.Prof.Dr. Mine MISIRLISOY
Co-Supervisor : Dr. Bilge SAY

June 2011, 143 pages

Situation Awareness (SA), is defined as perception of environmental entities, comprehension of their meaning, and estimation of their status in the near future (Endsley, 1995a). The general aim of the study is to investigate the relationship between SA and individual cognitive differences. Specifically, the predictive value of working memory and attentional capacity measures on SA measures, taken from pilots of different expertise levels, is of interest. In the literature, SA has mostly been studied from an applied perspective. The present study therefore aims at providing the necessary cognitive underpinnings of these more applied studies.

Two experiments were conducted. In Experiment 1, individual differences and SA measures have been taken from thirty-six pilots. Automated Operation Span, Stroop Task, and Choice Reaction Time Task with Dichotic Listening were used for measuring working memory capacity (WMC), inhibition, and divided attention, respectively. Online and offline SA measurements were employed together for tapping on different aspects of SA in a cognitively demanding flight scenario. Results showed that WMC and expertise explain 58% of variability in offline scores while inhibition, divided attention, and expertise explain 52% of variability in online scores. In Experiment 2, the aim was to find correlates of eye movements in terms of individual differences. Scan patterns were studied across four SA-related visual tasks with ten expert pilots. Results showed that more expert pilots produced less fixation durations but no other effects of individual differences on the eye movements were observed. It was also observed that expert pilots deploy some scan strategies while performing these tasks.

Keywords: situation awareness, working memory, attention, individual cognitive differences, expertise

ÖZ

ÇALIŞMA BELLEĞİ, DİKKAT VE DENEYİMİN PİLOT DURUMSAL FARKINDALIĞI ÜZERİNE ETKİLERİ

Çak, Serkan
Doktora, Bilişsel Bilimler Bölümü
Tez Yöneticisi : Y.Doç.Dr. Mine MISIRLISOY
Ortak Tez Yöneticisi : Dr. Bilge SAY

Haziran 2011, 143 sayfa

Durumsal Farkındalık, çevresel nesnelerin algılanması, anlamlandırılması ve yakın gelecekteki durumlarının tahmin edilmesi olarak tanımlanır (Endsley, 1995a). Bu çalışmanın genel amacı, Durumsal Farkındalığın bireysel bilişsel farklılıklarla ilişkisinin araştırılmasıdır. Spesifik olarak, farklı deneyim seviyelerindeki pilotlardan alınan çalışma belleği ve dikkat kapasitesi ölçümlerinin Durumsal Farkındalığı tahminleme gücü çalışmanın ilgi alanını oluşturmaktadır. Literatürde Durumsal Farkındalık, çoğunlukla uygulamalı bilimlerin yaklaşımları kullanılarak çalışılagelmiştir. Bu çalışma, söz konusu uygulamalı çalışmalara gerekli bilişsel temelleri sağlamayı hedeflemektedir.

Çalışmada iki deney gerçekleştirilmiştir. İlk deneyde, otuzaltı pilottan bilişsel kapasite ve Durumsal Farkındalık Ölçümleri alınmıştır. Çalışma belleği kapasitesinin ölçülmesi için Otomatize İşlem Erim Görevi, baskılama gücünün ölçülmesi için Stroop ve bölünmüş dikkat kapasitesi ölçümü için de Çift Dinleme görevi, Seçim Reaksiyon Zamanı göreviyle birlikte kullanılmıştır. Çevrimiçi ve çevrimdışı olarak bilinen ölçüm teknikleri, bilişsel açıdan zorlayıcı bir uçuş senaryosunda Durumsal Farkındalığın farklı bileşenlerini ölçümleyebilmek için birlikte kullanılmıştır. Sonuçlar, çalışma belleği kapasitesi ve deneyimin % 58 oranında çevrimdışı skorlarındaki değişimi açıklayabildiğini, baskılama gücü, bölünmüş dikkat kapasitesi ve deneyimin % 52 oranında çevrimiçi skorlarındaki değişimi açıklayabildiğini göstermiştir. İkinci deneyde temel hedef, bireysel farklılıklar açısından göz hareketlerinin korelasyonlarının bulunmasıdır. On deneyimli pilot örneklemi kullanılarak Durumsal Farkındalık ile ilintili dört farklı görsel görev gerçekleştirilirken bireysel farklılıklarla ilişkilendirilebilecek görsel tarama örüntüleri araştırılmıştır. Sonuçlar göstermiştir ki, deneyimli pilotlar, deneyimsizlere göre daha kısa fiksasyon süreleri üretmektedirler. Diğer bireysel farklılıkların bir etkisi gözlemlenememiştir. Ek olarak, deneyimli pilotların görsel görevleri gerçekleştirirken bazı tarama stratejileri kullandıkları gözlemlenmiştir.

Anahtar Kelimeler: durumsal farkındalık, çalışma belleği, dikkat, bireysel bilişsel farklılıklar, deneyim

dedicated to my beloved wife and daughter..

ACKNOWLEDGMENTS

I am heartily thankful to my supervisor Mine Mısırlısoy and my co-supervisor Bilge Say whose encouragement, supervision and support enabled me to develop an understanding of the subject.

I owe my deepest gratitude to my thesis examining committee members Kürşat Çağıltay for his guidance, Didem Gökçay for her suggestions and Oliver Wright for his insightful questions. I also thank Annette Hohenberger for her valuable comments on the manuscript.

I wish to express my appreciation to Türker Özkan, Mehmet Harma and Murat Perit Çakır for their suggestions on the statistical analyses conducted.

I also would like to make a special reference to Müjgan Doğan who was the instructor pilot throughout my first experiment. Without her cooperation I could not have completed the experiment with sufficient participation. I would like to thank all the participant pilots for their time and Subject Matter Experts for their help to create a flight scenario.

This work was supported by METU BAP Funding and partially supported by Modeling and Simulation R& D Center (ODTU-TSK Modsimmer) of METU, Ankara, Turkey. In this study, two Eye Tracking Systems have been used. One belonged to ODTU-TSK Modsimmer and the other belonged to Human-Computer Interaction Laboratory of METU. I wish to acknowledge their staff for making this study possible and eyetracking researchers Cengiz Acartürk, Türkan Karakuş, Özge Alaçam, and Nihan Ocak for their support.

I should acknowledge my department's administrative staff Sibel Gülnar, Necla Işıklar, Ali Kantar, and Hakan Güler who were always friendly, helpful and constructive throughout my study.

I would like to thank my mother and my father raising me as self-confident and supporting me for whatever I did in my life. I wish to thank my mother-in-law for preparing tasty teas and coffees in the sleepless nights and my father-in-law for encouraging me in the peaks of stress.

The last three years of my dissertation were very difficult for me. My daughter Zeynep was born and unfortunately, I missed invaluable moments of her development during the final stages of this PhD. She recognized me as "studying father" but not "loving" or "playing" one. Indeed, I love her very much. I will be recovering this as fast as I can.

Last but not least, my special thanks to my wife, Aslı. Without her, it would have been much harder to complete this study. Loving her and receiving her love makes me a better person.

TABLE OF CONTENTS

CHAPTERS

APPENDICES

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACC	Accuracy
AI	Attitude Indicator
ALT	Altimeter
AOI	Area of Interest
AOSPAN	Automated Operation Span
ASI	Airspeed Indicator
ATC	Air Traffic Control
ATIS	Automated Terminal Information Service
BSPAN	Backward Span
CARS	Crew Awareness Rating Scale
CRT	Choice Reaction Time with Dichotic Listening
C-SAS	Cranfield Situation Awareness Scale
DA	Decision Altitude
DME	Distance Measurement Equipment
EXP	Expertise
FFS	Full Flight Simulator
FSPAND	Forward Span Dissimilar
FSPANS	Forward Span Similar
HCI	Human-Computer Interaction
HDC	High Divided Attention Capacity Pilot
HIS	High Inhibition Strength Pilot
HSI	Horizontal Situation Indicator
HWMC	High Working Memory Capacity Pilot
IFR	Instrumented Flight Rules
ILS	Instrument Landing System
LCD	Liquid Crystal Display
LDC	Low Divided Attention Capacity Pilot
LE	Low Expert Pilot
LIS	Low Inhibition Strength Pilot
LTM	Long-Term Momory
LWMC	Low Working Memory Capacity Pilot
MARS	Mission Awareness Rating Scale
ME	More Expert Pilot
METU	Middle East Technical University
MIDAS	Man-Machine Design and Analysis System
OSPAN	Operation Span
PC	Personal Computer
PL	Phonological Loop
PSAQ	Participant Situation Awareness Questionnaire
RMI	Radio Magnetic Indicator
RT	Response Time
SA	Situation Awareness
SABARS	Situation Awareness Behaviorally Anchored Rating Scale

SACRI	Situation Awareness Control Room Inventory
SAGAT	Situation Awareness Global Assessment Technique
SAMPLE	Situation Awreness Model for Pilot-in-the-Loop Evaluation
SARS	Situation Awareness Rating Scale
SART	Situation Assessment Rating Technique
SAQ	Situation Awareness Query
SA-SWORD	SA Subjective Workload Dominance Metric
SME	Subject Matter Expert
SPAM	Situation-Present Assessment Method
STM	Short Term Memory
STRP	Stroop
UHF	Ultra High Frequency
VFR	Visual Flight Rules
VHF	Very High Frequency
VSI	Vertical Speed Indicator
VSSP	Visuospatial Sketchpad
WMC	Working Memory Capacity
WM	Working Memory

CHAPTER 1

INTRODUCTION

Situation Awareness (SA), a term used since the beginning of 1980's, can be informally defined as "knowing what is going on around". Cognitively, it is not far from the truth that SA is an "umbrella" term related to many cognitive processes such as perception, long-term memory (LTM), working memory (WM), attention, reasoning, learning, and decision-making (Horswill and McKenna, 2004; Sohn and Doane, 2004; Johannsdottir, 2004; Kokar, 2004; McCarley, Wickens, Goh, and Horrey, 2002; Sukthankar, 1997; Endsley, 1997). SA, generally speaking, is a cognitive state beginning with perception and ending with decision-making. SA was first conceptualized in the military aviation domain for solving some applied problems like improving performance of operators interacting with mission-critical systems[1].

Looking back at aviation history, SA has been operationally very important. There were many SA-related catastrophic accidents. According to the statistics, 80.2% of the accidents were caused by perceptual factors, 16.9% of them corresponded to failure in comprehension and 2.3% of them were caused by wrong predictions and decisions (Jones and Endsley, 1996). Perceptual factors consist of unavailable data for performing a task, undetectable data, failure in monitoring data, data misperception, and memory loss. Comprehension issues consist of lack of poor mental representations, incorrect mental representations, and over-reliance on default values. There are many other physiological and psychological factors causing the operators to lose SA. Fatigue, boredom, time pressure and anxiety are some of them (Hockey, 1986; Sharit

---

[1] Mission-critical systems are the systems whose failure results in the failure of business operations.

and Salvendy, 1982). Small amounts of stress can positively affect (increase) the level of SA although operators lose SA under intense stress (Janelle, Singer, and Williams, 1999).

The terrible accident of Boeing 737/800, Turkish Airlines flight on Feb. 25, 2009 was an example of how loss of SA affects the safety of lives in flight. In the accident, nine people (5 passengers and 4 crew members) were killed and 80 passengers were injured. In the final analysis, the aircraft manufacturer, the airport tower, the airliner and the pilots have all been found to play a partial role in the accident. The below paragraph which is quoted from the accident report prepared by the Dutch Safety Board explains SA-related part of the accident.

> "*The Board concludes that the improper functioning of the left-hand radio altimeter system led to the thrust from both engines being reduced by the autothrottle to a minimal value too soon, ultimately causing too big a reduction in speed. The airspeed reached stall speed due to a failure of monitoring the airspeed and pitch attitude of the aircraft and a failure to implement the approach to stall recovery procedure correctly. This resulted in a situation where the wings were no longer providing sufficient lift, and the aircraft crashed*"(p.7).

The fact that the flight crew did not effectively monitor the aircraft's flight parameters indicates loss of SA. Considering this example, monitoring activities which seem to be very important for flight safety can be investigated through analysing eye movement data related to required cross-checks of readings in flight displays. There were also contributing factors that led to this accident such as system design, flight manuals, training and maintenance procedures which are out of scope of this study. SA studies have been inevitably linked to the applied problems like the accident in the example. Because those studies have some major weaknesses in terms of theoretical research paradigms, there is a need to bridge the gap between the applied research of SA and theoretical research paradigms in order to find out cognitively plausible solutions to the applied problems.

One of the problems with past SA studies in applied areas is the implicit idea of an "ideal SA". The literature is laden with assumptions of how an "ideal SA" can be attained; however, in fact characterizations of cognitive characteristics may imply there is no "ideal" SA but only "actual" SA and individual differences. Dekker and Lützhöft (2004) made a distinction between actual and ideal SA by stressing accuracy in mapping from the objective outside world to the operator's inner representation of that world. If the operator's mental representations have significant differences from the "real" world (i.e. inaccurate mapping), the operator would be unaware of the current situation. They made an analogy between mind-matter and actual-ideal pairs. Matter (ideal SA) represents all available material outside related to SA and mind (actual SA) represents the scope and contents of the mental mirror of the outside world that the operator has. The assumption in Dekker and Lützhöft's study of mapping actual SA in its entirety to mental representations and processes can be fine grained. Some portion of SA variations could be explained with the operators' individual differences including expertise. Cognitive processes associated with working memory (WM) influence acquiring and maintaining SA. Limited capacity of attention draws the boundaries of our cognitive abilities, consequently, of the operator's SA. Therefore, WM and attention can be evaluated as important predictors of operator's SA level.

In the scope of this study, WM and attention are taken as two of many SA's constructs in SA's cognitive umbrella. Working Memory Capacity (WMC), inhibition, divided attention, and level of expertise are taken as individual differences. Although there are some studies on individual differences and SA (Endsley and Bolstad, 1995; Bolstad and Hess, 1995; Dillon, 1996), effects of the individual differences on acquiring and maintaining SA are not well-defined in the literature. Unlike others, Durso, Bleckley, and Dattel (2006) put forward a cognitively valuable approach questioning the contribution of SA in the validation of cognitive capacity tests. They showed a connection between SA and some cognitive capacity tests measuring individual differences. In that sense, this study has similar characteristics to Durso *et al.* (2006)'s approach to investigate possible relationships between SA measurements and cognitive capacity tests. Durso *et al.* (2006) focused on the difference between what an SA measurement technique measures and what a set of cognitive capacity tests measure. On the other hand, this study tries to reveal the commonalities among them because it is argued that a set of cognitive capacity tests could measure SA as much as it covers SA's cognitive constructs. Investigation of the relationship between an applied concept and cognitive capacity tests would give us an opportunity to define, model, and measure this applied concept better.

Regarding the applications of SA, aviation companies have many products that have positive contributions to the operator's SA and help prevent such losses of SA. System designers mostly use some ergonomics guidelines in designing those products in order to achieve "good" SA. How good is "good"? Before coming to a conclusion, we have to investigate underlying cognitive processes to achieve good SA and we have to measure it. Defining, modeling, and even measuring SA have been mostly studied by cognitively shallow methodologies without investigating its relevance to cognition. This investigation requires not only an applied perspective but also use of theoretical research paradigms. There is a limited number of less well-known cognitively-deep studies in the literature about SA. For instance, Johannsdottir (2004) stated that maintaining SA in a dynamic environment involves three WM components, namely the phonological loop, the visuospatial sketchpad and the central executive proposed by Baddeley (2000). To our best knowledge, apart from these two studies cited above, none of the studies on SA or eye movements in pilots as reviewed in Chapter 2 are based on a cognitively inclined approach to SA.

As mentioned above, perception is the entry to SA processes. Any failure in perception negatively affects SA and some visual search strategies such as skill to perceive environmental information without directly steering at the information and ability to suppress task-irrelevant information positively affects it. An operator having high SA is expected to perform three processes related to eye movements: deploying correct visual search strategies, quickly getting task-relevant information and precisely analyzing them to make correct decisions. In this regard, eye movement data deserves to be investigated as to whether they indicate level of expertise and consequently SA status in the aviation domain.

Another need for studies that bridge SA and cognitive science is enhancement of ecological validity. In some studies that were valuable for bridging SA and cognitive science, simplified tasks were carried out by university students (Johannsdottir, 2004), in yet others that were on the applied side of SA, actual professionals were investigated in their professional environments without using theoretical research paradigms. In this dissertation, two high fidelity flight simulators were used to create a realistic operational environment. The flight scenario was designed via adding realistic events to the scenario and actual professional pilots participated in the experiments. To sum up, this study intends to investigate individual cognitive differences including WMC, strength of inhibition, divided attention capacity and expertise as to whether they could explain, to some extent, the variability in acquiring and maintaining SA in the aviation domain. This study also intends to use eyetracking methodology to relate the pilots' eye movements to their individual differences.

Research questions are as follows:

- Which individual cognitive differences are better predictors of acquiring and maintaining SA?

- In the context of aviation, are there some eye movement scan patterns indicating the pilot's individual differences, specifically level of expertise?

There are two experiments in this study. In Experiment 1, a combined version of two major SA measurement techniques is used for measuring pilot participants' SA in the environment of two high fidelity flight simulators. WM, inhibition, and divided attention as individual differences are measured through three cognitive capacity tests. Regression analysis is used to predict SA level by using individual difference measures and level of expertise. Experiment 1 provides empirical evidence for the relationship between SA and individual cognitive differences. In Experiment 2, four visual tasks on the recorded videos of the flight displays in a simulated flight are used. Participants' eye movements and individual differences are collected. Non parametric tests are conducted to investigate the effects of individual differences on the eye movements.

In the context of this research, there are some limitations. Participants performed SA measurement and cognitive capacity tests under some personal and environmental conditions which cannot be always perfectly controlled by the researchers. For instance, unlike operational conditions, participants may have no stress in performing cognitive tasks during experiments. Additionally, the simulated environment may not give a realistic sense to the participants as in real flights. Consequently, their experimental responses may have some deviations from the operational ones. Another limitation was the relatively small number of participants due to the difficulties in accessing the pilot population and in flight simulator arrangements.

Although the domain of this study is limited to aviation, some of the results are generalizable to some other domains such as Air Traffic Control (ATC) operators, nuclear power plant operators, race drivers, etc - more inclusively, all operators supposed to execute high level cognitive functions like decision making, planning within a complex and dynamic environment under hard time-constraints - based on further confirmation from comparable studies in such domains.

This dissertation contributes to the need for an accumulation and an increase in coverage of studies in SA and cognition by investigating the relationship between attention, WM, and level of expertise and SA in the aviation domain. Consolidation of cognitively-deep SA studies will help us to find out better solutions to the applied SA problems.

The next chapter examines the relevant literature on SA, summarizes WM and attention studies and positions this study with respect to other studies in the literature. Chapter 3 describes the experiments in detail. The designs of the experiments are discussed and the information about the participants, apparatus, experimental design, procedure and measurements are given in this chapter. Chapter 4 includes the results, a discussion including the interpretation of the results, limitations, possible future work on SA and conclusions.

CHAPTER 2

LITERATURE REVIEW

It is the fundamental assumption of this dissertation that studies investigating the roles of cognitive mechanisms in achieving and maintaining SA would give us distinctive understandings of what SA really is. In this regard, this chapter covers SA definitions, SA research methods (modeling and measurement efforts), the relationship between SA and individual differences such as WM, attention and expertise.

2.1 Situation Awareness

Along with the rapid technological development, complicated systems used in safety-critical systems such as aviation and nuclear power plants have emerged in the last two decades. Increasing physical and functional complexity of the systems raises difficulties in perceiving and responding timely to mission-critical data changes. An operator has to intensively use many man-machine interfaces. Situation Awareness (SA), which is defined as perception of elements in the environment, comprehension of their meanings, and projection of their status into the near future (Endsley, 1995a), is an important cognitive state that strongly affects the mission success through the effective use of complex systems.

In the literature, there is no single definition of SA. In one of several attempts given in the literature, SA is seen as "accessibility of situation representation which is continuously being updated" (Sarter and Woods, 1991, p.45). The focus of this definition is automation and loss of SA caused by automation. For instance, mode transitions of aircraft subsystems[1] without the pilot's control may cause pilots to lose information about the subsystems' status and consequently SA. As time goes on during the flight, pilots cannot maintain SA without being sensitive to the minor changes in the elements of the environment. Dennehy and Deighton (1997) describe SA as ability to see the "big picture" and think forward. In another definition, SA is described as a cognitive state related to evaluating many environmental cues in a dynamic situation (Isaac, 1997).

Considering the SA definitions above, a time and context dependent nature of SA emerges. For instance, Endsley (2000) emphasizes temporal aspects of SA by describing temporal dynamics in SA context. Pew (2000) posits that SA has context-dependent characteristics. The task-irrelevant information the operator may know while performing a critical task does not support the operator for acquiring or maintaining SA. Only context-dependent information supports the operator in terms of SA. In addition to context and time dependent nature of SA, SA researchers define SA in such a way that the current definitions can create a duality. In other words, it is not clear whether SA is taken as a single cognitive process or a state as an outcome of some cognitive processes[2]. In the literature, the state account of SA is widely accepted (Endsley, 1995a). Endsley uses the term "state of knowledge" as representative of this cognitive phenomenon. This approach emphasizes the distinction between SA-related information and SA-related cognitive processes. In parallel with Endsley's view, Tenney, Adams, Pew, Huggins, and Rogers (1992) propose that the state of awareness is a state. In opposition of the state view, Sarter and Woods (1991) advocate that SA is an ongoing cognitive process. The duality in SA definitions is reminiscent of the criticism of cognitive science by dynamical systems theory. Whereas "classical" Cognitive Science holds a more static view, dynamical approaches stress the process characteristic of cognition (Van Gelder, 1998; Beer, 2000). In this study, there is no SA definition adopted from the current literature although SA measurements used are based on Endsley's SA definition.

---

[1] Some aircraft subsystems may automatically change their operational modes such as idle, operative, emergency modes in accordance with automation rules.

[2] In the literature, the cognitive state account of SA is predominantly called "product view of SA", however, the term "cognitive state" is preferred to be used for the same account in this study.

Regardless of whether SA is a cognitive state or a process, it can be easily said that SA is the basis for successive decision making and actions in operations of complex and dynamic systems. Perception of SA-related data and their integration by the operator are the first cognitive stages for attaining SA and estimation of future events and system states is the final stage (Endsley, 1995a). In Endsley's model depicted in the Figure 2.3, projection and decision making are separate but this separation is cognitively controversial: Where does the projection end? Where does the decision making start?

SA has been mostly investigated to find solutions to applied problems such as testing complex systems for their SA support to the operators or selecting operators for those complex systems. The SA literature dominantly includes these studies that belong to the area of applied research (Brill, Gilson, Mouloua, Hancock, and Terrence, 2004; Hauss and Eyferth, 2003; Matthews, Pleban, Endsley, and Strater, 2000; Hogg, Folleso, Strand-Volden, and Torralba, 1995). On the contrary, a limited number of studies address what SA really is (Sarter and Woods, 1991), what cognitive mechanisms underlie SA (Sohn and Doane, 2004), and how SA can be modeled (Shively, Brickner, and Silbiger, 1997; Zacharias, Miao, Illgen, Yara, and Siouris, 1995). There is a considerable difference between formal modeling methods in basic research and SA modeling efforts in applied research. In the next section, the modeling efforts concentrating on a few SA "semi-formal" modeling studies will be reviewed while summarizing the plethora of applied studies for SA's "conceptual modeling" in the current literature.

2.1.1 Modeling SA

The modeling approaches can be classified into two groups: Information-processing account and perception/action cycle account. As mentioned above, the modeling efforts in applied research is significantly different from the one in basic research. One predominant view of modeling in basic cognitive research includes the followings: developing a conceptual theoretical framework, converting modeling assumptions into mathematical or computational descriptions, estimating model parameters from the empirical data, and finally comparing model predictions with the empirical data (Busemeyer and Diederich, 2010). Except for Shively *et al.* (1997) and Zacharias *et al.* (1995)'s models which are still not cognitive models, there is no cognitive model of SA based on a cognitively plausible theory, experimental evidence and predictive power to the best of the author's knowledge.

Two models proposed by Zacharias *et al.* (1995) and Shively *et al.* (1997) are computational models whereas in a strict sense, they are not based on a plausible theory of cognition even though some of the cognitive mechanisms are implemented on the basis of a cognitive theory. Zacharias *et al.* (1995) develop an SA model called SAMPLE. Figure 2.1 depicts the overall architecture of SAMPLE.

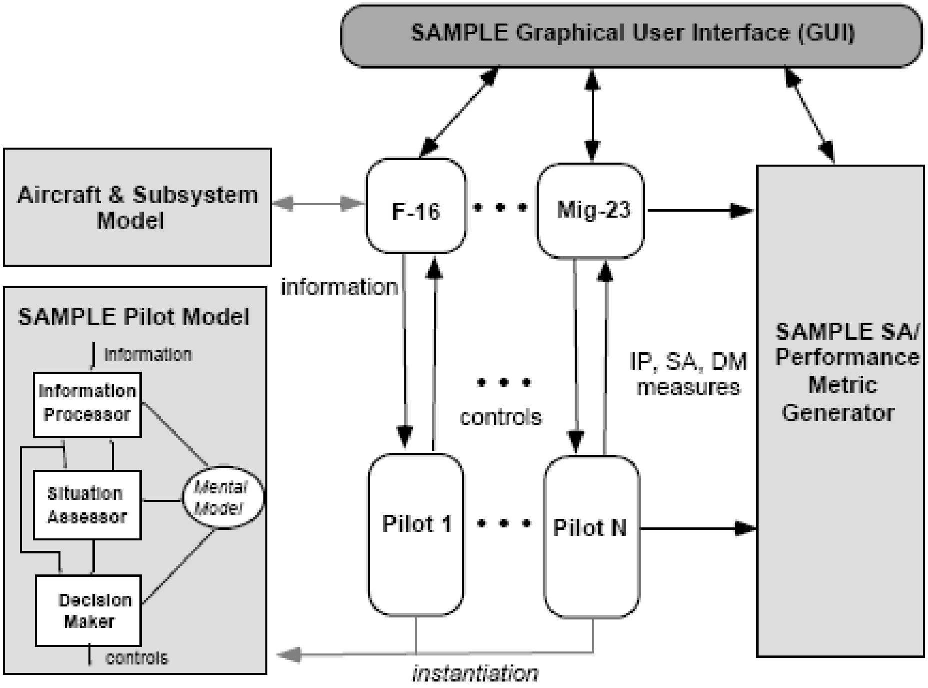


Figure 2.1: Architecture of SAMPLE adopted from Zacharias *et al.* (1995, p.4)

The pilot model is in the center of SAMPLE. Basically, the pilot model consists of Information Processing, Situation Assessment and Decision Making modules. These modules have access to short-term memory (STM) and long-term memory (LTM). Information processing is handled by fuzzy inference, situation assessment is performed by a Bayesian Belief Network and decision making is conducted by a rule-based expert system. Information on the current situation and the environment reside on STM. LTM includes expert knowledge. The pilot model describes the information as a set of event/situation relations in form of if-then rules (if event E1 then situation S1). Information Processing module processes the inputs by giving fuzzy membership grades and identifying them as known events. There is an input filtering mechanism for simulating human attentional capacity. Detected events are sent to the Situation Assessment module to make reasoning about the detected events by using deductive and abductive reasoning[3]. After assessing the situation, the assessment (current situation) is transferred to the Decision Making module to choose necessary actions required for the current situation.

In a computationally similar way, Shively *et al.* (1997) model SA by using the operator model of MIDAS architecture proposed by Smith and Tyler (1997). MIDAS is a general cognitive architecture used for modeling man-machine interfaces. The proposed SA model depicted in Figure 2.2 is based on two concepts: Situation elements and situation-sensitive nodes. Each situation element is associated to a higher-order node. For instance, consider an attack helicopter flying in an enemy zone where some ground-to-air missile launchers are located. The model processes the missile launchers as situation elements and threat as a higher order node which is a more general semantic concept of the missile launchers. Each node is weighted in accordance with its given priority in the situation. These two models are developed for designing man-machine interfaces and creating a model-based metric for the system's evaluation. Although the models have promising features in terms of cognitive sciences, they still have shortcomings in the above mentioned modeling process that Busemeyer and Diederich (2010) identified. For instance, it is barely possible to say that computational structures used in these models meet the computational principles in human cognition.

---

[3] A reasoning process of achieving an explanatory hypothesis

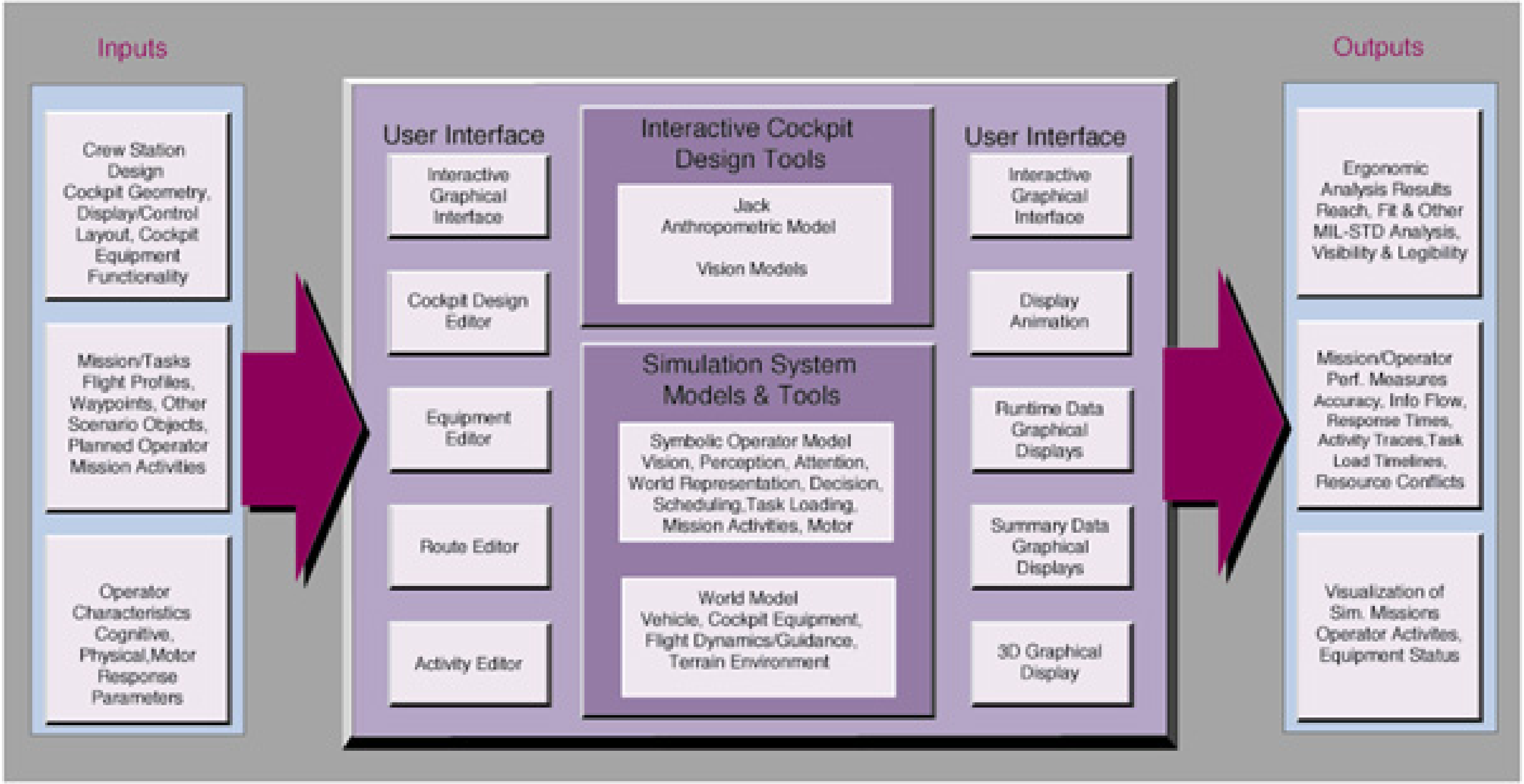


Figure 2.2: Architecture of MIDAS adopted from http://hsi.arc.nasa.gov /groups /midas /design /architecture.html

Having given examples of information-processing accounts involving computational implementation in the previous paragraph, it is noteworthy that there is another type of SA models in information-processing account which could be defined as "conceptual" involving no computational implementation. The typical example of that is the approach proposed by Endsley (1995a). Endsley's approach to modeling SA consists of three successive levels as shown in Figure 2.3. Consider the case in which a military pilot is inside a threat zone (missile launcher, radar sites etc). According to Endsley's view, the threat is a task factor representing the state of the environment. If the pilot sees the threat in avionics displays, perception of the threat is completed, and if the pilot recognizes the threat is dangerous, the pilot comprehends the situation. If the pilot is able to estimate the time at which the aircraft would be inside the threat coverage (danger zone) and determine when a maneuver is necessary, the pilot projects the future status of the situation. Endsley's multi-level model dictates a hierarchical and linear processing system. In the following paragraphs, each level will be explained in that linear order.

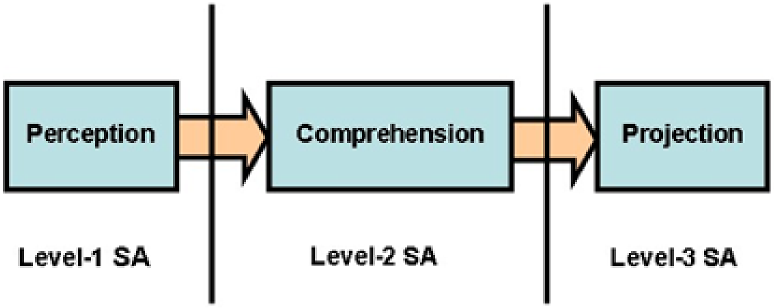


Figure 2.3: Three Levels of SA

The first step to attain SA covers perception of status, properties, and dynamics of the elements in the environment. For instance, the operator (pilot) is required to precisely perceive data (airspeed, position, altitude, route, direction, etc.) about the aircraft and its subsystems. Additionally, weather condition, Air Traffic Control clearances, emergency information are also important for attaining SA (Endsley, Farley, Jones, Midkiff, and Hansman, 1998). According to the model, perception of data required for performing the tasks is the initial stage in the process of acquiring and maintaining SA. As an example, the pilot knows the flight route and looks at the weather radar screen. The weather radar shows a meteorologically unsafe region on the flight route.

Assume that the weather radar screen has flares preventing the pilot to see the unsafe region on the screen. In this case, perception fails and level-1 SA is said to be poor.

Comprehension is based on synthesis of individual pieces of level-1 perceptual information. Level-2 SA can be summarized as understanding the importance of elements in the scope of the operator's objectives rather than being aware of the existence of those elements. The whole picture of the environment is constructed in order to generate patterns of the elements over level-1 perceptual information. The effects and the meanings of events, the information within the picture are comprehended (Endsley *et al.*, 1998). In the previous example, assume that pilot is novice and s/he has never used the weather radar and has some problems to interpret its outputs. If the pilot could see but could not report a meteorologically unsafe region on the flight route, the pilot is said to fail in level-2 SA because the pilot could not comprehend what s/he sees on the weather radar screen.

Level-3 SA is a capability of foreseeing the future actions of the elements in the environment. Level-3 SA could be attained through both level-1 and level-2 SA (i.e. status information, dynamics of elements and comprehension of the status) (Endsley *et al.*, 1998). Assume that the pilot in the example sees and reports the unsafe region on the weather radar screen. The pilot evaluates the wind direction and projects the future position of the unsafe region so that s/he could fly inside this unsafe region for the next few minutes, then she changes the flight route. In this situation, the pilot could correctly predict the future position of the unsafe region and level-3 SA is said to be high. Prediction of the future status of the elements in the environment gives the necessary inputs to the decision making process.

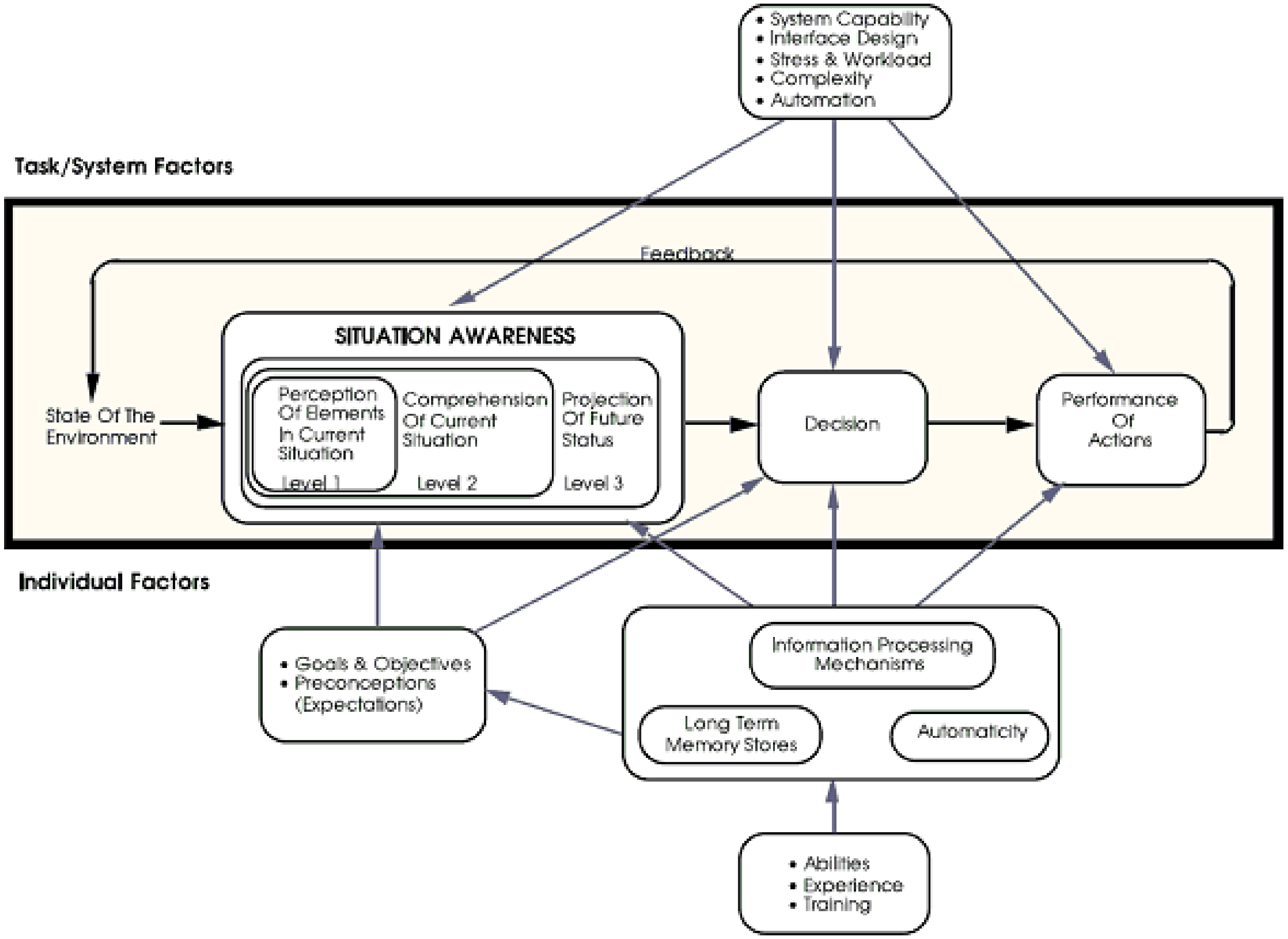


Figure 2.4: Endsley's SA Model adopted from Endsley (1995b, p.35)

In Endsley's SA Model depicted in Figure 2.4, perception is assumed to belong to cognition, and attention is implicitly identified in her information-processing mechanisms. Although Endsley and Rodgers (1996) study air traffic controllers who allocate their limited attention across multiple aircraft, they do not revise the model to explicitly show attentional mechanisms such as inhibition, attention switching and focusing. Additionally, schemata located in long-term stores are not psychologically well-defined abstract constructs (Uhlarik and Comerford, 2002). Endsley's model is based on short-term stores, implicit attention mechanisms and schemata. There is an attentional pool which divides the operator's attentional capacity to perception, comprehension, decision making and action guidance. In addition to schemata in LTM, Endsley (1995b) put forward scripts that are assumed to be used by the operators in decision making and action guidance.

In a continuation of the model development within the same approach, Zhang and Hill (2000) investigate SA in two sub-levels: situation template (representation) and situation assessment. Acquiring SA is possible through the processes of identifying the elements in the environment, their relations, constructing their structural representation of the situation, and assigning it to the mostly matched template in the memory. Although the proposed architecture includes Endsley's three-level SA processes, perception and cognition are separate in the latter model. Situation templates can be thought of as representations generated in LTM by the experience and rehearsal. Pilot forms a hypothesis about the situation by grouping perceptual elements and evaluating the relations between them. A pilot tries to match the hypothesis with the representations in the situation template and concludes that the closest one is the real situation. The relationship between SA and focused attention is explained by using a zoom lens metaphor as previously proposed by Eriksen and Yeh (1985). When the zoom lens magnifies at a low level, the field of view is greater, but with a low level of detail. As the magnification level of the lens increases, the amount of detail increases, but the field of view decreases. If we apply the zoom lens metaphor to focused attention, focused attention works like zoom lens with level of magnification. Considering the pilot's operational environment (i.e cockpit), monitoring the flight displays seems to be the similar to the case of increased field of view and decreased level of detail. On the other hand, reading a value from one of the displays seems to be the similar to decreased field of view and increased level of detail. Zhang and Hill (2000)'s SA model could be taken as a variant of Endsley's SA model.

Since Endsley's SA model and its variants predominantly affected the studies on modeling and measuring SA in the literature, it is explained in detail. Endsley's model that is being used today for solving applied problems about SA is important for this dissertation to the extent that it gives an opportunity to search for a cognitively plausible ground of an applied concept which is mostly defined in Endsley's way.

Apart from Endsley's approach to SA modeling, some models are developed using the concept of perception/action cycle (Adams, Tenney, and Pew, 1995; Smith and Hancock, 1995). The perception/action cycle is based on the actual world (available information), the cognitive map (knowledge and experiences stored in LTM), and finally locomotion and action (perceptual exploration). In perception/action cycle, the actual world modifies the cognitive map; the cognitive map directs locomotion and action that bring about sampling the actual world. The process is cyclical and it emphasizes SA's dynamic characteristics. Adams *et al.* (1995) describe SA by using this concept, a central feature underlying of which is that SA can be conceived as both a state and a process. In other words, they state that SA is a state which can be described as a "state of the currently activated schema"and is a process as "the current state of the entire perceptual cycle". They divide the psychological construct schema into explicit focus and implicit focus. Explicit focus represents working memory (WM) and implicit focus represents the activated schema. Long-term episodic memory and long-term semantic memory are also adapted to the proposed SA model. On the other hand, Adams and colleagues do not elaborate on the psychological constructs in their model such as schemas and semantic memory. Smith and Hancock (1995) add the concept of "Invariant"for SA to the perception/action cycle account of SA in order to link the actual world, the cognitive map, and the action. This link supports the competing behavior of the operator. That is, level of pilot's aviation knowledge which is *invariant* affects how much SA could be attained. For that reason, they suggest that SA can be evaluated by considering the operator's competing behavior as well as the current situation.

The two perception/action cycle accounts of SA explained above can be conceived to fit into a cognitively promising ground by Hommel, Müsseler, Ashersleben, and Prinz (2001)'s study proposing a framework for perception and action planning; The theory of event coding (TEC). TEC is based on common coding structure for sensory and motor events in such a way that actions can be represented by perceivable events. Hommel *et al.* (2001) suggest that event representations are not invariant but modified in accordance with the task demands and the operator's intentions. Unfortunately, it is difficult to validate the common coding events in TEC as criticized by Wolters and Raffone (2001). In this framework, scripts, schemata, invariant concepts defined in the current perception/action cycle accounts of SA are mapped into events in a cognitively plausible way. TEC could be an alternative SA framework especially when a quick and smooth action is required to a current perception. Instead of the linear succession of perception - comprehension - projection - action and against the traditional information processing acount with its translation process between perceptual and action codes, they posit a common coding or common representation of perceptual and action features. That is, at some point in the cognitive process, perception and action share the same representation and action automatically follows from perception. These actions that are most strongly related to a certain perception are automatically activated. This can be advantageous if quick responses are needed, however, also disadvantageous if automated responses need to be inhibited and more deliberate actions to be taken.

SA models based on an information processing account are prone to not capturing the dynamic nature of SA. For instance, an aircraft flying within the enemy zone may be locked by a ground-to-air missile which has approximately 9 seconds travel duration. Aircraft's subsystems can detect the locked missile within the last 3 seconds. Before detection, SA seems to be lost. Just few seconds later, the pilot has a high level of SA but s/he has just 3 seconds to make maneuvers to escape from the missile. This example dramatically shows time dimension of SA and any model proposed should capture this dimension. It is also important to note that high level of SA does not always guarantee the necessary actions operator has to perform.

The information processing account of SA dictates a sequential model similar to Endsley's model. Our experience about situation awareness conflicts with this nature of the model. Consider the case that each day on our way to work we may avoid the same holes and bumpers on the road and we can drive our car without passing over the holes on the road not only by perceiving them but by using our past knowledge about the road without any (conscious) perception. In Endsley's terminology, level-1 SA (perception) fails but level-3 SA (projection) is high. This clearly violates the model's property of being sequential. The Perception /Action Cycle account of SA is generally based on some psychological constructs such as semantic memory, and schemata which are not explicitly defined in the models. As mentioned above, this account also considers SA as a state and as a process at the same time since both the state of the active schema (state), and the state of the perceptual cycle (process) exist in this account. Apart from state and process views of SA, the concept of SA is based on three aspects: SA-related information, SA-related cognitive processes and resources. Despite the deficiency of the theoretical and empirical aspects of the existing models of SA, SA measurement methods are more extensively worked on because of the importance of the applied nature of SA. SA-related information could be investigated by offline SA measurements and SA-related cognitive processes could be probed by online SA measurements and finally, SA-related cognitive resources could be monitored by workload surveys as explained in section 3.1.3.1. Offline and online SA measurements, workload survey used in the scope of this study are expected to cover these three aspects of SA. The following subsection details the SA measurement techniques in the literature, their classification and comparison.

2.1.2 Measuring SA

Researchers working in applied areas such as aviation have developed some measurement techniques under the assumption that an operator's task-specific knowledge (i.e. SA-related information) during the operation indicates the level of SA or that the operator's task performance is a good indication for judging the quality of SA. These two assumptions results in knowledge-based and performance-based measurement techniques respectively.

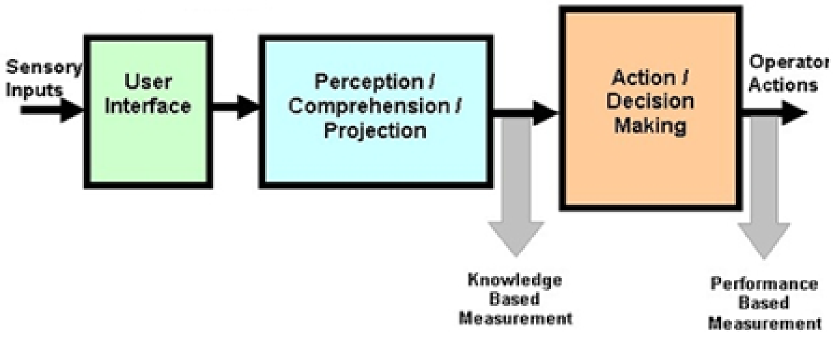


Figure 2.5: SA Measurement Points adapted from Pritchett and Hansman (2000)

Knowledge-based and performance-based measurement techniques have different measurement points as shown in Figure 2.5. These two measurement points correspond to different levels in the human cognitive system. Knowledge-based techniques probe the operator's knowledge after serial cognitive processes such as perceptual processes, memory processes and reasoning. Performance-based techniques probe the operator's performance after decision making and operator's actions. In knowledge-based techniques, SA is assumed to be poor when the operator knows less SA-related information than required for performing the tasks. In performance-based techniques, SA is assumed to be poor when the operator could not completely perform the required tasks in terms of accuracy, timing, etc.

There are some advantages and disadvantages of SA measurement techniques in terms of type of research question, reliability and accuracy requirements. Performance-based measurement techniques have advantages over knowledge-based techniques in evaluating the areas in which SA is poor because SA-related task performance mostly varies in parallel with SA. If an SA researcher wants to investigate characteristics of events and conditions where operators frequently suffer from losing SA, performance-based techniques would be suitable for this kind of investigation. In supporting this idea, it is also argued that performance-based techniques have the capability to question timing and reliability of the operator's reactions much better than knowledge-based techniques (Pritchett and Hansman, 2000). In case of exploration of SA as a cognitive phenomenon, knowledge-based measurement would be suitable because performance-based measurement techniques are not much informative about the human cognitive system but about the systems with which the operator interacts.

Considering the reliability requirements in SA measurements, performance-based techniques are more prone to unreliable measurements compared to the others due to the fact that scenario generation and evaluation of an operator's performance mostly are not standardized. Accuracy is another requirement for SA measurements. Since knowledge-based techniques directly probe the SA-related information, they are more accurate than performance-based techniques in the theoretical evaluation of the operator's SA because performance-based techniques can only make inferences dependent on the operator's performance and how the knowledge is interpreted. In the following paragraphs, self-rating[4], observer-rating[5] techniques as performance-based measurements and Freeze Online Probe[6], Real-time Probe[7] as knowledge-based measurements will be shortly explained.

In the literature, there are many SA measurement techniques and different classifications from different perspectives. For instance, self-rating (Matthews, Beal, and Pleban, 2002; Matthews, Pleban, Endsley, and Strater, 2000; Dennehy, 1997; Waag and Houck, 1994; Taylor, 1989; Vidulich, 1989) and observer-rating techniques (Neal, Griffin, Paterson, and Bordia, 1998; Dennehy, 1997) are two subcategories which can be called subjective methods. Another distinction among the measurement techniques in the literature is whether the simulator is frozen or not during the query administration (Freeze Online Probe vs. Real-time Probe). In this dissertation, a simpler classification is used for SA measurement techniques in the literature. If SA measurement is conducted as a post-trial administration or conducted while freezing the simulator screen, this type of SA measurements is classified as offline. If SA measurement is conducted during task performance, it is classified as online. Some measurement techniques use self-rating and observer-rating at the same time (Dennehy, 1997) and some others are based on merging the results of online and offline query administration (Jeannot *et al.*, 2003). Table 2.1 lists the major SA measurement

---

[4] Operator rates his/her own performance at the end of the operational scenario.

[5] Subject Matter Expert rates the participant's performance in accordance with a set of criteria.

[6] SA queries are administered to operator of a system such as an aircraft, a nuclear power plant, a mission-critical software while system's simulation is frozen. Accuracy of response is recorded.

[7] SA queries are administered to operator of a system during operation. Response times to queries are recorded.

techniques, their classification, and their authors in the literature.

Table 2.1: SA Measurement Techniques

SA Measurement Techniques	**Type**	**Author/Source**
SAGAT - Situation Awareness Global Assessment Technique	Offline	Endsley (1995a)
SART - Situation Awareness Rating Technique	Offline (Self-rating)	Taylor (1989)
SA-SWORD - SA Subjective Workload Dominance Metric	Offline (Self-rating)	Vidulich (1989)
SALSA	Offline	Hauss and Eyferth (2003)
SACRI - Situation Awareness Control Room Inventory	Offline	Hogg *et al.* (1995)
SARS - Situation Awareness Rating Scale	Offline (Self-rating)	Waag and Houck (1994)
SPAM - Situation-Present Assessment Method	Online	Durso and Dattel (2004)
SASHA - SA for Solutions for Human-Automation Partnerships in European ATM	Online and Offline	Jeannot *et al.* (2003)
SABARS - Situation Awareness Behaviourally Anchored Rating Scale	Online (Observer-rating)	Neal, Griffin, Paterson, and Bordia (1998)
MARS - Mission Awareness Rating Scale	Offline (Self-rating)	Matthews *et al.* (2002)
CARS - Crew Awareness Rating Scale	Offline	McGuinness and Foy (2000)
C-SAS - Cranfield Situation Awareness Scale	Online and Offline (Self/Observer Rating)	Dennehy (1997)
PSAQ - Participant SA Questionnaire	Offline (Self-rating)	Matthews, Pleban, Endsley, and Strater (2000)

Strater, Endsley, Pleban, and Matthews (2001) use SAGAT, SABARS, and PSAQ as measurement techniques. SAGAT is based on measuring accuracy of responses. SABARS is an observer-rating technique and PSAQ is based on post-trial administration of SA questionnaire. These three techniques have distinct features in terms of how and when they measure SA. They show that both SAGAT and SABARS are sensitive to experience levels of platoon[8] leaders in military operations. Endsley, Sollenberger, and Stein (2000) compare SAGAT, SPAM, and SART. SPAM is a real-time probe measurement which uses response times to SA queries and SART is a self-rating technique which is based on post-trial administration of SA questionnaire including situation familiarity, attentional focus, information quantity, information quality, instability of situation, concentration of attention, situational complexity, situation variability, arousal, and spare mental capacity. SART uses a seven point rating scale for these dimensions. They claim that SAGAT measures SA with high sensitivity, SPAM correlates with workload which they called "spare mental capacity" and SART has a high correlation with observer's performance ratings. However, this comperative study is weak in terms of its statistical power because they used only 10 ATC operators. It is also interesting to note that the concept of "spare mental capacity" used for workload in the study resembles the spare capacity proposed by Kahneman (1973) who directly links it to attention. This resemblance is in parallel with Wickens (2002)'s study that describes workload together with attentional demands exceeding the limited resources. In addition to this, Jones and Endsley (2000) conduct a comparative study for SAGAT, SPAM, and SART with 20 operators who were surveillance technicians, identification technicians and weapon team members working in an air operation center. They use two 60-minute scenarios for peace and war time. They indicate that SAGAT and SPAM had a weak correlation.

Characteristics of SA measurement techniques like SAGAT are clear indications of an application dominant perspective regarding SA research in the literature. SAGAT uses accuracy of reporting under the assumption that accuracy of reporting is positively correlated with the SA-level. SAGAT strongly focuses on the SA-related information, not the cognitive processes or resources. In their review of the book "Situation Awareness Analysis and Measurement" by Garland and Endsley (1995), Durso,

---

[8] a military unit

Crutchfield, and Batsakes (2002) criticize SAGAT in some aspects. They suggest that although WM is a crucial component of SA, it is not always right to conclude that operator's SA level is low whenever s/he could not report the SA-related information required for performing the task. There is no need for the operators to store the required information in WM if this information is continuously displayed and easily accessed by the operator. Therefore, memory probe could not measure all aspects, all components of SA. This is one of the strong criticisms on SAGAT and its application-dominant perspective regarding SA. Alternatively, SPAM uses response times to the SA-related queries administered under the assumption that response time is negatively correlated with SA-level. Using response times in SA measurements facilitates the researchers to evaluate SA-related cognitive processes beyond SA-related information. In SPAM, workload is removed from the measurement by giving the query-reject option to the participants in high workload condition but workload is an important component for evaluating the cognitive resources in the context of SA. In order to investigate which type of operators in terms of individual differences are good at acquiring and maintaining SA, a combined use of offline and online SA measurement techniques as explained in section 3.1.3.1 could be used to probe SA-related information and SA-related cognitive processes together with the evaluation of required cognitive resources through a workload survey.

In the scope of this study, apart from measuring SA, some individual cognitive differences are explored within the context of SA: WM, inhibition, divided attention and expertise. The below paragraphs unfold how WM and attention capacities are defined, what WM/attention approaches there are in the literature, and their relevance to SA and finally effects of expertise on SA.

2.2 Individual Differences

2.2.1 Working Memory Capacity

WMC is one of the major research areas in the WM literature. It is agreed that WM has a limited capacity. There are some approaches explaining this limited capacity. Just (1992) propose that there is a cognitive resource pool for storing activated representations in WM and performing cognitive processes. Towse, Hitch, and Hutton (2000) explain the limited capacity by putting forward time constraints for rehearsal. They argue that memory decay can only be prevented through rehearsal and rehearsal takes some time. This time constraint required for rehearsal brings about the limited capacity in WM. Waugh and Norman (1965) suggest that representations in WM interfere with each other. Interference results in limited use of resources. In parallel with this view, Saito and Miyake (2004) state that forgetting during a complex span task is not time-based, but interference-based.

In complex operational environments like an aircraft's cockpit, the limited capacity of WM pushes operators to use their executive cognitive abilities while integrating information from different sources (Endsley, 2000) because this kind of integration heavily loads WM. Considering WM-SA relation, limitations on WM bring about constraints on acquiring and maintaining SA (Fracker, 1988). Experts use different cognitive mechanisms from novices in order to overcome limitations on WM. In the same vein, Durso and Gronlund (1999) suggest that operators use four different strategies for reducing WM load: prioritization of information, chunking, encoding information changes in required levels, and restructuring the environment for constructing external memory cues. It can be said that high cognitive abilities developed through deploying such strategies have a key role in understanding the WM-SA relation. The question of what contribution WM as an individual cognitive difference makes in acquiring and maintaining SA could be answered through some empirical evidence supporting the WM-SA relation. Is it possible to say "pilots having high WMC are better at acquiring and maintaining SA compared to pilots having low WMC"?

There is only a limited number of studies in which the relationship between WM and SA is examined from a basic experimental perspective. Johannsdottir (2004)'s dissertation is one of the experimental studies to support the proposition that WM mechanisms have a crucial role in maintaining SA. After perceiving information, this information is stored in WM which serves as a medium to activate the related mental representations stored in LTM . Johannsdottir studies this relationship and concludes that maintaining SA in a dynamic environment selectively involves the phonological loop (PL), the visuospatial sketchpad (VSSP) and the central executive which are components of WM by Baddeley (2000). She uses a dual task paradigm in examining the relation between WM and SA. The primary task is a tracking task and the secondary task is WM load tasks for different WM components. She relates tracking performance to the SA status although SA is a meta-cognitive state rather than a cognitive function like tracking. It is not possible to completely equate SA status with tracking performance. This weakens the generalizability of the study. Nevertheless, Johannsdottir tries to explain an applied concept with basic research paradigms. Gonzales and Wimisberg (2007) use a water purification tank task to investigate the relationship between WMC and SA. Thirty-six volunteers are trained for performing this task. During the task performance, two different SA measures (online and offline) were taken. They find that WMC significantly predict offline SA scores. One way to experimentally study the relationship of WM and SA is through WMC. Therefore we must understand how memory span tasks measure WMC.

Complex Span Tasks (Operation Span, Reading Span, Counting Span) which are based on the dual task paradigm (the primary task is a memory span measure and the secondary task is a processing task) help us use WMC as a predictor of higher-level cognitive abilities (Daneman and Carpenter, 1980) such as SA. Studies in the literature have shown that simple span measures like Backward Span (BSPAN), Forward Span Dissimilar (FSPAND), and Forward Span Similar (FSPANS) do not have a high correlation with higher level cognitive abilities (Engle, Tuholski, Laughlin, and Conway, 1999). Unlike these simple span measures, more complex span measures like reading span, operation span, counting span have much higher correlations with reasoning, intelligence, language abilities, etc. These three span measures are widely used for measuring WMC.

In the Operation Span task, participants are required to solve series of algebraic operations and at the same time they are asked to remember set of unrelated words. Participants are required to read aloud the algebraic operations and the words to be remembered. After reading the word, the next algebraic operation is displayed on the monitor. At the end, participants try to recall all words in the presented order (Turner and Engle, 1989). The Reading Span Task (Daneman and Carpenter, 1980) is based on presenting series of sentences one at a time. Participants are supposed to read each sentence and remember the last words of the sentences. Then they are required to recall the last words of all the sentences in the presented order. The Counting Span Task (Case, 1985) involves counting shapes and remembering count totals and it is especially used for participants with a relatively lower education level or children. These tasks have moderately correlated results with WMC (Unsworth, Heitz, Schrock, and Engle, 2005).

Unsworth *et al.* (2005) propose a practical version of operation span task, namely, the automated operation span task (AOSPAN). The task is performed in three stages. In the first stage, participants practice with simple letter span. A letter appears on the screen, and the participants are asked to remember the letter presented. Then the participants perform mathematical operations. The participants are required to solve the operation as quickly as possible and then advance to the next screen. On the next screen, a result is presented and the participants are expected to click on true/false boxes according to their calculation. After each operation, the participants are informed on the accuracy of their answers. AOSPAN produces scores ranging between 0 and 72. High scores correspond to high WMC span and low scores correspond to low WMC span. Unsworth *et al.* (2005) state that the suitable WM task depends on the research aim. Regarding the aim of this study, AOSPAN is well-suited since it measures complex WMC, has similar results as the Counting Span and the Reading Span Tasks, yet it is not affected by participants' cultural differences or language.

In exploring characteristics of the relationship between SA and WM, WMC as an individual difference could be used by using complex span tasks because limited capacity of WM determines these characteristics. It would be fruitful if we knew how pilots compensate their low WM spans in acquiring SA in comparison with the ones having high WM span. Maybe, pilots with low WMC are not capable of acquiring SA like pilots with high WMC. Can attention capacity and expertise help pilots overcome memory bottlenecks in acquiring SA? All those questions bring up a matter that WMC together with other individual cognitive differences could be indicators of the level of pilots' SA during flight.

2.2.2 Attention Capacity

Attention is the ability to allocate cognitive resources to perform specific functions by filtering out functions-irrelevant stimuli. Attention is the concentration of mental effort on sensory and mental events. Attentional processes facilitate, enhance or inhibit other cognitive processes. Sarter and Woods (1991) stress the importance of attention in supporting SA but attention processes have not been linked with experimental work in SA. SA models in the literature are based on cognitive processes requiring attentional mechanisms. For instance, In Endsley's SA model (Endsley, 1995a), attention is implicitly identified in her information-processing mechanisms. Eriksen and Yeh (1985) explain the relation between SA and focused attention in their SA model by using the zoom lens metaphor as explained in section 2.1.1.

In the literature, attention is not a unitary mechanism. Rather, it includes different mechanisms including inhibition of task-irrelevant information, selective and divided attention (Cherry, 1953; Moray, 1959). Selective attention is the mental process that helps us select relevant information from a bulk of information. Because human processing capacity is limited, selective attention is needed to filter out irrelevant information while performing cognitive functions. Divided attention allows us to respond simultaneously to more than one cognitive task. Operationally, pilots selectively attend to the auditory information by ignoring task-irrelevant ones (inhibition). During all phases of flight, pilots are required to perform many tasks at the same time (divided attention). It can be said that pilots intensively use inhibition and divided attention mechanisms in flights. For that reason, these two could be taken as individual differences related to SA. Therefore, it is justifiable to investigate whether there is any effect of these individual differences on acquiring and maintaining SA during flight.

It is not possible to completely describe attention mechanisms without proposing an approach to attention control. As highlighted by Pashler, Johnston, and Ruthruff (2001), attention control can be invoked in two ways: bottom-up and top-down mechanisms. The former occurs after an unexpected and strong signal. For instance, a sudden and strong noise can capture one's attention involuntarily. The latter occurs voluntarily while accomplishing a high-level goal. Searching for a familiar object among unfamiliar objects could be an example of top-down mechanism. In this regard, the pilots mostly control their attention voluntarily to accomplish high-level goals such as directing their attention voluntarily more to the glideslope and localizer[9] than to the other displays while approaching. Needless to say, bottum-up attention control is also possible during flight but as frequently as unexpected events.

Attention theories are mostly classified in the literature as Bottleneck Theories and Capacity Model Theories. Bottleneck Theories are based on a filtering mechanism to filter out task-irrelevant information due to the limited-capacity of information processing (Broadbent, 1958; Moray, 1959; Treisman, 1960; Deutsch and Deutsch, 1963), Capacity Model Theories emphasize the processing capacity of the human cognitive system (Posner and Boies, 1971; Kahneman, 1973; Norman and Bobrow, 1975; Navon and Gopher, 1979).

Broadbent (1958) proposes a filtering mechanism for attention. According to the theory, the human perceptual system can attend to only one stimulus at a time. The limited capacity of our perceptual system requires a filtering mechanism which is metaphorically represented by a Y-shaped tube. Broadbent tries to explain the basic attention phenomena such as competing stimuli, intense stimulus, buffering, etc with falling balls into the upper legs of the Y-shaped tube. Balls represent the sensory information, the upper legs represent the multiple channel of information flow and the base leg of the tube represents the bottleneck in the perceptual system. In this theory, this bottleneck causes the unattended messages not to be perceived. On the other

---

[9] Glideslope and Localizer are integrated flight displays that show the vertical and lateral positions of aircraft within the Instrument Landing System (ILS) envelope.

hand, Moray's study (Moray, 1959) shows that the perceptual system processes the unattended message in certain situations. After Moray's findings, Treisman (1960) revises Broadbent's Filter Theory. In a dichotic listening task, Treisman directs the attended message to one ear and the unattended one to the other. In the middle of the experiment, the two messages were switched and Treisman finds that participants unexpectedly directed their attention to the words in the unattended message semantically related with the context. To this end, Treisman advocates that task-irrelevant information is not completely filtered-out but attenuated. Although both Broadbent and Treisman think filtering (selection) and attenuation mechanisms operate early in the perceptual process against the limitations in the perceptual system, Deutsch and Deutsch (1963) argue that filtering and discriminatory mechanisms operate late in the process. Deutsch and Deutsch highlight that all sensory inputs are processed in the same way upto the perceptual identification and the processing of the identified sensory inputs is through the level of arousal and discriminatory processes.

Recently, Lavie, Hirst, de Fockert, and Viding (2004) combine early and late selection approaches by proposing a dual filtering mechanism. They emphasize the discrimination of perceptual and cognitive load concepts. In the case of high perceptual load, early selection occurs and in the case of low perceptual load, late selection occurs.

Moray (1967) states that two filter theories (Broadbent, 1958; Treisman, 1969) do not apply to all tasks requiring divided or selective attention. The theories do not explain the performance of well-trained observers in divided attention tasks. Posner and Boies (1971) put forward an approach for attention based on limited capacity resource. In this approach, attention is composed of alertness, selection, and limited processing capacity. In their experiment, letters are displayed to the participants in a certain time-interval. After each letter display, an auditory warning signal is generated and the participants are asked to report whether the previously displayed letter is the same with the current one. Changing the onset of the warning signal and time-interval gives them an opportunity to discriminate the three components they proposed.

In support of capacity model theories, Kahneman (1973) points out that shared mental effort is not caused by the performer's will but by task demand. In the case of inadequacy of the available resources for performing a task, task performance decreases or task completion is not possible. For the tasks that do not require any performance, they may still need mental effort. According to Kahneman, this type of mental effort is supported by "spare capacity". Norman and Bobrow (1975) modify Kahneman's approach by studying limitations on central processing and sharing of common resources for attention. They argue that task performance is dependent on two factors: amount of available central processing resources and quality of data available. From SA's standpoint, both of these two factors are very important regarding SA-related information and SA-related resources.

Navon and Gopher (1979) propose the multiple resource theory which is based on the assumption that the human cognitive system has multiple-channel processing capability and each channel has its own resource. Later on, Wickens, Mountford, and Schreiner (1981) emphasize a multiple resource view for identification of factors supporting optimum performance and mental workload. They use four different tasks including "critical tracking", "number classification", "visual spatial line judgement", and "auditory running memory" tasks. In accordance with this approach, these four tasks are assumed to utilize different resources because of their different modalities (auditory and visual for inputs, discrete and analog for outputs), different stages of information processing (encoding, central processing and responding) and different "code" of central processing (spatial, verbal).

Apart from theorizing attention mechanisms and resources as explained above, multitasking provides researchers with an environment for exploring attention mechanisms experimentally. For instance, dichotic listening is a good representation of the audio environment that pilots experience in flights. Pilots are required to listen to various auditory stimuli coming from Air Traffic Controller (ATC) but only respond to the stimulus including their specific call signs because ATC speech begins with a call sign and the pilot directs his/her attention to ATC speech if his/her call sign is spoken; otherwise ATC speech becomes unattended. Different messages, ATC communications coming from different audio channels like VHF/UHF radio, intercom during flight are perceived and interpreted by the pilots. In a typical dichotic listening experiment, participants use a headphone for listening to two different sounds in different ears and attend to only one sound. Completing the task, the participants are asked about the content of the unattended sound. Dichotic listening is broadly used in measuring selective and divided attention capacity.

In addition to dichotic listening, there are some other tasks based on executive functioning mechanisms in the context of attention. Stroop and antisaccade tasks are among the most widely-used tasks that measure executive functioning including inhibition. It is important to note that inhibitory mechanisms are not unified. There are three types of inhibitory mechanisms in the literature as identified by Friedman and Miyake (2004): Prepotent response inhibition (suppressing habitual response), resistance to distracter interference (suppressing task-irrelevant information), and resistance to proactive interference (suppressing past memories). They find that prepotent response inhibition and resistance to distractor interference are related and resistance to proactive interference is clearly separate. Regarding the pilots' operational environment, it can be said that prepotent response inhibition and resistance to distractor interference are commonly used inhibitory mechanisms during flight simulation. For instance, checking only necessary set of flight displays is a good example of suppressing task-irrelevant flight information (i.e. resistance to distractor interference). Proactive interference could have a role when -let's say- an expert pilot tries to fly a new airplane having an unfamiliar cockpit design because the familiar one the pilot flew for several thousands of flight hours probably causes an interference during the flight. In the scope of this study, prepotent response inhibition is investigated in relation to SA.

The Stroop task is a color identification task. This task is used for measuring the strength of prepotent response inhibition. Color-words such as "RED", "BLUE", "YELLOW" are presented in different colors from the ones indicated by these words. In the critical condition, participants are asked to name the colors in which the words are printed. They experience delayed reaction times because interference occurs between participants' automatic reading abilities (prepotent response) and their effortful naming of the colors of the words. This effect is known as Stroop Effect (MacLeod, 1991). Reaction times represent the participant's strength of inhibition against a habitual response (tendency to read aloud the words instead of their colors) preventing the participants from performing the task.

In the antisaccade task, before the information is presented, a cue comes out on the opposite side of the display. Participants visually fixate a central stimulus and this central stimulus is replaced by an onset target that is displayed at non-central locations. The antisaccade task requires the inhibition of a habitual response (directing gaze towards the cue) and generating a correct saccade in the opposite direction. The antisaccade task requires participants to inhibit the tendency to look at an onset target and direct their gaze to the opposite hemifield. This task is used for measuring the prepotent response inhibition.

People can attend to a particular auditory stimulus while ignoring other distracter auditory stimuli. Cherry (1953) put forward the dichotic listening tasks as a way of investigating selective and divided attention in the auditory system. In the dichotic listening task, different messages are presented to each ear and participants are asked to attend to one of them (selective attention) or both of them (divided attention). In the former case, participants are asked to suppress unattended auditory stimulus (inhibition of task irrelevant information).

In this study, attention capacity as individual difference is taken as inhibition[10] and divided attention which are operationally crucial factors for the pilots in order to aviate, navigate, and communicate properly. Pilot's attention capacity can be expected to influence how fast and how deep SA is attained during flight. Parallel to this expectation, Gugerty (2011) posits that online SA scores could be used for assessing attention mechanisms in traffic domain. Do pilots with high inhibition strength better at acquiring and maintaining SA than the ones having low inhibition strength? Is there any significant effect of divided attention on SA status? These questions can be investigated in relation of SA with attention mechanisms.

It is possible to argue that expertise has a potential to extend the operator's cognitive capacities including the above-mentioned memory and attention capacities in the operational environment. It is possible that expertise helps operators overcome

---

[10] The term “inhibition“is used for the prepotent response inhibition throughout this study

the bottlenecks in memory and attention capacities. At least, we know for sure that expertise creates automaticity which reduces attentional demands in performing operational tasks. Expertise, to be explained in the next section, is therefore taken as another individual difference in the scope of this study.

2.2.3 Expertise

Some cognitive functions and structures underlying SA such as LTM, mental models, decision making, pattern matching and automaticity are related to expertise. In this regard, expertise is an important individual difference that can be investigated in the context of SA. Along with experiencing operational domain, most of the operational activities turn out to be automatized. Automaticity brings about a decrease in attentional demands which are normally required in order to perform these activities. Consequently, operators' performance in multi-tasking environment increases. Endsley (2006) emphasizes the roles of the mental models, pattern matching and automaticity in the context of the relation between SA and expertise.

In the light of how Ericsson and Kintsch (1995) explain expert performance, as detailed in the next paragraph, it can be said that expert operators establish mental models and patterns in their LTM. These cognitive structures related to the context are activated while doing operational activities. They match the patterns with environmental entities or cues. At the end of this process, the operator assesses the current situation. Situation assessment is very important in acquiring and maintaining SA. The following quotation from Endsley (2006) reveals the scope and the importance of mental models:

> "*A pilot develops not only a mental model of how the aircraft operates, including its many subsystems and its aerodynamic performance in the physical environment, but also a mental model of flight operations, including Air Traffic Control (ATC) procedures and expected behaviors associated with interacting with ATC and other pilots*"(p.638).

Expertise in aviation is investigated in many studies (Endsley, 2006; Burke *et al.*, 2004; Kasarskis *et al.*, 2001; Prince and Salas, 1998; Bellenkes *et al.*, 1997; Morrow and Altieri, 1992). Prince and Salas (1998) make a comparison among various pilot groups having different levels of expertise. They find that expert pilots spend much more time for pre-flight briefings and preparations. They also conclude that expert pilots focus on understanding situations more than novices. Considering the other studies cited above, expertise could be a significant predictor for the SA level the pilots acquire and maintain during flight. For that reason, expertise is taken as individual difference that is expected to predict a certain portion of variability in the SA level.

2.2.4 Interrelations among Individual Differences

Effects of WM, attention, and expertise on acquiring and maintaining SA, which form the scope of this study, can be well-grounded only if interrelations and interactions among these individual differences are studied very well. Regarding the relationship between WM and attention, Engle and Kane (2004) propose a controlled attention theory of WM. The theory is based on the fact that individuals with high WMC are better able to control and focus their attention than individuals with low WMC. Parallel to this theory, Colflesh and Conway (2007) argue that individuals with high WMC are able to zoom-in or zoom-out depending on WM task demands. Similarly, Bleckley, Durso, Crutchfield, Engle, and Khanna (2003) suggest that attentional allocation of low WMC individuals is not flexible. That is, individuals with low WMC are not capable of zooming in or out depending on WM task demands. On the contrary, high WMC individuals show flexible allocation. In case of performing a cognitive task requiring selective attention, the attentional focus should be narrowed (zooming-in) and only task relevant information should be accessed. Colflesh and Conway (2007) point out that a cognitive task requiring divided attention should be performed by expanding the attentional focus. High WMC individuals are able to not only narrow the attentional focus but also expand it better than low WMC individuals. These findings show a strong relationship between WM and attention mechanisms in the human cognitive system.

Inhibitory mechanisms are important for describing and measuring WMC. Conway, Cowan, and Bunting (2001) study the relationship between WMC and inhibition for task-irrelevant information by using a dichotic listening task. Brewin and Beaton (2002) state that WMC is positively correlated with suppression ability in the tasks requiring resistance to peripheral visual cues. Similarly, Unsworth, Schrock, and Engle (2004) investigate the relationship between WMC and inhibition of a habitual response.

In the WM literature, there are some attempts to put WM mechanisms into a model from the perspective of attentional and individual cognitive differences (Kane and Engle, 2000; Conway and Engle, 1996; Ericsson and Kintsch, 1995). These studies show a strong relationship between WM and other individual differences, inhibition and expertise, which are in the scope of this study. Regarding divided attention, there is a study showing that dual-task performance and WM functions are dissociable (Asloun, Soury, Couillet, Giroire, Joseph, Mazaux, and Azouvi, 2008). An opposing view is that performance of high WMC individuals in a dichotic listening task is significantly better than low WMC individuals as proposed by Colflesh and Conway (2007). Kane and Engle (2000) report similar findings about the relationship between WMC and divided attention as well.

There are some other studies on the relationship between WMC and important individual differences such as general fluid intelligence, inhibitory mechanisms, processing efficiency. Kane and Engle (2002) argue that WMC is a general measure of cognitive abilities. They propose that WMC and general fluid intelligence are correlated. The basic idea of this proposition is that WMC measures capture domain-general information-processing capabilities. Domain knowledge and expertise are important for improving WMC and they can compensate negative effects of the low WMC.

Cowan (1995) considers WM as a part of LTM. Representations in WM are taken as a subset of the representations in LTM. He states that WM is organized in two parts: Activated representations in LTM and focus of attention. The focus of attention represents limited capacity of WM. Engle *et al.* (1999) takes the activated representations in Cowan (1995)'s view as STM. It is known that high WMC individuals produce less Stroop interference than low WMC individuals (Kane and Engle, 2003). This finding reveals the interrelation between WMC and inhibition as one of the attention mechanisms. Parallel to this finding, Engle, Tuholski, Laughlin, and Conway (1999) describe WM as STM and controlled attention. Inhibitory mechanisms limit the information to enter WM. Conway and Engle's findings (Conway and Engle, 1994) put forward a general attention system which controls the inhibitory mechanisms for task-irrelevant information and the activation mechanisms for task-relevant information.

Ericsson and Kintsch (1995) propose a theory that emphasizes the relationship between Long-term Working Memory (LT-WM)[11] and expertise. The theory which is called "skilled memory theory" describes how participants are able to attain memory skills and to develop LTM with reasonable performance compared to Short-Term Memory (STM). Ericsson and Kintsch Model (Ericsson and Kintsch, 1995) is derived from this theory which is based on some assumptions. The relevant assumption related to this study is that experts have a capability of using their existing knowledge and expertise in semantic memory to store information during skilled performance of a given task. Ericsson and Kintsch (1995) modify the skilled memory theory as the LT-WM Theory. They argue that traditional models of human memory cannot explain expanded WMC of experts. They also question the role of LTM in WM mechanisms. The role of LTM in Ericsson and Kintsch WM Model provides a way to explain how participants store and access domain-specific information. Ericsson and Kintsch (1995) show that skilled performers can expand STM capacity through domain-specific knowledge and control processes facilitating them to efficiently encode and retrieve information from LTM. This finding emphasizes the importance of expertise on WM mechanisms.

Higher WMC is also claimed to bring about better ability to block or suppress task-irrelevant information. Conway, Tuholski, Shisler, and Engle (1999) observe that

---

[11] LT-WM is defined as the activated portion of LTM mediated by retrieval schema in which information is encoded and stored in LTM.

only high WMC participants show a negative priming effect which is demonstrated by a delay in performing a task involving a suppressed item in the previous task. Negative priming (slow response to previously ignored stimulus) could be used to investigate the processes involved in attention mechanisms. For instance, selective attention is performed through inhibition of task-irrelevant information and negative priming reveals this inhibition mechanism.

The findings in the related studies show that WMC, inhibition, divided attention, and expertise as individual differences have significant interrelations and interactions although there are some conflicting studies in the literature on the relationship between WM and divided attention. This dissertation, which is mainly a behavioral study, is conducted for the investigation of the possible relationship between SA and individual differences, however, some complementary methods including eyetracking could also be used in exploring an applied concept of SA to relate it to cognition. The next section introduces the use of eyetracking as a method for investigation of this applied concept.

2.3 Eyetracking as a method for studying SA-related cognition

It is known that human eyes span 200°of the visual area but the focal point (fovea) in the macula region of retina spans only 2°. Saccadic movements of the eye at a rate of 3-4 times/sec are important for processing visual information. Scanpaths that belong to the saccadic eye movements, gaze and fixations have been investigated by researchers of various disiplines through eyetracker systems. Such studies provide insights into the human visual processing system (Richardson, Dale, and Spivey, 2007).

Observation of eye movements could be a suitable method to study the bridge between perception and higher-level cognitive processes because eye movements are sensitive to some cognitive processes such as decision making and memory. It is also useful for probing into cognitive and perceptual processes having hidden connections with the external world or existing in the subconscious mind. For instance, Johansson, Holsanova, and Holmqvist (2006) state that eye movements dependent on mental imagery occur in accordance with the context. Some researchers investigate the relationship between attention, memory, and perception. In one of these studies, Hollingworth and Luck (2009) explore the relationship between attention and visual WM in a visual search task by using an eyetracker system. They find that attention distribution caused by a distractor having the same properties with the target affects saccade targeting mechanisms. This supports the idea of visual WM effects on directing attention.

Researchers who conduct basic and applied studies in aviation psychology commonly use the method of eyetracking. Diez, Boehm-Davis, Holt, Pinney, Hansberger, and Schoppek (2001) examine the effects of more automatized and more complex avionics systems on SA through an eyetracking study. They conducted a study in a PC-based flight simulator with 5 participants performing a task based on free and cued recall of readings on avionics displays. They conclude that frequency of saccades is more decisive in performing this task than fixation duration. Additionally, they find that fixations for monitoring are different in nature than fixations for acquiring information. Nevertheless, they could not explain some between-pilot variance due to two factors: They use a low-fidelity flight simulator and only a small number of participants.

In another study on how pilot's monitoring strategies change in response to highly automated cockpits, a full flight simulator is used (Sarter, Mumaw, and Wickens, 2007). Specifically, it is found through the evaluation of behavioral and eyetracking data that flight mode changes in automated systems could not be fully detected even by the expert pilots.

Some studies using eyetracking data in flight focused on the distinction of expert-novice pilots. For instance, Kasarskis, Stehwien, Hickox, Aretz, and Wickens (2001) conduct an eyetracking study in a PC-based simulator with 16 pilots. They reveal that expert pilots have more identifiable scan behaviors, more fixation durations and more Areas of Interest than novices. Some other eyetracking studies focus on cognitive parameters such as mental workload. Di Nocera, Camilli, and Terenzi (2007) report that eye movements in flight phases requiring high mental workload are different than the ones in flight phases requiring relatively low mental workload. They also point out that spatial distribution of eye movements caused by the mental workload gives some important clues about how attention is directed during the flight phases.

In novice-expert pilot distinction, Kim, Palmisano, Ash, and Allison (2010) investigate student and certified pilots' eye movements in a simulated aircraft landing task under day and night conditions. They suggest that pilots utilize the visual cues which are supplementary in the landing task through peripheral vision without directly steering at them. In another study by Ottati, Hickox, and Richter (1999), pilots' eye movement data are used to investigate the usability of an electronic map display. Novice and expert pilots' eye movements are recorded while they fly in VFR (Visual Flight Rules) conditions[12] which require the pilots to significantly look out the window. They find that novice pilots produce greater fixation durations than expert pilots while flying in VFR conditions. Parallel to this finding, Bellenkes, Wickens, and Kramer (1997) also report shorter fixations in expert pilots while they investigate the visual scanning strategies.

As can be seen in the literature, eyetracking studies in the aviation domain are richer in terms of cognitive aspects than the current behavioral SA studies. On the other hand, eyetracking studies are mostly performed in PC-based flight simulators and with a limited number of participants due to the fact that eyetracking devices have some operational constraints. This reduces the ecological validity of those studies.

We know that some scan patterns are taught to the pilots in their regular pilot training programs. Can SA-derived tasks uncover the deviations in scan patterns of pilots based on the individual differences? Apart from these scan patterns, do expert pilots create their own scan strategies which have different characteristics from what is taught to them to some extent? Eye movement data that may be collected from a small but focused and expert group of pilots can be qualitatively analyzed for obtaining insights about the relationship between SA and individual differences.

---

[12] Meteorological conditions allowing pilots to visually track where the aircraft flies

CHAPTER 3

METHODS

SA measurements, cognitive capacity tests and an eyetracking experiment were designed for exploring and validating working memory and attention mechanisms underlying SA. In this chapter, the design of the experiments will be explained. Experiment 1 consisted of SA measurements and cognitive capacity tests and Experiment 2 consisted of four visual tasks and cognitive capacity tests.

3.1 Experiment 1

3.1.1 Participants

The sample of this study was pilots flying fixed-wing aircraft. Participation was voluntary. All participants were native speakers of Turkish. The total number of participants for SA measurement and cognitive capacity tests was 36 (35 male, 1 female)[1]. Their mean age was 34.5, their average total flight hours was 2096[2] and their average simulator usage taken as level of expertise was 46.8. Table A.1 shows the biographical data of the participants in Experiment 1.

---

[1] The six participants flew on Cessna-172 while thirty others on Cessna-182. Cessna-172 and Cessna-182 Flight Simulators were quite similar in terms of their avionics displays. Because SMEs evaluated that effects of these two flight simulators on the participants would be similar, those six cases were added to the analysis.

[2] Integrated Commercial Pilot Training is completed after approximately 180 flight hours. The participants can be regarded as relatively experienced.

3.1.2 Apparatus

This section includes the descriptions of materials, software or systems used in Experiment 1.

The SA measurement part was conducted using a high fidelity flight simulator of Cessna-182, ALX FSTD manufactured by ALSIM. Figure 3.1 depicts the external view of ALX FSTD and Figure 3.2 shows the cockpit's displays including Air Speed Indicator (ASI), Altimeter (ALT), Radio Magnetic Indicator (RMI), Vertical Speed Indicator (VSI), Clock, Omni Bearing Indicator, Horizontal Situation Indicator (HSI), Attitude Indicator (AI), and other knobs, switches etc. The simulator has 208°panaromic field and supports the full range of weather scenarios. The instructor used software to control the flight simulation (freezing the simulator at certain points, adding weather conditions, adding malfunctions, etc). The computer in which the instructor software ran has a touch screen interface complemented by conventional mouse/keyboard systems.

Figure 3.1: ALX FSTD Flight Simulator (www.alsim.com)

An Ink-jet printer was used to print the flight route flown throughout the experiment. In order to calculate the response times (RTs) to the online SA queries[3], a video recorder was utilized. For the offline SA queries[4], the participants used a sheet of

---

[3] Online SA queries are the queries administered while the simulation runs.

[4] Offline SA queries are the queries administered when the simulation is frozen.

paper to answer them.

Figure 3.2: Cockpit View (adapted from ALSIM FSTD Flight Manual)

Figure 3.3: ALX FSTD Flight Simulator Setup

The cognitive capacity tests were controlled by using E-prime Version 2.0 software. E-prime ran on a notebook having 15" high resolution color LCD screen. The visual display was presented by this screen. The cognitive capacity tests were conducted in an office environment.

3.1.3 Design

This section explains the design of the experiment in two parts. In the first part, the design of the SA measurement test and in the second part, the design of the cognitive capacity tests will be explained.

3.1.3.1 SA Measurement

The flight scenario was designed in two flight simulators (Evans and Sutherland's UH-1 Flight Simulator and Frasca's Cessna-172 Flight Simulator) to create a realistic operational environment during SA measurement. Subject Matter Experts (SMEs) reported that using two different simulators did not affect the validity of the scenario because the scenario was developed both for fixed-wing and rotary-wing aircraft. The scenario was designed by adding a sequence of rare aviation-related events to the scenario to generate novel situations for the participants.

Specific events in the flight scenario were developed in conjunction with SMEs. The outline of the flight scenario was developed and then it was detailed by the SMEs. Four SMEs[5] were used for assistance in standardizing scenario events to participants' operating procedures. Changes in environmental conditions throughout the flight route were realistically added to the flight scenario.

Scenario duration, workload on the participants during the flight scenario, specific events added to the scenario, replicability, and ecological validity are the main factors for evaluating a flight scenario in terms of the SA measurement. The following subparagraphs explicate these factors in detail.

Scenario Duration **-** Short durations do not create a feeling of real flight and long durations may cause significant decrease in cognitive abilities. Optimum scenario

---

[5] They were two instructor pilots for UH-1 Flight Simulator and two instructor pilots for Cessna-172 Flight Simulator.

duration is long enough to create a realistic environment as in real flights, and short enough to have cognitive abilities (attention and working memory capacities) at limits. In terms of our research objectives, 60-75 minutes were seen optimum after the pilot study. In the workload survey administered after the experiment, it was found that the participants felt time pressure at a rate of 3.55 out of 6 which can be interpreted as moderate to high workload. As Hendy (1995) stated, any increase in time pressure reduces the resource capacity available to perform the cognitive tasks. Since the flight scenario was designed to enforce participants to use their cognitive abilities at the limit, the subjective evaluation of time pressure exposed on the participants are very informative about whether the flight scenario served this purpose as designed.

Workload - A typical flight consists of take-off, climb, cruise, descent, approach, and landing phases. In normal conditions (without novel events inserted to the scenario), take-off and landing phases are cognitively the most demanding phases of the flight. Due to the added events, this workload distribution was altered. At the beginning of the scenario, a relatively low workload, a routine segment (approximately 10 minutes long) was planned in order to allow participants to relax so that their behavior would more closely approximate actual flight behavior. Online (SPAM) and offline (SAGAT) SA queries[6] were planned to be administered after the end of this low workload segment. Online queries were scored by recording RTs for the correct (i.e accurate) responses and offline queries were scored by recording accuracy. In order to cover SA's WM and attention constructs, two types of SA queries were simultaneously applied in SA measurement as explained in section 2.1.2. Original SPAM measures remove workload component by taking accept/reject decision from the participants for the queries. In high workload conditions, SPAM gives participants an option to reject to answer the query. In the application, workload was a component that was desired to be taken into consideration in the scope of individual differences-SA relation. For online SA queries, no reject option was given to the participants as a modification to SPAM. Instead, segmentation on the flight scenario was made in terms of workload and workload was regarded as informative for the analyses on acquiring and maintaining SA because participants' individual differences can change their workload positively or negatively. After each flight, a workload survey

---

[6] Written consent was taken for using SAGAT and SPAM via email correspondence with their owners.

proposed by Hart *et al.* (1984) was applied. In the workload survey, overall workload rating in average was 4.75 out of 6. Additionally, average workload ratings that belong to each flight phase (take-off, climb, cruise, holding, approach and landing) showed that workload increases as time passes and has a peak while directing to an unknown point[7]. Workload remained more or less steady after this unknown point. Having decided to terminate the flight, the instructor directs the participant to the unknown point which has no navigation aid[8]. It means that the participant should find a solution to reach this point. After reaching the unknown point, Radio Magnetic Indicator (RMI) compass, Airspeed Indicator (ASI), and Vertical Speed Indicator (VSI) is frozen one by one. Accordingly, workload remains high until the aircraft lands.

Specific Events Added - The flight simulator was configured to set aircraft and environmental conditions (icing, rain, turbulence, crosswind[9] , low visibility, low ceiling[10]) and to generate planned malfunctions in VSI, RMI/HSI Compass, and ASI. These events create novel situations and change the psychological state of the participants in terms of stress, fatigue, frustration and workload as described above. The workload survey showed that participants felt stress at a rate of 3.25 out of 6 and fatigue at a rate of 3.8 out of 6[11].

Meteorological changes during the flight may strongly influence the pilot's current flight route or procedures s/he follows. Icing conditions may result in flight termination depending on the aircraft's anti-icing systems. The visibility can be restricted by rain and low visibility may threaten a safe landing. Turbulence makes the flight controls not easily controllable. This creates another restriction for the pilots in terms of maneuverability and speed. Crosswind is a frustrating factor for the pilots in setting the aircraft's heading[12] right. Low ceiling may make the landing impossible and enforce the pilots to follow a missed approach procedure[13] while landing. This is supported by the fact that the participants felt relatively high workload (4.2 out of 6)

---

[7] A holding point which has no navigation aid It is unknown because avionic displays do not show it.

[8] Any form of device that guides the pilot and her aircraft from one area to another

[9] A wind passing from the right or left side of the aircraft

[10] The height above the ground of the base of the lowest layer of clouds

[11] Rating scale (0-6) used in the workload survey is assumed to be linear. The scale score "0" represents very low and the scale score "6" represents very high.

[12] The direction in which the longitudinal axis of the aircraft is pointing according to a compass

[13] Flight procedures prescribed when an aircraft fails to land after completing an approach

during approaching and following the missed approach procedure.

Pilots should precisely know the aircraft's speed, heading, and altitude at any time in flight. Any loss or misperception of these data may threaten the flight safety. VSI indicates aircraft's vertical speed and this indicator is especially used while climbing and descending. Pilots are required to use certain climb and descent rates due to the aircraft performance limits and regulations. VSI Failure enforces the pilots to calculate their climb/descent rates by using changes in altitudes and time in their minds. The RMI compass precisely indicates the aircraft's heading. The RMI compass failure requires the use of a magnetic compass. The magnetic compass provides reliable heading information only during straight and level flights[14] with constant speed. Therefore, this failure brings about difficulty in getting reliable heading information especially while turning. Again, the pilots are supposed to calculate the turn duration in their minds (i.e. timed turn) by using the difference between the current heading and desired heading. Parallel to our expectations, difficulty level of the flight was evaluated by the participants as high (5.05 out of 6 in average). In addition to this finding, the participants felt mental and physical demand at scores of 4.7 and 4.75 out of 6 in average respectively.

One of the most important parts of the scenario was the part between the decision point for the flight termination and the unknown point which is designed as an unexpected and novel situation. The unknown point has no navigation aid that can help the participants to navigate to this point. There are two possibilities for the participants to directly navigate to the point. In the first one, they can request a radar vector which means that Air Traffic Controller (ATC) gives clear heading information to the participant to fly to the point. In the second one, they can use the navigation aids of the airport for making a DME arc[15] to the point. In high workload condition, it is difficult for the participants to find one of these two solutions. Endsley (1997) posed that working memory and limited attention are both important factors for someone who has to deal with novel situations to acquire SA.

---

[14] Coordinated flight with constant altitude and heading

[15] A technique that allows a pilot to fly a curved course a fixed distance from a given point.

***Replicability* -** The flight scenario may have many dynamic characteristics so that the participants' options can significantly change the flight profile[16]. The replicability allows us to have a controlled simulation environment and then to accurately compare SA-related outcomes. The scenario generation during simulation was automated on the instructor's side by preparing well-defined, unambiguous test instructions (see Appendix B) for the sake of replicability.

***Ecological Validity* -** Simulation fidelity is decisive in creating a similar psychological state as in real flights from the SA measurement's point of view. Both Evans and Sutherland's UH-1 Flight Simulator used for scenario generation and ALSIM's Cessna-182 and FRASCA's Cessna-172 Flight Simulators used in Experiment 1 have acceptable visual and functional realism. According to the workload survey, the participants felt workload, fatigue, and stress at a rate of 4.75, 3.8, 3.25 out of 6 in average respectively. A separate voice scenario or dedicated personnel could have been useful for ATC communication and Automated Terminal Information Services (ATIS)[17]. Although the instructor pilot has the role of ATC and ATIS unrealistically throughout the flight scenario, SMEs[18] advocated that having the instructor pilot play the multiple roles in the flight scenario has little or no effect on the SA measurements. The important point in this issue is to divide the participant's attention regardless of whether ATIS information comes from the radio or from the flight instructor. In both situations, the participants should allocate some of the attentional resources to understanding the ATIS information.

***Flight Scenario* -** A cognitively demanding flight scenario detailed in Appendix C was generated. The scenario begins with taking off from an airport (depicted as Take-off Airport in Figure 3.4) and flying to Waypoint#1 (depicted as Waypoint#1 in Figure 3.4). There is no novel situation between the Take-off Airport and Waypoint#1. It takes about 10-15 minutes. While navigating from Waypoint#1 to Waypoint#2, meteorological condition gets worse and the participant pilot decides to terminate the flight at Flight Termination Point depicted in Figure 3.4. ATC gives instructions to the participant for directing to Holding Point in Figure 3.4 for landing back on Take-

---

[16] A flight profile is the lateral and vertical pattern of the flight route including aircraft's angular position.

[17] ATIS is a continuous broadcast of recorded information such as weather information.

[18] Flight instructors who work in Cessna-182 flight simulator

off Airport. The holding point could not be displayed on the flight displays because it is not on the database. This enforces the participant to find a solution to reach to Holding Point. After reaching to Holding Point, the participant makes one racetrack-shaped turn at Holding Point and then starts to approach to Take-off Airport. During holding, displays showing the directional information turn off. When the aircraft is very close to Take-off Airport, the participant does not establish visual contact with the ground and goes around[19]. In the second approach, displays showing the airspeed and vertical speed freeze one by one. The participant lands the aircraft without using the frozen displays.

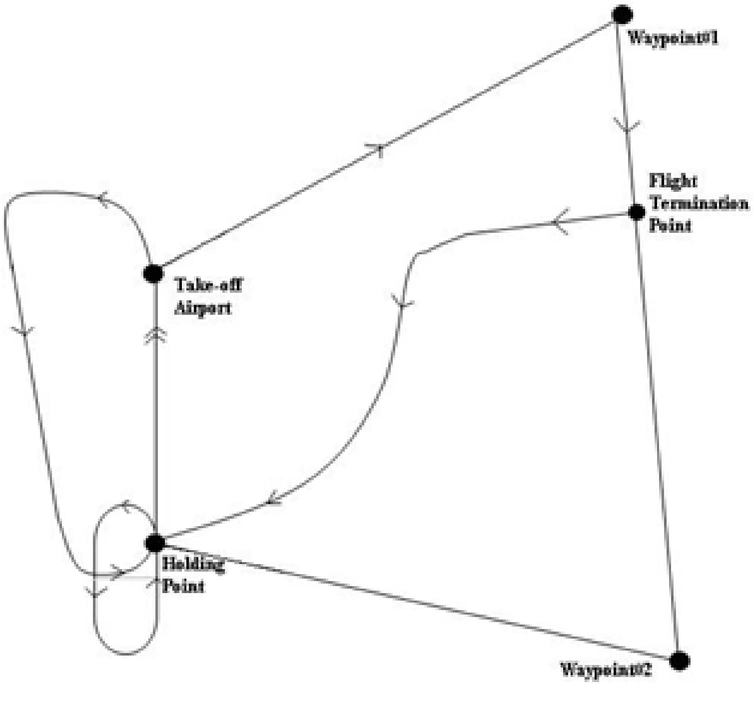


Figure 3.4: Flight Route in the Scenario

SA Queries - SAGAT and SPAM procedures are not an algorithmic recipe but only guidelines to SA researchers. In the study, a cognitive task analysis was applied at first. Having obtained SA-related tasks for a flight, information required for performing those tasks were determined and then that information was transformed into an SA query. For instance, waiting in a hold waypoint by making turns as a shape of racetrack is a task before starting to approach to landing. One piece of information

---

[19] The term "Go-around" is used in aviation for following missed approach procedure.

required for this task is heading information (angular direction of aircraft). During holding, the query "What is your heading?" could be administered. If the simulation runs at the time of administration, this query is called "online". If the simulation is frozen, the query is called "offline". For online SA queries, response accuracies and RTs are recorded. For offline queries, only response accuracies are recorded.

Endsley (1995a) suggested that SA queries should be classified as perception, comprehension and projection queries. Parallel to Endsley's three-level SA model, Endsley *et al.* (2000) proposed that the combined score could not be used for the analysis of perception, comprehension and projection queries individually. In this dissertation, SA queries are not developed in accordance with this classification. For that reason, a combined score is used as the summation of all offline query scores. For the online query scores, it is only possible to use a combined score after z-transformation because some questions were answered in few seconds while the others took tens of seconds. Averaging these questions without z-transformation would not be right, since the weight of the questions that take a longer time would be higher and averaging makes the valuable information lost.

The flight scenario was segmented into low workload and high workload phases. The scenario events were designed to create high workload in certain parts. Event timings in the scenario were refined in accordance with the workload surveys administered in the pilot study and they were validated by the ones administered in the Experiment 1. After identifying the SA requirements to complete the SA-related tasks in the flight scenario, SA queries were extracted from the SA requirements and classified as online or offline dependent on the research questions. As explained in the above paragraphs, SA measurement methodology allows researchers to study the SA's WM and attention constructs if it includes both online (see Appendix B) and offline queries (see Appendix D and Appendix E). Some queries may be both offline and online. Eight online queries were administered orally and thirteen offline queries were administered on a sheet of paper. They were multiple-choice type and administered in two predefined freeze points in which the flight simulation was stopped and participants could not see the frozen flight displays. Freeze points were selected on the flight scenario by considering the scenario phases in which the offline queries were meaningful because some information could only be requested at certain critical times in the scenario. Number of freeze points was restricted by the participant's expected level of disturbance because short durations between the freeze points increase the level of disturbance which negatively affects on accuracy of the offline SA measurements.

3.1.3.2 Cognitive Capacity Tests

Three cognitive capacity tests were selected for measuring participants' WM and attention capacities. Two of them existed in the literature and were not modified. The third one which is a choice reaction time task with dichotic listening was developed as explained below.

Automated Operation Span Task

Operation Span Task is used for measuring WMC. Unsworth *et al.* (2005) developed a practical version of the operation span task called "automated operation span task (AOSPAN)". In AOSPAN task, a letter appears on the screen and the participants are supposed to memorize the letter displayed. After that, a simple mathematical operation like "(2/2)-1" is displayed. Upon participant's intervention, the result of the operation with true/false boxes is displayed. Participants are required to select true/false boxes. After checking the outcome of the mathematical operation, another letter appears on the screen and the same steps are followed until the letter sequence including three to seven letters is finished.

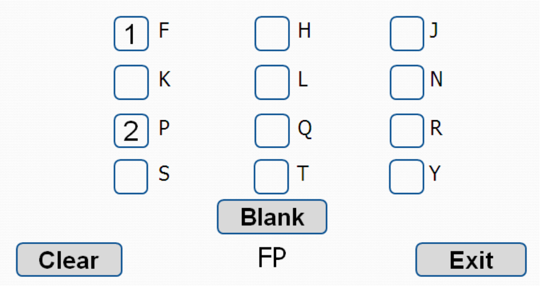


Figure 3.5: Recall Phase in AOSPAN

At the end, participants recall letters in the presented order (see Figure 3.5). For 80 different letter sequences having various lengths from three to seven, the procedure is repeated. After each recall, the computer provides feedback about the number of the letters correctly recalled in the current letter sequence. Validity of the results is attained by accepting only results having a minimum of 85% of mathematical operations correctly answered (Unsworth *et al.*, 2005). The dependent variable is OSPAN that is calculated as the sum of all perfectly recalled sets at the end of the experiment.

Stroop Task

Another test was one of the well-known cognitive tests in the literature which is Stroop Task (MacLeod, 1991). The Stroop task is used to measure strength of the prepotent response inhibition. This inhibitory mechanism, which can be described as inhibiting a habitual response which is conflicted with the task goal while performing a task, is very essential for working memory and attention processes. The colors the participants see in the experiment are KIRMIZI (red), MAVI (Blue), YESIL (Green), SARI (Yellow), SIYAH (Black). The language of the test is Turkish. The task is designed as participant-paced. The congruent ratio is 75% which maximizes the observed interference effect. Participants are expected to experience delayed RTs due to the interference between color naming and word reading. It is known that word reading is the prepotent response and participants are required to inhibit this prepotent response while color naming. It is possible to score strength of inhibition by using response times and error rates (Kane and Engle, 2003). In this study, strength of inhibition is scored by subtracting average RT in neutral case from average RT in incongruent words, where the color in which the word is written is different from its meaning. Error rates (subtracting the number of errors in neutral case from the number of errors in the incongruent case) are not used due to the low variability in the data.

Choice Reaction Time Task with Dichotic Listening

The last cognitive capacity test for the research purposes was developed by using a dual-task paradigm. In the dual-task paradigm, participants perform both primary and secondary tasks at the same time. The primary task is the choice RT task, in which more than one stimulus are presented. Participants generate a different response for each stimulus (Lysaght *et al.*, 1989).

The secondary task is dichotic listening which is a procedure where two different auditory signals (or speech) are presented to the participants, one to each ear. In the selective attention set-up of dichotic listening, participants are required to attend to only one of their ears. In the divided attention set-up, they are required to attend to both ears. A selective attention set-up of dichotic listening is used as the pilots deploy their attention only to the left sound which is Automated Terminal Information Service (ATIS) message[20]. Right sound is a cockpit communication recorded in a real flight.

Dichotic listening is generally performed by the pilots during their flights. Pilots are required to simultaneously interpret many cockpit communications and weather reports received from different communication channels. Using real ATIS broadcast and cockpit communications in dichotic listening task make the experimental set-up similar to the real flight environment. A one-minute stereo sound is prepared from cockpit communications recorded in a flight of a commercial airliner. The left sound includes ATIS message, which gives the pilot weather information in the vicinity of the reporting airport; and right sound includes ATC/cockpit communications during take-off in a commercial airliner's flight. Speakers of the right and left sounds are of opposite genders.

According to the test procedure, participants continuously make choices (pressing dedicated keys "F" and "J" with stickers showing the Turkish response words on them in response to the word "EVET", and "HAYIR" respectively) in the screen as the primary task while dichotic listening as the secondary one. The participants are asked to pay attention only to their left ear in dichotic listening. Before the experiment, three questions that are relevant to the left sound's content are presented to guide the participants in their listening effort. At the end of the task, these three questions are administered in order to be sure that the participants allocate sufficient attentional

---

[20] ATIS is a prerecorded broadcast message which includes meteorological information around an airport.

resources to understanding the left-ear sound. After answering the multiple-choice questions, the test ends. Mean RTs for correct responses in the primary task are measured for divided attention capacity.

3.1.4 Pilot Study for Experiment 1

In order to verify the feasibility of the experimental procedures, reliability and validity of the apparatus, a pilot study was developed and administered with the participation of 10 pilots[21]. After the pilot study, necessary adjustments to the experimental procedures and the experimental apparatus were made in accordance with the results of the pilot study.

After the first iteration of the flight scenario, it was realized that the scenario duration was significantly longer (50%) than the participants are used to encounter in flight simulators. This brought about fatigue, stress, and sometimes frustration more than expected. For the objectives of this study, 60-75 minutes of flight seemed to be optimum. On the other hand, the scenario duration is not the only factor that causes this overload. Participants, more or less, attached importance on their performances although they were informed on the fact that the conducted tests were not designed to test their personal performances. This issue was alleviated with frequent reminders both from the instructor pilot and the experimenter before the flight.

In Choice Reaction Time task with dichotic listening, only one pilot out of 10 was successful in meeting the criteria of correctly answering 8 out of 10 questions. ATC communications during take-off was used as the right sound and ATC communications during landing was used as the left sound. Regarding participants' WMC, it was observed that 10 questions were quite hard to be answered. Because two sounds had belonged to the same speaker, participants could not differentiate the sounds. Consequently, the sound set used was replaced such that speaker of the ATIS message

---

[21] Considering the pilot study of Experiment 1, the results of six participants who flew in Cessna-172 Flight Simulator in the pilot study were used in the analysis. SA measurements, AOSPAN and Stroop tasks were not modified for the last six participants but Choice Reaction Time Task with dichotic listening was significantly revised after the pilot study so the revised task was readministered to these six participants.

became female and the other became male. Additionally, the sound duration was reduced to one minute and the number of queries administered was reduced to 3. As well as increased applicability of the test, using cockpit communications and ATIS broadcast as left and right sounds respectively increased the ecological validity of the test.

During the pilot study, two different simulators were used: Evans and Sutherland's UH-1 Full Flight Simulator (FFS) and Frasca's Cessna-172 Flight Simulator. One was for fixed wing (Cessna-172) and the other was for rotary wing (UH-1) aircraft. Therefore, the "common" flight scenario had to include common events which are applicable for both simulators. Although common special events and meteorological conditions in the flight scenario were finally obtained, different flight routes for the simulators were used due to the fact that navigational databases (airports, airways, waypoints, etc) were not identical. SMEs[22] confirmed that these two flight routes were quite identical for the pilots in terms of using cognitive resources.

At the end of each flight, participants answered some survey questions proposed by Hart *et al.* (1984) about the workload required, self performance assessment, attention required, complexity of the scenario, time pressure, mental effort required, his/her motivation, his/her psychological mood at the end. By using this, evaluation and validation of the flight scenario was possible. At the end of this validation process, it was observed that 60-75 minutes of flight was optimal for the research objectives. The query set (see Appendix D, Appendix E), instructor's instructions (see Appendix B), participants' instructions (see Appendix F) were updated. It was found that the participants had difficulty in remembering the missed approach procedure when needed due to the fact that they performed a solo flight without a co-pilot. In real flights, missed approach procedure is briefed with the co-pilot. For that reason, a query for the missed approach procedure - which was not used in the analysis- was added to the query list (see Appendix B) and participants were allowed to read it from the approach chart[23]. Although pilot-copilot configuration is very common in flights, small aircraft

---

[22] Flight instructors in Turkish Army Aviation School (*tur.* Kara Havacılık Okulu) and Sindel Aviation Company

[23] Approach chart is a paper chart including necessary information for pilots to approach and land on the airport such as Missed Approach Procedure.

such as Cessna-172 and Cessna-182 can be flown without a co-pilot. The experiment was designed on the basis of solo flight because SMEs confirmed that ecological validity still remains for Cessna-172 and Cessna-182 Full Flight Simulators in solo flights.

It was observed that the pilots did not spend the required amount of time on pre-flight briefing session. This not only reduced the ecological validity (they spend much more time on pre-flight briefing for a real flight) but also prevented us to clearly observe the effects of experience on acquiring and maintaining SA. Before a real flight, a pilot fills a flight log including the flight routes, alternate airports in case of emergency, time and fuel calculations. The flight log helps pilots to be mentally ready for the flight. Consequently, it was decided to enforce the participants to fill the flight log prior to the flight simulation. Therefore, it was guaranteed that each participant would make the same effort for the flight.

Before the pilot testing, the ideal conditions of time separation in two types of tests (SA measures and cognitive capacity tests) were not clear. However, it was found out in the pilot study that long time intervals between these tests can be a problem. It was concluded that only instant measures for individual differences just before or after the SA measures can help us to investigate possible effects of individual differences on acquiring and maintaining SA. Although WM and attention are two main research areas for the SA study, there are many factors like stress, sleep, fatigue that are out of scope for this study. If SA measures and cognitive capacity tests are conducted within a long time interval, the status of the factors may be different in the SA measures and the cognitive capacity tests. In that case, it is not clear whether SA variations occur in response to individual differences or in response to variations in stress, sleep or fatigue conditions. For practical reasons, it was decided that participants take the cognitive capacity tests just after the SA measures with a break in between[24].

---

[24] As an exception, six participants repeated the Choice Reaction Time Task with dichotic listening after ten months from the SA measurement.

3.1.5 Procedure

The below paragraphs explain the test procedures for the SA measurement and cognitive capacity tests.

3.1.5.1 SA Measurement

When participants arrived at the simulation facility, the instructor introduced the flight to them. This introduction covered the aim of the flight, the departure airport, the landing airport and the waypoints inbetween. The experimenter gave the test instructions (see Appendix F) to the participants and asked the participants for signing the Informed Consent Form (see Appendix G). The entire flight from take-off to landing was recorded by a video recorder.

After the participants prepared the flight simulator and received clearance from the instructor for starting the simulation, Instructor as ATC issued a special weather observation to the participants in the vicinity of the airport indicating that the weather would deteriorate. The operational implication was that the current aircraft simulated by ALX FSTD could not fly under the icing conditions. This meant the participants are required to quickly return for landing. At the beginning of the scenario, a relatively low workload segment (1.9 out 6 in average) was planned in order to allow the participants to relax so that their behavior would more closely approximate actual flight behavior. Most of the test measures were planned to be taken after this low workload segment.

In two predefined freeze points, simulation was frozen and the participants turned their directions in such a way that they could not see the information displays. Offline queries were administered on a sheet of paper and the participants answered the queries by using the information that remained in their working memory. At the end of the offline administration, the simulator ran from where it was frozen. For the online queries, a video recorder was used for recording the query-answer pairs and then

RTs for each pair was calculated.

At the end of the flight, the participants answered the written queries in workload survey (see Appendix H) and finally, the debriefing sheet (see Appendix I) was given to the participants.

3.1.5.2 Cognitive Capacity Tests

The cognitive capacity tests were conducted in an office environment by using a notebook computer after enabling participants to relax approximately for an hour to remove fatigue and stress experienced during the flight simulation. The cognitive capacity tests were conducted in the following order: Stroop Task, Choice Reaction Time Task with dichotic listening and AOSPAN.

Since AOSPAN is a completely automated test, there was no experimenter intervention. AOSPAN took approximately 20-25 minutes in average. In the Stroop task, a checklist was used for the answers and the wrong answers were removed from the analysis. The Stroop task took about 5 minutes in average. The Choice reaction time task with dichotic listening was also a fully automated test which took about 5 minutes in average. The results of all three tests were recorded by E-prime Version 2.0 in the notebook.

3.1.6 Hypotheses

Experiment 1 tests the following hypotheses:

H_1: WMC is positive predictor of SA status.

H_2: Strength of inhibition is a positive predictor of SA status.

H_3: Divided attention capacity is a positive predictor of SA status.

H_4: Expertise is a positive predictor of SA status.

H_5: Offline SA score is appropriate for measuring WM component of SA.

H_6: Online SA score is proper for measuring attentional component of SA.

3.2 Experiment 2

3.2.1 Participants

The total number of participants was 10. The research sample was pilots who fly a fixed-wing aircraft. Participation was voluntary. Average age was 39.6 and average flight hours were 3472. Table A.2 shows the biographical data of the participants in Experiment 2. SMEs pointed out the importance of flight hours spent in a certain flight simulator/aircraft where the measurement is done rather than total flight hours. Consequently, number of flight hours flown in the Flight Simulators (Cessna-172 and Cessna-182) of Experiment 1 was taken as level of expertise.

3.2.2 Apparatus

The eyetracking experiment was conducted at the Human-Computer Interaction Research and Application Laboratory at Middle East Technical University. Tobii 120 Eye Tracker depicted in Figure 3.6 and Tobii Studio Version 2.2.8 software was used for analysing the eyetracking data.

Figure 3.6: Tobii 120 Eyetracker in HCI Lab of METU

3.2.3 Design

Before designing Experiment 2, there were several attempts to use a mobile eyetracker in the Cessna-172 Flight Simulator. It was not possible to use it due to some technical and organizational difficulties and a test set-up based on PC-based Eyetracker System (Tobii 120) was designed. Participants were required to perform the visual tasks while watching the associated videos. During video displays, PC-based Eyetracker System recorded the participant's eye movement. Some calibrated portions of eye movements recorded with a mobile eyetracker in Cessna-172 Flight Simulator were qualitatively analyzed to extract actual areas of interest and scan patterns. After this analysis, visual tasks listed below were constructed from the SA-related flight scenario used in Experiment 1.

There are four visual tasks on recorded videos (approx. 2 minutes long):

- Visual Task-1 Estimating the aircraft's position on the flight track

While watching a video of the flight displays in a simulated flight, participant estimates and marks the aircraft's position on the flight track which is given on a sheet of paper after the video display.

- Visual Task-2 Estimating wind direction and magnitude

While watching a video of the flight displays recorded in a simulated flight, participant estimates wind direction and magnitude. Four flight parameters (heading, course, ground speed, airspeed) are assummed to be used in estimating wind direction and magnitude.

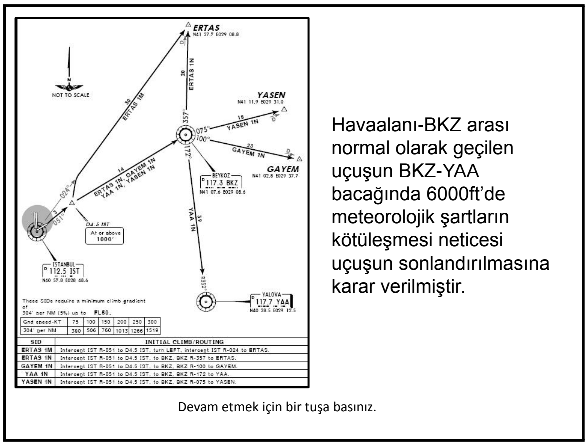
- Visual Task-3 Looking for abnormal values or events

While watching a video of the flight displays recorded in a simulated flight, the participant looks for abnormal values and events. The video includes failure in RMI and HSI indicating heading information, failure in VSI indicating vertical speed.

- Visual Task-4 Judging level of expertise

While watching a video of the flight displays recorded in a simulated flight, the participant tries to judge the pilot's level of expertise by evaluating trends in the readings of the flight displays. The initial expectation before the experiment is that small fixation durations (focusing on the trends instead of the numbers) would be observed and stability of airspeed, vertical speed, altimeter, and heading during turning and maneuvering would be important while judging the expertise level of a pilot.

In the experiment, all the events occuring in the flight scenario of Experiment 1 were presented to participants as shown in Figure 3.7.



Havaalanı-BKZ arası normal olarak geçilen uçuşun BKZ-YAA bacağında 6000ft'de meteorolojik şartların kötüleşmesi neticesi uçuşun sonlandırılmasına karar verilmiştir.

Devam etmek için bir tuşa basınız.

Figure 3.7: A Sample Screen giving Scenario Details

Having given the scenario details to participants as in Figure 3.7, cockpit displays were introduced as in Figure 3.8.

The flight track was displayed and in the next screen, participant was expected to perform visual task-1 while watching a related video of cockpit displays recorded in the simulated flight. The participants were expected to find out the correct answer by acquiring necessary information from the readings on the cockpit displays. The same procedure proceeds for the other three visual tasks. Slide transitions are upon participant's request. There is no backward transition in slide presentation.

Hava Hızı Göstergesi (Airspeed Indicator), hava hızını knots cinsinden gösterir. Gösterge skalası, 10 knots'lık artışlarla 40 knots'dan 210 knots'a kadar olan hızları içerir.

Dikey Hız Göstergesi (Vertical Speed Indicator), sıfır yatay referansın üstünde ve altında derecelenmiş skalaya sahiptir. 100 ft'lik artışlarla 0 ft/dakika'dan 20,000 ft/dk ve -20,000 ft/dk'ya kadar olan hızları gösterir.

Devam etmek için bir tuşa basınız.

Figure 3.8: Screen introducing Cockpit Displays

Figure 3.9 depicts the Areas Of Interest (AOIs) designed for the visual tasks. There are totally 13 AOIs but only six of them are of utmost importance for navigation. For that reason, Airspeed Indicator (ASI), Attitude Indicator[25] (AI), Altimeter (ALT), Radio Magnetic Indicator (RMI), Horizontal Situation Indicator (HSI) and Vertical Speed Indicator (VSI) are used in the qualitative analysis.

Figure 3.9: Areas of Interest

Participants entered the results of the visual tasks to a sheet of paper. An informal interview was performed about pilot's scanning strategies but no record was taken. Six out of ten participants reported a scan strategy called T-scan that is depicted in Figure 3.10. Three out of ten participants reported another scan strategy called selective radial scan that is depicted in Figure 3.10 and only one participant reported circular scan strategy depicted in Figure 3.10[26]. The scan strategies reported by the participants are well-known strategies in aviation that can be reached through the training manuals of some aviation schools such as Selkirk College (Selkirk, 2009). The following quote from this manual shows the importance of the scan strategies in favor

---

[25] A flight display which indicates the aircraft's rotation about the lateral axis, and about the longitudinal axis

[26] It is known that these scan strategies are taught in pilot training programs. Participants did not report a different scan strategy created by themselves.

of selective radial scan.

> "*Controlling an airplane by instruments alone requires a scan. The recommended procedure is called selective radial scan. It is covered in the Transport Canada Flight Training Manual.. All this was covered in the first year of the Professional Pilot Program.*"(p.9).

Some aircraft manufacturers locate ASI, AI, ALT, and HSI for building a T-shape and pilots are expected to primarily scan those flight displays in a T-shaped form. Selective radial scan is based on primarily gazing at AI and selectively at other peripheral displays. Circular scan is a scan strategy which is based on scanning the flight displays in a circular form. Currently, it is not known whether expertise results in perfection in deploying these scan strategies or results in deviations from the scan strategies or facilitates pilots to learn how to interchangeably use these scan strategies even in a single task. Important outcomes of the interviews are listed below:

- T-scan including Airspeed Indicator (ASI), Attitude Indicator (AI), Altimeter (ALT), Horizontal Situation Indicator (HSI) is the best scanning strategy among the others dependent on the installation configuration of the displays.
- AI is the central display in the scan path.
- Each fixation should not exceed 2-3 seconds.
- Each scan begins with AI.
- Fixations on AI should be about one third of total fixations.

3.2.4 Procedure

When the participant arrived at the HCI-lab in METU, the researcher introduced the eyetracking experiment to the participant. The researcher asked the participant for

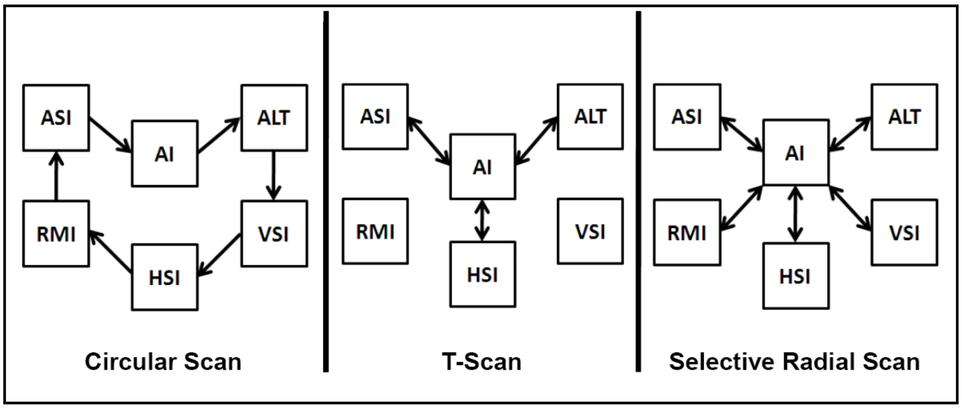


Figure 3.10: Three Scan Strategies

signing the Informed Consent Form (see Appendix G). Test instructions were embedded in the screens.

The eyetracker system in Figure 3.6 recorded the eye movements of the participant. The experiment took about 20 minutes. Average time to first fixation, average fixation count, and average fixation duration were calculated for the analysis. Average time to first fixation is expected to indicate the importance of AOIs from participants'view. Average fixation count and average fixation duration are used to verify that participants follow the rules listed in the above section. At the end of the experiment, Stroop Task, Choice Reaction Time Task, and AOSPAN were administered to participants and then the debriefing sheet (see Appendix I) was given.

This experiment is an exploratory effort to test some hypotheses about the relation between individual differences and eye movement in aviation at the end of the study. The SA status and eyetracking could be linked through individual differences including especially expertise. For gaining hypotheses, participants are grouped in terms of their individual differences. Mean values are used to create two groups for each individual difference. The following section includes the hypotheses tested in Experiment 2.

3.2.5 Hypotheses

Experiment 2 tests the following hypotheses:

***H_7*:** More Expert Pilots (MEs) significantly differ from Less Expert Pilots (LEs) in terms of fixation count produced while performing the four visual tasks.

***H_8*:** MEs significantly differ from LEs in terms of fixation duration produced while performing the four visual tasks.

***H_9*:** High WMC Pilots (HWMCs) significantly differ from Low WMC Pilots (LWMCs) in terms of fixation count produced while performing the four visual tasks.

***H_{10}*:** HWMCs significantly differ from LWMCs in terms of fixation duration produced while performing the four visual tasks.

***H_{11}*:** Pilots with high inhibition strength (HISs) significantly differ from pilots with low inhibition strength (LISs) in terms of fixation count produced while performing the four visual tasks.

***H_{12}*:** HISs significantly differ from LISs in terms of fixation duration produced while performing the four visual tasks.

***H_{13}*:** Pilots with high divided attention capacity (HDCs) significantly differ from pilots with low divided attention capacity (LDCs) in terms of fixation count produced while performing the four visual tasks.

***H_{14}*:** HDCs significantly differ from LDCs in terms of fixation duration produced while performing the four visual tasks.

CHAPTER 4

RESULTS AND DISCUSSION

This chapter includes the results of the study, analysis of the results, general discussion, limitations and avenues for future work. In Table 4.1, Experiment 1 and Experiment 2 are summarized in terms of research questions, participants, environments, and main findings.

Table 4.1: Brief Summary of the Experiments

	Experiment 1	**Experiment 2**
Research Question	Can SA status be predicted by WMC, inhibition, divided attention, or expertise?	Exploratory experiment to gain insight on the relationship between eye movement data and individual differences
Participants	36 Pilots	10 Expert Pilots
Environment	Cessna-172 and Cessna-182 Full Flight Simulators for SA Measurements and E-prime Software for measuring Individual Differences	T120 Tobii PC-based Eyetracker System
Main Finding	WMC and expertise contribute to WM constructs of SA measured through offline SA queries while inhibition, divided attention, and expertise contribute to attention constructs of SA measured through online SA queries	Some scan patterns were observed while performing the four visual tasks. Confirmatory study is needed with a larger group of participants

4.1 Results

Appendix J shows the results from the SA measurement in Experiment 1. The results (RTs) for the online queries that the participant could not answer or wrongly answered are shown as N/A (Durso and Dattel, 2004). The results of the cognitive capacity tests in Experiment 1 are also shown in Appendix J.

4.1.1 Experiment 1

Online SA scores are z-transformed and averaging is performed to obtain one online SA score. Offline SA scores are calculated by taking the sum of the results of 13 Offline queries. OSPAN scores in AOSPAN are used as WMC. No transformation is made in the results of the Stroop task and the Choice Reaction Time task with dichotic listening. Inhibition scores are calculated by subtracting average RTs of the neutral case in the Stroop from average RT of incongruent words. Divided attention scores are calculated by subtracting average RTs in Choice Reaction Time from average RTs in Choice Reaction Time with dichotic listening. For expertise scores, participants' number of flight hours spent in the flight simulator, where SA measurements were administered, were taken. Some predictor variables (divided attention and expertise) are moderately skewed such that they violate the assumption of normality (Table L.1). To a certain extent, skewness of RTs recorded for divided attention and skewness of expertise could be taken as normal because they cannot take the values less than zero (positive skewness). Square root transformations of these variables were computed and a regression analysis was conducted using the transformed scores. This procedure was not found to make any significant difference to the variance explained or the regression coefficients. Only the analyses using nontransformed scores were reported.

First, a correlation analysis was carried out among the six variables. As shown in Table 4.2, there is a non-significant correlation of -.28 (p=n.s) between online and offline SA scores. WMC and divided attention are significantly negatively correlated, r= -.39, p<.05. WMC and offline SA scores are highly significantly positively correlated, r= .744, p<.00. Expertise and online SA scores are significantly negatively correlated, r= -.58, p<.00. Expertise and offline SA scores are significantly positively correlated, r= .46, p<.00. Divided attention and online SA scores are significantly positively correlated, r= .56, p<.00. Regression model assumptions are

Table 4.2: Correlations for Experiment 1

		WMC	STRP	CRT	EXP	OFFLINE	ONLINE
WMC	Pearson's	1	.171	-.397	.284	.744	-.344
	Sig. (2-tailed)		.317	.016	.093	.000	.040
	N	36	36	36	36	36	36
STRP	Pearson's	.171	1	.219	.049	.136	.283
	Sig. (2-tailed)	.317		.200	.778	.430	.094
	N	36	36	36	36	36	36
CRT	Pearson's	-.397	.219	1	-.299	-.326	.562**
	Sig. (2-tailed)	.016	.200		.077	.052	.000
	N	36	36	36	36	36	36
EXP	Pearson's	.284	.049	-.299	1	.462	-.588**
	Sig. (2-tailed)	.093	.778	.077		.005	.000
	N	36	36	36	36	36	36
OFFLINE	Pearson's	.744	.136	-.326	.462	1	-.283
	Sig. (2-tailed)	.000	.430	.052	.005		.095
	N	36	36	36	36	36	36
ONLINE	Pearson's	-.344	.283	.562	-.588	-.283	1
	Sig. (2-tailed)	.040	.094	.000	.000	.095	
	N	36	36	36	36	36	36

assessed through percent variability explained by the model, scatter plots for linear relationship, checking for multi-collinearity and normality.

Two linear multiple regression analyses were performed on offline and online SA scores as outcomes and four individual differences: WMC, Inhibition (STRP), Divided Attention (CRT), and Expertise (EXP). The analysis was performed using SPSS Version 17. There were thirty-six participants in the analysis and no missing values. Because there is no a priori hypothesis about which predictor variable individually has more predictive power over online and offline SA scores, *Enter* method was used for the regression analysis.

In order to investigate whether the path relating divided attention to online SA scores is mediated by expertise and whether the path relating WMC to offline SA scores is mediated by inhibition, the linear models were modified to include possible interaction terms. The basic argument behind this investigation was that Choice Reaction Time task with dichotic listening (CRT) is an ecologically valid test which is based on participants' expertise as well as divided attention performance. Additionally, some studies in the literature such as Unsworth *et al.* (2004); Brewin and Beaton (2002); Conway *et al.* (2001) theorize the relationship between WMC and inhibition. It was observed that regression analyses did not show any significant effect of interaction terms and then these terms were removed from the analyses.

The four predictor variables (WMC, inhibition, divided attention, and expertise) produced an adjusted R^2 of .58 ($F(4,35) = 12.81$, $p = .00$) for the prediction of offline SA scores (Table L.2). The same predictor variables produced an adjusted R^2 of .52 ($F(4,35) = 10{,}29$, $p = .00$) for the prediction of online SA scores (Table L.5).

For offline SA Scores, WMC and expertise explained 58% of variance. Inhibition and divided attention did not have any predictive power. The strongest predictor for offline SA Scores was WMC ($\beta = .675$, $t(31)= 5.31$, $p<.00$), followed by expertise ($\beta = .278$, $t(31)= 2.35$, $p<.05$) (Table L.4). Each of these two predictors had a positive relationship with offline SA Scores. Consequently, the greater WMC and the higher level of expertise as individual differences, the higher offline SA scores collected through offline SA query administration.

Considering online SA scores, inhibition, divided attention, and expertise explained 52% of variance. WMC did not have any predictive power. The strongest predictor for online SA scores was expertise ($\beta = -.470$, $t(31)= -3.73$, $p<.001$), followed by divided attention ($\beta = .313$, $t(31)= 2.25$, $p<.05$) and inhibition ($\beta = .260$, $t(31)= 2.058$, $p<.05$) (Table L.7). Since online SA scores were computed by averaging z-transformed RTs that belonged to the online SA queries, the lowest value in online SA scores represents the highest SA level. In this regard, the greater expertise, inhibition and divided attention, the higher the SA level measured through online SA query administration.

In accordance with the above results of the two regression analyses, all hypotheses defined in section 3.1.6 are accepted. In other words, WMC, inhibition, divided attention and expertise have predictive power over the SA status measured by means of online and offline query administration. Specifically, WMC and expertise have predictive power over offline SA scores while inhibition, divided attention and expertise have a predictive power over online SA scores.

4.1.2 Experiment 2

Data from a small and focused sample of expert pilots were used to quantitatively and qualitatively analyze the results of the eyetracking experiment for the investigation of how expert pilots move their eyes while performing some visual tasks. It is known that scanning strategies are taught in pilot's training programs but currently it is not known if pilots use these strategies after getting experience or if they find their own strategies in accordance with their individual differences. All visual tasks designed in the experiment are interpretation-type and only the importance of the flight displays slightly changes for each visual task. These slight changes could be observed in the heat maps[1] (see Figure 4.1) that belong to ten participants across the four visual tasks.

4.1.2.1 Observations about Scan Strategies

Table 4.3 through 4.6 show average time to first fixation, average fixation count and average fixation duration for each AOI in four visual tasks. In visual task-1, participants tried to estimate the aircraft position by watching and analyzing a short video (around 2 minutes) that was captured during a simulated flight. Judgements about the aircraft position, generally speaking, can be made by interpreting the course, heading, distance to a navigation aid site (navaid), vertical speed etc. These parameters could be obtained by using readings on the six flight displays, namely Airspeed Indicator (ASI), Altimeter (ALT), Attitude Indicator (AI), Horizontal Situation Indicator (HSI), Radio Magnetic Indicator (RMI), and Vertical Speed Indicator (VSI). Considering the three eye movement parameters shown in Table 4.3 recorded while estimating the aircraft position, AI and RMI seem to be the most important flight displays in estimating aircraft position because the average times to first fixation for AI and RMI were significantly smaller than the others'. Average fixation counts in Table 4.3 showed that

---

[1] Heat maps were created in accordance with all participants' eye movements

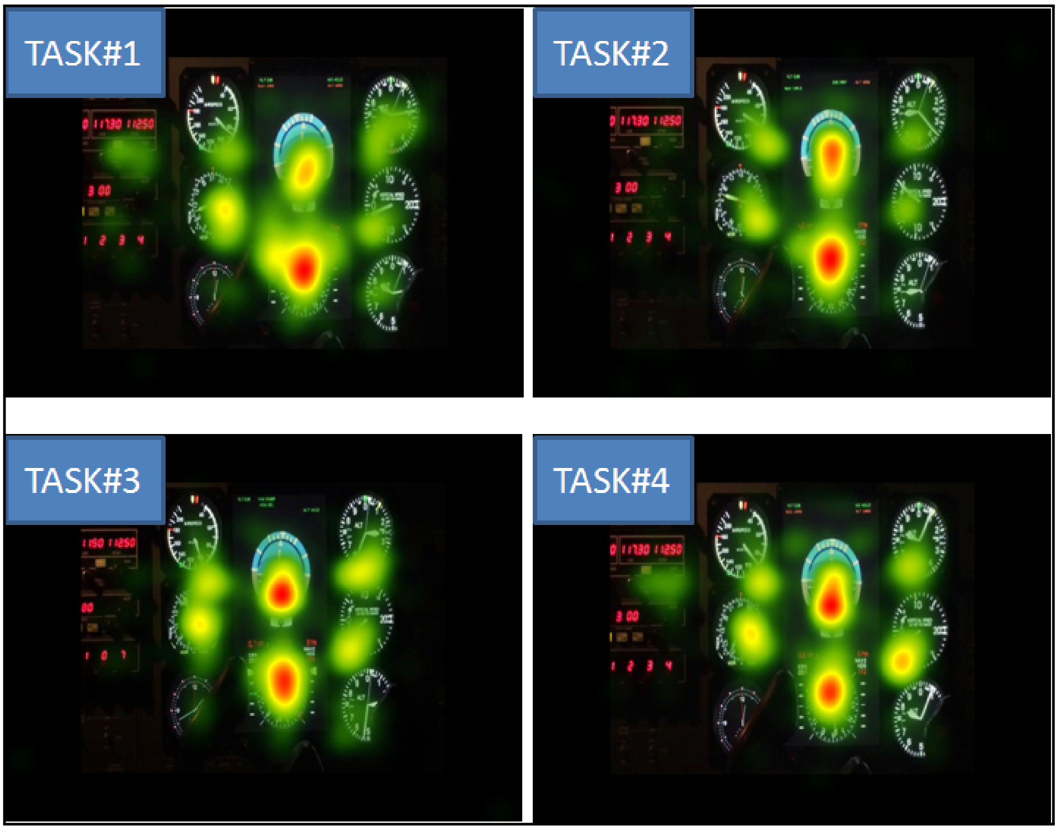


Figure 4.1: Heat Maps

HSI was visited more than AI and RMI. This finding is interesting and a little bit unexpected because both HSI and RMI indicate heading information and RMI is a simpler display than HSI. Consequently, the duration of mental process required for interpreting the readings of RMI is smaller than the one for HSI. Average fixation durations in Table 4.3 are also opposing this view. One of the possible interpretations of this result is that HSI which indicates more than one flight parameters is not interpreted holistically but partially.

In visual task-2, participants estimated wind direction and magnitude and Table 4.4 shows the recorded eye movement parameters. Estimating wind direction and magnitude requires a significant cross-check (i.e going back and forth) between heading and course information and between airspeed and ground speed. Because of this fact, it is expected that average fixation counts for the flight displays indicating these flight parameters are higher than the ones for the other flight displays.

Table 4.3: Eye movement data while estimating aircraft position

AOI	Average Time to First Fixation (sec.)	Average Fixation Count	Average Fixation Duration (sec.)
ASI	7.12	13.17	0.47
ALT	7.86	21.83	0.57
AI	2.29	42.00	0.41
HSI	7.43	76.83	0.41
RMI	2.66	33.67	0.42
VSI	7.74	17.50	0.29

It was observed that HSI, AI, RMI, and ASI were the most visited flight displays as seen in Table 4.4. Theoretically, HSI-ASI pairs are visited for checking airspeed and ground speed. DME Indicator on top of HSI indicates the ground speed and ASI indicates the airspeed. Difference between the ground speed and the airspeed is an indicator of wind magnitude. Wind direction is estimated by cross-checking heading information in RMI and course information in HSI. In performing this task, HSI, RMI and ASI give necessary information to participants; however, AI displays none of the course, heading, ground speed or airspeed.

Average fixation counts in Table 4.4 showed that unexpectedly AI was among the mostly visited flight displays. This may support the idea that AI is central in scanning strategies regardless of the visual tasks as reported by the participants in the interviews (see section 3.2.3). The same interpretation could be made for average time to first fixation because the value that belongs to AI was the second smallest one which indicated the importance of AI. It can be thought that this occurs due to the ecological validity of the task. In real flight, as opposed to video display, some indication of aircraft's attitude might not be available from physiological processes (e.g. vestibular system). The fact that participants looked at AI more than expected might be the reflection of this. In fact, pilots' vestibular system is the major cause of spatial disorientation[2]. For that reason, they are trained for trusting the flight displays but not their senses. This reduces the possibility of this alternative interpretation.

---

[2] Spatial disorientation is a condition in which an aircraft pilot's perception of direction (proprioception) does not agree with reality (en.wikipedia.org).

Table 4.4: Eye movement data while estimating wind direction and magnitude

AOI	**Average Time to First Fixation (sec.)**	**Average Fixation Count**	**Average Fixation Duration (sec.)**
ASI	2.07	24.17	0.64
ALT	7.62	20.83	0.44
AI	2.52	60.17	0.59
HSI	4.17	87.50	0.59
RMI	2.94	40.67	0.63
VSI	4.92	16.50	0.46

Participants looked for abnormal events (malfunctions in the flight displays, conflicting information, etc) in visual task-3. Table 4.5 shows the eye movement parameters recorded while performing this task. Prior to the experiment, it was expected that each flight displays would be more or less equally visited to judge whether they worked correctly or indicated conflicting information. On the displayed video, heading information indicated by RMI and HSI, vertical speed indicated by VSI were conflicting. In this task, participants were expected to find these three malfunctions. Regarding average time to first fixation shown in Table 4.5, it was found that VSI, which was one of the malfunctioned flight displays, was visited more rapidly than in the other three visual tasks. This finding could be interpreted in such a way that expert pilots performed attention focusing without a direct steer at VSI even in deploying very well-defined and well-structured scanning strategies and expert pilots could draw their attention (consequently their gaze) to this abnormality. AI was again found to be central in this visual task (see average time to first fixation and average fixation count in Table 4.5) although AI was not malfunctioned. This might be evidence that expert pilots did not violate the basic rules reported in section 3.2.3 for the scanning strategies regardless of which type of visual tasks were performed. This also supports the above hypothesis for the existence of peripheral vision without a direct steer while deploying a scan strategy.

Visual task-4 was about the trends, not about the numbers indicated by the flight displays. In this task, participants tried to judge the pilot's expertise level by monitoring the trends in the flight displays. Generally speaking, smoothness in turning, maneuvering, climbing and descending are the indicators of expertise. For that reason, the

Table 4.5: Eye movement data while looking for abnormal events

AOI	**Average Time to First Fixation (sec.)**	**Average Fixation Count**	**Average Fixation Duration (sec.)**
ASI	18.08	15.67	0.44
ALT	9.62	23.67	0.46
AI	1.42	56.33	0.55
HSI	2.37	63.17	0.60
RMI	1.29	39.83	0.60
VSI	3.91	31.17	0.49

initial assumption before the experiment was that small fixation durations (focusing on the trends instead of the numbers) would be observed.

In visual task-4, stability of airspeed, vertical speed, altimeter, and heading during turning and maneuvering would be important while judging the expertise level of a pilot. As seen in Table 4.6, AI had the smallest average time to first fixation, the highest average fixation count, and one of the highest average fixation duration. This finding is consistently observed in four visual tasks indicated the central role of AI in the scanning strategies even though the four visual tasks had different scanning requirements as explained above.

One of the unexpected results was that ASI had the smallest average fixation count. This conflicts with the initial assumption about the importance of airspeed stability for judging the level of expertise. Although HSI was not one of the rapidly visited flight displays (see average time to first fixation in Table 4.6), HSI was the second frequently visited one. Apart from AI, HSI was seen as second important flight display in the scanning strategies deployed in the visual tasks. It could be said that displaying multiple information in a single display such as HSI could create effectiveness in the scanning strategies.

Considering the results of Experiment 2, there were three types of observations relevant to the interviews' outcomes. These observations are listed below:

Table 4.6: Eye movement data while judging level of expertise

AOI	Average Time to First Fixation (sec.)	Average Fixation Count	Average Fixation Duration (sec.)
ASI	3.24	16.83	0.53
ALT	7.84	19.17	0.46
AI	2.00	56.17	0.53
HSI	7.57	45.50	0.49
RMI	4.56	29.33	0.43
VSI	7.46	29.83	0.50

- Regarding the importance of AI in the scanning strategies, Average Time to First Fixation (ATFF) for AI (M_{ATFF}=2.05 sec) shows that expert pilots fixate on AI at first compared to the other displays. This supports the interview's outcomes.

- Average fixation count on AI (22.7 % of total fixations) was found to be slighly less than expected. It is assumed that this could be the difference between what is taught to the pilots and what is applied by them.

- None of the average fixation duration for AOIs has greater than 0.64 sec and this supports the idea of having fixations of less than 2-3 secs on AOIs.

4.1.2.2 Testing Hypotheses H_7 through H_{14}

Eight Mann-Whitney U tests were conducted to evaluate the hypotheses about the possible relationships between individual differences and the eye movement characteristics represented by fixation count mean and fixation duration mean. According to the results shown in Appendix-L, all hypotheses except H_8 were rejected. It was found that More Expert Pilots (MEs) significantly differ from Less Expert Pilots (LEs) in terms of fixation duration produced while performing the four visual tasks. It means that MEs produced less fixation durations compared to LEs. WMC, strength of inhibition and divided attention capacity were found to have no significant effect on fixation durations and fixation counts.

4.2 Discussion

The primary aim of this study was to investigate the possible effects of working memory, attention, and expertise on pilots' situation awareness. Two experiments were designed for explanatory and exploratory parts of the study. The following paragraphs include important discussion points obtained in Experiment 1 and Experiment 2.

4.2.1 Experiment 1

Experiment 1 included a combined use of online and offline SA measurements and three cognitive capacity tests (namely AOSPAN, Stroop, and Choice Reaction Time with Dichotic Listening). Thirty-six participants took part in Experiment 1. SA Measurements were conducted in a cognitively demanding flight scenario in Cessna-172/182 Flight Simulators. Offline SA queries were administered while the simulation was paused. During query administration, participants could not see the flight displays. Online SA queries were administered while the simulation was running. Accuracy of reporting for offline SA queries and RTs for online SA queries were recorded under the assumption that accuracy of reporting is positively and RTs are negatively correlated with the SA level. After the flight simulation, three cognitive capacity tests were conducted in the office environment to measure the participants' individual differences including WMC, inhibition, and divided attention. Regression analysis was applied to investigate how much of variability in the SA level could be explained by pilots' individual differences including expertise. It was found that WMC and expertise are significant predictor variables for the offline SA scores and for the online SA scores, inhibition, divided attention and expertise were found to be significant predictors.

Regarding the significant correlation found between WMC and offline SA scores, is it possible to think that offline SA scores are just the same thing as WMC? It would be reasonable to conclude that offline SA scores and WMC have something in common rather than concluding that these two are the same. In accordance with the results of Experiment 1, offline and online SA measures were found to be too simple measurements to measure a highly complex cognitive phenomenon which is SA. The situation could be explained by making an anology about the relationship between STM and WM. Consider the case in which a researcher tries to investigate WM through a simple span task which is known as measure of STM. As we all know, WM requires more complex span measures. There is a theoretical ground for that. The same theoretical ground may necessitate a relationship between STM and WM. In that case, STM measure and WM measure represent something in common to some extent. This does not mean that a simple span measure is just the same as a complex span measure.

There are also some past studies supporting the idea that offline SA measure and WMC correlate (Durso *et al.*, 2006; Gonzales and Wimisberg, 2007). Durso *et al.* (2006) selected a set of cognitive, personality and demographic variables including two WM task scores in order to predict the operator performance in an ATC task. They conducted both offline and online SA measures. Their aim was to investigate a possible prediction capability of offline and online SA measures on the task performance above and beyond the selected variables. They conducted a few regression analyses and the results showed that offline SA scores increased the predictibility by only 2% over the task performance after adding offline scores to the set of predictor variables. It is because the predictor variables include WM scores which correlate with offline SA scores. In another study(Gonzales and Wimisberg, 2007), online and offline SA measures were taken from thirty-six volunteers in water purification tank task which requires resource allocation and scheduling. Participants' WMC scores were collected by means of a visual span task. It was found that WMC scores could only predict offline SA scores. This finding is consistent with the correlation found between WMC and offline SA scores in this study.

One of the interesting findings was the nonsignificant correlation between WMC and inhibition. The predominant view in the literature is that there is a relationship between WM and inhibiton (Conway *et al.*, 2001; Brewin and Beaton, 2002; Kane and Engle, 2003). There are also different approaches but not excluding this view. As Unsworth *et al.* (2004) stated, "*we do not consider the limiting function of WM span* *to neccessarily be an inhibitory one but, rather, an attention one, which could be oriented to maintenance or suppression*"(p.1317). It is largely accepted that inhibitory mechanisms are not unified. Friedman and Miyake (2004) reported three types of inhibitory mechanisms, namely, prepotent response inhibition (e.g. Stroop task), resistance to distractor interference (e.g. word naming with distractors) and resistance to proactive interference (e.g. cued recall task). Friedman and Miyake (2004) argued that "*inhibition has been overextended and that researchers need to be more-specific when discussing and measuring inhibition-related functions*"(p.101). Consequently, the possible explanations for the nonsignificant correlation found between WMC and inhibition should be made by comparing the result of this study with the ones of the past studies focusing on the prepotent response inhibition as to be discussed next.

Considering the Stroop task, there are also three ways for measuring strength of inhibition. The first one is subtracting RTs of congruent trials from RTs of incongruent trials. The second one is subtracting RTs of neutral trials from RTs of incongruent trials. The last one is to use error rates. Another problematic issue is about the experimental design. In case of a large group of participants, it is possible to choose extreme design[3] in order to maximally observe the effects of working memory on the strength of inhibition. Weak or moderate correlations found in the literature could disappear depending on the method of calculation and the experimental design chosen. All in all, it seems to be viable to compare and contrast this nonsignificant correlation with the result of Kane and Engle (2003)'s study although there are many others related to the relationship between WMC and inhibition. Kane and Engle (2003) suggested that error rates are highly sensitive to WM span differences compared to RTs[4]. The nonsignificant correlation found between WMC and the prepotent response inhibition is consistent with the results of two participants-paced experiments in Kane and Engle (2003)'s study[5]. One of the possible explanations is that WM mechanisms are not susceptible to the prepotent response inhibition but to the other inhibitory mechanisms including the resistance to distractor interference and the resistance to proactive

---

[3] selecting upper quartile and lower quartile participants for WMC and using 75% congruency ratio

[4] Kane and Engle (2003) used both RTs and error rates for measuring the prepotent response inhibition. They stated that span differences were observed in accuracy, not in response times.

[5] There were five experiments in the study. First two experiments were participant-paced as in this study. The authors attributed the nonsignificant correlation found between WMC and inhibition to the large error rates obtained.

interference. Other possible explanation is that WM span differences as suggested by Unsworth *et al.* (2004) may be oriented to maintenance component of attention.

In some of the previous studies, SA measurement techniques were assumed to measure all aspects of SA (Endsley, 1995a; Durso and Dattel, 2004). In some others, there were attempts to combine existing measurement techniques to cover SA's major constructs (Jeannot *et al.*, 2003). The results of this study indicated that offline SA scores could only represent the contribution of WM and expertise to acquiring and maintaining SA. The results also showed that online SA scores represent the contribution of attention mechanisms (inhibition and divided attention) and expertise to acquiring and maintaining SA. This can be interpreted such that inhibition and divided attention are two attention processes that underlie SA and expertise makes SA-related cognitive processes more effective on acquiring and maintaining SA. These results are in contradiction with some previous work suggesting that an SA measurement technique in the literature such as SAGAT or SPAM could measure all aspects of SA. On the contrary, the finding about the relationship between attention mechanisms and online SA scores provides Gugerty (2011)'s hypothesis with an empirical evidence. Gugerty (2011) hypothesized that online SA scores could be used for exploring attention mechanisms in traffic domain.

It is suggested in this study that offline SA measurements based on query administration while freezing the simulator to hide the flight displays have a weakness in capturing SA's dynamic characteristics during flight. They are completely based on the content of WM. Not in all but many situations, expert pilots do not have to memorize readings of flight displays because they know where to look at to get the required information to navigate. Probing contents of WM can measure memory and expertise aspects of SA but not the others. The other suggestion in this study is that attention and expertise related aspects of SA could be measured by means of online SA measurement based on query administration while navigating.

In this study, both measurement techniques were used to cover the SA constructs including WM, attention, and expertise. Since online and offline SA scores are not correlated in Experiment 1, it is possible that offline and online measurements measure SA's different constructs. In this regard, there is also a need for an interpretation of nonsignificant correlation found between offline and online SA scores. Offline and online SA measurements are claimed to measure the same construct. Parallel to this view, Durso *et al.* (2006) stated that

"*SPAM and the off-line measure were intended to measure the same underlying construct, namely SA, and in fact they were highly correlated with each other. However, the two measures differed in their incremental validities, suggesting that typical cognitive measures already capture much of what off-line measures contribute. It remains for future research to identify the cognitive components of SA that are captured by online measures such as SPAM.*"(p.731).

One problem with this interpretation is that if offline and online SA measures are highly correlated with each other, it is not easy to conclude that these two measurements capture different components of SA. Durso *et al.* (2006) used a set of predictor variables consisting of cognitive, demographic and personality variables in order to predict the ATC task performance. It is noteworthy that none of the twenty-three variables was attention-related. This is the reason why online SA scores have a predicting power over and beyond the selected predictor variables. As it is hypothesized in this study (see $\boldsymbol{H_6}$ in section 3.1.6), online SA measures represent the attention-related constructs of SA. Three out of five predictor variables remained after the factor analysis were memory-related. For that reason, Durso *et al.* (2006) did not report a significant predicting capability of offline SA scores after adding them to the set. This is consistent with the hypothesis $\boldsymbol{H_5}$ tested and the results obtained in this study.

The nonsignificant correlation between online and offline SA scores calls into question the idea of definition of what SA is. In particular, we should refer it back to the consideration of the theoretical approaches to SA and ways of measuring it. Offline and online SA measurements used in this study are based on information processing account. For this account of SA, it is quite reasonable to say that the most accepted definition of SA proposed by Endsley (1995a) brings some practical advantages from measurement's standpoint beyond the characteristics of this account. It seems also that the idea of measuring this complex cognitive construct somehow guided the applied researchers to define it in accordance with this practical aim, that is, measuring SA. Classification of the SA queries in terms of perception, comprehension, and projection helps the SA researchers measure and evaluate SA in different levels of information processing. Freezing the simulator and blanking the flight displays during the offline query administration results in changing the operational environment to a large extent hence offline query results reflect only the content of pilots' WM. In fact, pilots probabily use the operational environment as a memory store that reduces the cognitive load during the flight because the flight displays constantly provide necessary information for navigation.

The situation can be interpreted by using the notion of *Extended Cognition* in which cognition is not confined to an individual's skull but extends into his/her environment (Clark and Chalmers, 1998; Hutchins, 1995). In that sense, cognitive performance of a pilot might decrease after blanking the flight displays during the query administration. On the other side, online SA scores seem to represent something different from the content of WM which is the attention component of SA because WMC was not found to be one of the predictor variables for online SA scores while both inhibition and divided attention capacity were found to be the significant predictors.

Although the nonsignificant correlation between offline and online SA scores is not in line with the current literature, the empirical evidence provided by this study suggest that offline and online SA measurements capture different components of SA. To be specific, offline SA scores represent memory-related components of SA and online SA scores represent attention-related components of SA.

The Stroop task was used in this study to measure participants' strength of inhibition. As Unsworth *et al.* (2004) pointed out, inhibition is important for WM mechanisms. For that reason, it is reasonable to expect strength of inhibition to be the predictor of offline SA scores together with WMC. However, this was not observed. The one possible explanation is that *Enter* method in regression analysis gives priority to the variable (WMC in this case) that explains most of the variance. It could happen that attention mechanisms do play a role in offline SA scores but not over and above WMC and expertise. It was observed that strength of inhibition significantly predicts some portion of the variance in online SA scores. This makes a lot of sense since inhibition is not a general ability but pays off in concrete situations where pilots have to manage too much information in WM. As an alternate interpretation to the one explained above, the same situation might happen for WMC in predicting online SA scores and the contribution of WMC was not observed due to the fact that the role of WMC in online SA measurement was not over and above attention mechanisms.

It was observed that expertise significantly predicted both offline and online SA scores. $\boldsymbol{H_4}$ is accepted in accordance with this emprical evidence. Considering the unique contribution of expertise, the current results are in parallel with the previous work (Endsley, 2006; Durso, Truitt, Hackworth, Crutchfield, Nikolic, Moertl, Ohrt, and Manning, 1995). This finding could also be interpreted such that expertise is multi-dimensional including familiarity in flying an aircraft and procedural knowledge relevant to navigation. It seems that familiarity has a role in attention mechanisms mostly captured by online SA scores. On the other side, procedural knowledge has a role in memory mechanisms mostly captured by offline SA scores.

WMC and expertise predicted 58% of variability in offline SA scores and WMC was found to be the strongest predictor followed by expertise over offline SA scores. Comparing this result of the present study to the result obtained by Gonzales and Wimisberg (2007) presents a consistency to a large extent. Gonzales and Wimisberg (2007) found that WMC measured by a visual span task predicted 45% of variability in offline SA scores. The amount of variability predicted by WMC and expertise implied that offline SA scores do not represent all aspects of a very complex cognitive construct related to perception, attention, LTM, reasoning as well as WM and expertise.

Inhibition, divided attention capacity, and expertise predicted 52% of variability in online SA scores and expertise was found to be the strongest predictor followed by divided attention capacity and inhibition. Unfortunately, the SA literature is not rich about the studies on the relationship between online SA measurement and attention mechanisms. In a recent review, Gugerty (2011) surveyed the SA measurement studies and suggested that online SA measurement techniques can be used to assess the attention mechanisms in driving tasks. Regarding the effects of expertise on online

SA scores, Durso *et al.* (1995) reported the results of a pilot study on chess players' SA. They observed that the expert chess player produced higher online SA score compared to the intermediate and novice chess players. Since the participants were three chess players, it is difficult to make strong conclusions about the relationship between expertise and online SA scores. Nevertheless, it can be safely said that the two studies mentioned above support the finding in this study. Apart from these two studies, there is also a consideration about the administration of online SA measurement. Online SA queries were administered while the participants were flying. They were flying in a familiar environment and it is highly probable that their level of expertise did affect positively how fast and how accurate they responded to the SA queries. It is also likely that effects of inhibition and divided attention capacity, on the other side, could maximally observed in dynamically changing environment.

4.2.2 Experiment 2

Experiment 2 was designed in order to test the hypotheses about the relationship between pilots' individual differences and the eye movements. Experiment 2 was also utilized to investigate as to whether pilots deploy some scan strategies while performing four different SA-derived visual tasks. Ten expert pilots participated in Experiment 2. Four videos recorded during a simulated flight were presented to the participants and the participants were asked to perform a visual task during each video display. As the explanatory part of the analysis, the eye movement data were quantitatively analyzed by using non parametric tests in order to investigate as to whether there is any effect of individual differences on pilots' eye movement data. On the exploratory part of the experiment, participants' scan patterns were visually inspected.

The results of the quantitative analyses showed that More Expert Pilots (MEs) produced less fixation durations compared to Less Expert Pilots (LEs) and there was no observable effects of other individual differences, namely WMC, inhibition, and divided attention capacity on the eye movement data characterized by fixation duration mean, and average fixation count. This is not to say that such effects do not exist. Increasing the number of participants might reveal some of the possible effects hypothesized in section 3.2.5.

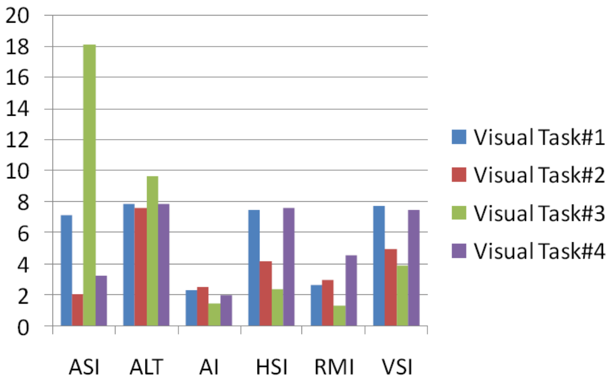


Figure 4.2: Time To First Fixation Means Across the Four Visual Tasks

Having small fixation durations for MEs is very much in line with Kasarskis *et al.* (2001)'s suggestion that expert pilots do not need excessive fixation durations and they could extract information from their peripheral vision. Kasarskis et al's study focused on the scan behaviors of expert and novice pilots. Since the participants taken part in Experiment 2 were all experts, they were classified as More Expert (MEs) and Less Expert Pilots (LEs). Even in this situation, significant effect of expertise on pilots' fixation durations was observed. The second part of Kasarskis et al.'s suggestion is the reason why expert pilots do not need excessive fixation durations. Effects of using peripheral vision is also observed in visual task-3. In this task, participants were required for abnormal events under the condition that 3 out of 6 displays, namely, RMI, HSI and VSI were mulfunctioned. Time To First Fixation on ASI was unexpectedly high as seen in Figure 4.2. One and obvious interpretation is that malfunctioned displays were detected and this detection took some mental processing. The amount of this mental processing could be seen on the time when ASI was first visited. The higher level interpretation is that expert pilots use their peripheral vision although they deploy certain scan strategies. Parallel to this finding, Kim *et al.* (2010) provided empirical evidence for the certified pilots using their peripheral vision. It is noteworthy that using peripheral vision is not a part of the scan strategies taught. It seems that as pilots acquire experience, they could use peripheral vision to increase the efficiency in the scan strategies. Figure 4.4 showed fixations duration means in very narrow band ranging between 0.28 sec. and 0.64 sec. This finding is also supported by the past two studies (Ottati *et al.*, 1999; Bellenkes *et al.*, 1997) showing shorter fixations in expert pilots. Additionally, Fixation Duration Means were found to be far less than the target values 2-3 seconds[6] which are taught in the pilot training programs. This seems to be a suitable feedback to the current scan strategies taught.

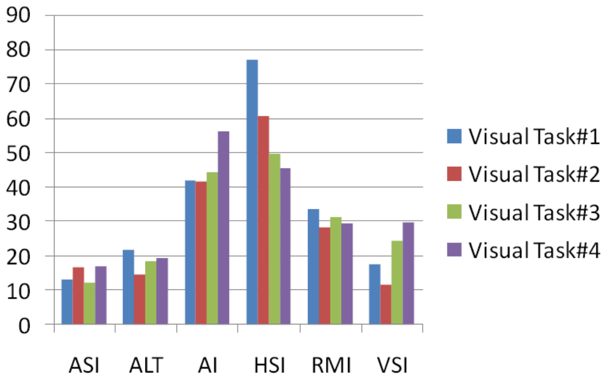


Figure 4.3: Average Fixation Counts Across the Four Visual Tasks

Average Fixation Counts depicted in Figure 4.3 show that AI and HSI were the mostly visited flight displays across the four visual tasks. Analyzing the scan paths also provides evidence for the transitions from AI to HSI and from HSI to AI more than expected transitions in T-scan. This indicates a deviation from the T-scan strategy. One way to think about this indication is that HSI as a complex display indicating more than one information is used by the expert pilots more frequently than in a typical T-scan. It is also suggested that expertise comes with efficiency because different types of information displayed in a single body of display, that is HSI, is preferred than simple displays. This suggestion is also parallel to the current design trends in the flight displays such as glass cockpits[7].

The eye movement data were qualitatively analyzed to extract some scanning strategies and compare them with what is taught to the pilots in the pilot training programs. Three-types of scanning strategies were observed in the eye movement data: T-scan, selective radial scan and circular scan with some small deviations. Figure 4.5 shows

---

[6] It was reported in the informal interviews with the participants.

[7] Glass cockpit is a term for an aircraft cockpit including electronic flight displays rather than analog ones.

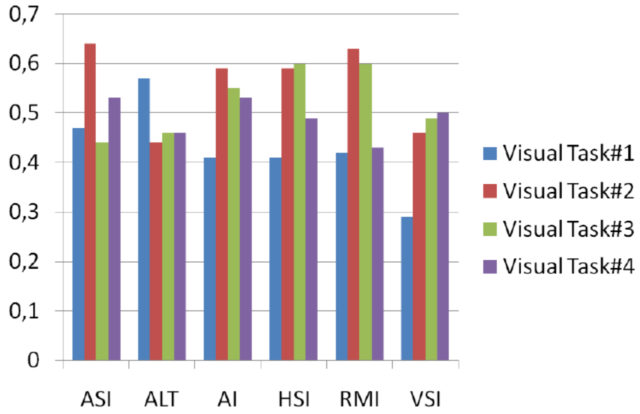


Figure 4.4: Fixation Duration Means Across the Four Visual Tasks

sample scan strategies[8] observed.

T-scan observed (P#08 in Task#1)

Sel. Radial Scan observed (P#09 in Task#4)

Circular Scan observed (P#10 in Task#3)

Figure 4.5: Sample Scan Strategies Observed

In addition to the scanning strategies, it was observed that the participants focused on avionics displays in slightly different ratios from the rules that are taught in the training programs. It was also observed that scan strategies were not continuously deployed. Nevertheless, there was no evidence found for interchangeably deploying these scan strategies during the task performance.

---

[8] Numbers displayed in Figure 4.5 indicate the fixation sequence

4.2.3 Considerations about the relationship between Experiment 1 and Experiment 2

Experiment 1 is designed to investigate the relationship between pilots' SA status measured by means of offline and online SA scores and pilots' individual differences, namely, WMC, inhibition, divided attention capacity, and expertise. Experiment 1 is considered as a bridge between the applied world of SA represented by two SA measurement methods and the basic research represented by the three cognitive capacity tests. Experiment 2 is designed to investigate the relationship between the same individual differences and the pilots' eye movements characterized by fixation duration mean and average fixation counts.

Considering the relationship between Experiment 1 and Experiment 2, the basic idea was to relate the eye movements to the processes of acquiring and maintaining SA. It is assumed that individual differences having effects on both SA status and the eye movements will create a new opportunity to link SA status and the eye movements. Considering the results of Experiment 1 and Experiment 2, this link could be utilized in two ways. First, the eye movements can be used for developing a more complex SA measurement technique. Second, scan strategies can be improved in order to maximize the pilots' SA status.

4.3 Limitations

Due to the difficulties in accessing the pilot community, finding empty slots in the simulator schedule, and budget limitations, thirty-six pilots could participate in Experiment 1 and ten pilots in Experiment 2. In Experiment 1, a regression analysis was applied. The statistical power in a regression analysis generally suffers from the small number of participants. Because of the high predictive power of predictors, however, the regression analysis gave significant results both for online and offline SA scores in spite of the small number of participants. Although the results seem to be promising, it is necessary to be cautious in drawing some conclusion on the relationship between the SA status and individual differences. Experiment 2 was intended to be both exploratory and explanatory study for investigating as to whether there is any effect of individual differences on pilots' eye movement and for creating some hypothesis about the relationship between individual differences and the eye movements. In that sense, a larger confirmatory study is needed for the eyetracking experiment.

All participants in the experiments were fixed-wing aircraft pilots. There are various types of aircraft including rotary-wings. They have different flight characteristics and they require different skills for the pilots in some sense. It is not possible to generalize the outcomes of this study to a wide classification of pilots. The other limitation that needs to be acknowledged is the gender issue since all participants were male except one. This restricted us to make broad generalization in terms of gender. In fact, gender distribution in Experiment 1 more or less represents the actual gender distribution of pilots.

Although participants were trained for performing a solo flight in the flight simulators, solo flights are rare events in commercial aviation. In the real world, pilot-copilot pairs share the workload in the cockpit and continuously communicate with each other. This workshare and communication increases individual SA. This concept is called "team SA" in the literature (Sulistyawati, Wickens, and Chui, 2009; Saner, Bolstad, Gonzales, and Cuevas, 2009; Gorman, Cooke, and Winner, 2006). In this study, individual SA was examined for the sake of feasibility.

Online SA scores (RTs) were calculated by using the videos recorded during the simulated flights. After recording, the scenes covering query-answer pairs were clipped from the recorded videos. The time difference between the query's endpoint and the answer's startpoint were taken as RT. It was not always clear to decide on where the answers started. Hesitation, murmuring, thinking aloud brought about some vague startpoints for the answers. This created some unreliability in the data to some extent. It is realized that query structures may reduce or increase the effect of this reliability problem. Generally, yes/no type queries reduce this effect. Nevertheless, it is not practicable to create all queries as yes/no type. For instance, significant difference in terms of what researcher gains from the query may occur if the query for the wind direction is administered in several ways as in below:

- In what direction do you think the wind is blowing?
- Do you think the wind is blowing?
- Do you think there is a crosswind?

Wind direction can be calculated by the difference between the course and the heading of the aircraft. First query requires significant mental effort than the other two. Answering this query needs precise interpretation of the values indicated by the related flight displays. On the other hand, the other two queries could be easily answered by the participant who takes into account whether s/he makes the wind corrections at the time of query administration. Semantically and syntactically simple queries reduce the variability in RTs and give little information about the participant's SA status. Regarding this obvious trade-off, Online SA queries were generated in balance and not in favor of simple or complex queries.

It was also observed that pilot communities occupied in different organizations use some different aviation terminology in Turkish language and this resulted in some misunderstandings in experimental instructions. Especially for the online SA queries, misunderstandings in query administration might have resulted in a reliability problem although such terminological alternations were rare and corrected to a great extend during the pilot studies.

Regarding Experiment 2, there was self-reported data about the scanning strategies, rules of thumbs which could not be independently verified. This data may include some sources of bias, however, it was observed that the data was mostly in parallel with what is taught to the pilots in the pilot training programs.

4.4 Future Work

A larger group of participants can take part in Experiment 1 in order to increase the statistical power of the analysis and in order to reveal some relationships between SA and individual differences which might not be observable due to the limited number of participants. Additionally, the same experiments can be replicated in similar domains such as ATC and driving in order to extract domain-specific characteristics of SA.

This dissertation covers WMC, inhibition, divided attention and expertise as individual differences in relation with SA. Other cognitive constructs including perception and LTM that are supposed to underlie SA can be added to the scope of a future study. Durso and Dattel (2004) argued that SA measurements measure something more than a limited set of cognitive capacity tests. Increasing the coverage of cognitive capacity tests can give us additional information about the cognitive processes that underlie SA.

In Experiment 1, all malfunctions and events are designed to occur relatively at the same scenario points and in the same manner. It is not possible to observe individual effects of these malfunctions and events on acquiring and maintaining SA. In fact, knowing which flight display is more effective in adverse flight conditions than the ones that display the same information is very informative and decisive in creating a scan strategy for avionics displays. The same situation exists for meteorological conditions. Consequently, another experiment could be designed in the future such that different participant groups can be used for different set of malfunctions and events. This set-up provides an opportunity to compare the individual effects of the malfunctions and events on SA status.

A larger confirmatory study for eyetracking is needed to make quantitative analysis on expert pilots' scanning strategies and to establish a link between their eye movements and individual differences. It is also possible to conduct an eyetracking experiment by using a mobile eyetracker while measuring SA in the flight simulator. This will produce ecologically more valid data than the PC-based eyetracker. In that way, direct analysis of the relationship between the SA status and characteristics of eye movement data (fixation count, fixation duration, time to first fixation, etc) could be possible.

Most of the pilots flying in Turkey are not native-English speakers. They have to use English language as second language. Effects of second language on acquiring and maintaining SA is a fruitful research area from aviation, psychology and linguistics standpoints. Two participant groups (one group consists of participants speaking native language in the SA measurement and the other consists of participants speaking English as a second language) can be compared to observe the possible effects of speaking English as a second language in aviation.

4.4.1 Suggestions to Pilot Training Process

High-fidelity flight simulators have become the integral part of today's pilot training. Effective use of the flight simulators increases candidate pilots' familiarity to operational environment (i.e. cockpit) and their tolerance against unexpected events including emergency conditions. In order to achieve this, complex and cognitively demanding flight scenarios should be used in the high-fidelity flight simulators. It was observed that commercial airliners follow the process as explained. On the other hand, aviation schools, where aviation fundamentals are first taught to candidate pilots, use more standard flight scenarios and these scenarios are not customized in accordance with candidates' technical profile. For instance, it is possible to create a flight scenario requiring certain procedures or knowledge that a candidate pilot shows weakness about.

To some extend, evolving flight displays in the current aviation world probably changed the scan strategies which are taught in pilot training programs. Revising the scan strategies is necessary in accordance with the eye movements collected from the expert pilots. Continuous effort is also needed for analysing pilots' eye movements in terms of theoretical considerations and practical way of deploying the scan strategies.

Eye movement studies can be conducted by means of Eyetracker Systems. PC-based Eyetracker Systems bring about a decrease in the ecological validity of the pilots' eye movement characteristics while Mobile Eyetracker Systems have calibration problems due to the lighting conditions and relatively long operational periods. It would be fruitful to design and develop a full flight simulator in which there is an embedded eyetracker system installed on the flight deck.

4.5 Conclusions

In this study, it is shown that WMC and expertise can predict SA status measured by means of Offline SA measurement while inhibition, divided attention and expertise can predict SA status measured by means of Online SA Measurements. In addition to this, expert pilots' eye movements were qualitatively analyzed and important insights regarding the relationship between individual differences and eye movements were gained in order to establish a ground for a future work in which SA measurements and eyetracking will be simulataneously conducted in a high-fidelity simulator environment.

This study provides empirical evidence for the conceptual models of WM, attention and expertise that underlie the phenomenon of SA. The study could also be extended to examining the role of other cognitive processes such as LTM, perception, problem solving, and decision making involved in acquiring and maintaining SA.

There is a considerable gap in the literature between models of SA cognitive constructs and their measurement methods. These two research areas seem to develop in theoretical isolation from each other. In fact, they have to be drawn to the same theoretical context. It is observed that there is a difficulty in adopting certain theoretical models of WM and attention while studying a complex cognitive construct in the applied domain such as SA. This study tries to bridge this gap by using basic research methods for explaining the relationship between SA and individual differences. The study also gives an opportunity to criticize SA measurements in terms of their theoretical relevance to SA constructs.

Measuring the phenomenon of SA requires covering all its aspects and constructs. The relationship between SA and the cognitive capacity tests appear to facilitate the researchers to develop SA measurement techniques suitable for a wide range of research questions. Measuring and modeling SA still needs to be carefully investigated by adding all cognitive processes that underlie SA to the scope of the future studies.

References

Adams, M. J., Tenney, Y. J., and Pew, R. W. (1995). Situation awareness and the cognitive management of complex systems. *Human Factors*, **37**(1), 85–104.

Asloun, S., Soury, S., Couillet, J., Giroire, J., Joseph, P., Mazaux, J., and Azouvi, P. (2008). Interactions between divided attention and working-memory load in patients with severe traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology*, **30**(4), 481–490.

Baddeley, A. D. (2000). Episodic buffer: New component of working memory? *Trends in Cognitive Science*, **4**, 417–423.

Beer, R. D. (2000). Dynamical approaches to cognitive sciences. *Trends in Cognitive Sciences*, **4**(3), 91–99.

Bellenkes, A., Wickens, C. D., and Kramer, A. (1997). Visual scanning and pilot expertise: The role of attentional flexibility and mental model development. In *Annual Scientific Meeting of the Aerospace Medical Association*, volume 68, pages 569–579.

Bleckley, M. K., Durso, F. T., Crutchfield, J. M., Engle, R. W., and Khanna, M. M. (2003). Individual differences in working memory capacity predict visual attention allocation. *Psychonomic Bulletin & Review*, **10**, 884–889.

Bolstad, C. A. and Hess, T. M. (1995). Situation awareness and older workers. In *International Conference on Experimental Analysis and Measurement of Situation Awareness*, Daytona Beach, FL.

Brewin, C. R. and Beaton, A. (2002). Thought suppression, intelligence, and working memory capacity. *Behavior Research and Theraphy*, **40**, 923–930.

Brill, J. C., Gilson, R. D., Mouloua, M., Hancock, P. A., and Terrence, P. I. (2004). Increasing situation awareness of dismounted soldiers via directional cueing. In D. A. Vincenzi, M. Mouloua, and P. A. Hancock, editors, *Human performance, situation awareness and automation: Current research and trends*, pages 130–132. Lawrence Erlbaum Associates, Mahvah, NJ.

Broadbent, D. (1958). *Perception and Communication*. Pergamon, London.

Burke, C., Salas, E., Wilson-Donnelly, K., and Priest, H. (2004). How to turn a team of experts into an expert medical team: guidance from the aviation and military communities. *Qual Safe Health Care*, **13**, 96–104.

Busemeyer, J. R. and Diederich, A. (2010). *Cognitive Modeling*. SAGE Publications.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, **25**(5), 975–979.

Clark, A. and Chalmers, D. J. (1998). The extended mind. *Analysis*, **58**, 10–23.

Colflesh, G. J. and Conway, A. R. (2007). Individual differences in working memory capacity and divided attention in dichotic listening. *Psychonomic Bulletin & Review*, **14**(4), 699–703.

Conway, A. R. and Engle, R. W. (1994). Working memory and retrieval: Resource-dependent inhibition model. *Journal of Experimental Psychology: General*, **123**(4), 354–373.

Conway, A. R. and Engle, R. W. (1996). Individual differences in working memory capacity: More evidence for a general capacity theory. *Memory*, **4**, 577–590.

Conway, A. R., Tuholski, S. W., Shisler, R. J., and Engle, R. W. (1999). The effect of memory load on negative priming: An individual differences investigation. *Memory & Cognition*, **27**, 1042–1050.

Conway, A. R., Cowan, N., and Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review*, **8**(2), 331–335.

Cowan, N. (1995). *Attention and memory: An integrated framework*. Oxford University Press, New York.

Daneman, M. and Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, **19**, 450–466.

Dekker, S. and Lützhöft, M. (2004). Correspondence, cognition and sensemaking: A radical empricist view of situation awareness. In S. Banbury and S. Tremblay, editors, *A Cognitive Approach to Situation Awareness: Theory and Application*, pages 22–24. Ashgate.

Dennehy, K. (1997). Cranfield situation awareness scale, user manual. Technical Report (COA Report No.9702), Applied Psychology Unit, Cranfield University, Bedford.

Dennehy, K. and Deighton, C. D. (1997). The development of an interactionist framework for operationalising situation awareness. In D. Harris, editor, *Engineering psychology and cognitive ergonomics Volume One: Transportation systems*, pages 283–290. Ashgate.

Deutsch, J. A. and Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, **70**, 80–90.

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

Diez, M., Boehm-Davis, D. A., Holt, R. W., Pinney, M. E., Hansberger, J. T., and Schoppek, W. (2001). Tracking pilot interactions with flight management systems through eye movements. In *11th International Symposium on Aviation Psychology*, Columbus, Ohio.

Dillon, A. (1996). User analysis in hci: The historical lesson from individual differences research. *International Journal of Human Computer Studies*, **45**(6), 619–637.

Durso, F. and Gronlund, S. (1999). Situation awareness. In F. Durso, R. Nickerson, R. Schvaneveldt, S. Dumais, D. Lindsay, and M. Chi, editors, *Handbook of Applied Cognition*. John Wiley & Sons Ltd.

Durso, F., Truitt, T., Hackworth, C., Crutchfield, J., Nikolic, D., Moertl, P., Ohrt, D., and Manning, C. (1995). Expertise and chess: a pilot study comparing situation awareness methodologies. In D. Garland and M. Endsley, editors, *Experimental Analysis and Measurement of Situation Awareness*. Embry-Riddle Aeronautical University Press.

Durso, F. T. and Dattel, A. R. (2004). SPAM: The real-time assessment of SA. In S. T. S. Banbury, editor, *A Cognitive Approach to Situation Awareness: Theory, Measures and Application*, pages 137–154. Ashgate.

Durso, F. T., Crutchfield, J. M., and Batsakes, P. J. (2002). Situation awareness: Should we stop for directions? *Contemporary Psychology, APA Review of Books*, **47**(1), 61–63.

Durso, F. T., Bleckley, M. K., and Dattel, A. R. (2006). Does situation awareness add to the validity of cognitive tests? *Human Factors*, **48**(4), 721–733.

Endsley, M. (1997). The role of situation awareness in naturalistic decision making. In C. Zsambok and G. Klein, editors, *Naturalistic Decision Making*. Lawrence Erlbaum Associates.

Endsley, M. (2006). Expertise and situation awareness. In K.A.Ericsson, N. Charness, P. Feltovich, and R. Hoffman, editors, *The Cambridge Handbookof Expertise and Expert Performance*, chapter 36, pages 636–651. Cambridge University Press.

Endsley, M. R. (1995a). Measurement of situation awareness in dynamic systems. *Human Factors*, **37**(1), 65–84.

Endsley, M. R. (1995b). Toward a theory of situation awareness in dynamic systems. *Human Factors*, **37**(1), 32–64.

Endsley, M. R. (2000). Theoretical underpinnings of situation awareness: A critical review. In M. R. Endsley and D. J. Garland, editors, *Situation awareness: Analysis and Measuremen*. Lawrence Erlbaum Associates.

Endsley, M. R. and Bolstad, C. A. (1995). Individual differences in pilot situation awareness. *International Journal of Aviation Psychology*, **4**(3), 241–264.

Endsley, M. R. and Rodgers, M. D. (1996). Attention distribution and situation awareness in air traffic control. In *Proceedings of the 40th Annual Meeting of the Human Factors and Ergonomics Society*, pages 82–85.

Endsley, M. R., Farley, T. C., Jones, W. M., Midkiff, A. H., and Hansman, R. J. (1998). Situation awareness information requirements for commercial airline pilots (icat-98-1). Technical report, Massachusetts Institute of Technology International Center for Air Transportation, Cambridge, MA.

Endsley, M. R., Sollenberger, R., and Stein, E. (2000). Situation awareness: A comparision of measures. In *Proceedings of the Human Performance,Situation Awareness and Automation: User Centered Design for the New Millenium Conference*, Savannah, GA. SA Technologies.

Engle, R. W. and Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. H. Ross, editor, *The psychology* *of learning and motivation Volume 44*, chapter 145-199. Elsevier, New York.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., and Conway, A. R. (1999). Working memory, short-term memory, general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology:General*, **128**, 309–331.

Ericsson, K. A. and Kintsch, W. (1995). Long-term working memory. *Psychological Review*, **102**, 211–245.

Eriksen, C. and Yeh, Y. (1985). Allocation of attention in the visual field. *Journal of Experimental Psychology, Human Perception and Performance*, **11**, 583–597.

Fracker, M. (1988). A theory of situation assessment: Implications for measuring situation awareness. In *Proceedings of the Human Factors Society 32nd Annual Meeting*, pages 102–106, Santa Monica, CA. Human Factors Society.

Friedman, N. P. and Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology*, **133**(1), 101–135.

Garland, D. J. and Endsley, M. R. (1995). *Experimental Analysis and Measurement of Situation Awareness*. Embry-Riddle Aeronautical University Press.

Gonzales, C. and Wimisberg, J. (2007). Situation awareness in dynamic decision making: Effects of practice and working memory. *Journal of Cognitive Engineering and Decision Making*, **1**(1), 56–74.

Gorman, J., Cooke, N., and Winner, J. (2006). Measuring team situation awareness in decentralized command and control environments. *Ergonomics*, **49**, 1312–1325.

Gugerty, L. (2011). Situation awareness in driving. In D. Fisher, M. Rizzo, M. K. Caird, and J. D. Lee, editors, *Handbook of Driving Simulation for Engineering,* *Medicine and Psychology*, chapter 19. CRC Press / Taylor and Francis, Boca Raton, FL.

Hart, S. G., Battiste, V., and Lester, P. T. (1984). Popcorn: A supervisory control simulation for workload and performance research. In *Proceedings of the Twentieth Annual Conference of Manual Control*, pages 431–453, Washington DC. NASA.

Hauss, Y. and Eyferth, K. (2003). Securing future atm-concept's safety by measuring situation awareness in atc. *Aeorspace Science Technology*, **7**(6), 417–427.

Hendy, K. C. (1995). Situation awareness and workload: Birds of a feather? In *AGARD AMP Symposium on Situation Awareness: Limitations and Enhancements in Aviation Environment.*, Brussels.

Hockey, G. R. J. (1986). Changes in operator efficiency as a function of environmental stress, fatigue and circadian rythms. In K. Boff, L. Kaufman, and J. Thomas, editors, *Handbook of Perception and Performance (2)*, chapter 44, pages 1–49. John Wiley, New York.

Hogg, D. N., Folleso, K., Strand-Volden, F., and Torralba, B. (1995). Development of a situation awareness measure to evaluate advanced alarm systems in nuclear power plant control rooms. *Ergonomics*, **38**(11), 2394–2413.

Hollingworth, A. and Luck, S. J. (2009). The role of visual working memory (vwm) in the control of gaze during visual search. *Attention, Perception, & Psychophysics*, **71**(4), 936–949.

Hommel, B., Müsseler, J., Ashersleben, G., and Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, **24**(5), 849–878.

Horswill, M. and McKenna, F. (2004). Driver's hazard perception ability: Situation awareness in the road. In S. Banbury and S. Tremblay, editors, *A Cognitive* *approach to situation awareness: theory and application*. Ashgate.

Hutchins, E. (1995). *Cognition in the wild*. MIT Press, Cambridge, MA.

Isaac, A. R. (1997). Situation awareness in air traffic control: Human cognition and advanced technology. In D. Harris, editor, *Engineering Psychology and Cognitive Ergonomics Volume One: Transportation Systems*. Ashgate.

Janelle, C. M., Singer, R. N., and Williams, A. M. (1999). External distraction and attentional narrowing: visual search evidence. *Journal of Sport & Exercise Psychology*, **21**(1), 70–91.

Jeannot, E., Kelly, C., and Thompson, D. (2003). The development of situation awareness measures in atm systems. Technical report, Eurocontrol, Brussels.

Johannsdottir, K. R. (2004). *Situation Awareness and Working Memory: An Integration of an Applied Concept with a Cognitive Fundamental Process*. Ph.D. thesis, Carleton University.

Johansson, R., Holsanova, J., and Holmqvist, K. (2006). Pictures and spoken descriptions elicit similar eye movements during mental imagery, both in light and in complete darkness. *Cognitive Science: A Multidisciplinary Journal*, **30**(6), 1053–1079.

Jones, D. G. and Endsley, M. R. (1996). Sources of situation awareness errors in aviation. *Aviation, Space and Environmental Medicine*, **67**(6), 507–512.

Jones, D. G. and Endsley, M. R. (2000). Examining the validity of real-time probes as a metric of situation awareness. In *Proceedings of 14th Triennial Congress of the International Ergonomics Association*.

Just, M. A.and Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, **99**, 122–149.

Kahneman, D. (1973). *Attention and Effort*. Prentice-Hall, New York.

Kane, M. J. and Engle, R. W. (2000). Wm capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **26**, 336–358.

Kane, M. J. and Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual differences perspective. *Psychonomic Bulletin & Review*, **9**, 637–671.

Kane, M. J. and Engle, R. W. (2003). Working memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to stroop interference. *Journal of Experimental Psychology: General*, **132**, 47–70.

Kasarskis, P., Stehwien, J., Hickox, J., Aretz, A., and Wickens, C. D. (2001). Comparison of expert and novice scan behaviors during vfr flight. In *11th International Symposium on Aviation Psychology*, Columbus, OH. The Ohio State University.

Kim, J., Palmisano, S. A., Ash, A., and Allison, R. S. (2010). Pilot gaze and glideslope control. *ACM Transactions on Applied Perception*, **7**(3).

Kokar, M. (2004). Situation awareness: Issues and challenges. In *The Seventh International Conference on Information Fusion*, pages 533–534.

Lavie, N., Hirst, A., de Fockert, J. W., and Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General*, **133**, 149–181.

Lysaght, R. J., Hill, S. G., Dick, A. O., Plamondon, B. D., Linton, P. M., Wierwille, W. W., Zaklad, A. L., Bittner, A. C., and Wherry, R. J. (1989). Operator workload: Comprehensive review and evaluation of operator workload methodologies. Technical report, Army Research Institude for the Behavioral and Social Sciences, Alexandria, VA.

MacLeod, C. M. (1991). Half a century of research on the stroop effect: An integrative review. *Psychological Bulletin*, **109**(2), 163–203.

Matthews, M. D., Pleban, R. J., Endsley, M. R., and Strater, L. D. (2000). Measures of infantry situation awareness for a virtual mout environment. In *Proceedings of the Human Performance, Situation Awareness and Automation: User Centered Design for the New Millenium Conference*.

Matthews, M. D., Beal, S. A., and Pleban, R. J. (2002). Situation awareness in a virtual environment: Description of a subjective measure. Research (Research Report 1786), U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA.

McCarley, J., Wickens, C., Goh, J., and Horrey, W. (2002). A computational model of attention/situation awareness. In *Perception and Performance*, pages 1669–1673. Human Factors and Ergonomics Society Annual Meeting, Human Factors and Ergonomics Society.

McGuinness, B. and Foy, L. (2000). A subjective measure of SA: The crew awareness rating scale (CARS). In *Proceedings of the First Human Performance, Situation Awareness, and Automation Conference*, Savannah, Georgia.

Moray, N. (1959). Attention in dichotic listening: affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, **11**, 56–60.

Moray, N. (1967). Where is capacity limited? A survey and a model. In A. Sanders, editor, *Attention and Performance*. North-Holland, Amsterdam.

Morrow, D.G.and Leirer, V. O. and Altieri, P. (1992). Aging, expertise, and narrative processing. *Psychology and Aging*, **7**(3), 376–388.

Navon, D. and Gopher, D. (1979). On the economy of the human processing system. *Psychological Review*, **86**, 214–253.

Neal, A., Griffin, M., Paterson, J., and Bordia, P. (1998). Human factors issues: Performance management transition to a cns/atm environment. Technical report, University of Queensland, Brisbane.

Norman, D. A. and Bobrow, D. G. (1975). On data-limited and resource limited processes. *Cognitive Psychology*, **7**, 44–64.

Ottati, W. L., Hickox, J. C., and Richter, J. (1999). Eye scan patterns of experienced and novice pilots during visual flight rules (vfr) navigation. In *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting*.

Pashler, H., Johnston, J. C., and Ruthruff, E. (2001). Attention and performance. *Annual Review of Psychology*, **52**, 626–651.

Pew, R. W. (2000). The state of situation awareness measurement: Heading toward the next century. In M. Endsley and D. Garland, editors, *Situation awareness: Analysis and Measurement*. Lawrence Erlbaum Associates.

Posner, M. I. and Boies, S. J. (1971). Components of attention. *Psychological Review*, **78**(5), 391–408.

Prince, C. and Salas, E. (1998). Situation assessment for routine flight and decision making. *International Journal of Cognitive Ergonomics*, **1**(4), 315–324.

Pritchett, A. R. and Hansman, R. J. (2000). Use of testable responses for performance-based measurement of situation awareness. In M. Endsley and G. Garland, editors, *Situation Awareness: Analysis and Measurement*. Lawrence Erlbaum Asssociates.

Richardson, D., Dale, R., and Spivey, M. (2007). Eye movements in language and cognition. In M. Gonzalez-Marquez, I. Mittelberg, S. Coulson, and M. J. Spivey, editors, *Empirical Methods in Cognitive Linguistics*, pages 323–344. John Benjamins, Amsterdam/Philadelphia.

Saito, S. and Miyake, A. (2004). On the nature of forgetting and the processing-storage relationship in reading span performance. *Journal of Memory and Language*, **50**, 425–443.

Saner, L. D., Bolstad, C. A., Gonzales, C., and Cuevas, H. M. (2009). Measuring and predicting shared situation awareness in teams. *Journal of Cognitive Engineering and Decision Making*, **3**(3), 280–308.

Sarter, N. B. and Woods, D. D. (1991). Situation awareness: A critical but ill-defined phenomenon. *International Journal of Aviation Psychology*, **1**, 45–57.

Sarter, N. B., Mumaw, R. J., and Wickens, C. D. (2007). Pilots' monitoring strategies and performance on automated flight decks: An empirical study combining behavioral and eye-tracking data. *HUMAN FACTORS*, **49**(3), 347–357.

Selkirk, C. (2009). *IFR for Professional Pilots*. Selkirk College.

Sharit, J. and Salvendy, G. (1982). Occupational stress: Review and reappraisal. *Human Factors*, **24**(2), 129–162.

Shively, R. J., Brickner, M., and Silbiger, J. (1997). A computational model of situational awareness instantiated in midas. In *the Ninth International Symposium on Aviation Psychology*, Columbus, Ohio.

Smith, B. R. and Tyler, S. (1997). The design and application of midas: A constructive simulation for human-system analysis. In *the Ninth International Symposium on Aviation Psychology*, Canberra, Australia.

Smith, K. and Hancock, P. A. (1995). Situation awareness is adaptive, externally directed consciousness. *Human Factors*, **37**(1), 137–148.

Sohn, Y. W. and Doane, S. M. (2004). Memory processes of flight situation awareness: Interactive roles of working memory capacity, long-term working memory, and expertise. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, **46**(3), 461–475.

Strater, L. D., Endsley, M. R., Pleban, R. J., and Matthews, M. D. (2001). Measures of platoon leader situation awareness in virtual decision-making exercise. Technical Report 1770, U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA.

Sukthankar, R. (1997). *Situation Awareness for Tactical Driving*. Ph.D. thesis, Carnegie Mellon University.

Sulistyawati, K., Wickens, C. D., and Chui, Y. P. (2009). Exploring the concept of team situation awareness in a simulated air combat environment. *Journal of Cognitive Engineering and Decision Making*, **3**(4), 309–330.

Taylor, R. M. (1989). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In *Proceedings of the AGARD AMP Symposium on Situational Awareness in Aerospace Operations*, Seuilly-sur Seine. NATO AGARD.

Tenney, Y. J., Adams, M. J., Pew, R. W., Huggins, W. F., and Rogers, W. H. (1992). A principled approach to the measurement of situation awareness in commercial aviation. NASA Contractor Report 4551, NASA, Washington DC.

Towse, J. N., Hitch, G. J., and Hutton, U. (2000). On the interpretation of working memory span in adults. *Memory & Cognition*, **28**, 341–348.

Treisman, A. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, **12**, 242–248.

Treisman, A. (1969). Strategies and models of selective attention. *Psychological Review*, **76**, 282–299.

Turner, M. L. and Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, **28**, 127–154.

Uhlarik, J. and Comerford, D. (2002). A review of situation awareness literature relevant to pilot surveillance functions. Technical DOT/FAA/AM-02/3, Office of Aerospace Medicine, Washington DC.

Unsworth, N., Schrock, J. C., and Engle, R. W. (2004). Working memory capacity and the antisaccade task: Individual differences in voluntary saccade control. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 1302–1321.

Unsworth, N., Heitz, R. P., Schrock, J. C., and Engle, R. W. (2005). (2005). an automated version of operation span task. *Behavioral Research Methods*, **37**, 498–505.

Van Gelder, T. J. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences*, **21**, 615–665.

Vidulich, M. A. (1989). The use of judgement matrices in subjective workload assessment: The subjective workload dominance (sword) technique. In *Proceedings of the Human Factors Society 33rd Annual Meeting*, pages 1406–1410, Santa Monica, CA. Human Factors and Ergonomics Society.

Waag, W. L. and Houck, M. R. (1994). Tools for assessing situational awareness in an operational fighter environment. *Aviation, Space and Environmental Medicine*, **65**(5), A13–A19.

Waugh, N. C. and Norman, D. A. (1965). Primary memory. *Psychological Review*, **72**, 89–104.

Wickens, C. D. (2002). Situation awareness and workload in aviation. *Current Directions in Psychological Science*, **11**(4), 128–133.

Wickens, C. D., Mountford, S. J., and Schreiner, W. (1981). Multiple resources, task-hemispheric integrity, and individual differences in time-sharing. *Human Factors*, **23**(2), 211–229.

Wolters, G. and Raffone, A. (2001). How are events represented? *Behavioral and Brain Sciences*, **24**(5), 908–909.

Zacharias, G. L., Miao, A. X., Illgen, C., Yara, J. M., and Siouris, G. M. (1995). Sample: Situation awareness model for pilot in-the-loop evaluation. In *First Annual Conference on Situation Awareness in the Tactical Air Environment*, Patuxent River, MD. Naval Air Warfare Center.

Zhang, W. and Hill, R. W. (2000). A template based and pattern-driven approach to situation awareness and assessment in virtual humans. In *Proceedings of the Fourth International Conference on Autonomous Agents*, Barcelona.

APPENDICES

APPENDIX A: PARTICIPANTS' BIOGRAPHICAL DATA

A.1 Experiment 1

Table A.1: Participants' Biographical Data (Experiment 1)

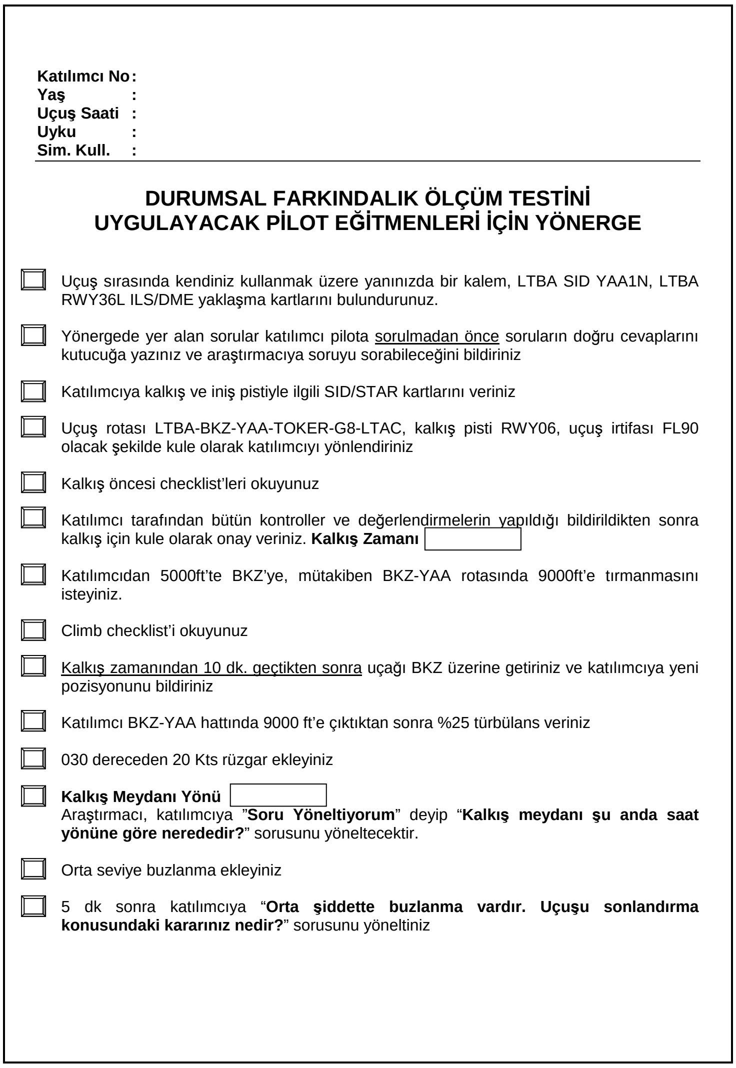
Participant	Age	Total Flight Hours
1	31	2900
2	32	3500
3	32	2700
4	30	1600
5	46	506
6	46	4500
7	32	200
8	27	190
9	28	205
10	60	10000
11	33	1754
12	49	650
13	32	205
14	31	1000
15	30	203
16	30	205
17	31	1100
18	41	4200
19	31	2100
20	43	4200
21	37	1950
22	43	4675
23	27	850
24	37	2598
25	31	1250
26	32	675
27	27	255
28	30	320
29	21	610
30	26	514
31	35	1935
32	39	5100
33	36	3700
34	33	4530
35	34	2150
36	35	2440

A.2 Experiment 2

Table A.2: Participants' Biographical Data (Experiment 2)

Participant	Age	Total Flight Hours
1	35	1935
2	39	5100
3	36	3700
4	33	4530
5	34	2150
6	35	2440
7	46	3570
8	47	4340
9	45	3346
10	46	3610

APPENDIX B: TEST INSTRUCTIONS FOR INSTRUCTOR PILOTS



Katılımcı No :
Yaş :
Uçuş Saati :
Uyku :
Sim. Kull. :

DURUMSAL FARKINDALIK ÖLÇÜM TESTİNİ UYGULAYACAK PİLOT EĞİTMENLERİ İÇİN YÖNERGE

☐ Uçuş sırasında kendiniz kullanmak üzere yanınızda bir kalem, LTBA SID YAA1N, LTBA RWY36L ILS/DME yaklaşma kartlarını bulundurunuz.

☐ Yönergede yer alan sorular katılımcı pilota sorulmadan önce soruların doğru cevaplarını kutucuğa yazınız ve araştırmacıya soruyu sorabileceğini bildiriniz

☐ Katılımcıya kalkış ve iniş pistiyle ilgili SID/STAR kartlarını veriniz

☐ Uçuş rotası LTBA-BKZ-YAA-TOKER-G8-LTAC, kalkış pisti RWY06, uçuş irtifası FL90 olacak şekilde kule olarak katılımcıyı yönlendiriniz

☐ Kalkış öncesi checklist'leri okuyunuz

☐ Katılımcı tarafından bütün kontroller ve değerlendirmelerin yapıldığı bildirildikten sonra kalkış için kule olarak onay veriniz. **Kalkış Zamanı** ☐

☐ Katılımcıdan 5000ft'te BKZ'ye, mütakiben BKZ-YAA rotasında 9000ft'e tırmanmasını isteyiniz.

☐ Climb checklist'i okuyunuz

☐ Kalkış zamanından 10 dk. geçtikten sonra uçağı BKZ üzerine getiriniz ve katılımcıya yeni pozisyonunu bildiriniz

☐ Katılımcı BKZ-YAA hattında 9000 ft'e çıktıktan sonra %25 türbülans veriniz

☐ 030 dereceden 20 Kts rüzgar ekleyiniz

☐ **Kalkış Meydanı Yönü** ☐
Araştırmacı, katılımcıya "**Soru Yöneltiyorum**" deyip "**Kalkış meydanı şu anda saat yönüne göre neredir?**" sorusunu yöneltecektir.

☐ Orta seviye buzlanma ekleyiniz

☐ 5 dk sonra katılımcıya "**Orta şiddette buzlanma vardır. Uçuşu sonlandırma konusundaki kararınız nedir?**" sorusunu yöneltiniz

☐ Katılımcı uçuşu sonlandırma yönünde karar verirse "**ATA'ya iniş için ERMAN'a devam ediniz. 3000ft'te alçalış serbest, QNH 1013**" deyiniz. Katılımcı uçuşa devam etmeye karar verirse buzlanma seviyesini artırınız ve 1 dk. sonra uçuşu iptal etmesini isteyiniz.

☐ 10 dk. boyunca katılımcının ERMAN noktasına nasıl ulaşabileceği konusunda ürettiği çözümleri uygulamasını sağlayınız. Talep edildiği durumda bu süre içerisinde GCA'in gayri faal olduğunu bildiriniz

☐ Katılımcıyı - talep etmese de- GCA ile ERMAN noktasına yönlendiriniz.

☐ **Erman'a Mesafe**
Araştırmacı, katılımcıya "**Soru Yöneltiyorum**" deyip "**ERMAN'a olan mesafeniz nedir?**" sorusunu yöneltecektir.

☐ **Rüzgarın Yönü**
Araştırmacı, katılımcıya "**Soru Yöneltiyorum**" deyip "**Rüzgarın yönünü ne olarak tahmin ediyorsunuz?**" sorusunu yöneltecektir.

☐ Simülasyonu dondurunuz ve katılımcıdan göstergelere ve kullandığı kartlara bakmadan sorulara (Ek-A) hafızasında kaldığı kadarıyla cevap vermesini isteyiniz
IAS **MSL** **Heading** **ERMAN'A SÜRE**

☐ ATC olarak hava durum bilgisini "**Görüş 830m, Bulut Alttavanı 300ft'tir**" olarak veriniz

☐ Katılımcı alçalma için klerans istediğinde ATATURK Kule olarak "**ERMAN üstünde 2700 ft'de bir beklemeye müteakip RWY36L ILS/DME yaklaşması için klerans verilmiştir**" deyiniz

☐ **Beklemeye Giriş Usulü**
ERMAN'a gelmeden araştırmacı katılımcıya "**Soru Yöneltiyorum**" deyip "**Beklemeye giriş usulünüz ne olacaktır?**" sorusunu yöneltecektir.

☐ ERMAN'da bekleme sırasında Inbound-Outbound dönüşü yapılırken RMI pusulasını ve HSI'ı dondurunuz

☐ **Baş Açısı**
Araştırmacı, katılımcıya Inbound-Outbound dönüşü sırasında "**Soru Yöneltiyorum**" deyip "**Baş açınız nedir?**" sorusunu yöneltecektir.

☐ ERMAN'da bekleme sırasında Dikey Hız Göstergesi'ni dondurunuz

☐ Katılımcının görsel temas sağlayamayıp pisti pas geçmesi için bulut alttavanını limit dışına çekiniz

☐ Simülasyonu dondurunuz ve katılımcıdan göstergelere ve kullandığı kartlara bakmadan sorulara (Ek-B) hafızasında kaldığı kadarıyla cevap vermesini isteyiniz
GS **Bir Sonraki ATC Mesajı** **MSL**

☐ Katılımcı ILS zarfına oturduktan sonra "**Soru Yöneltiyorum**" deyiniz ve "**Pas geçme usulünüz nedir?**" sorusunu yöneltiniz

☐ Approach & Landing Checklist’i okuyunuz.

☐ Katılımcının pisti pas geçmesini müteakip Climb Checklist’i okuyunuz

☐ 1-2 dakika sonra havaaracı pozisyonunu değiştirip ERMAN üzerine getiriniz

☐ Katılımcıya “**Süre kazanmak adına havaaracı ERMAN üzerine getirilmiştir**” deyiniz

☐ Bulut Alttavanını inişin gerçekleşebilmesi için 300ft’e çıkarınız

☐ Katılımcının alçalma için klerans istemesini müteakip ATATURK Kule olarak ILS/DME yaklaşması için klerans veriniz

☐ Approach & Landing Checklist’i okuyunuz.

☐ **İrtifa** ☐
Araştırmacı, katılımcıya ”**Soru Yöneltiyorum**” deyip “**İrtifanız nedir?**” sorusunu yöneltecektir

☐ Hız göstergesi’ni dondurunuz

☐ **Süzülüş Oranı** ☐
Araştırmacı, katılımcıya ”**Soru Yöneltiyorum**” deyip “**Süzülüş oranınız nedir?**” sorusunu yöneltecektir.

☐ İnişten sonra katılımcıya teste katılımdan ötürü teşekkür ediniz

APPENDIX C: DETAILED FLIGHT SCENARIO

C.1 Flight Scenario

Flight Scenario includes both pilot and instructor actions. Approach charts are used by the pilots during approach and landing phases. Take-off airport's YAA 1N departure chart[1] in Figure C.1 shows the waypoints and navigation aids (BKZ, YAA, and IST) mentioned in the following scenario developed.

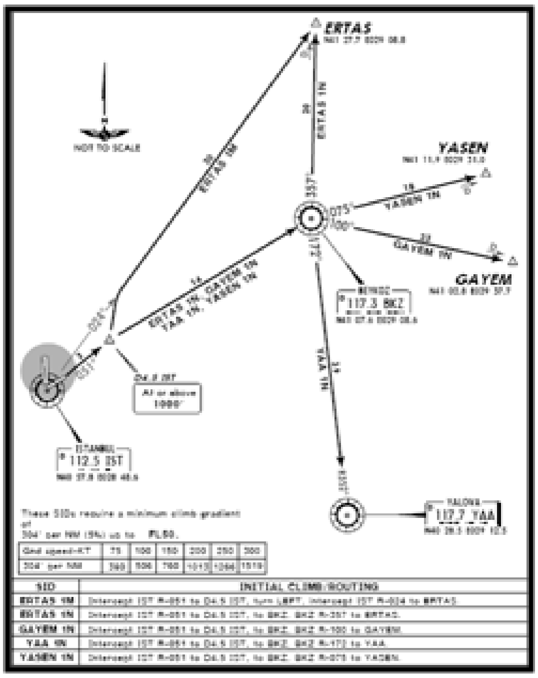


Figure C.1: Airport's Departure Chart

---

[1] A chart defining a pathway out of an airport and onto the airway structure

The participants take off from Runway 06 of ATATURK Airport (LTBA - international code of the airport). They climb to 1,000 ft on runway heading to proceed to BKZ by climbing to 5,000ft. A few minor events are inserted after this low-workload segment. Moderate turbulance (25%) is inserted by the instructor. Rate of climb[2] and Indicated Air Speed[3] are expected to decrease significantly because of the turbulance. Crosswind is inserted by the instructor. Crosswind and moderate turbulance are used to make sure the participants pay attention to the environmental conditions. The instructor sets the cloud ceiling to 150ft (low ceiling) and cloud top to 10,000ft. At BKZ, the participants turn right to proceed to YAA by climbing to 9,000ft. In-between BKZ-YAA line, the participants may decide to terminate the flight due to the bad weather conditions. Moderate icing condition is inserted by the instructor to guarantee the flight termination.

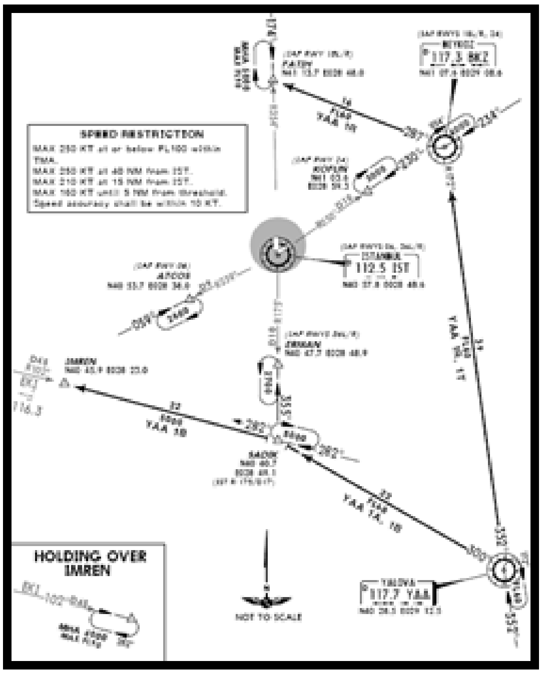


Figure C.2: Airport's Arrival Chart

After recognizing the icing condition, the instructor as Air Traffic Controller asks the pilot about their "Go/No-Go" decision[4]. The weather is poor because of the icing condition and they are supposed to terminate the flight regarding the deficiency of

---

[2] The speed at which an aircraft is gaining (or losing) altitude

[3] The airspeed indicator reading uncorrected for instrument, position, and other errors

[4] In case of emergency conditions including deteriorated weather and important malfunctions in subsystems, pilot makes a decision about flight termination.

anti-icing system of the aircraft. The participants decide to terminate the flight and return to take-off airport LTBA. Otherwise the instructor enforces the participants to terminate the flight by making the weather conditions worser.

When the participants receive normal weather condition from the LTBA ATIS, they are supposed to discover that they can land on LTBA. The tower gives clearance for ILS/DME approach[5] via ERMAN depicted in Figure C.2. The tower requests one hold at ERMAN. The participants establish the holding pattern in a suitable way.

The compass of RMI is frozen. The participants can use only old-fashion magnetic compass which is not user-friendly. Later on, VSI is frozen and the participants are unable to get the vertical speed. Immediately after that failure, the pilot is supposed to report it. The instructor sets cloud ceiling to 100ft to make sure that the participants cannot land on LTBA and follow the Missed Approach Procedure. The tower warns the pilot about the possibility to follow Missed Approach Procedure because of the low ceiling.

The participants are supposed to reach the Missed Approach Point and to follow the Missed Approach Procedure to pass the runway because they cannot have visual contact with the ground at the decision height above the surface. The participants climb on runway heading to 1,500ft on IS, then turn left to proceed to CEK depicted in Figure C.3. The instructor sets cloud ceiling to 300ft to make LTBA available for landing and the participants are directed to ERMAN again to establish on ILS envelope. The tower gives clearance for descent. They establish on ILS envelope. Airspeed Indicator (ASI) is frozen and the pilot is supposed to report it immediately. The participants are supposed to keep the Indicated Air Speed around 90 knots by using artificial horizon display with acceptable vertical and horizontal errors and successfully land the aircraft.

Figure C.4 depicts the events points on the flight route. The flight instructor generated below events more or less at the same point and at the same time of the flight for all participants in order to administer the queries in the same workload and fatigue

---

[5] Lateral and vertical guidance as well as distance information to airport for approaching

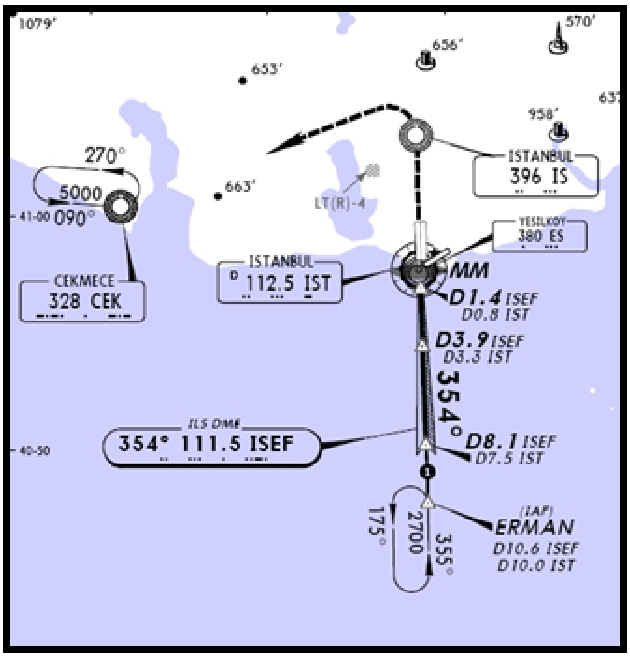


Figure C.3: Airport's Approach Chart

conditions. Since level of workload and fatigue are changing throughout the flight, occurences of the events in the same workload and fatigue conditions important factor for validity of measurements. It is clear that RTs would be different under low and high workload conditions for the same query because it is difficult to answer the query in high workload condition.

E1: Instructor adds 25% turbulence,

E2: Instructor adds 20 Kts wind from 030,

E3: Instructor administers the query "Where is the take-off airport in the clock direction?",

E4: Instructor adds moderate icing condition,

E5: Pilot terminates the flight due to the deteriorated weather conditions,

E5: Instructor directs the participant to ERMAN for landing on LTBA,

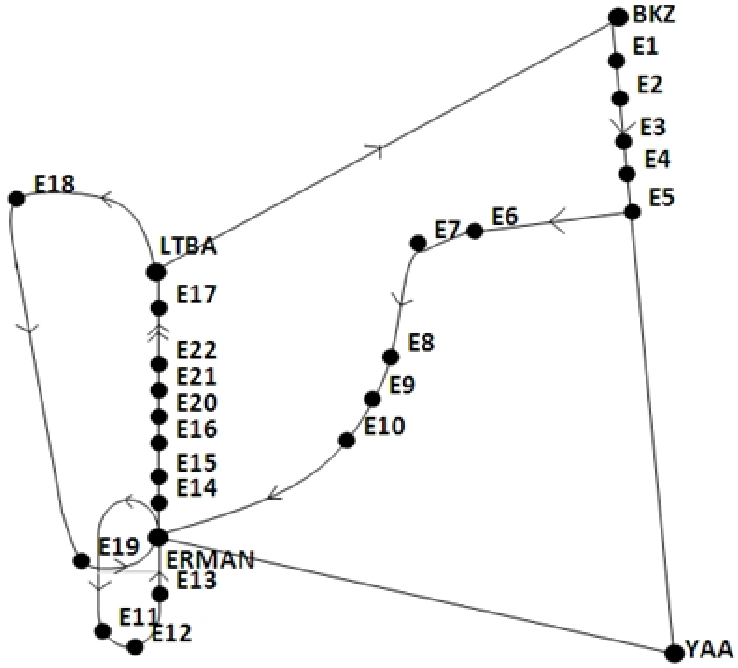


Figure C.4: Flight Route and Designed Event Points in the Scenario

E5: Pilot flies toward LTBA for closing 10 NM to make 10 DME arc to ERMAN,

E6: Instructor administers the query "What is your distance to ERMAN?",

E7: Pilot starts to make 10 DME arc to reach ERMAN

E8: Instructor administers the query "What is the wind direction?",

E9: Instructor administers the query "What is your entry method for holding at ERMAN?",

E10: Instructor freezes the simulator and administers the offline queries in Appendix-D,

E11: Instructor freezes the RMI/HSI compass

E12: Instructor administers the query "What is your heading?",

E13: Instructor freezes the VSI,

E14: Instructor decreases the ceiling to enforce the participant for a missed approach,

E15: Instructor freezes the simulator and administers the offline queries in Appendix-E,

E16: Instructor administers the query "What is the Missed Approach Procedure?",

E17: Pilot decides to follow the missed approach procedure,

E18: Instructor puts the aircraft near ERMAN,

E19: Instructor increases the ceiling to allow the participant to safely land,

E20: Instructor administers the query "What is your altitude?",

E21: Instructor freezes the ASI,

E22: Instructor administers the query "What is your descent rate?"

C.2 SA Queries

RT-SAQ1

RT-SAQ1 is the response time to the online query "Where is the take-off airport in terms of clock direction?". The participant's response to the the online query depends on the aircraft position where the query is administered. Clock direction implies a fuzziness in the answer. Correct answer (for instance, 3 in clock direction) is evaluated as 100% accurate. 2 and 4 are evaluated as 75% accurate and correct answers as well. Less than 75% accuracy is not taken into account and is regarded as wrong response.

RT-SAQ2

RT-SAQ2 is the response time to the online query "In what direction do you think the wind is blowing?". The participant's response to this online query includes a fuzziness as well. Correct answer (for instance, 3 in clock direction) is evaluated as 100% accurate. 2 and 4 are evaluated as 75% accurate and correct answers as well. Less than 75% accuracy is not taken into account and is regarded as wrong response.

RT-SAQ3

RT-SAQ3 is the response time to the online query "What is your distance to next waypoint?". Correct answer to this online query can be calculated by the participants and it depends on the aircraft position where the query is administered. Distance Measurement Equipment (DME) is a transponder-based radio navigation technology that measures distance by timing the propagation delay of radio signals. Participants are supposed to use relevant DME site by setting its frequency in the DME receiver. The query will be administered after the participants decide to terminate the flight and direct to next waypoint in response to ATC guidance.

RT-SAQ4

RT-SAQ4 is the response time to the online query "What will your entry method be for holding pattern?". This online query is a procedural question. There are three types of entry method to a holding pattern: Teardrop, direct and parallel entry. Selection of the correct type depends on the orientation of holding pattern, heading of aircraft, and turn direction while holding.

RT-SAQ5

RT-SAQ5 is the response time to the online query "What is your heading?".This online query is normally a straight forward question and the participants can easily answer this question by looking at the RMI compass display. In case of RMI compass (or directional gyro) failure, participants should use the magnetic compass.

The magnetic compass can only provide reliable heading information while the flight is straight and unaccelerated. While turning, reporting the correct heading requires some mental effort. Correct answer could be actual heading plus or minus 5 degree and depends on the aircraft position where the query is administered.

RT-SAQ6

RT-SAQ6 is the response time to the online query "What is the Missed Approach Procedure?". It is designed for preparing the participants about following the missed approach procedure. In case of pilot/copilot configuration, pilot can brief the missed approach procedure with the copilot before starting the approach. Since the copilot is not used in the experiment, the participants could not have a chance to brief the missed approach procedure. For that reason, the participants are allowed to read the procedure on the relevant chart. This online query is not used in the analysis.

RT-SAQ7

RT-SAQ7 is the response time to the online query "What is your altitude in MSL?". Altitude is one of the few flight variables pilots should continuously check. This online query can be answered easily by looking at the altimeter. If the current altitude is in the participants' working memory at the time the query is administered, they can answer the query a little bit faster than others who look at the altimeter. The answer is correct if it is in the range of (current altitude plus or minus 50ft).

RT-SAQ8

RT-SAQ8 is the response time to the Descent rate is displayed in VSI (Vertical Speed Indicator). In case of VSI failure, the participants have to calculate and stabilize the descent rate by using the altimeter changes in a minute and the artificial horizon. The online query "What is your descent rate?" is administered after the VSI failure occurs. The answer is correct if it is in the range of (current descent rate plus or minus 100ft/min).

ACC-SAQ9

Participants are required to check their airspeed continuously on the speed indicator. They usually fly at a preplanned airspeed compliant to the aviation regulations for Instrument Flight Rules (IFR) but some meteorological events as in the flight scenario may create a deviation from the preplanned airspeed. The offline query "What is your indicated Air Speed?" questions whether the participants continuously check it or they are unaware of what airspeed the aircraft flies at. The answer is regarded as 100% accurate if it is correct. Accuracy for the other answers is 0%.

ACC-SAQ10

The offline query "What is your altitude?" questions another dimension of aircraft's position information that has to be checked continuously by the pilots. Generally speaking, cruise altitude is constant throughout a certain flight leg except climbing and descending. Frequency of checking the altimeter increases while climbing or descending. During the level flight, some meteorological conditions such as icing, turbulence may bring about a difficulty in maintaining the cruise altitude. The pilots are also expected to frequently check the altimeter display in such cases. The answer is regarded as 100% accurate if it is correct. Accuracy for the other answers is 0%.

ACC-SAQ11

Directional gyro which displays the heading information is regarded as one of the most observed flight displays during the flight. The offline query "What is your heading?" simply questions "in which direction the participant is going". Crosswind as dependent on its magnitude, may make the course heading deviated from what it is desired. In case of crosswind, the pilot is always in the process of making corrections for maintaining the desired heading by turning the aircraft's head towards the direction to which the wind is blowing. The answer is regarded as 100% accurate if it is correct. Accuracy for the other answers is 0%.

ACC-SAQ12

The offline query "How much time will it take to reach to the next waypoint?" questions the positional awareness of the participants. In the flight plans, the pilots calculate the arrival times for the waypoints before take-off and they are supposed to roughly know the time required to reach to the next waypoint. If the next waypoint supports DME, this query turns out to be a working memory question. Otherwise, it requires a little computation as well. The answer is regarded as 100% accurate if it is correct. Accuracy for the other answers is 0%.

ACC-SAQ13

The offline query "What is your position on the map?" is administered for measuring the positional awareness of the participants. The map includes the positions of the scenario related waypoints and the participant is supposed to mark aircraft's relative position to the waypoints shown. Accuracy is calculated as:

Accuracy (%) = 100 - [(Dist (marked position, actual position) / Leg Distance] * 100

where Dist is a function that calculates euclidean distance between marked position and actual position.

ACC-SAQ14

One of the indicators in high SA is estimating the next event or status of aircraft during flight. The offline query "What will be the next ATC call?" questions whether participant is able to correctly anticipate the next ATC call. The query has one correct answer. Accuracy is 0% for the other answers.

ACC-SAQ15

Before starting to approach to the landing airport, the pilots are expected to brief the necessary procedures including Missed Approach Procedure. The offline query "What will be the first altitude to climb in Missed Approach Procedure?" questions information pertaining to that procedure. The query has one correct answer. Accuracy is 0% for the other answers.

ACC-SAQ16

Ground speed and indicated airspeed are different in case of windy weather. If the wind is headwind, ground speed is less than the indicated airspeed. If the wind is tailwind, ground speed is greater than the indicated airspeed. The offline query "What is your ground speed?" questions the meteorological awareness of the participants. The query has one correct answer. Accuracy for the other answers is calculated as follows:

Accuracy (%) = 100 - [(abs(reported ground speed- actual ground speed)) / actual ground speed] * 100

ACC-SAQ17

The approach charts the pilots are supposed to brief before starting to approach cover necessary information including the frequencies. The offline query "What is the approach ILS frequency?" questions the frequency that the participants use just before starting to conduct ILS approach. There is only one correct answer. Accuracy for the other answers is 0%.

ACC-SAQ18

Decision altitude (DA) is written in the airport's approach chart. DA is required for the Missed Approach Procedure. The offline query "What is the decision altitude (DA) for the landing airport?" questions whether the participant knows what he/she should know during the approach. There is only one correct answer. Accuracy for the other answers is 0%.

ACC-SAQ19

The offline query "What is the current altitude?" questions the positional awareness of the participant. Correct answer is [actual altitude plus or minus 100ft]. Accuracy of other answers is calculated as follows:

Accuracy (%) = 100 - [(abs(reported altitude- actual altitude)) / actual altitude] * 100

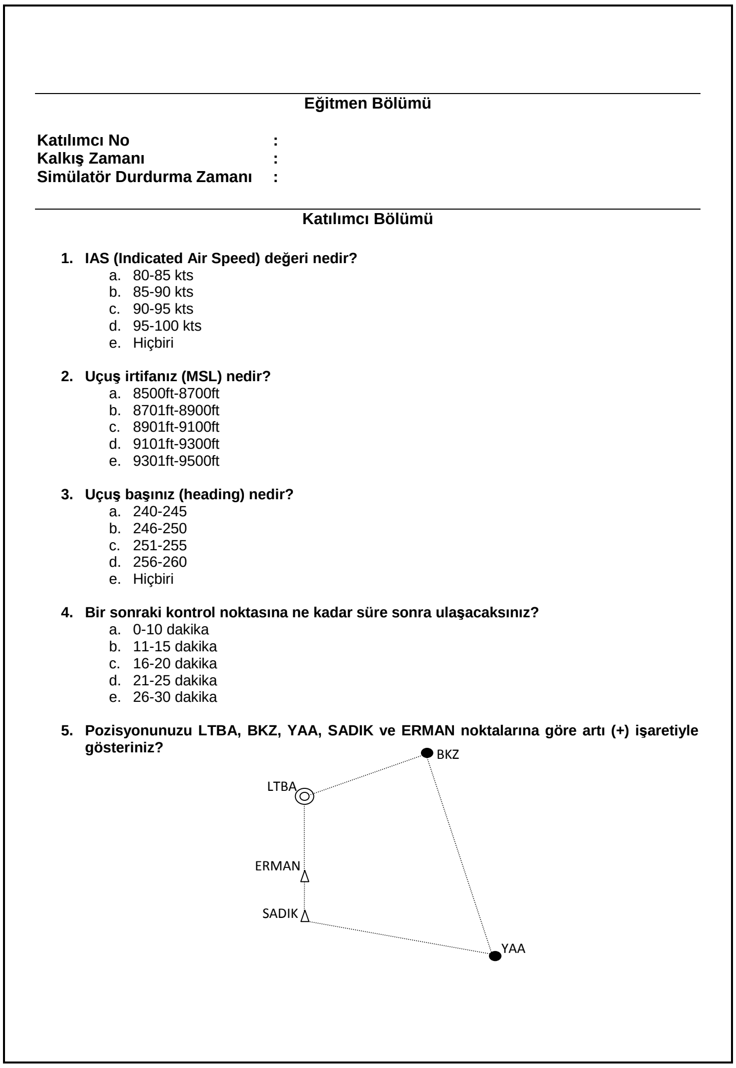
ACC-SAQ20

The offline query "What is the airport elevation?" questions information on the approach charts. There is only one correct answer. Accuracy for the other answers is 0%.

ACC-SAQ21

Barometric altimeter should be calibrated by the pilots in accordance with the local barometric pressure. Before approaching, Airport tower gives the barometric reference (QNH) to the pilots. The offline query "What is the value of QNH?" questions that information. There is only one correct answer. Accuracy for the other answers is 0%

APPENDIX D: FORM FOR OFFLINE SA QUERIES (Set #1)



Eğitmen Bölümü

Katılımcı No :
Kalkış Zamanı :
Simülatör Durdurma Zamanı :

Katılımcı Bölümü

1. **IAS (Indicated Air Speed) değeri nedir?**
   a. 80-85 kts
   b. 85-90 kts
   c. 90-95 kts
   d. 95-100 kts
   e. Hiçbiri

2. **Uçuş irtifanız (MSL) nedir?**
   a. 8500ft-8700ft
   b. 8701ft-8900ft
   c. 8901ft-9100ft
   d. 9101ft-9300ft
   e. 9301ft-9500ft

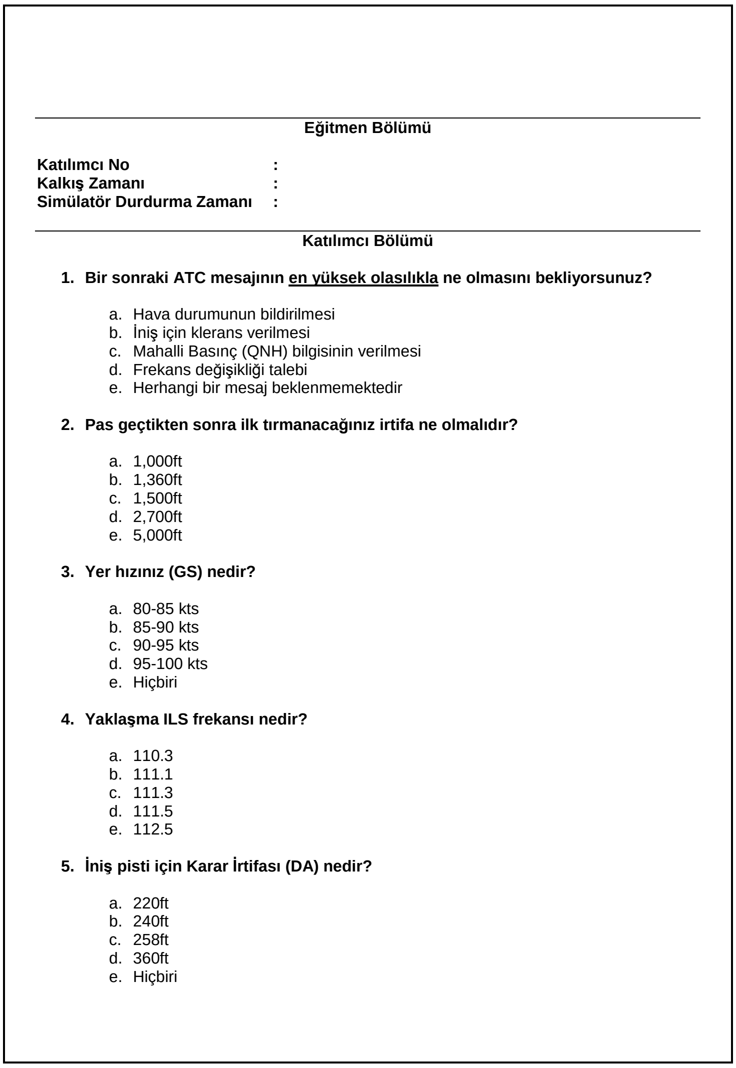
3. **Uçuş başınız (heading) nedir?**
   a. 240-245
   b. 246-250
   c. 251-255
   d. 256-260
   e. Hiçbiri

4. **Bir sonraki kontrol noktasına ne kadar süre sonra ulaşacaksınız?**
   a. 0-10 dakika
   b. 11-15 dakika
   c. 16-20 dakika
   d. 21-25 dakika
   e. 26-30 dakika

5. **Pozisyonunuzu LTBA, BKZ, YAA, SADIK ve ERMAN noktalarına göre artı (+) işaretiyle gösteriniz?**

()

APPENDIX E: FORM FOR OFFLINE SA QUERIES (Set#2)



Eğitmen Bölümü

Katılımcı No :
Kalkış Zamanı :
Simülatör Durdurma Zamanı :

Katılımcı Bölümü

1. **Bir sonraki ATC mesajının <u>en yüksek olasılıkla</u> ne olmasını bekliyorsunuz?**
    a. Hava durumunun bildirilmesi
    b. İniş için klerans verilmesi
    c. Mahalli Basınç (QNH) bilgisinin verilmesi
    d. Frekans değişikliği talebi
    e. Herhangi bir mesaj beklenmemektedir

2. **Pas geçtikten sonra ilk tırmanacağınız irtifa ne olmalıdır?**
    a. 1,000ft
    b. 1,360ft
    c. 1,500ft
    d. 2,700ft
    e. 5,000ft

3. **Yer hızınız (GS) nedir?**
    a. 80-85 kts
    b. 85-90 kts
    c. 90-95 kts
    d. 95-100 kts
    e. Hiçbiri

4. **Yaklaşma ILS frekansı nedir?**
    a. 110.3
    b. 111.1
    c. 111.3
    d. 111.5
    e. 112.5

5. **İniş pisti için Karar İrtifası (DA) nedir?**
    a. 220ft
    b. 240ft
    c. 258ft
    d. 360ft
    e. Hiçbiri

6. **Geçilen İrtifa nedir?**

   a. 2500-2700ft
   b. 2300-2499ft
   c. 2100-2299ft
   d. 1900-2199ft
   e. Hiçbiri

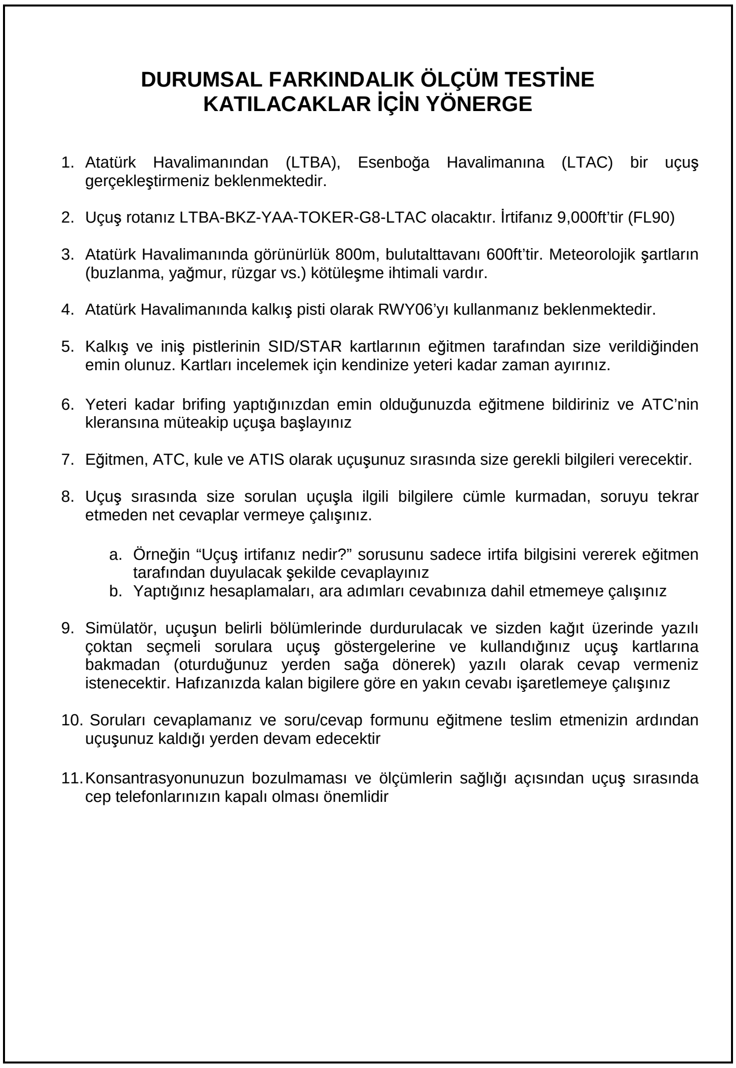
7. **Meydan İrtifası nedir?**

   a. 90ft
   b. 93ft
   c. 102ft
   d. 163ft
   e. Hiçbiri

8. **Mahalli Basınç (QNH) nedir?**

   a. 996 mb
   b. 1010 mb
   c. 1013 mb
   d. 1020 mb
   e. Hiçbiri

APPENDIX F: TEST INSTRUCTIONS FOR PARTICIPANTS



DURUMSAL FARKINDALIK ÖLÇÜM TESTİNE KATILACAKLAR İÇİN YÖNERGE

1. Atatürk Havalimanından (LTBA), Esenboğa Havalimanına (LTAC) bir uçuş gerçekleştirmeniz beklenmektedir.

2. Uçuş rotanız LTBA-BKZ-YAA-TOKER-G8-LTAC olacaktır. İrtifanız 9,000ft'tir (FL90)

3. Atatürk Havalimanında görünürlük 800m, bulutalttavanı 600ft'tir. Meteorolojik şartların (buzlanma, yağmur, rüzgar vs.) kötüleşme ihtimali vardır.

4. Atatürk Havalimanında kalkış pisti olarak RWY06'yı kullanmanız beklenmektedir.

5. Kalkış ve iniş pistlerinin SID/STAR kartlarının eğitmen tarafından size verildiğinden emin olunuz. Kartları incelemek için kendinize yeteri kadar zaman ayırınız.

6. Yeteri kadar brifing yaptığınızdan emin olduğunuzda eğitmene bildiriniz ve ATC'nin kleransına müteakip uçuşa başlayınız

7. Eğitmen, ATC, kule ve ATIS olarak uçuşunuz sırasında size gerekli bilgileri verecektir.

8. Uçuş sırasında size sorulan uçuşla ilgili bilgilere cümle kurmadan, soruyu tekrar etmeden net cevaplar vermeye çalışınız.

    a. Örneğin "Uçuş irtifanız nedir?" sorusunu sadece irtifa bilgisini vererek eğitmen tarafından duyulacak şekilde cevaplayınız
    b. Yaptığınız hesaplamaları, ara adımları cevabınıza dahil etmemeye çalışınız

9. Simülatör, uçuşun belirli bölümlerinde durdurulacak ve sizden kağıt üzerinde yazılı çoktan seçmeli sorulara uçuş göstergelerine ve kullandığınız uçuş kartlarına bakmadan (oturduğunuz yerden sağa dönerek) yazılı olarak cevap vermeniz istenecektir. Hafızanızda kalan bigilere göre en yakın cevabı işaretlemeye çalışınız

10. Soruları cevaplamanız ve soru/cevap formunu eğitmene teslim etmenizin ardından uçuşunuz kaldığı yerden devam edecektir

11. Konsantrasyonunuzun bozulmaması ve ölçümlerin sağlığı açısından uçuş sırasında cep telefonlarınızın kapalı olması önemlidir

APPENDIX G: INFORMED CONSENT FORM

Gönüllü Katılım Formu

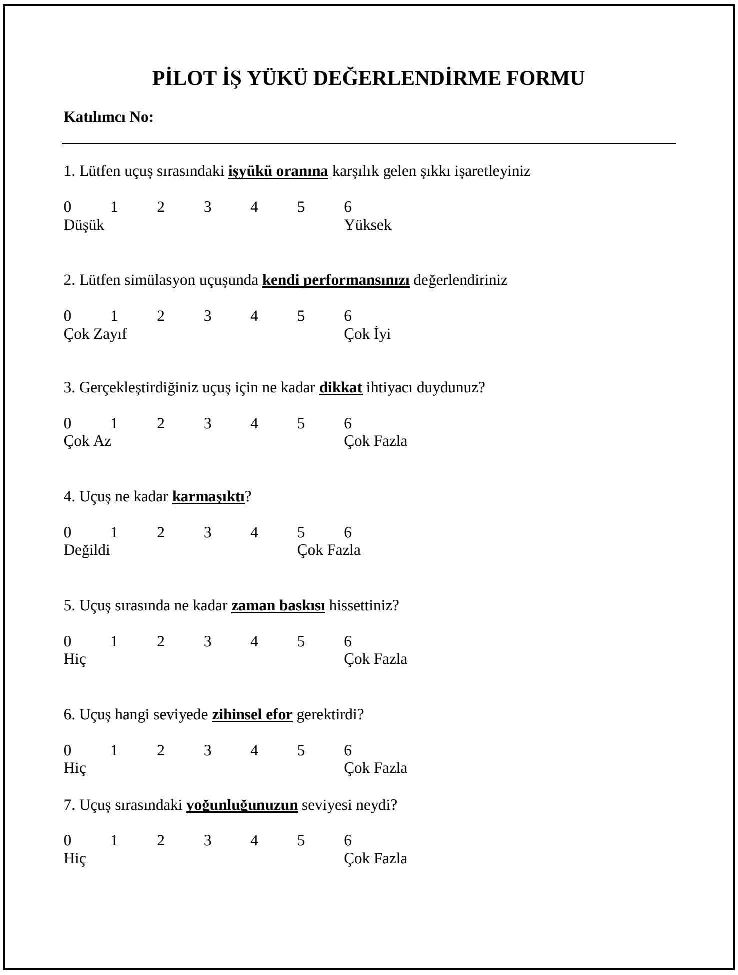
Bu çalışma, Serkan ÇAK, Yrd.Doç.Dr. Mine MISIRLISOY ve Dr. Bilge SAY tarafından durumsal farkındalığın çalışma belleği ve dikkat mekanizmalarını ortaya çıkarmaya yönelik bir çalışmadır. Bu kapsamda askeri/sivil pilotlar ve kontrol grubundan bilgi toplanması hedeflenmektedir. Çalışmaya katılım tamamen gönüllülük temelinde olmalıdır. Deney öncesi, sizden kimlik belirleyici hiçbir bilgi istenmemektedir. Cevaplarınız tamamen gizli tutulacak ve sadece araştırmacılar tarafından değerlendirilecektir; elde edilecek bilgiler bilimsel yayınlarda kullanılacaktır.

Deney, genel olarak kişisel rahatsızlık verecek sorular, aktiviteler içermemektedir. Ancak, katılım sırasında sorulardan ya da herhangi başka bir nedenden ötürü kendinizi rahatsız hissederseniz cevaplama işini veya deneyi yarıda bırakıp çıkmakta serbestsiniz. Böyle bir durumda deneyi uygulayan kişiye, deneyi tamamlamadığınızı söylemeniz yeterli olacaktır. Deney sonunda, bu çalışmayla ilgili sorularınız cevaplandırılacaktır. Bu çalışmaya katıldığınız için şimdiden teşekkür ederiz. Çalışma hakkında daha fazla bilgi almak için Psikoloji Bölümü öğretim üyelerinden Yrd. Doç. Dr. Mine MISIRLISOY (Tel: 210 51 07; E-posta: mmine@ii.metu.edu.tr), Dr. Bilge SAY (Tel: 210 37 48; E-posta: bsay@ii.metu.edu.tr) ya da Serkan ÇAK (Tel: 266 37 50; E-posta: scak@stm.com.tr) ile iletişim kurabilirsiniz.

***Bu çalışmaya tamamen gönüllü olarak katılıyorum ve istediğim zaman yarıda kesip çıkabileceğimi biliyorum. Verdiğim bilgilerin bilimsel amaçlı yayımlarda kullanılmasını kabul ediyorum*.** (Formu doldurup imzaladıktan sonra uygulayıcıya geri veriniz).

İsim Soyad	Tarih	İmza
	----/----/-----	

APPENDIX H: PILOT WORKLOAD SURVEY



PİLOT İŞ YÜKÜ DEĞERLENDİRME FORMU

Katılımcı No:

1. Lütfen uçuş sırasındaki **işyükü oranına** karşılık gelen şıkkı işaretleyiniz

0 1 2 3 4 5 6
Düşük Yüksek

2. Lütfen simülasyon uçuşunda **kendi performansınızı** değerlendiriniz

0 1 2 3 4 5 6
Çok Zayıf Çok İyi

3. Gerçekleştirdiğiniz uçuş için ne kadar **dikkat** ihtiyacı duydunuz?

0 1 2 3 4 5 6
Çok Az Çok Fazla

4. Uçuş ne kadar **karmaşıktı**?

0 1 2 3 4 5 6
Değildi Çok Fazla

5. Uçuş sırasında ne kadar **zaman baskısı** hissettiniz?

0 1 2 3 4 5 6
Hiç Çok Fazla

6. Uçuş hangi seviyede **zihinsel efor** gerektirdi?

0 1 2 3 4 5 6
Hiç Çok Fazla

7. Uçuş sırasındaki **yoğunluğunuzun** seviyesi neydi?

0 1 2 3 4 5 6
Hiç Çok Fazla

8. Uçuşun **zorluk derecesi** neydi?

0 1 2 3 4 5 6
Kolay Zor

9. Lütfen bu uçuştaki **motivasyonunuzu** değerlendiriniz?

0 1 2 3 4 5 6
Düşük Yüksek

10. Uçuştan sonra kendinizi nasıl **hissettiniz**?

0 1 2 3 4 5 6
Dinç Yorgun

0 1 2 3 4 5 6
Sakin Gergin

11. Aşağıdaki uçuş fazlarında maruz kaldığınız **işyükünü** notlayınız?

Kalkış
0 1 2 3 4 5 6
Düşük Yüksek

LTBA-BKZ Arası
0 1 2 3 4 5 6
Düşük Yüksek

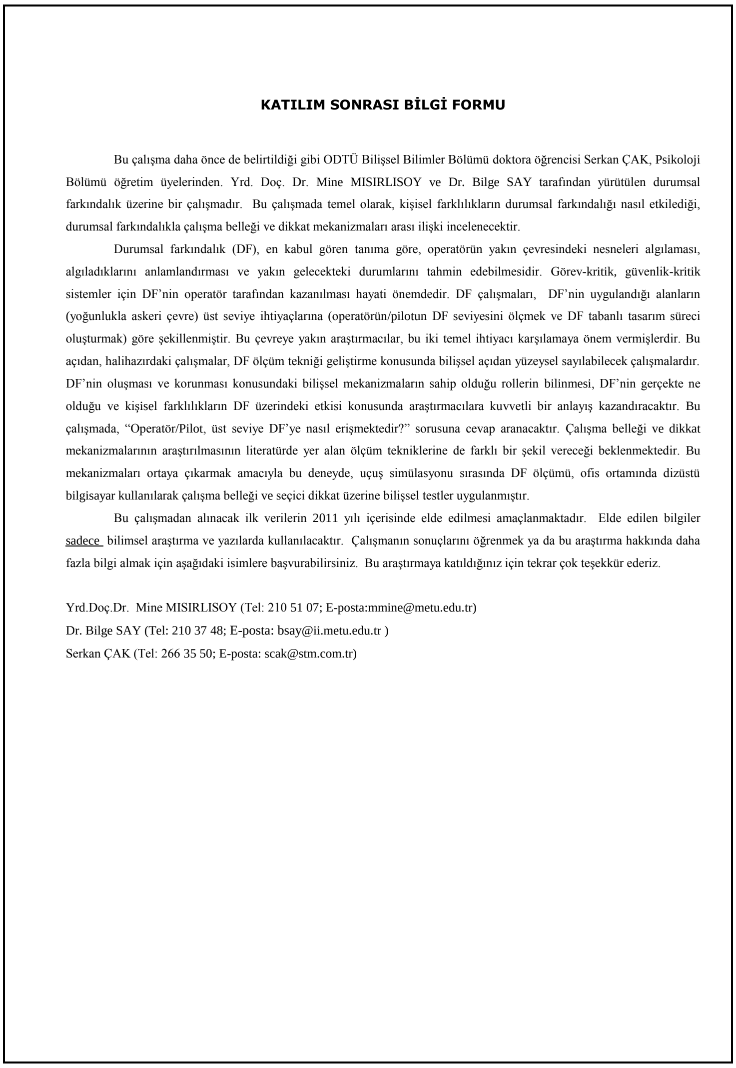
BKZ-YAA Arası
0 1 2 3 4 5 6
Düşük Yüksek

BKZ-YAA’dan ERMAN’a dönüş
0 1 2 3 4 5 6
Düşük Yüksek

ERMAN’da bekleme
0 1 2 3 4 5 6
Düşük Yüksek

Yaklaşma ve Pas Geçme
0 1 2 3 4 5 6
Düşük Yüksek

APPENDIX I: DEBRIEFING SHEET



KATILIM SONRASI BİLGİ FORMU

Bu çalışma daha önce de belirtildiği gibi ODTÜ Bilişsel Bilimler Bölümü doktora öğrencisi Serkan ÇAK, Psikoloji Bölümü öğretim üyelerinden. Yrd. Doç. Dr. Mine MISIRLISOY ve Dr. Bilge SAY tarafından yürütülen durumsal farkındalık üzerine bir çalışmadır. Bu çalışmada temel olarak, kişisel farklılıkların durumsal farkındalığı nasıl etkilediği, durumsal farkındalıkla çalışma belleği ve dikkat mekanizmaları arası ilişki incelenecektir.

Durumsal farkındalık (DF), en kabul gören tanıma göre, operatörün yakın çevresindeki nesneleri algılaması, algıladıklarını anlamlandırması ve yakın gelecekteki durumlarını tahmin edebilmesidir. Görev-kritik, güvenlik-kritik sistemler için DF'nin operatör tarafından kazanılması hayati önemdedir. DF çalışmaları, DF'nin uygulandığı alanların (yoğunlukla askeri çevre) üst seviye ihtiyaçlarına (operatörün/pilotun DF seviyesini ölçmek ve DF tabanlı tasarım süreci oluşturmak) göre şekillenmiştir. Bu çevreye yakın araştırmacılar, bu iki temel ihtiyacı karşılamaya önem vermişlerdir. Bu açıdan, halihazırdaki çalışmalar, DF ölçüm tekniği geliştirme konusunda bilişsel açıdan yüzeysel sayılabilecek çalışmalardır. DF'nin oluşması ve korunması konusundaki bilişsel mekanizmaların sahip olduğu rollerin bilinmesi, DF'nin gerçekte ne olduğu ve kişisel farklılıkların DF üzerindeki etkisi konusunda araştırmacılara kuvvetli bir anlayış kazandıracaktır. Bu çalışmada, "Operatör/Pilot, üst seviye DF'ye nasıl erişmektedir?" sorusuna cevap aranacaktır. Çalışma belleği ve dikkat mekanizmalarının araştırılmasının literatürde yer alan ölçüm tekniklerine de farklı bir şekil vereceği beklenmektedir. Bu mekanizmaları ortaya çıkarmak amacıyla bu deneyde, uçuş simülasyonu sırasında DF ölçümü, ofis ortamında dizüstü bilgisayar kullanılarak çalışma belleği ve seçici dikkat üzerine bilişsel testler uygulanmıştır.

Bu çalışmadan alınacak ilk verilerin 2011 yılı içerisinde elde edilmesi amaçlanmaktadır. Elde edilen bilgiler sadece bilimsel araştırma ve yazılarda kullanılacaktır. Çalışmanın sonuçlarını öğrenmek ya da bu araştırma hakkında daha fazla bilgi almak için aşağıdaki isimlere başvurabilirsiniz. Bu araştırmaya katıldığınız için tekrar çok teşekkür ederiz.

Yrd.Doç.Dr. Mine MISIRLISOY (Tel: 210 51 07; E-posta:mmine@metu.edu.tr)
Dr. Bilge SAY (Tel: 210 37 48; E-posta: bsay@ii.metu.edu.tr)
Serkan ÇAK (Tel: 266 35 50; E-posta: scak@stm.com.tr)

APPENDIX J: EXPERIMENT 1 RESULTS

Table J.1: Experiment 1: SA Measurements and Cognitive Capacity Tests

Participant	WMC	STRP	CRT	EXP	OFFLINE	ONLINE
1	59	866	459	210	12.76	-0.593
2	57	147	503	72	8.14	-0.627
3	21	113	612	68	5.17	-0.577
4	3	329	465	48	4.99	0.164
5	57	324	504	25	9.20	-0.275
6	46	594	470	63	8.40	-0.248
7	64	812	469	78	9.27	-0.480
8	36	308	565	3	6.25	1.038
9	19	353	465	66	5.03	-0.619
10	19	418	864	29	6.26	0.397
11	23	206	438	58	6.66	-0.798
12	55	633	601	60	7.48	-0.602
13	31	131	416	47	5.60	-0.499
14	48	1141	512	33	6.63	-0.114
15	38	530	389	42	5.60	-0.117
16	41	298	456	47	7.36	-0.358
17	29	296	487	38	7.46	0.540
18	56	-104	536	18	9.61	-0.061
19	38	334	497	21	5.10	-0.202
20	14	739	1119	1	2.99	2.036
21	62	560	431	28	9.91	0.075
22	9	510	789	31	6.72	0.195
23	43	661	671	67	6.85	-0.319
24	43	138	585	81	7.95	-0.516
25	37	598	430	41	7.03	0.394
26	24	788	1063	3	6.38	0.560
27	43	909	531	23	10.06	0.298
28	23	392	429	16	4.75	0.055
29	33	315	441	69	6.25	-0.457
30	33	195	475	61	6.18	-0.460
31	39	971	534	45	5.80	0.134
32	42	236	639	65	6.95	-0.585
33	39	465	514	9	5.60	0.268
34	50	377	407	49	10.83	0.241
35	49	579	483	10	8.25	-0.008
36	43	342	490	63	7.69	-0.305

APPENDIX K: EXPERIMENT 2 RESULTS

Table K.1: Experiment 2: Eye Movement Characteristics and Cognitive Capacity Tests

Participant	**WMC**	**STRP**	**CRT**	**EXP**	**Fixation Count**	**Fixation Duration**
1	37	598	430	3346	77,58 fixs/min	0,70 sec
2	39	971	534	1935	98,05 fixs/min	0,51 sec
3	42	236	639	5100	113,58 fixs/min	0,43 sec
4	38	530	389	3570	99,38 fixs/min	0,46 sec
5	41	298	456	4340	108,93 fixs/min	0,42 sec
6	39	465	514	3700	90,68 fixs/min	0,48 sec
7	50	377	407	4530	74,90 fixs/min	0,61 sec
8	49	579	483	2150	84,58 fixs/min	0,62 sec
9	43	342	490	2440	89,00 fixs/min	0,59 sec
10	44	475	468	3610	104,25 fixs/min	0,53 sec

APPENDIX L: STATISTICAL ANALYSIS RESULTS

L.1 Experiment 1

Table L.1: Descriptive Statistics

Variable	Mean	St.Dev.	Skewness	Kurtosis
WMC	37.944	15.183	-.336	-.397
CRT	548.306	167.148	2.260	5.089
STRP	459.833	275.122	.493	-.016
EXP	46.889	36.232	2.542	11.099
OFFLINE	7.143	1.964	.693	.877
ONLINE	-.067	.555	1.724	4.737

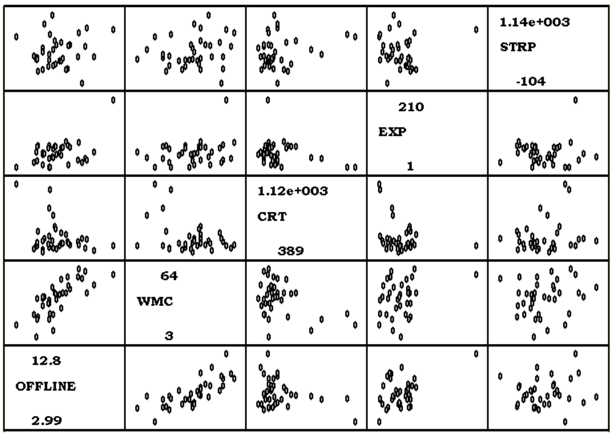


Figure L.1: Scatterplots of Offline SA Scores vs. Explanatory Variables

Table L.2: Model Summary for Offline SA Score

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.789	.623	.575	1.28083

Table L.3: ANOVA for Offline SA Scores

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	84.088	4	21.022	12.814	.000
Residual	50.856	31	1.641		
Total	134.945	35			

Table L.4: Coefficients for Offline SA Score

Model	Unstandardized Coeffs(B)	Standardized Coeffs (β)	t	Sig.
1 (Constant)	2.958		2.427	.021
WMC	.087	.675	5.313	.000
STRP	5.646E-6	.001	.007	.995
CRT	.000	.025	.193	.848
EXP	.015	.278	2.356	.025

L.2 Experiment 2

Table L.5: Model Summary for Online SA Scores

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.755	.571	.515	.38660

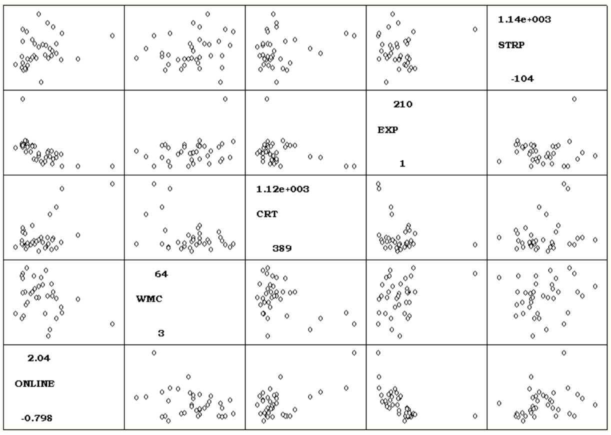


Figure L.2: Scatterplots of Online SA Scores vs. Predictor Variables

Table L.6: ANOVA for Online SA Scores

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	6.154	4	1.539	10.294	.000
Residual	4.633	31	.149		
Total	10.787	35			

Table L.7: Coefficients for Online SA Score

Model	Unstandardized Coeffs(B)	Standardized Coeffs (β)	t	Sig.
1 (Constant)	-.359		-.977	.336
WMC	-.005	-.131	-.964	.342
STRP	.001	.260	2.058	.048
CRT	.001	.313	2.256	.031
EXP	-.007	-.470	-3.730	.001

Table L.8: Mann-Whitney U Test for H_7

Fixation Count	EXP	Rank	MEs<LEs	LEs<MEs
74,90	ME	1		0
77,58	LE	2	1	
84,58	LE	3	1	
89,00	LE	4	1	
90,68	ME	5		3
98,05	LE	6	2	
99,38	ME	7		4
104,25	ME	8		4
108,93	ME	9		4
113,58	ME	10		4
			U=5	U’=19

Table L.9: Mann-Whitney U Test for H_8

Fixation Duration	**EXP**	**Rank**	**MEs<LEs**	**LEs<MEs**
0,42	ME	1		0
0,43	ME	2		0
0,46	ME	3		0
0,48	ME	4		0
0,51	LE	5	4	
0,53	ME	6		1
0,59	LE	7	5	
0,61	ME	8		2
0,62	LE	9	6	
0,70	LE	10	6	
			U'=21	U=3

Table L.10: Mann-Whitney U Test for H_9

Fixation Count	**WMC**	**Rank**	**HWMCs<LWMCs**	**LWMCs<HWMCs**
74,90	HWMC	1		0
77,58	LWMC	2	1	
84,58	HWMC	3		1
89,00	HWMC	4		1
90,68	LWMC	5	3	
98,05	LWMC	6	3	
99,38	LWMC	7	3	
104,25	HWMC	8		4
108,93	LWMC	9	4	
113,58	HWMC	10		5
			U'=14	U=11

Table L.11: Mann-Whitney U Test for H_{10}

Fixation Duration	**WMC**	**Rank**	**HWMCs<LWMCs**	**LWMCs<HWMCs**
0,42	LWMC	1	0	
0,43	HWMC	2		1
0,46	LWMC	3	1	
0,48	LWMC	4		1
0,51	LWMC	5	1	
0,53	HWMC	6		4
0,59	HWMC	7		4
0,61	HWMC	8		4
0,62	HWMC	9		4
0,70	LWMC	10	5	
			U=7	U'=18

Table L.12: Mann-Whitney U Test for ***H_{11}***

Fixation Count	**STRP**	**Rank**	**HISs<LISs**	**LISs<HISs**
74,90	HIS	1		0
77,58	LIS	2	1	
84,58	LIS	3	1	
89,00	HIS	4		2
90,68	HIS	5		2
98,05	LIS	6	3	
99,38	LIS	7	3	
104,25	HIS	8		4
108,93	HIS	9		4
113,58	HIS	10		4
			U=8	U'=16

Table L.13: Mann-Whitney U Test for ***H_{12}***

Fixation Duration	**STRP**	**Rank**	**HISs<LISs**	**LISs<HISs**
0,42	HIS	1		0
0,43	HIS	2		0
0,46	LIS	3	2	
0,48	HIS	4		1
0,51	LIS	5	3	
0,53	HIS	6		2
0,59	HIS	7		2
0,61	HIS	8		2
0,62	LIS	9	6	
0,70	LIS	10	6	
			U'=17	U=7

Table L.14: Mann-Whitney U Test for ***H_{13}***

Fixation Count	**CRT**	**Rank**	**HDCs<LDCs**	**LDCs<HDCs**
74,90	HDC	1		0
77,58	HDC	2		0
84,58	LDC	3	2	
89,00	LDC	4	2	
90,68	LDC	5	2	
98,05	LDC	6	2	
99,38	HDC	7		4
104,25	HDC	8		4
108,93	HDC	9		4
113,58	LDC	10	5	
			U'=13	U=12

Table L.15: Mann-Whitney U Test for $\boldsymbol{H_{14}}$

Fixation Duration	**CRT**	**Rank**	**HDCs<LDCs**	**LDCs<HDCs**
0,42	HDC	1		0
0,43	LDC	2	1	
0,46	HDC	3		1
0,48	LDC	4	2	
0,51	LDC	5	2	
0,53	HDC	6		3
0,59	LDC	7	3	
0,61	HDC	8		4
0,62	LDC	9	4	
0,70	HDC	10		5
			U=12	U’=13