

Wright State University

CORE Scholar

[Browse all Theses and Dissertations](#)

[Theses and Dissertations](#)

2014

Investigation of Capabilities of Observers in a Watch Window Study

Ositadimma Nnanna Eziolisa
Wright State University

Follow this and additional works at: https://corescholar.libraries.wright.edu/etd_all



Part of the [Operations Research, Systems Engineering and Industrial Engineering Commons](#)

Repository Citation

Eziolisa, Ositadimma Nnanna, "Investigation of Capabilities of Observers in a Watch Window Study" (2014). *Browse all Theses and Dissertations*. 1227.
https://corescholar.libraries.wright.edu/etd_all/1227

This Thesis is brought to you for free and open access by the Theses and Dissertations at CORE Scholar. It has been accepted for inclusion in Browse all Theses and Dissertations by an authorized administrator of CORE Scholar. For more information, please contact library-corescholar@wright.edu.

INVESTIGATION OF CAPABILITIES OF OBSERVERS IN A WATCH WINDOW
STUDY

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Engineering

By

Ositadimma Nnanna Eziolisa
B.S., Wright State University, 2011

2014

Wright State University

WRIGHT STATE UNIVERSITY

GRADUATE SCHOOL

Date May 6, 2014

**I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY
SUPERVISION BY Osita Eziolisa ENTITLED Investigation of Capabilities of Observers in
a Watch Window Study BE ACCEPTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF Master of Science in Engineering**

Mary E. Fendley, Ph.D.
Thesis Director

Thomas N. Hangartner, Ph.D.
Chair, Department of Biomedical,
Industrial and Human Factors Engineering

Committee on
Final Examination

Mary E. Fendley, Ph.D.

Subhashini Ganapathy, Ph.D.

Alan Boydstun Ph.D.

Robert E.W. Fyffe, Ph.D.
Vice President for Research and
Dean of the Graduate School

ABSTRACT

Eziolisa, Ositadimma Nnanna. M.S.Egr., Department of Biomedical, Industrial and Human Factors Engineering, Wright State University, 2014. Investigation Of Capabilities Of Observers In A Watch Window Study.

Due to an abundance of data and dynamic nature of tasks, challenges with information retrieval in surveillance and target identification tasks have risen in today's Intelligence, Surveillance, and Reconnaissance (ISR) community. In this study, two variables, Area of Coverage and Amount of Activity (AOC/ACT), are manipulated to study their effects on the number of Watch Windows an observer can monitor. This research describes the analyst's task model, and explains how the level of AOC/ACT and number of Watch Windows affects the analyst's cognitive load. Results showed a significant difference in performance and physiological indicators of workload between high AOC/ACT conditions and low AOC/ACT conditions. Confidence levels were higher with low AOC/ACT conditions, while NASA-TLX ratings decreased. A linear correlation was exhibited between the number of Watch Windows and the number of fixations. The results show that these variables can be manipulated in tasking to maintain appropriate levels of cognitive workload.

TABLE OF CONTENTS

	Page
INTRODUCTION	1
Research objective	1
LITERATURE REVIEW	2
Intelligence Surveillance and Reconnaissance (ISR)	2
Imagery Collection	3
Motion Imagery	5
Watch Window	6
Workload Measurement	13
NASA-Task Load Index (NASA-TLX)	15
Visual Search	16
Eye tracking	22
RESEARCH COMPONENTS.....	27
Research Questions	27
Task Models	28
METHODS	30
Participants	30
Apparatus and Stimuli	30
Experiment Design	31
Independent Variables	33
Dependent Variables	33
Stimulus I	35
Stimulus II	36
Stimuli Design	37
Procedure	39
Data collection	40
RESULTS	43
Number of Watch Windows and Area of Coverage and Amount of Activity on Performance	43
Number of Watch Windows and Area of Coverage and Amount of Activity on Perceived Difficulty	45
Fixation	48
Sample Size	49

GENERAL DISCUSSION	50
RELEVANCE TO THE INTELLIGENCE SURVEILLANCE AND RECONNAISSANCE DOMAIN	55
FUTURE WORK.....	56
Appendix A: Tables of task Results.....	57
Appendix B: Performance.....	59
B-1: Effects of Number of Watch Window on Performance	60
B-2: Effects of level of AOC and ACT on performance.....	61
Appendix C: Confidence Ratings	65
C-1: Effects of number of watch windows on confidence rating	66
C-2: Effect of level of AOC/ACT on confidence ratings	67
C-3: Effects of Cross of both Number of Watch Windows and Level of AOC/ACT on Confidence Rating.....	67
Appendix D: Comparison of All Dependent Variables- Performance Versus Confidence and NASA-TLX.....	71
References	72

TABLE OF FIGURES

Figure 1: The ISR team and role of the intelligence Analyst. Double sided arrows represent communication pathways; dashed line represents flow of information.....	3
Figure 2: The intelligence analyst and methods for obtaining workload data	5
Figure 3: Decisions Made by an Observer Performing a Visual Search Task.....	7
Figure 4: Parallel vs. Serial Search Methodologies (based on Bruce & Tsotsos, 2009).....	10
Figure 5: Yerkes-Dodson graph showing curves of high medium and low stimulations	25
Figure 6: Multiple Reflections used to Calculate Gaze Direction	26
Figure 7: Central Task Model of an Intelligence Analyst.....	28
Figure 8: Six-screen Display and Labeling	31
Figure 9: Fishbone diagram showing the independent variables with respect to each dependent variable.....	35
Figure 10: Stimulus I- easy tasks (low AOC/ACT); one to six watch windows or screen displays..	36
Figure 11: Stimulus II- hard tasks (high AOC/ACT) one to six watch windows or screen displays	37
Figure 12: Confidence Questionnaire	41
Figure 13: General Behavior of Performance Versus Number of Watch Windows	44
Figure 14: Mean Score versus number of Watch Windows and Amount of AOC/ACT.....	44
Figure 15: Graph of Confidence versus Number of Watch Windows.....	46
Figure 16: Graph of Confidence Ratings versus Amount of AOC/ACT (High/Low). Each error bar is constructed using a 95% confidence interval of the mean.....	47
Figure 17: Graph of NASA-TLX versus Amount of AOC/ACT (High/Low). Each error bar is constructed using a 95% confidence interval of the mean.....	48
Figure 18: Graph of mean fixation versus number of watch windows	49
Figure 19: Graph of Performance Versus Confidence and NASA-TLX.....	71

ACKNOWLEDGMENTS

I would like to express my most sincere gratitude to my advisor Dr. Mary Fendley, whose relentless encouragement, patience, enthusiasm, and great knowledge guided me throughout this journey.

Besides my advisor, I would like to recognize my colleagues at Wright State's Cognitive Systems Laboratory for their support and guidance throughout this project. Finally I would like to recognize my family; especially Victoria Petit who stood beside me every step of the way.

INTRODUCTION

The ability to use aerial and satellite imagery to acquire visual data has increased human visual tasking load in surveillance and target identification. Therefore, in the Intelligence, Surveillance, and Reconnaissance (ISR) domain, the human bears the responsibility of signal recognition and initiation of best course of action (CoA). Due to the abundance of data and the nature of tasks in today's surveillance and reconnaissance, the human is often tasked with monitoring multiple displays of visual data simultaneously. Subject matter experts (SMEs) suggest that when an observer is tasked to monitor activities in a multiple-window display, two main elements- area of coverage and the amount of activity- affect difficulty perceived by the observer. These elements subsequently affect the number of windows the observers can successively monitor. SMEs are interested in the observer's performance in up to six watch windows. Studying these variables in an experiment that mimics a surveillance environment will be used to evaluate this.

Research objective

The objective of this study is to investigate the impact that the variables *Area of Coverage and Amount of Activity* have on the performance of an observer who is watching up to six watch windows. Also, the study hopes to discover strategies that would contribute to effectiveness of observers monitoring events in multiple displays.

LITERATURE REVIEW

In order to understand this study it is important to understand some of the related topics discussed in this section.

Intelligence Surveillance and Reconnaissance (ISR)

Intelligence Surveillance and Reconnaissance (ISR) is a critical community to the U.S Department of Defense (DoD) that account for about \$40 billion annually; and its functions involve various methods of information acquisition for national security decision makers (Best Jr., 2005). The ISR team consists of the Mission Operations Chief (or Commander) (MOC), pilot, sensor operator, and imagery/mission supervisor. The MOC coordinates execution of the mission and communicates with other members of the team. An imagery supervisor works as a mediator between the Intelligence Analyst (IA) and the MOC. Figure 1 shows the components of the ISR team and the communication pathways. This study will examine tasks that are issued to IAs. Because information processing currently involves observers watching multiple windows, the ISR community faces several challenges in the area of visual surveillance. These challenges involve information overload, information fusion, fleeting targets that appear within a short period of time, thus demanding a quick response, and relocateable targets (Barber, 2002; Duncan & Ayache, 2000; Jones, Shapiro & Roshon, 2007; Pham et al., 2008). The most common device for information acquisition for real time analysis in the ISR community is the Unmanned Aerial Vehicle (UAV).

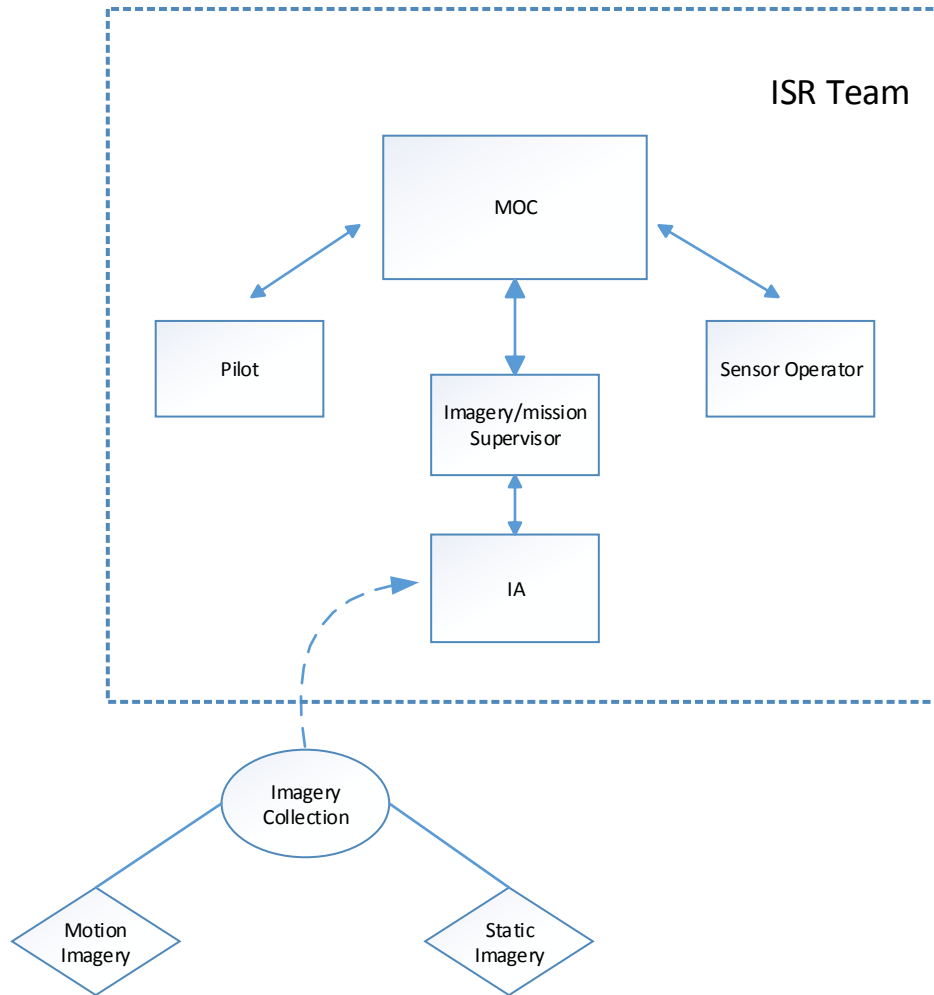


Figure 1: The ISR team and role of the intelligence Analyst. Double sided arrows represent communication pathways; dashed line represents flow of information.

Imagery Collection

The most common method of imagery collection in the ISR community are using UAVs, which range from hand-held devices to orbiting satellites; therefore, many studies involving visual tasks have used UAVs for aerial data acquisition (Trinh & Kuchar, 1999; Dixon et al., 2005; Freed et al., 2004; Srinivasan et al., 2004; Ruff et al., 2004; Wickens et al., 2003; Hickman et al., 2008; O’Kelly et al., 2005,).

Capabilities of UAVs increase with advances in wireless communication and computing power. The UAV's ability to transfer information is utilized by the military, Research and Development (R&D), and other companies. UAVs are unique for their ability to acquire simultaneous coverage of large areas and to perform mapping, subsequently relaying the information back to the operator. Current uses of UAVs range from surveillance and mapping of hostile territory to border patrol (Ryan et al., 2004). There are two methods by which data collected from UAVs are handled. One is by live streaming of data from the UAV to the control station where data is processed by the ISR team. In this approach, streaming data is of high importance. For instance in a battleground, the Troops-in-Contact (TIC) would communicate with personnel at the control station who obtain information from the UAVs. In the other method, the individual captures the videos collected by the UAV and relays them to the ISR team to be exploited by IAs. In this approach, data is not processed live, therefore, it is used when mission objective can afford to wait for some time. Data acquired from UAV's in a mission that involves transfer of information are relayed to and processed by Intelligence Analysts.

An Intelligence Analyst or Image Analyst (IA) is the person who specializes in performing tasks such as detection and recognition of various targets or objects. Their tasks can be performed using either still imagery or motion imagery. IAs can perform static or dynamic recognition activities. Static refers to detection and recognition of objects while dynamic refers to detection and recognition of activities (Irvine et al., 2006).

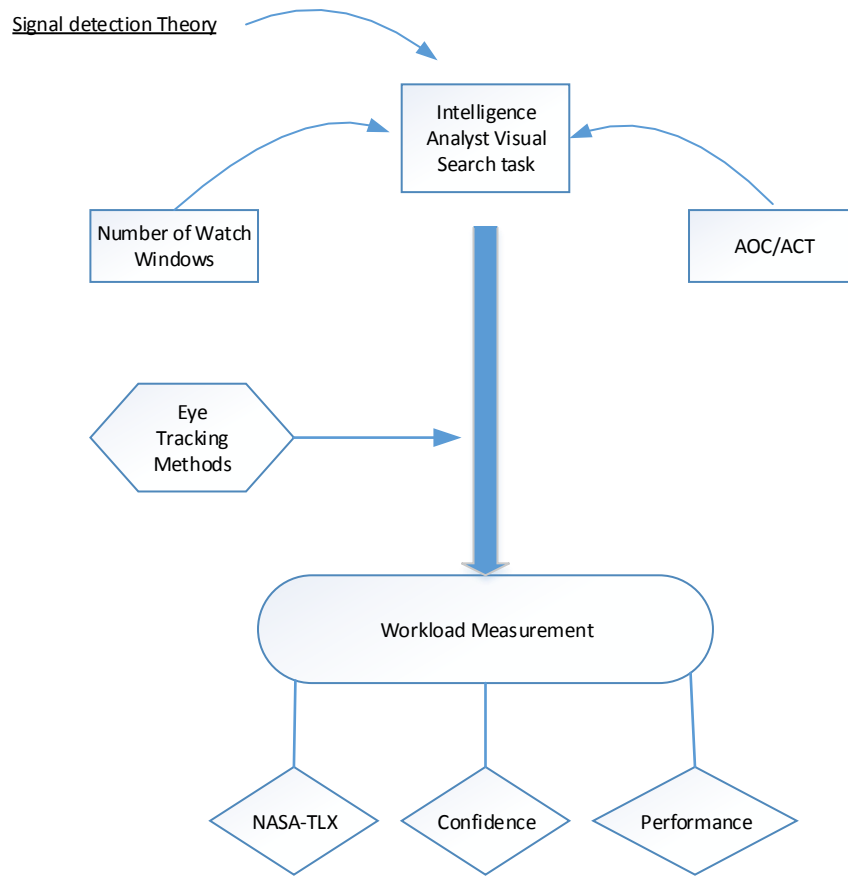


Figure 2: The intelligence analyst and methods for obtaining workload data

Motion Imagery

According to Ling et al., (2012), there are challenges with frame-rate limitations and fast camera action due to the camera being attached to a flying object. Wide Area Motion Imagery (WAMI) and Full Motion Video (FMV) are the two most commonly used forms of imagery in experiments that are of interest to IAs because they are the formats used for the majority of their work. Factors that may affect an IAs interpretation of motion imagery include (Irvine et al., 2006):

- Target motion
- Camera motion
- Ground Sample Distance or resolution
- Scene complexity
- Color
- Frame rate
- Image exploitation task category

WAMI is a rapidly developing sensing modality characterized by the collection of Electro-Optical or Infra-Red (EO/IR) images with very large spatial extents. WAMI's frame rates are one to two frames per second. FMVs are 24-30 frames per second (United States Air Force, 2012; Paul & Fendley, 2013). One can see that WAMI's frame rate is significantly lower than that of FMV. Higher frame rate of the FMV translates to enhanced resolution, visibility, and color compared to WAMI and those factors will play a role in the selection of stimuli for this study.

Watch Window

The term Watch window refer to a geographic area on a computer screen that an observer (usually an IA in the ISR community) is tasked to observe. Observers are often tasked to search for events such as suspicious activities, perform surveillance, and provide information that will be useful for troops on the ground or TIC in that area. As one might predict, signal detection plays a role in a watch window performance. In the case of a TIC task, one can see that the well-being of troops on the ground or TIC is dependent upon IAs making the right observations and relaying appropriate information to the MOC and the MOC taking appropriate CoA. Currently, IAs may watch up to four windows at a given time.

Accuracy of decision-making that involves signal detection is based on odds that favor certain possibilities of outcomes (McNicol, 2005; Abbot & Sherratt, 2013; Wixted, 2007; Hautus

et al., 2008; Pleskac & Busemeyer, 2010; O'Mahony & Hautus 2008; Verghese, 1994, 2001; Palmer et al., 1993; Eckstein, 2000; Ramos-Alvarez, 2012). Signal detection involves identifying an item in the midst of distractors and distraction. In signal detection there are two primary parties, the target and the observer. In a mission, a target is either present or not; and from the observer's perspective he/she either detects a target or not. Errors are associated with the observer's perspective. That is, the observer can make errors in detecting the target or not. Thus, there can be four possible outcomes:

1. A hit- observer rightfully detects the target,
2. A false alarm- the observer claims he/she detected a target while target is actually absent,
3. A miss- the observer fails to detect a (present) target
4. And, a correct no-call- the observer rightfully claims there is no target present or provides no response.

	Target Present	Target absent
Yes	Hit	False alarm
No	Miss	Correct rejection

Figure 3: Decisions Made by an Observer Performing a Visual Search Task

Heeger (2007) suggested that two factors which influence accuracy of decision making are information acquisition and criterion. First, the more information present, the better chance of the observer making a correct decision (hit or correct rejection). The second factor, criterion, pertains to the fact that depending on the nature of the task, not all errors are weighted the same. That is, some errors are more serious than others. The higher the priority of the task, the more likely it is to generate an accurate output. Abbott and Sherratt (2013) found that assigning cognitive resources to one task reduces resources available to another task, thereby reducing accuracy and/or speed of the additional task. Greater accuracy in a given task may correspond to slower performance.

Factors that may hinder a decision maker in making a correct decision (a hit or correct rejection) are called uncertainties or noise. Wichchukit & O'Mahony (2010) described noise as "random unpredicted signals produced by natural processes, both internal and external to the system." External noise is noise associated with experiment environment that may hinder the operator's performance, such as keyboard malfunctions, computer freezes or a slow running system, temperature of environment, and experiment instructions. Neural noise (internal noise) also affects decision variables and is associated with inhibition of the operator's thought process due to factors such as sleep deprivation, meal deprivation, cognitive stress, and inadequate training (Stanislaw and Todorov, 1999; Heeger, 2007). Of course it is ideal for experimenters to minimize external noise as much as possible. However, there is usually some form of noise in the system and this must be accounted for. Signal detection theory does not address questions on how an operator might filter noise, or how one might combine two noisy signals to cancel the noise out (Pleskac & Busemeyer, 2010). However, Heeger (2007) suggested that the experiment designer can minimize noise by increasing signal strength and/or making targets easier to detect in which case the ROC curves are narrower, which is ideal. Thus, as Wichchukit & O'Mahony (2010) mentioned, signal to noise ratio is an important measure to take into consideration when

conducting experiments involving SDT. According to Heeger (2007), providing more information is a technique to improve signal strength and this is more applicable to a visual search task. Wichchukit & O'Mahony (2010) illustrated the effects of information in signal strength and discriminability (d'), which serves as an index of sensitivity for signal/noise differentiation (Johnson et. al, 2006). For instance, in a blind-folded soup tasting experiment, a judge might be tasked to tell the difference between soup Y and soup Z. In this initial test, the judge might find the two soups very similar in taste and decide to go with a slightly spicy sensation as the differentiating factor between the two soups. Assuming that he believes soup Z to be the one with spicy sensation, giving soup samples that are spicy in a follow up test would cause the judge to declare more soups "soup Zs," causing him to have both more hits and more false alarms. Similarly, giving him soup samples that are not spicy would cause him to be less willing to declare that the soup is "Z"; resulting in fewer hits and false alarms. However, making the flavor of the spicy soups more pronounced in the initial test (where soups Y and Z were differentiated) would make the judge more confident that the spicy soup he is tasting is actually soup Z. Similarly increasing d' increases the probability of hits while also reducing probability of false alarms in a signal detection task. Consequently, it is ideal to maximize d' in a detection task. It is important to note that some external noise is necessary for a complete signal detection experiment therefore experimenters usually implement them as distractors. Implementation of distractors is common in visual search tasks because they exist in real IA's tasks.

Cameron et. al (2003) defined visual attention as the process by which one grants priority among visual information. Bruce and Tsotsos (2009) performed a study in which the aim was to improve visual attention by directing the eyes into areas of the screen that contained relevant information to the search task. In the study of visual attention, set size refers to the amount of distractors per relevant material. Serial vs. parallel search concepts imply that some targets are found with minimal effort when a parallel search methodology is used while others (usually more

difficult tasks) require a serial search methodology. Search time in serial search tasks typically increases when the number of distracting elements is increased. On the other hand, addition of elements of distraction usually has no significant effects on the speed of parallel search tasks. Figure 4 was inspired by Bruce and Tsotsos' (2009) work, and it will be used to illustrate parallel and serial search methods. The box on the left is an example of a parallel search task while the one in the middle and on the right are examples of serial search tasks. Detecting the oddly shaped rectangles is easier and more natural with a parallel search whereas detecting a target using a parallel search; as in the two boxes to the right, is more challenging due to the increased variance in the sample. Therefore, a serial search in those scenarios is ideal. The ultimate goal of a search method is to discriminate between targets and distractors. The tasks in this study are serial search tasks.

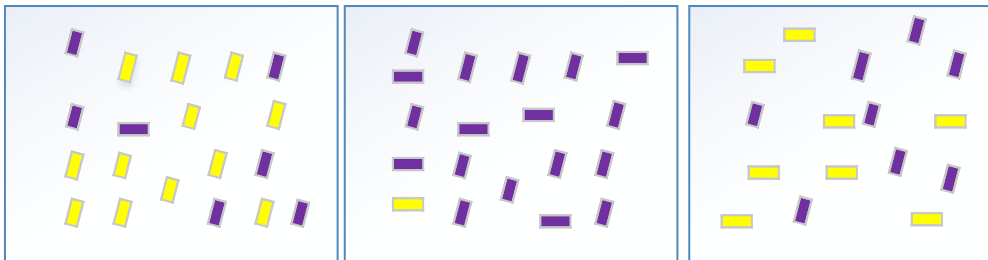


Figure 4: Parallel vs. Serial Search Methodologies (based on Bruce & Tsotsos, 2009)

In visual search tasks, SDT is the study of discriminability (d') between targets and distractors. Verghese (2001) argued that both the mean separation between target and distractors and the variability between their presentations are determinants of discriminability. For example, in the middle and right boxes in Figure 4, if experimenters were to have the items closer together, an observer's ability to discriminate between targets and distractors decreases; a similar result will be seen if color variability between objects was increased. Set size (how big an area the search task is and/or how much content is in the search task) is found to be inversely related to d' .

The larger the set size, the lower the d' and vice versa. According to Verghese, several past studies such as Eckstein et al. (2000), Verghese and Nakayama (1994) and Palmer et al. (1993) have effectively illustrated relationship between set size and d' . Therefore, in visual search tasks SDT may be used to make a prediction of the observer's accuracy as a function of d' .

Furthermore Palmer & McLean (1995) determined the existence of a relationship between response time and d' . They found that response time increases exponentially with decreasing d' . Verghese (2001) also stated that object attention might depend on degree of separation between the target object and other parts of the screen, as well as the observer's level of familiarity with the target object.

In order for an object to be correctly identified in a visual search task, the object that is being searched for must have a separate identity with characteristics that are distinguishable from the surroundings. In other words, perceptual grouping is essential. Perceptual grouping is a natural process involving texture segregation (Treisman 1982) that is mediated by differentiation of basic separable features in the area of the search. It is believed that perceptual grouping affects all successive stages of a visual search such as grouping, separation, identification, and recall. Treisman (1982) highlighted the significance of grouping by showing that objects differing either in color (such as yellow As and Bs versus green As and Bs) or shape (such as yellow and green As vs. yellow and green Bs) are easily grouped into perceptual groups while objects that differ in both color and shape (such as yellow As combined with green Bs in a location versus yellow As combined with green Bs in another location) are not easily separated by the human observer. Gale & Buynak's (1977) findings suggested that focused attention is necessary for object grouping.

Another event that is often witnessed in visual search tasks is when more than one target appears within the same time interval, and the observer is expected to detect all targets. Previous

studies have shown that in this case the probability of detecting the first target is much higher than the probability of detecting a second target in a different location at the same time. It is believed that once the first target is detected, it becomes a high level distractor, which hinders the observer's ability to detect the second target. Cain and Mitroff (2012) found that subtracting targets from the display as they are found increases likelihood of finding subsequent targets; thus increasing search accuracy. However adding distractors do not improve search accuracy. Furthermore, highlighting found targets helped reduce visual salience of the target as well as the and mental load on the observers. However, replacing found targets with other objects reduced the target's visual salience but did not alleviate mental load. In conclusion, Cain and Mitroff (2012) found that, when a target is found, working memory load has a larger effect on the effectiveness of finding subsequent targets than the observer's perceived salience.

Dickinson and Zelinsky (2013) suggested that humans tend to use less frequent eye movements in dynamic viewing conditions than in static viewing conditions. A static visual search task is to a visual task, in which items retained their positions throughout the entirety of the task, whereas a dynamic visual search task is a task where items change locations randomly throughout the display. Dickinson and Zelinsky's (2013) article suggested that observers employ a sit-and-wait strategy, in which their gaze is focused on a constrained region throughout viewing task waiting for the target to appear in their gaze range. The authors found that this strategy resulted in an increased in miss rate. An active search is the alternative to the sit-and-wait strategy. In this method the observer actively searches for the target by observing specific items or item groups using either a systematic or random method.

Workload Measurement

Workload is a measurement of the effort put in by the human operator to complete a given task. Vicera (2013) found a strong relationship between those elements that influence attention and those that effect perceived load. This relationship was deduced because the authors found that when the observer attempts to process information that is not relevant to the search task efficiency is negatively affected. Evidently, processing of task-irrelevant information adds to the amount of workload perceived by the observer and hence is taken into account in measuring perceived load. Therefore a measure of the degree of attention applied by the observer throughout a visual search task can play a role in determining perceived load.

Lavie's (1995) work is also important when discussing perceptual workload. Her research suggests that when perceptual load is low in a task, mental processing resources tend to "spill over" to events that are not relevant. On the other hand, when workload perceived by the observer is high, all mental resources are allocated to task relevant material. While Lavie's study illustrated the effects of high and low perceptual load tasks, Vicera defined the key differences between these task types that actually cause the observed effects on search efficiency. In a low load task, the targets are expected to "pop-out" whereas in a high load task, the targets are not expected to "pop out" at the observer; as a result, that the observer has to put more effort into identifying the target. In other words target and distractor features exhibit more of a camouflage in a high load task. Lavie (1995) added to this concept, identifying a low load task as one in which the target is visually different from homogenous distractors and a high load task as one in which the target looks much like heterogeneous distractors.

Hart and Steveland (1988) emphasized that workload is human centered and not task centered because it emerges from the interaction between the requirements of the task, the task environment, and the capabilities of the operator. Hart and Steveland's multi-year research study

included evaluation of ten workload-related factors obtained from 16 experiments. Experimental tasks included simple cognitive, manual control, complex laboratory, supervisory control, and aircraft simulation tasks. In their article they developed in detail a framework, which illustrates factors that influence performance and workload. According to the NASA Human Performance Research Group (1987), specific sources of load such as mental capacity and environmental factors imposed by different tasks are the most important determinants of workload experiences.

A factor that makes workload measurement so difficult is that there is no real standard, such as a task's "actual workload," with which the operator's results can be compared (Hart and Steveland, 1988). Therefore, workload measurement is relative and subjective. The overall goal of Hart and Steveland's study was to develop a sensitive workload rating scale that could both account for variation between and within tasks, as well as eliminate the influence of human perceptions and bias in judgment. To that end, the authors began their research by asking the following questions: what factors contribute to workload? What are the ranges, anchor points, and interval values? What subset of these factors contributes to the workload imposed by specific tasks? And what do individual operators take into account when experiencing and rating workload? Hart and Steveland (1988) also found in their research that task related sources of variability (such as task difficulty and amount of time available) between the operators were better predictors of workload experiences than biases.

The next step in developing a workload scale was to ask several groups of operators to evaluate their experiences during the experiments. Various concepts of workload were discovered by finding out which ratings were most consistent across all operators in all experiments. The paragraphs below will discuss the rating factors in Hart and Steveland's research and summarize those rating factors in a table. The factors form the framework from which the current NASA- Task Load Index (NASA-TLX) originated.

NASA-Task Load Index (NASA-TLX)

Since humans cannot be programmed to give one hundred percent accurate workload readings of a task that they completed, there is need of a creative means to acquire accurate unbiased workload information. A NASA-TLX is a multidimensional rating procedure used to obtain a weighted overall workload score (NASA Human Performance Research Group, 1987).

According to the NASA Human Performance Research Group (1987), the earlier version of NASA-TLX scale had nine subscales. Researchers later eliminated three subscales on the basis of redundancy and irrelevancy.

Table 1: Current Version of NASA-TLX

Title	Description
Mental demand	Amount of mental activity assigned to the tasks such as calculating, recall, and research
Physical demand	Amount of physical activity assigned to task such as dragging, clicking, and pushing
Temporal Demand	Extent of pressure felt due to time while performing task. Was the task rushed, slow, or comfortable pace?
Performance	Self-perception of success in task
Effort	Amount of physical and mental work designated to the task
Frustration	Extent of Irritation, aggravation or feelings of similar nature felt during task

Reid and Nygren (1988) narrowed the workload subscales further, defining three factors- performance, effort, and temporal demand and leaving out time load, mental effort load, and psychological stress load. The rational behind the change was that the authors felt that their term mental effort could simultaneously describe both mental demand and effort in the NASA-TLX, and that their factors of time load and physiological stress load replaced NASA-TLX's psychological demand and frustration respectively.

The NASA-TLX offers a combined score based on subscale ratings that are weighted according to their importance to the operator with regards to a specific task. There are multiple factors that contribute to workload such as physical and mental load and load from time pressure. NASA-TLX is sensitive because offers a more precise way to acquire not only information on workload, but the exact types of load present (physical, mental, and temporal) and in what parts of the experiment the different types of load were experienced.

The NASA-TLX is composed of six-subscale ratings, and aims to collect information on mental demand, physical demand, temporal demand, own performance, effort, frustration; and the operator perception of his/her own performance. Three of the measurements-mental, physical, and temporal demands-relate to demands imposed on the operator whereas the other three effort, frustration, and own performance describe the operator's interaction with the task (NASA Human Performance Research Group, 1987).

It is ideal to administer the NASA-TLX to operators upon completing each task in an experiment (as opposed to issuing it at the end of the entire experiment) to avoid recency bias. Recency bias is an order effect bias whereby information presented later has a greater influence on the subject's rating (Aquino, Culpepper & Pierce 2010; Peggy & Richard, 1997; Fabrigar & Wegener, 1994; Tversky & Kahneman, 1973).

Visual Search

Previous studies indicate that there are limitations to how many windows a human can effectively monitor at a time. The following discussion will focus on previous research that has been done on human capabilities in simultaneous multiple-window monitoring.

Sulman and Sanocki's (2008) article investigated the relationship between number of displays watched by the observer and the accuracy of target detection. According to the paper, there is a trade-off when the number of displays that an observer is monitoring is greater than four. The researchers found that when required to monitor nine displays, observers failed to detect when targets entered a forbidden region 60% of the time. In addition, when targets were identified, the probability of identifying another target within the same time frame was decreased. On the other hand, when monitoring only four displays, miss rates were reduced to 20%; significantly lower than observers responsible for nine displays. Given the significant difference in target detection efficiency, Sulman and Sanocki's (2008) results are relevant to the work of IAs and to the ISR community as a whole.

Sulman and Sanoki's (2008) conclusions imply that although adding cameras and monitors to improve security seems logical, such additions may actually defeat the purpose by deleteriously affecting the likelihood of detecting security concerns effectively. Some points to consider when deciding monitor-to-operator ratio in CCTV viewing tasks are: is the observer looking for behaviors that are easy or difficult to detect? How complex are the scenes and backgrounds? How easy is it to tell between normal and incident (suspicious) behavior? How many incidents can (or usually do) occur in a given time? Do incidents take place in the foreground, middleground, or background? What is the quality and brightness of video and resolution of camera? What is the chance of an event occurring in more than one monitor simultaneously? In addition to the aforementioned questions, previous studies have also suggested that monitor-to-operator ratio at a given facility also depends on risk factor of the area being monitored.

Swanson et. al (2013) conducted a study on 26 participants test their ability to detect the presence of assigned Targets of Interest. During the study observers monitored four Remotely Piloted Aircraft videos. The results showed that dwell time (time spent in waiting for an event to

occur), viewing angle, and inter-event time have a significant effect on the ability of an observer to detect targets simultaneously in four displays. Longer inter-event times seemed to help observers by allowing them the time to put more cognitive effort into watching the other displays. Response time was generally longer when viewing videos with short inter-event times. Furthermore, when multiple events are cued within a short time, observers must process each event individually thereby increasing response time.

Shafiullah, Gyasi-Agyei, & Wolfs (2007) addressed the impact of an increased number of CCTV feeds on train drivers. During the experiment, subjects were asked to determine whether a target item was present and if it is safe for the train to move. Results in the study agreed with results in previously discussed literature. That is, the time required by the train operator to reliably scan the images increased with number of images displayed. Furthermore, busier images tended to result in an increase in false alarm. Train drivers required more time to make more accurate scans in those occasions. Results in this train study showed no significant difference between day versus night videos.

During a visual search, the human registers a wide field of view with the eye's retina and various areas of the retina have various image resolutions (Najemnik & Geisler, 2005). The human eye uses quick movements to direct the fovea -the region of the eye with the greatest image resolution- to areas most likely to contain the target. Eye saccade patterns are assumed to be in the general direction of the target being sought; on the other hand, Araujo et al., (2001), suggested that that the human observer often finds this difficult. To test this hypothesis, they developed a saccadic plan to be used their visual task experiment. Most of their subjects failed to follow the plan and therefore did not optimize performance in the tasks. Many observers would often begin the task according to the saccadic plan, but eventually deviated to a more natural search method. The authors suggested that following a planned saccade is too difficult because it requires extra processing resources from both the eyes and brain. Factors that affected

effectiveness of following saccadic patterns include behavior of stimulus and spatial distance between one saccade to the next.

The most sensitive visual information is acquired at the fovea, which represents the center of a focused gaze, and the human observer usually keeps gaze proximities close to this location (Araujo et al., 2001, Najemnik & Geisler, 2005). In order to identify valuable pieces of information, it is imperative for the human observer to create an image of the world during a search task. Researchers are currently performing studies that can help observers perform accurate visual searches and identify targets with minimal number of eye fixations.

Morvan & Maloney (2012) conducted a study that encouraged the observers to perform visual search tasks with least minimum fixations by instructing observers to move their eyes according to a preset fixation pattern. They found that most observers failed to perform the tasks as instructed and therefore did not accomplish the tasks using minimal fixation. This suggests that such minimal fixation patterns are uncomfortable and unnatural to most people.

In the case of a map, a human searcher who has a certain target in mind will usually scan the map and identify the closest attribute that resembles the target. Najemnik and Geisler's (2008) visual search study found that observers tend to fixate their gaze in a donut shaped formation around the center of the display. This pattern was also one of four found in the work of Fendley (2009). Najemnik and Geisler's (2008) also found that the observers had higher gaze duration on top and bottom of the donut shaped perimeter. In their search task, they compared human observers to ideal observers in an attempt to discover areas for improving the human performance. The ideal observer was described by Najemnik and Geisler (2008) as an arbitrary character whose search performance is comparable to that of a computer system. According to the article, the ideal observer focuses gaze on areas that are most relevant to the information being sought for, and is aware of and takes into account visual fields that are less sensitive. This

means an ideal observer allocates less effort to items that are in regions of less visual sensitivity and more effort in locations of higher visual sensitivity. Najemnik and Geisler (2008) also formulated a framework that outlines the search strategy of a human observer. The observer begins with some initial beliefs about the target, which are represented as probabilities. These probabilities are mixed with assumptions and biases. In the first glance at search area, the observer obtains visual data from every likely target location. The observer then uses the data to update previous beliefs. If the observer's maximum belief exceeds the criterion, the search is stopped and the closest signal to the target is picked; otherwise, the mission is restarted with new belief (by obtaining more information) about the target or location. A high precision eye tracker was used to obtain gaze data in the study. Najemnik and Geisler (2008) found that the observers's visual acuity is highest at the center of the fovea and falls smoothly within the retina. The subject's visibility tended to decline fastest from top to bottom as opposed to from side to side; therefore visibility was poorest in the upper and lower regions of the search area. The authors found that humans implement search strategies that are similar to the ideal observer. That is, a search method that involves forming a donut-shaped area around the center of the display and allocating more gaze towards the shape's top and bottom. However, the human observer's susceptibility to bias is the main difference between them and the ideal. Biases in visual search tasks include contrast bias, Anchoring and Adjustment bias, order effects bias, availability bias, confirmation bias, representative bias, attentional bias, belief bias, conservatism bias, and empathy gap (Morgeson & Campion, 2010; Peggy Wegner & Fabrigar, 1997; Haugtvedt & Wegener, 1994; Tversky & Kahneman, 1973).

Recent studies have also discovered that the size of functional visual field decreases with increasing task difficulty, and an increase in fixation is an indication of high workload (Young & Hulleman, 2013; Dodonov & Dodonova, 2012; Levin, Angelone & Beck, 2011; Lieberman, Coffey, & Kobrick, 1998). The functional visual field addresses the amount of information an

eye's retina can obtain in a single fixation and depends on task difficulty (Young & Hulleman, 2012). Fixation is keeping the eye-gaze in one location and it is the point between two saccades. When an observer attempts to fixate on multiple items, the number of fixations needed to accomplish this will surpass memory capacity resulting in the observer being forced to revise previously fixated areas (Young & Hulleman, 2013). Obviously, revisiting previously fixated areas adds to the fixation count and increases the time-on-task.

The act of selecting particular areas to which attention is allocated is pertinent to an individual's interaction with the environment. However, this process can be infiltrated by biases. The converse of focused attention is distributed attention where the observer allocates attention to several different items in the search task. Enns & Girgus (1985) discovered in their study that in the foveal realm, differences in effectiveness between focused and distributed attention are minimal. In the peripheries, on the other hand, focused attention corresponds to more effective target detection (Ambler & Finklea, 1976). The feature-integration theory described in Treisman's (1982) article suggests that attention serves as a selective tool that selects both the features that are to be grouped together and a visual search task. The range of attention can vary in intensity or in dimensions. Current eye tracking equipment has been helpful in uncovering the degree of attention given to an area in a given time during a visual task. Treisman (1982) suggests that the observer achieves object grouping by focusing on one area at a time. Objects are obtained through features that occur within an observer's single fixation, and when these fixations are interrupted, false images may occur. Palmer (1992) showed in his study that proximity and movement of isolated elements are solid determinants of perceptual grouping. Proximity and movement of items assist the observer to group items with similar features and separate items with dissimilar features. A feature-integration theory mentioned in Treisman's (1982) article suggested that without prior knowledge of the target an observer scans items individually, whereas when the observer has some previous knowledge of the target, he/she

identifies target groups. In other words, the observer groups objects according to their resemblance to the target.

Vecera (2013) investigated the elements that affect the workload perceived by the observer by studying target-distractor similarity and distractor-distractor similarity. The author's goal was to examine whether elements that affect attention also affect perceptual load. His results were consistent with previous research, concluding that target identification was most effective when target-distractor similarity was at a minimum and less efficient when target-distractor similarity increased. On the contrary, target identification was least effective when target and distractor features were more similar and more effective when distractor-distractor features are more similar. Similarity, distractor-distractor features allows for grouping of distractors and increased salience of the target.

Vecera's (2013) study also found a relationship between effectiveness of completing the visual search task and workload perceived by the observer. The study's results also suggested that the difficulty of the task had no contribution to perceptual load and thus the two cannot be directly correlated.

Eye tracking

An eye tracker is ideal in order to carry out an experiment that evaluates an observer monitoring multiple visual displays. As shown in Figure 2, eye tracking methods can be used to acquire measurements of an IA's workload during in a given task. Eye tracking is a technique used to observe a person's eye movements on a display. In other words, eye tracking is used to visualize areas where someone allocates visual attention. The eye gaze provides a very efficient way of pointing just as we do with our hands during interaction. Eye tracking technology allows people to use this interaction method with computers and even other machines, because it is fast

and natural. Eye tracking has been used to study several different fields including psychology, cognitive science, disability rehabilitation research, medicine, and human- computer-interactions (HCI) (Cheng & Veregaal, 2004). Useful information can be gained from eye tracking in order to both understand human behavior and improve human computer interactions. Most eye trackers use principles of corneal reflection tracking. Corneal reflection cameras compare a video input of the user's pupil with highlighted reflections off the cornea usually from light sources that are invisible to humans. The center of the pupils are tracked in real time which provides information about the user's Point of Gaze (POG) (Morgante, Zolfaghari, & Johnson, 2012). Atkins, Moise, and Rohling (2006) suggested that a visual search is composed of two processes: search and comparison, and detection and verification. Search and comparison includes preliminary scanning of the topographic area; the process is usually quicker and it is used to initially detect suspicious occurrences, such as unusual movement of people and vehicles. The detection and verification process involves revisiting areas of suspicious occurrences and scanning the areas in more detail for verification. Clearly this verification phase is expected to be of higher cognitive load than the first detection phase. The areas on the map that require detection and verification are expected to consume more of the test subject's attention, resulting in increased fixation of gaze, gaze vector, and gaze duration is expected to be seen in those areas. "Indeed a fixation at a given location is strong evidence that attention has been there" (McCarley & Kramer, 2008). Renninger et. al. (2010) conducted a study that showed relationship between pupil dilation and detection and verification of the target in a visual search task. Also, one may suspect there would be a noticeable change in behavior measured with physio measures such as the EEG and the galvanized skin response- perhaps a higher heart rate and pulse due to high cognitive load during those times (detection and verification). Furthermore, targets can be missed even though the test subject has his eyes fixated on it for some time. Vachon et. al. (2012) suggest that this is due to a failure of "attentional processes." That is, failure to implement enough conscious cognitive

thought processes to successfully detect the target. Eye trackers are useful in detecting such occurrences, as they are significant factors of consideration in a visual search task.

Limitations of eye tracking tools are generally centered on accuracy and usability. Accuracy includes temporal and spatial accuracy. Temporal accuracy is the timing of the user's POG with the visual stimulus/stimulus events as a function of the computer's processing capacity, while spatial accuracy pertains to accuracy of the user's POG compared to where the user is actually looking (Morgante et al., 2012). The term usability describe the ease of calibration and flexibility during experiments (Morgante et al., 2012). Calibration of POG is accomplished by moving a stimulus across specific locations on x and y axis of the screen and instructing the user to visually follow the stimulus. Then the eye tracker computer records the user's corneal reflection on the screen coordinates. Cheng and Vertegaal's (2004) study hypothesized that workload and accuracy follow the Yerkes-Dodson framework, which states that depending on the nature of task a low, medium, or high arousal/stimulus is required for optimum performance. That said, the wrong amount of stimulus will be either too low to stimulate adequate cognitive performance or so high that it overwhelms the test subject; either case inevitably leads to poor performance.

The Yerkes-Dodson model (developed in 1908) states that there is a strong relationship between the severity of arousal and the level of performance in humans. When a subject is given a task to complete, the amount of stimulation can be affected by environmental factors such as noise, temperature, and visibility as well as cognitive factors such as time pressure, presence/absence of audience, presence/absence of a supervisor, hints, instructions, etc. The model suggests that arousal may improve performance, but this improvement ends at a certain level, and the intensity of stimulation is dependent on nature of the task.

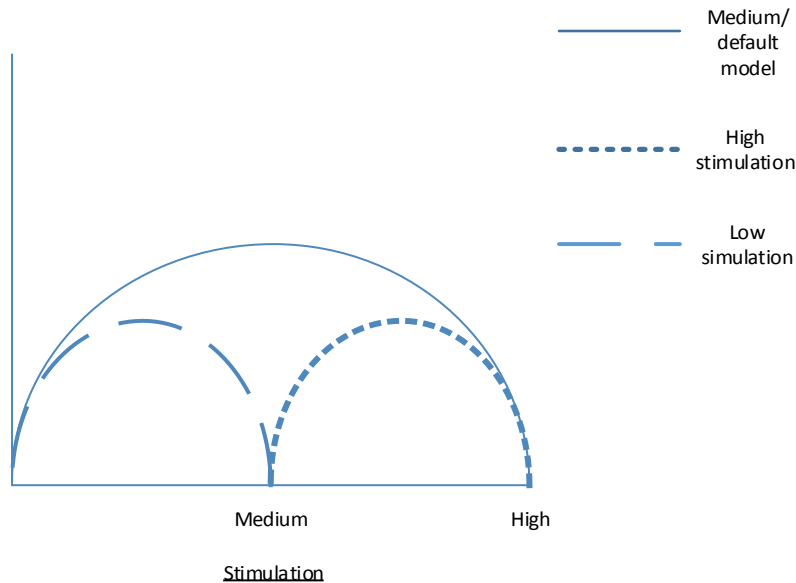


Figure 5: Yerkes-Dodson graph showing curves of high medium and low stimulations

Furthermore, the performance apex is also dependent on the task. For example, an extremely high stimulation, such as being chased by a grizzly bear, might be effective if a track coach wants to obtain the best performance from his athletes. Therefore, a high stimulation is required to accomplish optimal performance for this variety of task and the peak curve shifts to the right of graph, shown below.

Though high stimulation tends to improve performance, the stimulus that causes the performance boost may differ from task to task. For example, a soccer player's performance may become improved when there is a crowd cheering him/her on, whereas a crowd may hinder the performance of the chess player. Conversely, a soothing music may improve focus and performance of a chess player whereas the stimulation would be too low to help the soccer player. In other words, a soccer player's optimal performance is achieved at medium stimulation- so the peak curve is at medium; whereas a chess player's optimal performance is achieved by low to no stimulation resulting in a left shift of the peak curve (Duggan, 2012).

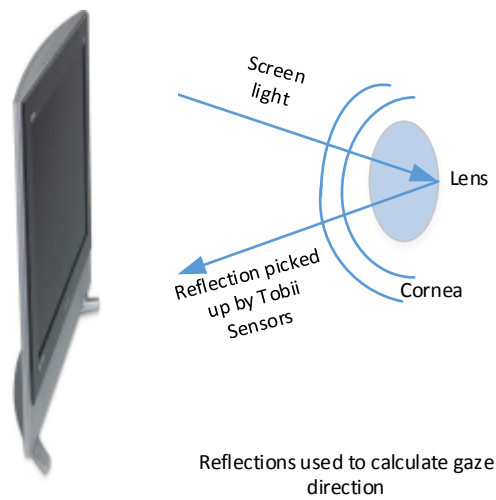


Figure 6: Multiple Reflections used to Calculate Gaze Direction

RESEARCH COMPONENTS

The research components- research questions and experimental outline-will assist in narrowing down the focus of this study.

Research Questions

As stated in the introduction, this study was interested in learning the impact that variables Area of Coverage and Amount of Activity have on the performance of an observer who watched up to six watch windows. The approach taken in this study was to assign *number of watch windows* and *level of AOC/ACT* as variables that would contribute to performance and perceived difficulty. According to stated objectives of the study, the following research questions were to be addressed:

- 1) Is there a difference in performance as an observer watches *one to six watch windows*, given high or low *levels of area of coverage and amount of activity*?
- 2) Is there a difference in perceived difficulty between an observer watching windows with a high *level of area of coverage and activity* and an observer watching windows with a *low level of area of coverage and activity*?

Given the above questions, the following hypotheses were generated:

H₀₁: There is no difference in the *number of watch windows* an observer can effectively monitor between high and low levels of area of coverage and amount of activity.

H₁₁: There is a difference in the *number of watch windows* an observer can effectively monitor between high and low levels of area of coverage and amount of activity.

H₀₂: There is no difference in perceived difficulty between observers watching windows with high and low levels of area of coverage and amount of activity.

H₁₂: There is a difference in perceived difficulty between observers watching windows with high and low levels of area of coverage and amount of activity.

Task Models

Because information obtained from this study is intended for use in the ISR domain, it is important to study the task model of IAs. Figure 7 illustrates the central task that an IA undertakes in a mission. The tasks proceeds as follows: 1) acquire information about target, 2) initiate target search, 3) detect target signal, 4) no target signal detected, 5) identify Target, 6) verify target, 7) initiate callout, 8) Mission End. These central tasks are simulated in the experiment

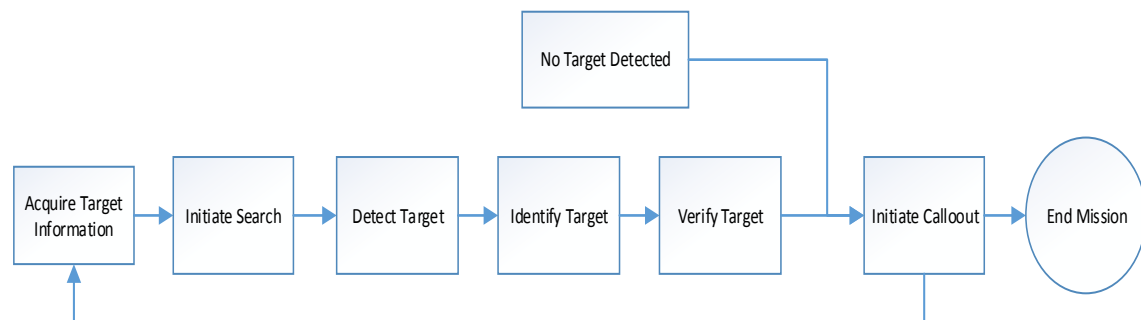


Figure 7: Central Task Model of an Intelligence Analyst

Figure 2 outlines factors that effect the IA's workload and methods for obtaining workload data. The ultimate goal of IAs and mission planners in a given visual search task is to answer questions that pertain to Essential Elements of Information (EEIs), which makes up a condensed version of the information required by the client (Paul, 2013).

Intelligence analysis carries challenges of effective individual and team perception (Trent, Patterson & Woods, 2007). Previous research has been done to investigate the challenges associated with the intelligence analyst's work. Paul (2013) discovered that the intelligence

analyst faces many challenges during a mission and identified various techniques to measure these challenges. They discovered difficulties associated with the cognitive demands required of analysts during a mission, cognitive heuristics that involve biases in decision making, and decision points that lead to errors. The findings from this study along with interviews with analysts have identified two specific demands that impact performance. These two demands are Area of Coverage (AOC) and Amount of Activity (ACT). This study is going to leverage their findings by manipulating *AOC/ACT* and measuring the difference in both performance and perceived difficulty.

METHODS

Participants

The Participants recruited for this study included 25 adults (12 males and 13 females) between the ages of 22 to 45 who had normal or corrected vision. Participants were recruited from the Wright State community and all had experience using a computer.

Apparatus and Stimuli

The participants used the Tobii T120 equipment, which consisted of a 17-inch LCD monitor with an integrated eye tracking system. Gaze directions were computed by capturing multiple reflections of the light source on the eye as shown in Figure 6. The vector between the pupil center and corneal reflection were mapped out on the screen during calibration procedures (Weigle & Banks, 2008). Cameras and light sources were affixed to the monitor. The Tobii system performs binocular tracking at sixty hertz. Generally, head movements were allowed within a 40-by-22-by-30 centimeter area with the user's head centered at about 70 centimeters from the camera (Weigle & Banks, 2008). Tobii uses "near-infrared, image-based corneal-reflection" technologies (Weigle & Banks, 2008). It is camera based so it is non-intrusive; that is, the cameras were not attached to the user's head. Subjects were instructed to keep their head movements to a minimum and attempt to maintain a head position relative to that of the center of the screen.

The Tobii screen displayed the video footage that was used as the stimuli in the experiment. Participants used a mouse as an input device. The screen numbers were labeled on the corners of each monitor, placing them out of the observer's line of sight when viewing the tasks, but immediately available for reference when initiating a callout. A Logitech USB desktop

microphone was used to capture the subjects' callouts. The auditory data was synced with the gaze tracking data from the Tobii.

Limitations to the study, such as the area of the Tobii monitor (17 inches in diagonal measurement) and the nature of available imagery, allowed for the use of a maximum of six simultaneous displays. A keyboard was not available because it was not needed. A mouse was available to the subjects. The experiment setup is illustrated on the figure below.

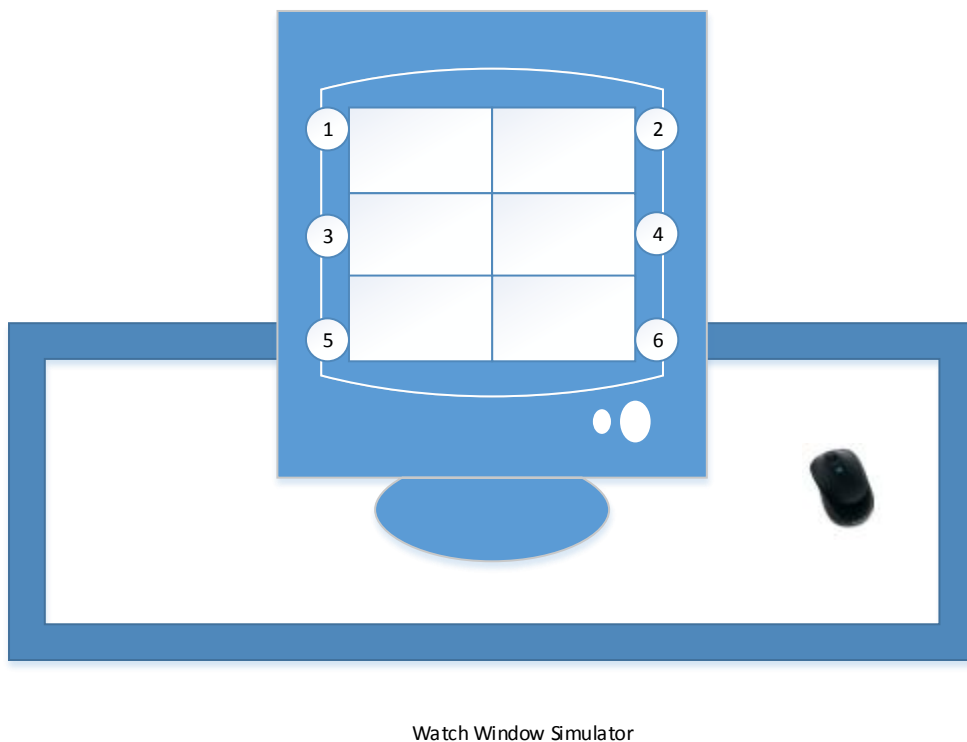


Figure 8: Six-screen Display and Labeling

Experiment Design

An informal pilot study was conducted using three observers in order to specify criterion used to define and assign tasks of high or low *levels of AOC/ACT*. The observers viewed a series of videos, a summary of which can be seen in and Table 2 and Table 3. For this study, AOC and ACT were not separated for two reasons. First, during the cognitive task analysis using real footages,

the IA never distinguished which one was more important. The second reason is that available data for this study did not allow for differentiation between AOC and ACT within an image sequence.

Table 2: Description of High and low levels of AOC and ACT

	AOC	ACT
Definition	Amount of area covered by UAV	Amount of activity witnessed in viewing area
Low	<u>Instructions:</u> subject views entities ≤ 2 on the display eg) one road, one building and a road... <i>This one entity could be across multiple screens</i>	<u>Instructions:</u> ≤ 8 occurrences of signal
High	<u>Instructions:</u> subjects view entities > 2 on the display eg) on the entire display...	<u>Instructions:</u> ≥ 18 occurrences of signal

Table 3: Nature of Tasks for Low AOC/ACT and high AOC/ACT

Task	
Low AOC/ACT	High AOC/ACT
Reading task	
Counting task ≤ 8	Counting task ≥ 18
Identification task	
Task that requests a screen number	

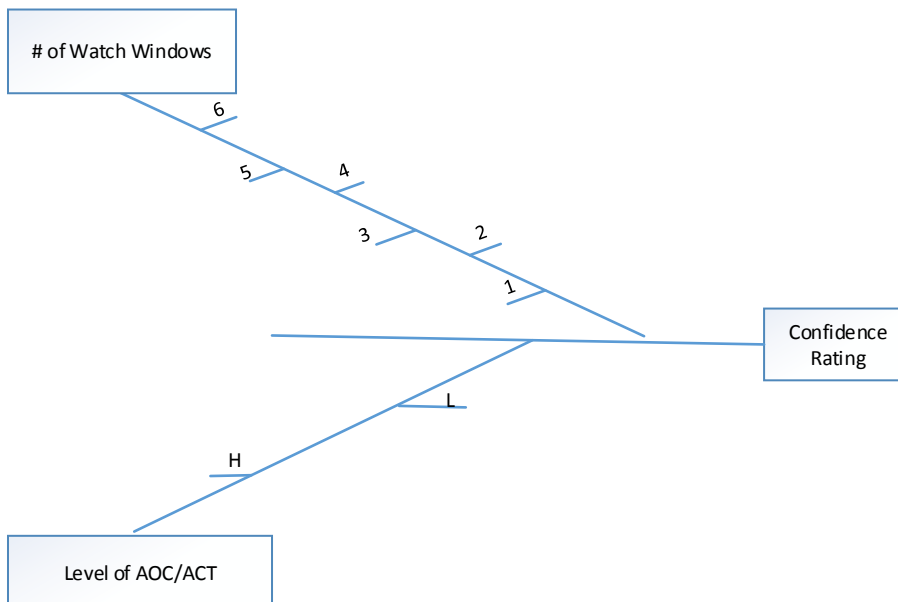
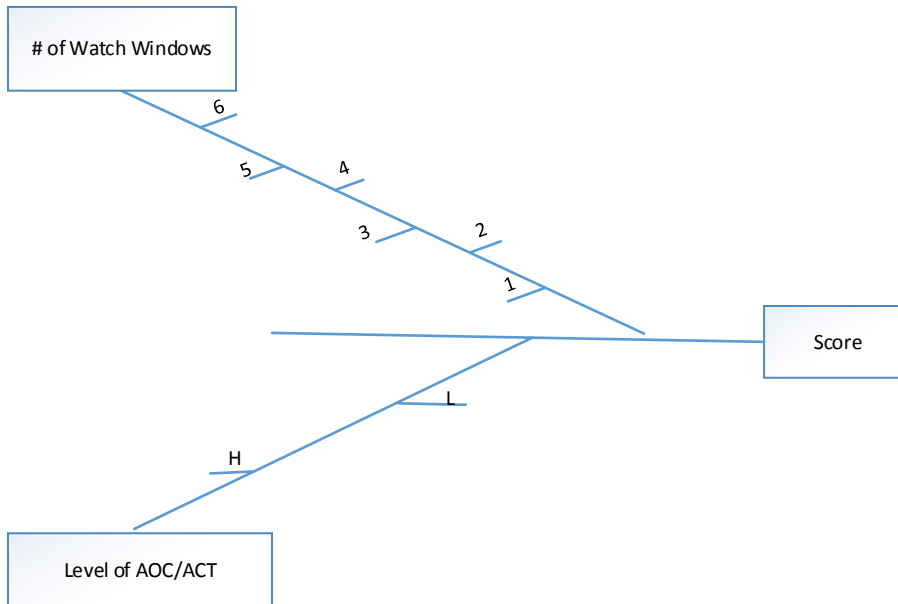
The experiment was a 2 X 6 within-subject method. All subjects were exposed to all levels of independent variables. Each subject was treated with more than one level of each factor. A within-subject method was ideal because since all subjects were assumed to have equal level of expertise, it allowed the investigators to detect other effects on performance that could have potentially gone unnoticed in a between-subject design (Tabachnick & Fidell, 2001). Also, the within-subject method has a smaller error variance compared to the between-subject method (Tabachnick & Fidell, 2001).

Independent Variables

The main independent variables in the experiment were a) the *level of area of coverage and amount of activity (AOC/ACT)* which were high or low, and b) *number of watch windows* as shown in the diagrams in Figure 9.

Dependent Variables

The foremost dependent variables collected in the study were a) experiment score and average score, b) confidence rating, and c) NASA-TLX rating. Each fish-bone diagram in Figure 9 corresponds to one of the dependent variables. Scores comprise of how many of the targets the participant was able to identify and were graded on a scale of 0 to 1 (1 = 100% of targets identified). Confidence ratings that were obtained after each task and were graded on a scale of 1 to 7 (7 = very confident), and NASA-TLX ratings that were obtained after each stimulus were scored from 0-100 (100 = high workload).



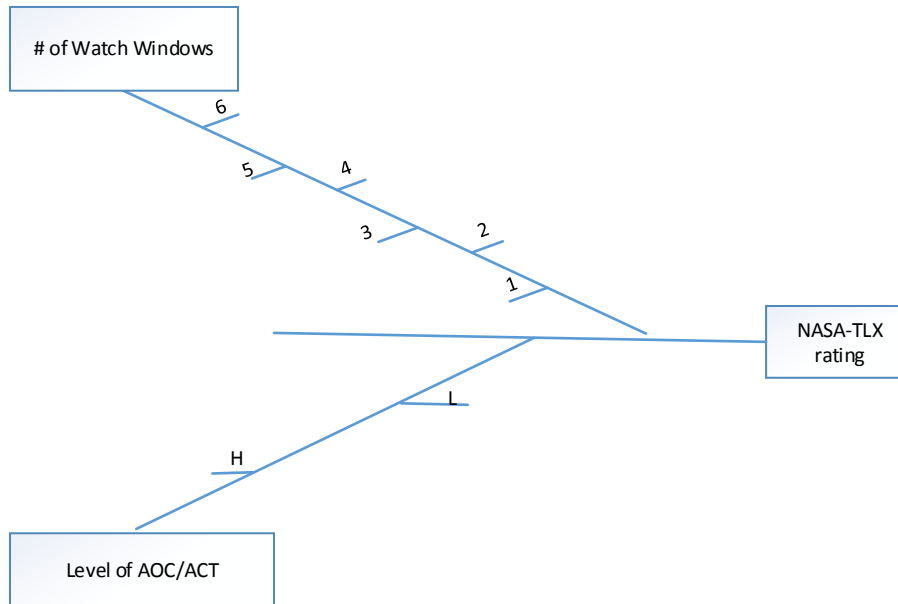
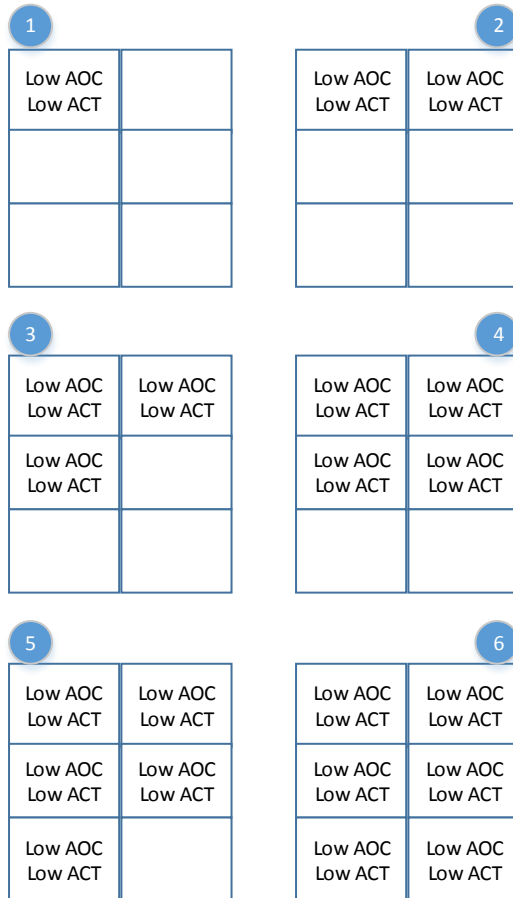


Figure 9: Fishbone diagram showing the independent variables with respect to each dependent variable

Stimulus I

The objective of stimulus I was to help the research team to understand the subjects' baseline capacity. It comprised of all easy tasks (low AOC and low ACT).



Stimulus I

Figure 10: Stimulus I- easy tasks (low AOC/ACT); one to six watch windows or screen displays

Stimulus II

In contrast to stimulus I, stimulus II was designed to obtain a top curve performance data from the subjects. In other words, stimulus II was expected to provide investigators with data that represents the subjects' maximum performance within the experiment constraints.

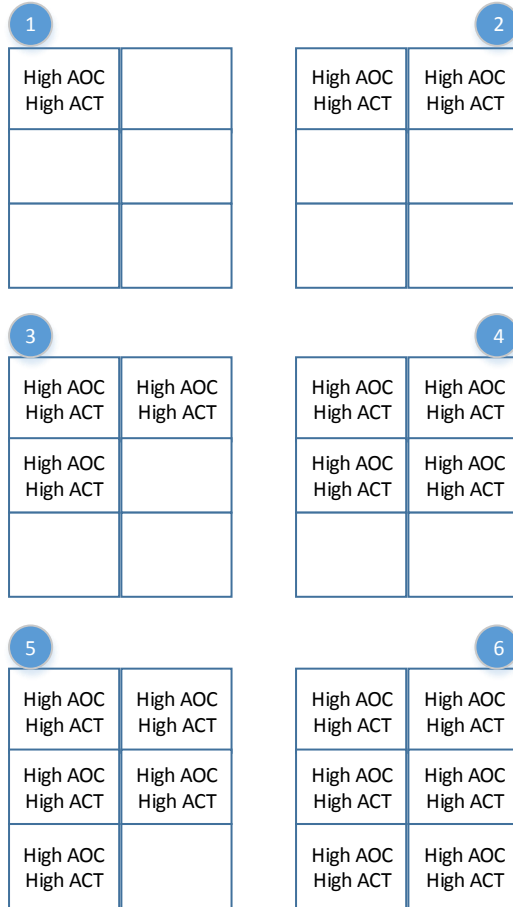


Figure 11: Stimulus II- hard tasks (high AOC/ACT) one to six watch windows or screen displays

Stimuli Design

Each of the stimuli was presented to all participants in a randomized complete block design. Both independent variables, *number of watch window* and *level of AOC/ACT*, were randomized. The order of stimuli presented to each subject was developed in JMP. Stimuli were divided into sets of high or low difficulty. The primary determinants of task difficulty are AOC and ACT. The secondary determinants of difficulty are instructions given to the subject and amounts of essential elements of information (EEIs) or signals in each task.

For this experiment, an easy task was a video that has a small area of coverage and a small number of activity occurrences and a hard task was a video that has a large area of coverage and activity level. When the subject was instructed to monitor a small area within a larger scale of coverage, the task difficulty was defined as easy, whereas when the subject was instructed to monitor a large area the task difficulty was hard. A small area was defined as two or less areas. A large area was defined as areas greater than two. In an easy task for example, the subject would be instructed to count the number of moving vehicles in roads X and Y only instead of looking at the entire display. In this task, the subject focuses attention on those two areas and ignores other signals elsewhere in the display. On the other hand, in a hard task the subject could be instructed to count the number of moving vehicles throughout the entire display. Obviously, this type of task would be more challenging because the subject will have to pay attention to more areas.

The next secondary variable that determines difficulty is the number of signals that the subject has to detect. In an easy task, the subject would be required to detect ten or less signals throughout a display. In a hard task, the subject would be required to detect signals numbering eighteen or greater.

There are two reasons for this two-way designation of difficulty (easy and hard). The first is because of the nature of the imagery, which is explained in detail in the difficulty level section of the paper, the second is to be able to highlight the effects of AOC and ACT. Assigning easy/hard difficulty to the watch windows experiment and avoiding an intermediate or medium difficulty criterion highlights effects of the variables in the results. For example, if task difficulty affects cognitive workload, which may affect error rates, then results would be expected to highlight a clear margin in observers' performance between high and low difficulty tasks. One may predict that error rates would be higher for observers viewing difficult displays, especially as number of displays is increased.

Due to nature of data acquired for stimuli, the variables AOC and ACT jointly represent an easy or a hard task. In other words, the variables AOC and ACT are not isolated; rather, they are used as a pair to represent an easy or a hard task.

Due to camera altitude in the video data available for the study, when the camera view is restricted to a smaller location out of the entirety of footage, the coverage area is small. The majority of footage is of a landscape or a cityscape. The camera's altitude allows for an aerial view of the specific part of the city in which traffic patterns range from moderate to low; however, due to availability of multiple venues to be viewed by the observer in the city, a task that uses city footage is designated a difficult task. On the other hand, when the camera is zoomed in to acquire a lower area of coverage the resulting view is that of a building, a field, a parking, etc. with little to no activity. Thus the use of AOC and ACT jointly to determine difficulty is due to design limitations.

- Easy task: An easy task is task that consists of a low AOC and low ACT
 - Subject is instructed to view two or less areas on the display
- Hard task: A hard task is a task that consists a high AOC and high ACT
 - Subject is instructed to view more than two areas on the display

Procedure

The experimental area was designed to mimic that of an intelligence analyst's workplace. The lighting was set to dim or dark (no lighting) depending on the comfort of the test subjects. The participants followed the following procedures during the experiment:

1. Subject was briefed on background and purpose of the study

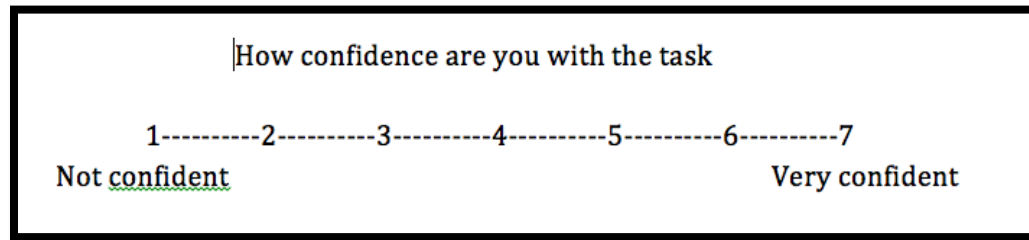
2. Document of consent was presented and summarized orally, then time was given to the subject to read, ask any questions, and sign the document
3. Subject was be given a background pre-test questionnaire to fill out
4. The subject was seated in front of the workstation, which includes the Tobii monitor, the keyboard and the mouse
5. Instructions were given
6. Subject proceeded with the tasks (subject did not have the ability to pause, rewind, and fast-forward the videos in the stimuli)
 - a. There were two stimuli
 - i. Subject answered a a confidence questionnaire after each task (12 total)
 - ii. Subject filled out a NASA-TLX after each stimulus (two total).
7. Post-test questionnaires were given to the subject
8. In an interview format the subject discussed the experiment with the experimenter
9. Subject was thanked for his/her contribution and dismissed

Data collection

Data was collected through NASA-TLX, interviews, pre- and post-test questionnaires and confidence rating. The pre-test questionnaire issued in this study acquired demographic data such as age and gender, as well as deficiencies such as color blindness or visual impairments.

Although the subject population was not separated on basis of expertise, the pre-test questionnaire attempted to obtain information about the subject's familiarity with the tasks. For example, if a subject is an IA who is exposed to rigorous visual search tasks on a daily basis, he/she might perform exceptionally well compared to the other subjects. The confidence questionnaire that was presented after each task asked the subject how confident they felt about the completed task.

On the Tobii monitor, a scale of one (not confident) to seven (very confident) was presented and the subject responded with a mouse click



How confidence are you with the task

1-----2-----3-----4-----5-----6-----7

Not confident Very confident

Figure 12: Confidence Questionnaire

The post-test interview was given to the subject at the end of the experiment. It was formatted in a seven-point Likert scale that allowed the subject to select a range from one- “strongly disagree” to seven- “strongly agree” to every question. Then, there was a space below each scale labeled “comment” for the subject to elaborate on his/her selection. The questionnaire contained 14 questions which covered areas such as easiness to learn, easiness to recover errors, mental workload, satisfaction etc.

The purpose of the interview was to allow the subject to orally provide input on overall experiment. Throughout the experiment some tasks would be more difficult than others. The subject was given the opportunity to comment on what made a task more difficult and what make same tasks easy, and what actions might be taken by the researchers to help make tasks more doable. The experimenter documented the subject’s inputs and suggestions.

A microphone was used to obtain auditory response. Synchronizing the microphone with the Tobii allowed the investigators to capture what the subject saw on the screen along with auditory responses. Also, using an auditory response mechanism captures the nature of the IA’s work environment; Subject Matter Experts (SMEs) suggest that callouts are the most common method of communication.

In addition to auditory responses, the subjects were instructed to write their answers down on a piece of paper that was provided for them. The written responses were provided after viewing each task so that they did not take their eyes off the monitor and potentially miss events. Because the experiment was focused on the observer's ability to monitor several screens, given that the eyes are fixed on the display, it minimized events where the subject gazed away from the screen.

RESULTS

Number of Watch Windows and Area of Coverage and Amount of Activity on Performance

In this experiment, The JMP 10 statistical analysis software: Analysis of Variance (ANOVA) was used to analyze data. Analysis tools that were utilized in JMP included Rsquare, Least Square Means (LSM), fit Y by X plots, LSMeans Tukey's HSD (honestly significant difference), and graph builder plots. Performing an Analysis of Variance (ANOVA)- fit model test, it was discovered that the *number of watch windows* ($p < 0.001$), level of AOC and ACT ($p < 0.001$), and a cross of both variables ($p < 0.001$) all have significant effect on performance (score) in a scale where a p-value of less than 0.05 is considered significant.

The next step was to examine the effects of each number of watch windows. The general trend occurred as expected. Overall, there was a decrease in performance as subjects went from one to six watch windows. However there were a few deviations from expected behavior. According to the Least Square Means (LSM) table, four screen-displays earned the highest score (LSM = 0.75) and performance in two-screen displays (LSM = 0.74) was slightly higher than performance in one-screen displays (LSM = 0.73). Further analysis was done using an LSMeans Tukey HSD (honestly significant difference): a connecting letters report in Table 10 in the appendix. It was discovered that score did not differ significantly from one to four screen displays, because all of the scores were categorized with the same letter.

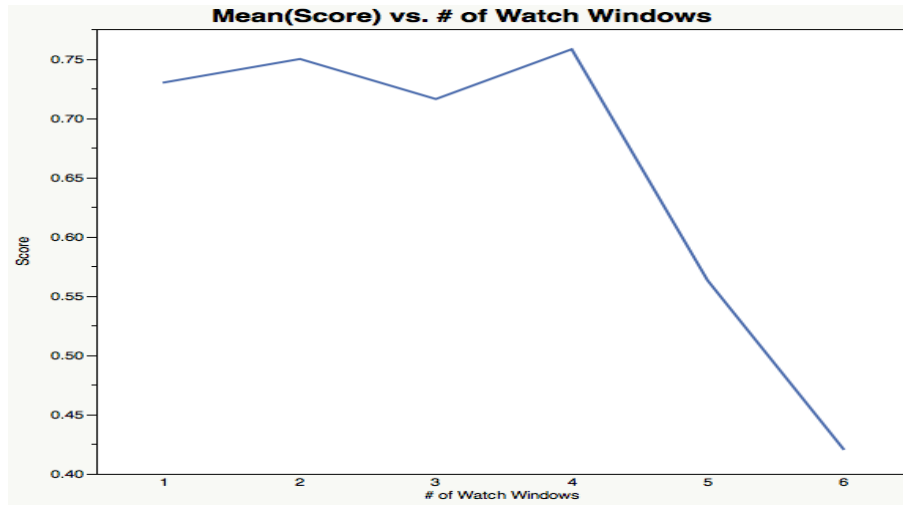


Figure 13: General Behavior of Performance Versus Number of Watch Windows

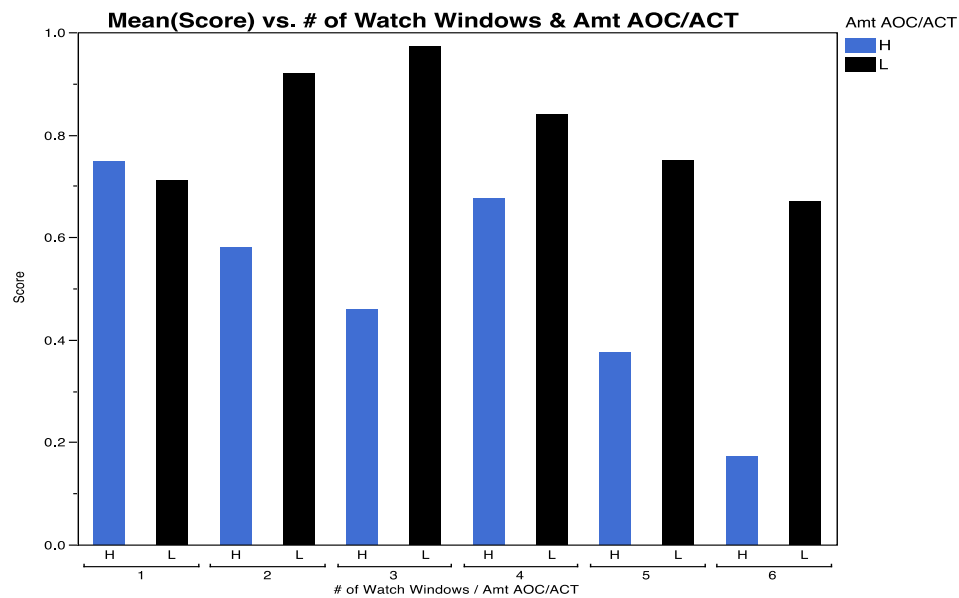


Figure 14: Mean Score versus number of Watch Windows and Amount of AOC/ACT

To identify the effects of *level of AOC/ACT*, an LSM table was generated. The data revealed that subjects earned higher scores on tasks of low AOC/ACT (LSM = 0.81) than on tasks of high AOC/ACT (LSM = 0.50).

An LSM table was also generated to observe the effects of the combined variables *number of watch windows* and *level of AOC/ACT*. Almost all *number of watch windows* (from two to six-screen displays) showed significantly higher scores in tasks of low AOC/ACT than tasks of higher AOC/ACT, with the exception of the one-screen display. One-watch window showed scores that were similar in LSM (high AOC/ACT = 0.74, low AOC/ACT = 0.71). Therefore, on average, the obtained results were consistent with predicted behavior.

Number of Watch Windows and Area of Coverage and Amount of Activity on Perceived Difficulty

A fit model test was performed comparing *number of watch windows* and *level of AOC/ACT* versus confidence ratings individually. The effects test revealed that *number of watch windows* had a significant effect on perceived difficulty ($p < 0.001$) while *level of AOC/ACT* ($p = 0.280$) and a cross of both variables ($p = 0.128$) did not have any significant effect.

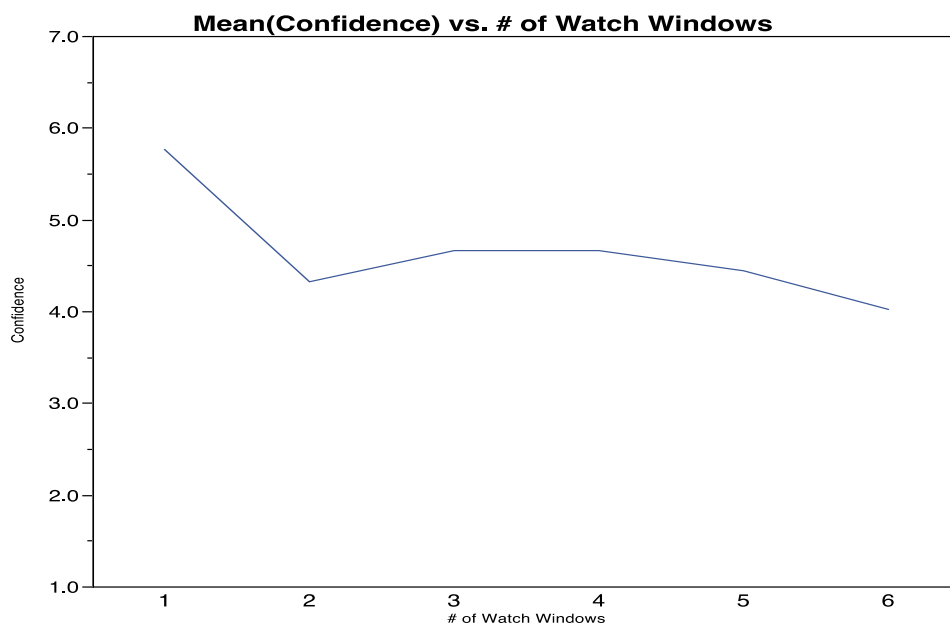


Figure 15: Graph of Confidence versus Number of Watch Windows

Observing the effects details revealed expected results: confidence ratings generally decreased continuously from one (LSM = 5.76) display to six (LSM = 4.02) displays. An LSMeans Tukey HSD: connecting letters report was performed to further study the effects of *number of watch windows*, with the results located in Table 19 of the appendix. The results showed that the confidence rating associated with the one-window tasks (classified under the letter A) was significantly higher than that of the rest (classified under Bs) and a significantly lower confidence rating was associated with two-, five- and six-window tasks (classified under Cs). Confidence ratings were appropriate for five- and six- window tasks but unexpectedly low on two-window displays.

Although the effects of *level of AOC/ACT* were found to be insignificant to confidence ratings ($p = 0.285$), an LSM table identified that subjects reported to be more confident in tasks of low AOC/ACT (LSM = 4.74) than tasks of high AOC/ACT (LSM = 4.56). The findings are consistent with predicted behavior. The other measurement of perceived difficulty is the NASA-TLX ratings that were obtained after each stimulus.

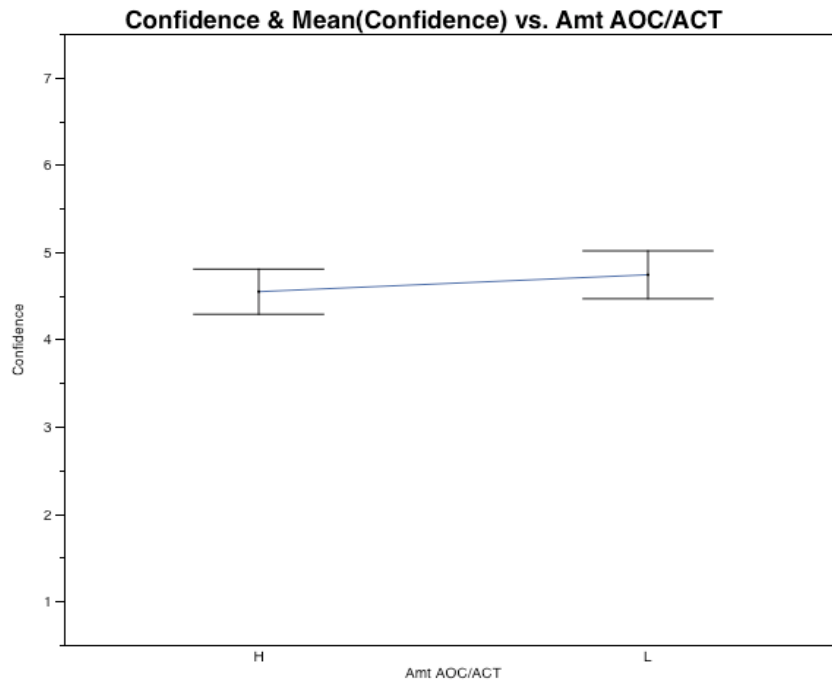


Figure 16: Graph of Confidence Ratings versus Amount of AOC/ACT (High/Low). Each error bar is constructed using a 95% confidence interval of the mean.

The effect of *level of AOC/ACT* was also found to be insignificant to NASA-TLX at a p-value of less than 0.05. Further observation of effects details revealed that subjects assigned lower TLX scores to tasks of low AOC/ACT (LSM = 60.27) than they did to tasks of high AOC/ACT (LSM = 63.07). The findings were consistent with predicted behavior.

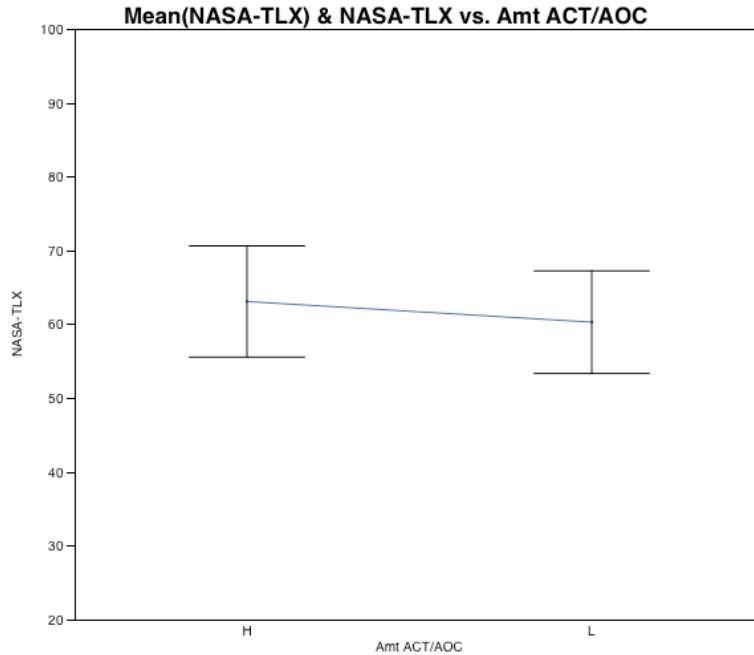


Figure 17: Graph of NASA-TLX versus Amount of AOC/ACT (High/Low). Each error bar is constructed using a 95% confidence interval of the mean.

Fixation

Analysis of Fixation data revealed that *number of watch windows* ($p < 0.001$) and *amount of AOC/ACT* ($p < 0.001$) had a significant effect on fixation. There was a linear relationship between *number of watch windows* and fixation: as the *number of watch windows* increased, so did the fixation count. This illustrates that fixations increase with cognitive load. These findings are consistent with previous research as discussed in the literature review section of this paper.

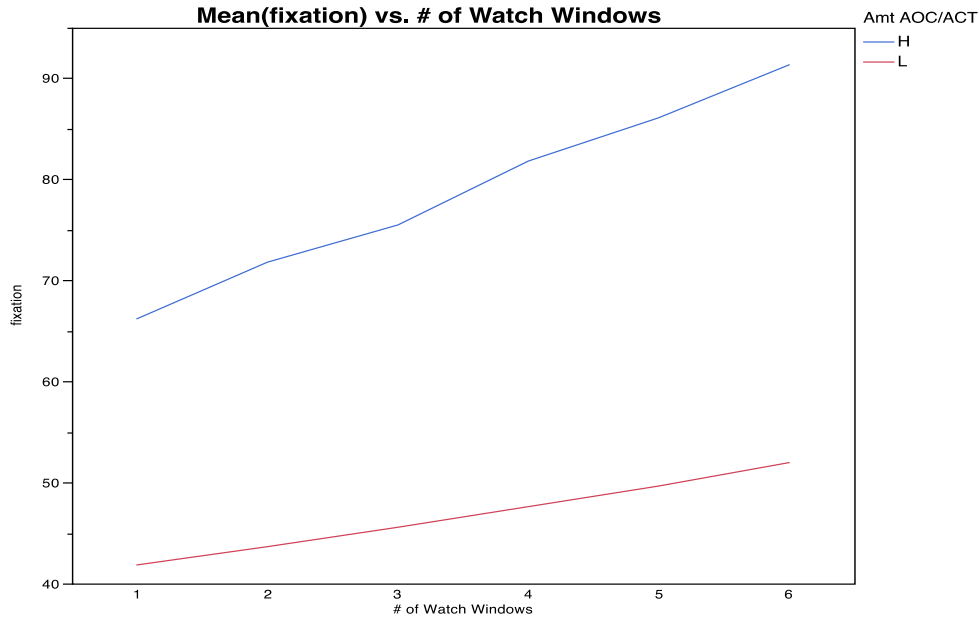


Figure 18: Graph of mean fixation versus number of watch windows

Sample Size

K-sample means (as opposed to one or two sample means) in JMP was used to calculate the power range or power-of-test (0-1) because the study was comparing difference in mean values across multiple (k) samples (greater than two). Power tests yielded adequate values for all of the following significant variables: effect of *number of watch windows* on score (Power = 0.70), *level of AOC/ACT* on score (Power = 0.99), and *number of watch windows* on confidence rating (Power = 0.7). *Level of AOC/ACT* yielded the strongest power value, which implies that level of area of coverage and amount of activity was the strongest determinant of performance. However, the fact that power values for all significant variables were suitable (≥ 0.7) justifies the sample size ($n = 25$).

GENERAL DISCUSSION

The results obtained from the experiment support the hypothesis that the *level of AOC and ACT* do have significant effect on the *number of watch windows* an observer can monitor in a visual search task. Results also support that there is a difference in perceived difficulty between observers watching displays of high AOC/ACT and those watching displays of low AOC/ACT.

However, in the visual search task, it is noticeable that performance does not vary significantly from one to four watch-windows in the mean score versus number of watch window graph (Figure 13). This pattern deviates from the expected behavior that performance should decrease as the number of watch windows increases. This observation suggests that the observer's performance from one to four watch windows can be less predictable than performance on watch-windows greater than four. For an observer who is watching one to four watch windows, internal noise factors such as tiredness, sleep deprivation and fatigue, or external noise such as distracting images or sounds in the experiment room may cause a decline in performance, whereas an observer who is not affected by the noise factors may perform better; these variables contribute to unpredictability in an observer's performance between one to four watch windows. On the other hand, performance drops when the observer is tasked to monitor more than four watch windows independent of the presence or absence of noise.

Similarly, in the mean (score) versus number of watch windows and amount of AOC/ACT graph (Figure 14), performance was significantly higher when the observer completed tasks of low AOC/ACT throughout one to six watch windows, with the exception of one watch window. This suggests that differences in difficulty level presented by high level of AOC/ACT and low levels of AOC/ACT are minimal for a one watch window display, while difference in difficulty becomes more pronounced when the tasks are presented in more than two watch windows.

P-values reported for the variables *number of watch windows*, *level of AOC/ACT*, and a combination of both were strong evidence that they were all significant factors that affected the participant's performance. Since analysis suggested that *level of AOC/ACT* and *number of watch windows* by themselves affected performance, workload can be defined as a counter balance of both variables. That is, an observer who is given visual tasks of low *levels of AOC/ACT* is expected to effectively monitor more watch windows than one who is given tasks of high *levels of AOC/ACT*. Conversely, an observer who is given visual task of a fewer *number of watch windows* is expected to effectively monitor higher *levels of AOC/ACT* than one who is given a larger *number of watch windows*. The other result that supports this counter-balance idea is the result obtained from performing a cross of *number of watch windows* and *level of AOC/ACT*. The pattern observed in comparison of means suggest that the observer is more likely to successfully complete visual search tasks with higher numbers of windows when tasks are of low AOC/ACT and vice-versa.

In addition, comparison of least square means revealed that performance was significantly higher at windows of low *level of AOC/ACT*, with *number of watch windows* of 4, 5, and 6 and significantly lower at *high level of AOC/ACT* with *number of watch windows* 1, 2, and 3.

In addition, findings on analysis of fixations provide evidence that fixation is an indication of workload. The within-subject test implies that fixation is subjective because each participant worked at different speeds and fixation counts reflected that. Therefore, in a visual search task, fixation is not a direct measure of workload; rather, it is an indication of the level of workload. There was no significant correlation between fixation counts and NASA-TLX ratings at $\alpha = 0.05$. However, the general pattern observed indicated that fixation increased as NASA-TLX ratings increased, which followed the expected trend. After data analysis, it was observed that as tasks went from low to high *level of AOC/ACT*, the frustration rating in TLX increases.

Though there was no noticeable change in frustration rating between one and four *watch windows*, an increase in frustration ratings was observed when the number of watch window exceeded four. Also, there was a correlation between gender and performance. A close to even distribution of male and female subjects for this study allowed experimenters to observe a clear difference in behavior between the two genders. Female subjects seemed to perform better, but reported less confidence in their performance than their male counterparts throughout the tasks.

Another significant finding was that performance seemed to peak at four window displays. This result supports the findings of Sulman and Sanocki (2008) that the human performance in visual search tasks increases up to four displays and drops with any additional displays. Furthermore, the finding that there was no significant change in performance between one- to four- screen displays supports Nillie Lavie's (1995) statement that low load may cause under-load and the observer's cognitive capacities may spill over. A certain load (in this case four windows) provides enough of a challenge for the observer to implement their full cognitive capacity and therefore perform better, and when this load is surpassed (in this case five and six window displays) the observer is overloaded and performance drops, as Sulman and Sanocki (2008) suggested. Therefore, this study further validates both Sulman and Sanocki (2008) and Lavie's (1995) findings.

When observing performance on tasks containing a low level of AOC/ACT there is rise from one to three watch windows, followed by a gradual drop in performance from three to six watch windows. In tasks with a high level of AOC/ACT, there is continuous drop in performance from one to three watch windows, a spike at four watch windows, then a decline from four to six watch windows. Furthermore, tasks with *high AOC/ACT*, Figure 14 exhibited a close fit to the Yerkes-Dodson model from three to six *watch windows*. Performance increased from three to four *watch windows* and then dropped from four to six *watch windows*. This pattern suggests that the task presented in four *watch window* and high *level of AOC/ACT* allowed the observer to be

more engaged compared to the three *watch windows*, that did not provide enough stimulus, and five and six *watch windows*, that provided too much stimulus, resulting in a drop in performance. The performance disparity between high and low *level of AOC/ACT* seen in three *watch windows* in Figure 14 suggests that out of all *number of watch windows* the three *watch windows* had the most significant effect on the difficulty of task.

Although ANOVA did not show a significant difference in confidence ratings between observers watching windows of low AOC/ACT and those watching high AOC/ACT windows, it did show a significant difference among observers watching different *numbers of watch windows*. The general pattern obtained from the LSM analysis suggests that as the participant's *number of watch windows* increased, their confidence level decreased. Therefore, behavior is as predicted. Studying the effects details of *level of AOC/ACT*, it was discovered that the participants reported higher confidence with tasks of low AOC/ACT than tasks of high AOC/ACT. NASA-TLX findings also suggest a higher perceived difficulty in tasks of high AOC/ACT than tasks of low AOC/ACT. Therefore, all findings support the conclusion that there are differences in perceived difficulty between observers watching different *numbers of watch windows*, and between observers watching windows of high and low *levels of AOC/ACT*.

Video quality varied throughout the experiment. At the same difficulty levels, participants tended to perform better on videos with better quality. The participants also struggled with mapping tasks. Mapping tasks are ones in which participants were instructed to count events that were portrayed in multiple screens, as opposed to one screen. Therefore, in addition to *levels of AOC/ACT*, a combination of mapping issues along with poor image quality contributed to poorer performance in the visual search tasks.

Investigation into the use of active and passive search strategies showed the predominant use of an active search strategy by observers. A serial search task has a higher workload than a

parallel search task because it involves active search. (Bruce and Tsotsos, 2009). The stimuli in this study were designed to be serial search tasks; therefore, most (23) of the participants reported that they employed an active search strategy as opposed to a sit-and-wait strategy. Two participants reported a combination of sit-and-wait and active search methods. They claimed that active search was used on tasks that seemed doable: usually in tasks of one to four *number of watch windows*, whereas the sit-and-wait strategy was method was employed in tasks with greater than four watch windows.

RELEVANCE TO THE INTELLIGENCE SURVEILLANCE AND RECONNAISSANCE DOMAIN

Findings in this study are of great relevance to understanding the IA's performance on visual search tasks in the ISR domain. These results provide enhanced knowledge of the effect of both AOC and ACT on the *number of watch window* an observer can effectively monitor. Quantitative analysis supports qualitative data gathered from SMEs. This provides workload recommendation for tasking IAs. The study also showed that demographics have an impact on performance.

The overall population showed a direct correlation between perceived difficulty and performance. Visual search tasks of low AOC and ACT minimized perceived difficulty, hence it can be reasoned that effectiveness in a visual search is fairly dependent on the observer's perceived workload. In other words, the observer is more likely to perform better when he/she is more confident in his/her ability to complete the task. This indicates that it is worthwhile to allocate resources to proper training.

The discovery that performance peaked on four-window displays which was supported by Sulman and Sanocki (2008), Lavie (1995) and the Yerkes Dodson framework is also useful information to the ISR domain. The performance results suggest that it is ideal to keep the *amount of watch windows* at four. The findings provide a solid platform for conducting experiments using IAs and assigning missions to analysts.

FUTURE WORK

The current study discovered that observers found challenges in tasks involving multi-tasking within multiple windows. For future studies, R^2 values that yield more correlation between dependent and independent variables can be obtained by taking other factors that may influence the observer's performance, such as amount of sleep obtained, computer literacy, state of mind, and time of the day, into account during analysis. Also, future studies may focus specifically on the area of multi-tasking in order to shed some light on the scopes and limitations of human performance in a visual search. Cognitive load measurements acquired during this study may be different if the tasks involved mapping. Studies that explore the field of visual mapping in multiple screen displays may be a useful extension of this study, and information obtained from such a study may also be useful to the ISR community.

Appendix A: Tables of task Results

Table 4: Table of performance for Stimulus I

	stimulus I						
Participant	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Average Score
1	1	1	1	1	1	0.83	0.971666667
2	1	1	1	1	1	1	1
3	1	1	1	1	0.85	0.5	0.891666667
4	1	1	1	1	0	0.66	0.776666667
5	1	1	1	1	0.85	0.83	0.946666667
6	0	0	1	1	1	0.66	0.61
7	1	1	1	0.5	0	0	0.583333333
8	1	1	1	1	0.5	0.66	0.86
9	0.6	1	1	0.5	0.3	1	0.733333333
10	1	1	1	0	1	1	0.833333333
11	0	1	1	1	1	0.5	0.75
12	0	1	1	0.5	0.85	1	0.725
13	1	1	1	1	1	0.66	0.943333333
14	1	1	1	1	1	0.83	0.971666667
15	0.3	1	1	1	1	0.66	0.826666667
16	1	1	1	1	1	0.83	0.971666667
17	0.3	1	1	1	0.71	1	0.835
18	1	1	1	1	0.14	0.66	0.8
19	1	1	1	0	1	0.83	0.805
20	0	1	1	1	0.71	0.66	0.728333333
21	1	1	1	0.5	0.71	0.33	0.756666667
22	0	1	1	1	1	0.66	0.776666667
23	1	0	1	1	1	0.66	0.776666667
24	0.6	1	1	1	0.14	0	0.623333333
25	1	1	1	1	1	0.33	0.888333333
Average	0.712	0.92	1	0.84	0.7504	0.67	

Table 5: Table of performance for stimulus II

	Stimulus II						
Participant	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Average Score
1	1	0.79	0	1	0.6	0	0.565
2	0	0.93	1	0.77	0	0	0.45
3	1	0	0	1	0.9	0.6	0.583333
4	0	0	0	0	0	0	0
5	0	1	0	0.6	0	0	0.266667
6	1	0.86	0.86	1	0.3	0	0.67
7	0	0	0	0	0	0	0
8	1	1	1	1	0.2	0.7	0.816667
9	0.7	0	0	0	0.2	0	0.15
10	0.8	0	0	0	0	0	0.133333
11	1	1	1	1	0.5	0.6	0.85
12	1	1	0	1	0.7	0	0.616667
13	1	1	1	0.92	0	0	0.653333
14	0.7	0	0	1	0.4	0	0.35
15	1	1	0.73	1	0.3	0	0.671667
16	0.5	0	1	0	0.9	0	0.4
17	0.8	0	0	0.74	0	0.6	0.356667
18	1	1	1	0.96	1	0	0.826667
19	0.8	0	0.8	0.81	0.9	0	0.551667
20	0.8	0.5	0	0.81	0.3	0	0.401667
21	1	1	0	0	0.9	0.5	0.566667
22	0.7	0.71	0.66	0.96	0.2	0	0.538333
23	0.9	1	1	0.81	0.5	0.5	0.785
24	1	1	0.73	0.92	0.6	0	0.708333
25	1	1	0.73	0.6	0	0	0.555
Average	0.748	0.5916	0.4604	0.676	0.376	0.14	

Appendix B: Performance

The table below shows the analysis of variance performed on the two main variables, *number of watch windows* and *level of AOC and ACT*, with respect to the dependent variable *performance*. R^2 represents the percentage of variance in the dependent variable that is predicted from the independent variables, or the degree to which the independent variables correlate to the dependent variables. The $R^2 = 0.23$ in the summary indicates that 23% of the variance is explained by the model. The other 77% of the variance can be explained by the response mean. $F < 0.0001$ indicates that the experiment model is a fit one.

Table 6: Summary of fit for variables number of watch windows and level of AOC and ACT.

Summary of Fit	
RSquare	0.233127
RSquare Adj	0.217424
Root Mean Square Error	0.36332
Mean of Response	0.656333
Observations (or Sum Wgts)	300

Table 7: ANOVA on the variables number of watch windows and level of AOC and ACT.

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	6	11.757539	1.95959	14.8452
Error	293	38.676428	0.13200	Prob > F
C. Total	299	50.433967		<.0001*

The table below represents a cross between the two independent variables and their effect on performance. At a 95% confidence level, it was determined that all three variables had a significant effect on the subjects' performance.

Table 8: Cross of variables number of watch windows and level of AOC and ACT.

Effect Tests					
Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
# of Watch Windows	5	5	4.6057307	7.3900	<.0001*
Amt AOC/ACT	1	1	7.1518080	57.3764	<.0001*
# of Watch Windows*Amt AOC/ACT	5	5	2.7780360	4.4574	0.0006*

The LSM table below indicates a general pattern of a drop in performance from one to six watch windows, although it is not a completely linear drop.

B-1: Effects of Number of Watch Window on Performance

Table 9: LSM of score for each number of watch windows (out of 1.00)

Least Squares Means Table			
Level	Least Sq Mean	Std Error	Mean
1	0.73000000	0.05138121	0.730000
2	0.74980000	0.05138121	0.749800
3	0.71600000	0.05138121	0.716000
4	0.75800000	0.05138121	0.758000
5	0.56320000	0.05138121	0.563200
6	0.42100000	0.05138121	0.421000

The table below represents the LSM of performance in tasks with high and low *level of AOC/ACT*. Scores were higher on tasks with low *AOC/ACT* than with high *AOC/ACT*.

Table 10: LSMeans Tukey HSD table showing the effects of number of watch windows on score

Level		Least Sq Mean
4	A	0.75800000
2	A	0.74980000
1	A	0.73000000
3	A	0.71600000
5	A B	0.56320000
6	B	0.42100000

B-2: Effects of level of AOC and ACT on performance

Table 11: LSM score for each level of AOC and ACT (out of 1.00)

Least Squares Means Table			
Level	Least Sq Mean	Std Error	Mean
H	0.50193333	0.02966496	0.501933
L	0.81073333	0.02966496	0.810733

The LSM table below shows the score gradient of a cross of the two the variables *number of watch windows* and *level of AOC/ACT*. The general pattern shows higher scores in low tasks and tasks with fewer watch windows though there were some random variations.

B-3: Effects of Cross of both *Number of Watch Windows* and *Level of AOC/ACT* on Performance

Table 12: Score for each number of watch window and level of AOC and ACT (out of 1.00)

Least Squares Means Table		
Level	Least Sq Mean	Std Error
1,H	0.74800000	0.07061082
1,L	0.71200000	0.07061082
2,H	0.57960000	0.07061082
2,L	0.92000000	0.07061082
3,H	0.46000000	0.07061082
3,L	0.97200000	0.07061082
4,H	0.67600000	0.07061082
4,L	0.84000000	0.07061082
5,H	0.37600000	0.07061082
5,L	0.75040000	0.07061082
6,H	0.17200000	0.07061082
6,L	0.67000000	0.07061082

Below is the LSM Differences Tukey HSD table and corresponding connecting letters report. Each cell contains four values: the difference in means, the standard errors of the difference and the lower and upper confidence limits (JMP®, Version <10.0.2>. SAS Institute Inc., Cary, NC, 1989-2007). A Connecting Letters Report appears by default below the LSM Differences Tukey HSD table. If levels are connected by the same letter, it implies that these levels do not differ statistically at a significance level of 0.05 ($\alpha = 0.05$). On the other hand, the levels that are not connected by the same letter are statistically different (JMP®, Version <10.0.2>. SAS Institute Inc., Cary, NC, 1989-2007.) .

Table 13: LSMeans Differences Tukey HSD table for the dependent variable score, and the independent variables number of watch window and level of AOC/ACT

		LSMean[j]											
Mean[i]-Mean[j]	1,H	1,L	2,H	2,L	3,H	3,L	4,H	4,L	5,H	5,L	6,H	6,L	
Std Err Dif													
Lower CL Dif													
Upper CL Dif													
1,H	0	0.036	0.1684	-0.172	0.288	-0.224	0.072	-0.092	0.372	-0.0024	0.576	0.078	
	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	
	0	-0.2931	-0.1607	-0.5011	-0.0411	-0.5531	-0.2571	-0.4211	0.04294	-0.3315	0.24694	-0.2511	
	0	0.36506	0.49746	0.15706	0.61706	0.10506	0.40106	0.23706	0.70106	0.32666	0.90506	0.40706	
1,L	-0.036	0	0.1324	-0.208	0.252	-0.26	0.036	-0.128	0.336	-0.0384	0.54	0.042	
	0.09986	0	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	
	-0.3651	0	-0.1967	-0.5371	-0.0771	-0.5891	-0.2931	-0.4571	0.00694	-0.3675	0.21094	-0.2871	
	0.29306	0	0.46146	0.12106	0.58106	0.06906	0.36506	0.20106	0.66506	0.29066	0.86906	0.37106	
2,H	-0.1684	-0.1324	0	-0.3404	0.1196	-0.3924	-0.0964	-0.2604	0.2036	-0.1708	0.4076	-0.0904	
	0.09986	0.09986	0	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	
	-0.4975	-0.4615	0	-0.6695	-0.2095	-0.7215	-0.4255	-0.5895	-0.1255	-0.4999	0.07854	-0.4195	
	0.16066	0.19666	0	-0.0113	0.44866	-0.0633	0.23266	0.06866	0.53266	0.15826	0.73666	0.23866	
2,L	0.172	0.208	0.3404	0	0.46	-0.052	0.244	0.08	0.544	0.1696	0.748	0.25	
	0.09986	0.09986	0.09986	0	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	
	-0.1571	-0.1211	0.01134	0	0.13094	-0.3811	-0.0851	-0.2491	0.21494	-0.1595	0.41894	-0.0791	
	0.50106	0.53706	0.66946	0	0.78906	0.27706	0.57306	0.40906	0.87306	0.49866	1.07706	0.57906	
3,H	-0.288	-0.252	-0.1196	-0.46	0	-0.512	-0.216	-0.38	0.084	-0.2904	0.288	-0.21	
	0.09986	0.09986	0.09986	0.09986	0	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	
	-0.6171	-0.5811	-0.4487	-0.7891	0	-0.8411	-0.5451	-0.7091	-0.2451	-0.6195	-0.0411	-0.5391	
	0.04106	0.07706	0.20946	-0.1309	0	-0.1829	0.11306	-0.0509	0.41306	0.03866	0.61706	0.11906	
3,L	0.224	0.26	0.3924	0.052	0.512	0	0.296	0.132	0.596	0.2216	0.8	0.302	
	0.09986	0.09986	0.09986	0.09986	0.09986	0	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	
	-0.1051	-0.0691	0.06334	-0.2771	0.18294	0	-0.0331	-0.1971	0.26694	-0.1075	0.47094	-0.0271	
	0.55306	0.58906	0.72146	0.38106	0.84106	0	0.62506	0.46106	0.92506	0.55066	1.12906	0.63106	
4,H	-0.072	-0.036	0.0964	-0.244	0.216	-0.296	0	-0.164	0.3	-0.0744	0.504	0.006	
	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0	0.09986	0.09986	0.09986	0.09986	0.09986	
	-0.4011	-0.3651	-0.2327	-0.5731	-0.1131	-0.6251	0	-0.4931	-0.0291	-0.4035	0.17494	-0.3231	
	0.25706	0.29306	0.42546	0.08506	0.54506	0.03306	0	0.16506	0.62906	0.25466	0.83306	0.33506	
4,L	0.092	0.128	0.2604	-0.08	0.38	-0.132	0.164	0	0.464	0.0896	0.668	0.17	
	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0	0.09986	0.09986	0.09986	0.09986	
	-0.2371	-0.2011	-0.0687	-0.4091	0.05094	-0.4611	-0.1651	0	0.13494	-0.2395	0.33894	-0.1591	
	0.42106	0.45706	0.58946	0.24906	0.70906	0.19706	0.49306	0	0.79306	0.41866	0.99706	0.49906	
5,H	-0.372	-0.336	-0.2036	-0.544	-0.084	-0.596	-0.3	-0.464	0	-0.3744	0.204	-0.294	
	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0	0.09986	0.09986	0.09986	
	-0.7011	-0.6651	-0.5327	-0.8731	-0.4131	-0.9251	-0.6291	-0.7931	0	-0.7035	-0.1251	-0.6231	
	-0.0429	-0.0069	0.12546	-0.2149	0.24506	-0.2669	0.02906	-0.1349	0	-0.0453	0.53306	0.03506	
5,L	0.0024	0.0384	0.1708	-0.1696	0.2904	-0.2216	0.0744	-0.0896	0.3744	0	0.5784	0.0804	
	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0	0.09986	0.09986	
	-0.3267	-0.2907	-0.1583	-0.4987	-0.0387	-0.5507	-0.2547	-0.4187	0.04534	0	0.24934	-0.2487	
	0.33146	0.36746	0.49986	0.15946	0.61946	0.10746	0.40346	0.23946	0.70346	0	0.90746	0.40946	
6,H	-0.576	-0.54	-0.4076	-0.748	-0.288	-0.8	-0.504	-0.668	-0.204	-0.5784	0	-0.498	
	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0	0.09986	
	-0.9051	-0.8691	-0.7367	-1.0771	-0.6171	-1.1291	-0.8331	-0.9971	-0.5331	-0.9075	0	-0.8271	
	-0.2469	-0.2109	-0.0785	-0.4189	0.04106	-0.4709	-0.1749	-0.3389	0.12506	-0.2493	0	-0.1689	
6,L	-0.078	-0.042	0.0904	-0.25	0.21	-0.302	-0.006	-0.17	0.294	-0.0804	0.498	0	
	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0.09986	0	
	-0.4071	-0.3711	-0.2387	-0.5791	-0.1191	-0.6311	-0.3351	-0.4991	-0.0351	-0.4095	0.16894	0	
	0.25106	0.28706	0.41946	0.07906	0.53906	0.02706	0.32306	0.15906	0.62306	0.24866	0.82706	0	

Table 14: LSM Differences Tukey HSD: Connecting Letters Report for dependent variable score and independent variables number of watch window and level of AOC/ACT

Level		Least Sq Mean
3,L	A	0.97200000
2,L	A	0.92000000
4,L	A B	0.84000000
5,L	A B C	0.75040000
1,H	A B C	0.74800000
1,L	A B C	0.71200000
4,H	A B C D	0.67600000
6,L	A B C D	0.67000000
2,H	B C D	0.57960000
3,H	C D E	0.46000000
5,H	D E	0.37600000
6,H	E	0.17200000

Appendix C: Confidence Ratings

In the summary of fit table below, the $R^2 = 0.11$ indicate that there is an 11% correlation between the independent variables and the dependent variables, and that 89% of the variance correlates to the response mean.. The effects test indicates that *# of watch windows* ($F < 0.0001$) had a significant effect on the subjects' confidence ratings, while *amount of AOC/ACT* (0.2880) did not.

Table 15: Summary of fit of effect for variables number of watch windows and level of AOC and ACT on confidence rating

Summary of Fit	
RSquare	0.112507
RSquare Adj	0.094333
Root Mean Square Error	1.572955
Mean of Response	4.643333
Observations (or Sum Wgts)	300

Table 16: Effects test for variables number of watch windows and level of AOC and ACT on confidence rating

Effect Tests					
Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
# of Watch Windows	5	5	89.096667	7.2021	<.0001*
Amt AOC/ACT	1	1	2.803333	1.1330	0.2880

The cross of the independent variables and their effect on confidence shows that the variable number of watch window ($F < 0.000$) had a significant effect on performance while neither amount of AOC/ACT ($F = 0.2851$) nor a cross of both variables (0.1284) had such an effect.

Table 17: Cross of variables number of watch windows and level of AOC/ACT on confidence

Effect Tests					
Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
# of Watch Windows	5	5	89.096667	7.2914	<.0001*
Amt AOC/ACT	1	1	2.803333	1.1471	0.2851
# of Watch Windows*Amt AOC/ACT	5	5	21.096667	1.7265	0.1284

C-1: Effects of number of watch windows on confidence rating

The LSM table below shows a nonlinear drop in observer confidence from one *watch window* to six *watch windows*.

Table 18: LSM table for confidence ratings of one to six watch windows (out of 7.0)

Least Squares Means Table			
Level	Least Sq Mean	Std Error	Mean
1	5.7600000	0.22244939	5.76000
2	4.3200000	0.22244939	4.32000
3	4.6600000	0.22244939	4.66000
4	4.6600000	0.22244939	4.66000
5	4.4400000	0.22244939	4.44000
6	4.0200000	0.22244939	4.02000

Table 19: LSMeans Tukey HSD table that shows effects of number of watch windows on confidence ratings

Level		Least Sq Mean
1	A	5.7600000
3	B	4.6600000
4	B	4.6600000
5	B	4.4400000
2	B	4.3200000
6	B	4.0200000

C-2: Effect of level of AOC/ACT on confidence ratings

Although it was determined that *level of AOC/ACT* did not have significant effect on confidence ratings, the LSM table below shows that confidence ratings were slightly higher in the tasks with low *AOC/ACT* (4.740) than tasks with high *AOC/ACT* (4.546)

Table 20: LSM table that shows effect of level of AOC/ACT on confidence ratings

Least Squares Means Table			
	Least		
Level	Sq Mean	Std Error	Mean
H	4.5466667	0.12843122	4.54667
L	4.7400000	0.12843122	4.74000

C-3: Effects of Cross of both Number of Watch Windows and Level of AOC/ACT on Confidence Rating

The LSM table below shows the confidence rating for a combination of the two the variables *number of watch windows* and *level of AOC/ACT*. The pattern shows a close to even distribution in confidence ratings between tasks with high and low *level of AOC/ACT*. Tasks 1, 2, and 4 had higher confidence ratings on tasks with low *level of AOC/ACT* while tasks 3, 5 and 6 had higher confidence ratings on tasks with higher *level of AOC/ACT*.

Table 21: LSM table show confidence rating for each number of watch window and level of AOC and ACT (out of 7.0)

Least Squares Means Table		
	Least	
Level	Sq Mean	Std Error
1,H	5.2800000	0.31265885
1,L	6.2400000	0.31265885
2,H	3.8800000	0.31265885
2,L	4.7600000	0.31265885
3,H	4.8400000	0.31265885
3,L	4.4800000	0.31265885
4,H	4.6400000	0.31265885
4,L	4.6800000	0.31265885
5,H	4.4800000	0.31265885
5,L	4.4000000	0.31265885
6,H	4.1600000	0.31265885
6,L	3.8800000	0.31265885

Below is the LSM difference student's t test. Levels that are not connected with the same letter are significantly different.

Table 22: LSMeans Differences Tukey HSD table for dependent variable confidence and independent variables number of watch window and level of AOC/ACT

		LSMean[j]											
Mean[i]-Mean[j]	1,H	1,L	2,H	2,L	3,H	3,L	4,H	4,L	5,H	5,L	6,H	6,L	
Std Err Dif													
Lower CL Dif													
Upper CL Dif													
LSMean[i]	2,H	-1.4	-2.36	0	-0.88	-0.96	-0.6	-0.76	-0.8	-0.6	-0.52	-0.28	0
		0.44217	0.44217	0	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217
		-2.857	-3.817	0	-2.337	-2.417	-2.057	-2.217	-2.257	-2.057	-1.977	-1.737	-1.457
		0.05704	-0.903	0	0.57704	0.49704	0.85704	0.69704	0.65704	0.85704	0.93704	1.17704	1.45704
	2,L	-0.52	-1.48	0.88	0	-0.08	0.28	0.12	0.08	0.28	0.36	0.6	0.88
		0.44217	0.44217	0.44217	0	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217
		-1.977	-2.937	-0.577	0	-1.537	-1.177	-1.337	-1.377	-1.177	-1.097	-0.857	-0.577
		0.93704	-0.023	2.33704	0	1.37704	1.73704	1.57704	1.53704	1.73704	1.81704	2.05704	2.33704
	3,H	-0.44	-1.4	0.96	0.08	0	0.36	0.2	0.16	0.36	0.44	0.68	0.96
		0.44217	0.44217	0.44217	0.44217	0	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217
		-1.897	-2.857	-0.497	-1.377	0	-1.097	-1.257	-1.297	-1.097	-1.017	-0.777	-0.497
		1.01704	0.05704	2.41704	1.53704	0	1.81704	1.65704	1.61704	1.81704	1.89704	2.13704	2.41704
3,L	-0.8	-1.76	0.6	-0.28	-0.36	0	-0.16	-0.2	0	0.08	0.32	0.6	
	0.44217	0.44217	0.44217	0.44217	0.44217	0	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217	
	-2.257	-3.217	-0.857	-1.737	-1.817	0	-1.617	-1.657	-1.457	-1.377	-1.137	-0.857	
	0.65704	-0.303	2.05704	1.17704	1.09704	0	1.29704	1.25704	1.45704	1.53704	1.77704	2.05704	
4,H	-0.64	-1.6	0.76	-0.12	-0.2	0.16	0	-0.04	0.16	0.24	0.48	0.76	
	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217	0	0.44217	0.44217	0.44217	0.44217	0.44217	
	-2.097	-3.057	-0.697	-1.577	-1.657	-1.297	0	-1.497	-1.297	-1.217	-0.977	-0.697	
	0.81704	-0.143	2.21704	1.33704	1.25704	1.61704	0	1.41704	1.61704	1.69704	1.93704	2.21704	
4,L	-0.6	-1.56	0.8	-0.08	-0.16	0.2	0.04	0	0.2	0.28	0.52	0.8	
	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217	0	0.44217	0.44217	0.44217	0.44217	
	-2.057	-3.017	-0.657	-1.537	-1.617	-1.257	-1.417	0	-1.257	-1.177	-0.937	-0.657	
	0.85704	-0.103	2.25704	1.37704	1.29704	1.65704	1.49704	0	1.65704	1.73704	1.97704	2.25704	
5,H	-0.8	-1.76	0.6	-0.28	-0.36	0	-0.16	-0.2	0	0.08	0.32	0.6	
	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217	0	0.44217	0.44217	0.44217	
	-2.257	-3.217	-0.857	-1.737	-1.817	-1.457	-1.617	-1.657	0	-1.377	-1.137	-0.857	
	0.65704	-0.303	2.05704	1.17704	1.09704	1.45704	1.29704	1.25704	0	1.53704	1.77704	2.05704	
5,L	-0.88	-1.84	0.52	-0.36	-0.44	-0.08	-0.24	-0.28	-0.08	0	0.24	0.52	
	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217	0	0.44217	0.44217	
	-2.337	-3.297	-0.937	-1.817	-1.897	-1.537	-1.697	-1.737	-1.537	0	-1.217	-0.937	
	0.57704	-0.383	1.97704	1.09704	1.01704	1.37704	1.21704	1.17704	1.37704	0	1.69704	1.97704	
6,H	-1.12	-2.08	0.28	-0.6	-0.68	-0.32	-0.48	-0.52	-0.32	-0.24	0	0.28	
	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217	0	0.44217	
	-2.577	-3.537	-1.177	-2.057	-2.137	-1.777	-1.937	-1.977	-1.777	-1.697	0	-1.177	
	0.33704	-0.623	1.73704	0.85704	0.77704	1.13704	0.97704	0.93704	1.13704	1.21704	0	1.73704	
6,L	-1.4	-2.36	0	-0.88	-0.96	-0.6	-0.76	-0.8	-0.6	-0.52	-0.28	0	
	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217	0.44217	0	
	-2.857	-3.817	-1.457	-2.337	-2.417	-2.057	-2.217	-2.257	-2.057	-1.977	-1.737	0	
	0.05704	-0.903	1.45704	0.57704	0.49704	0.85704	0.69704	0.65704	0.85704	0.93704	1.17704	0	

Table 23: LSM Differences Tukey HSD: Connecting Letters Report for dependent variable confidence and independent variables number of watch window and level of AOC/ACT

Level		Least Sq Mean
1,L	A	6.2400000
1,H	A B	5.2800000
3,H	A B	4.8400000
2,L	B	4.7600000
4,L	B	4.6800000
4,H	B	4.6400000
3,L	B	4.4800000
5,H	B	4.4800000
5,L	B	4.4000000
6,H	B	4.1600000
2,H	B	3.8800000
6,L	B	3.8800000

Appendix D: Comparison of All Dependent Variables- Performance Versus Confidence and NASA-TLX

Graphing the independent variables score versus confidence yielded a graph whose behavior shows a direct relationship between performance and confidence. Therefore, in general the more confident the subjects were, the better they performed. A plot of performance (score) versus NASA-TLX also yielded a direct relationship, which was predicted. The lower the NASA-TLX rating, the better the subjects performed. The graph produced from the analysis shows a drop in performance as NASA-TLX values increased (which signifies the subject experienced higher workload in the task).

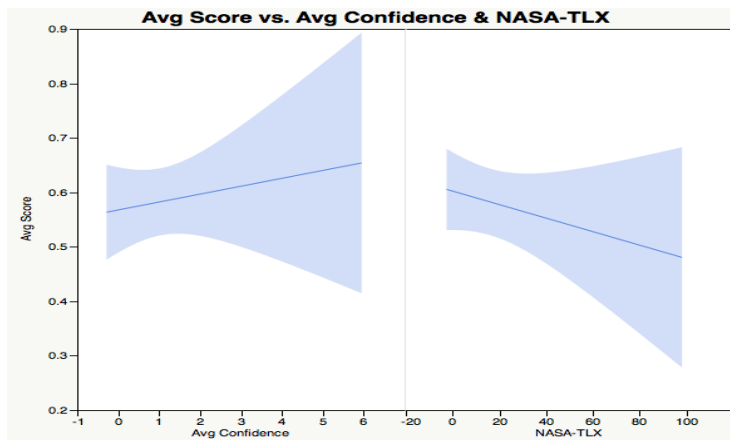


Figure 19: Graph of Performance Versus Confidence and NASA-TLX

References

- Abbott, K. R., & Sherratt, T. N. (2013). Optimal sampling and signal detection: unifying models of attention and speed–accuracy trade-offs. *Behavioral Ecology*, 24(3), 605-616.
- Aguinis, H., Culpepper, S. A., & Pierce, C. A. (2010). Revival of test bias research in preemployment testing. *Journal of Applied Psychology*, 95(4), 648.
- Araujo, C., Kowler, E., & Pavel, M. (2001). Eye movements during visual search: The costs of choosing the optimal path. *Vision Research*, 41(25-26), 3613-3625. doi:10.1016/S0042-6989(01)00196-1
- Atkins, M. S., Moise, A., & Rohling, R. (2006). An application of eyegaze tracking for designing radiologists' workstations. *ACM Transactions on Applied Perception (TAP)*, 3(2), 136-151. doi:10.1145/1141897.1141902
- Barber, C. J. (2001). An intelligence, surveillance and reconnaissance (ISR) vision for the canadian forces. *Canadian Military Journal*, 2(4), 2001-2002.
- Best Jr, R. A. (2005). *Intelligence, Surveillance, and Reconnaissance (ISR) Programs: Issues for Congress* (No. CRS-RL32508). DEFENSE ACQUISITION UNIV FORT BELVOIR VA DAVID D ACKER LIBRARY AND KNOWLEDGE REPOSITORY.
- Best, S. J., Johnson, N., & Sandford, A. M. (1992). *U.S. Patent No. 5,113,437*. Washington, DC: U.S. Patent and Trademark Office.
- Bruce, N. D., & Tsotsos, J. K. (2009). Saliency, attention, and visual search: An information theoretic approach. *Journal of Vision*, 9(3)

- Cain, M. S., & Mitroff, S. R. (2013). Memory for found targets interferes with subsequent performance in multiple-target visual search. *Journal of Experimental Psychology. Human Perception & Performance*, 39(5), 1398-1408. doi:10.1037/a0030726
- Cheng, D., & Vertegaal, R. (2004). An eye for an eye: A performance evaluation comparison of the LC technologies and tobii eye trackers. *Eye Tracking Research & Application: Proceedings of the 2004 Symposium on Eye Tracking Research & Applications*, , 22(24) 61-61.
- Dickinson, C. A., & Zelinsky, G. J. (2013). New evidence for strategic differences between static and dynamic search tasks: An individual observer analysis of eye movements. *Frontiers in Psychology*, 4doi:10.3389/fpsyg.2013.00008
- Dixon, S. R., Wickens, C. D., & Chang, D. (2005). Mission control of multiple unmanned aerial vehicles: A workload analysis. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 47(3), 479-487.
- Dodonov, Y. S., & Dodonova, Y. A. (2012). Response time analysis in cognitive tasks with increasing difficulty. *Intelligence*, 40(5), 379-394. doi:10.1016/j.intell.2012.07.002
- Duggan, S. (2012, 04-26-2012). Bruce Mayhew Blog: Business Communication. Retrieved September, 2013, from <http://brucemayhew.wordpress.com/tag/yerkes-dodson-law/>
- Duncan, J. S., & Ayache, N. (2000). Medical image analysis: Progress over two decades and the challenges ahead. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(1), 85-106.

- Eckstein, M. P., Thomas, J. P., Palmer, J., & Shimozaki, S. S. (2000). A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Perception & Psychophysics*, 62(3), 425-451.
- Enns, J. T., & Girgus, J. S. (1985). Developmental changes in selective and integrative visual attention. *Journal of Experimental Child Psychology*, 40(2), 319-337.
- Fendley, M. E. (2009). *Human cognitive biases and heuristics in image analysis* (Doctoral dissertation, Wright State University).
- Fendley, M. E. (2009). *Human cognitive biases and heuristics in image analysis* Wright State University / OhioLINK. Retrieved from http://ezproxy.libraries.wright.edu:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=ir00279a&AN=oletd.wright1257278185&site=eds-live; http://rave.ohiolink.edu/etdc/view?acc_num=wright1257278185
- Freed, M., Harris, R., & Shafto, M. G. (2004). Human vs. autonomous control of UAV surveillance. Paper presented at the *AIAA 1st Intelligent Systems Technical Conference*, 1-7.
- Gale, W. F., & Buynak, G. L. (1982). Fecundity and spawning frequency of the fathead minnow—a fractional spawner. *Transactions of the American Fisheries Society*, 111(1), 35-40.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Human Mental Workload*, 1(3), 139-183.
- Hautus, M. J., O'MAHONY, M., & LEE, H. (2008). Decision strategies determined from the shape of the same–different ROC curve: What are the effects of incorrect assumptions? *Journal of Sensory Studies*, 23(6), 743-764.

Heeger, D. (1997). Signal detection theory. *Dept.Psych., Stanford Univ., Stanford, CA, Teaching Handout*,

Hickman, M., Mirchandani, P., & Transportation Research Board. (2008). Airborne traffic flow data and traffic management. check out page 129. *The Fundamental Diagram for Traffic Flow Theory*, , 121.

Irvine, J. M., Fenimore, C., Cannon, D., Haverkamp, D., Roberts, J., Israel, S. A., . . . Brennan, M. (2006). Development of a motion imagery quality metric. Paper presented at the *Proceedings of the American Society for Photogrammetry and Remote Sensing (ASPRS) Annual Meeting*, 1-5.

Jones, A. E., Shapiro, N. I., & Roshon, M. (2007). Implementing early Goal- directed therapy in the emergency setting: The challenges and experiences of translating research innovations into clinical reality in academic and community settings. *Academic Emergency Medicine*, 14(11), 1072-1078.

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences*, 9(2), 75-82. doi:10.1016/j.tics.2004.12.004

Lee, H. C., Lee, A. H., & Cameron, D. (2003). Validation of a driving simulator by measuring the visual attention skill of older adult drivers. *The American Journal of Occupational Therapy : Official Publication of the American Occupational Therapy Association*, 57(3), 324-328.

Levin, D. T., Angelone, B. L., & Beck, M. R. (2011). *Visual search for rare targets: Distracter tuning as a mechanism for learning from repeated target-absent searches* Wiley-Blackwell. doi:10.1348/000712610X519503

- Liang, P., Teodoro, G., Ling, H., Blasch, E., Chen, G., & Bai, L. (2012). Multiple kernel learning for vehicle detection in wide area motion imagery. Paper presented at the *Information Fusion (FUSION), 2012 15th International Conference on*, 1629-1636.
- Lieberman, H. R., Coffey, B., & Kobrick, J. (1998). A vigilance task sensitive to the effects of stimulants, hypnotics, and environmental stress: The scanning visual vigilance test. *Behavior Research Methods, Instruments & Computers*, 30(3), 416-422. doi:10.3758/BF03200674
- McCarley, J. S., & Kramer, A. F. (2007). Eye movements as a window on perception and cognition. In R. Parasuraman, & M. Rizzo (Eds.), (pp. 95-112). New York, NY US: Oxford University Press.
- McNicol, D. (2005). *A primer of signal detection theory* Psychology Press.
- Morgante, J. D., Zolfaghari, R., & Johnson, S. P. (2012). A critical test of temporal and spatial accuracy of the tobii T60XL eye tracker. *Infancy*, 17(1), 9-32. doi:10.1111/j.1532-7078.2011.00089.x
- Morvan, C., & Maloney, L. T. (2012). Human visual search does not maximize the post- saccadic probability of identifying targets. *PLoS Computational Biology*, 8(2), 1-11. doi:10.1371/journal.pcbi.1002342
- Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, 434(7031), 387-391.
- Najemnik, J., & Geisler, W. S. (2008). Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision*, 8(3), 1-14. doi:10.1167/8.3.4

- NASA Human Performance Research Group. (1987). Task load index (NASA-TLX) v1. 0 computerised version. *NASA Ames Research Centre*,
- O'Kelly, M., Matisziw, T., Li, R., Merry, C., & Niu, X. (2005). Identifying truck correspondence in multi-frame imagery. *Transportation Research Part C: Emerging Technologies*, 13(1), 1-17.
- O'MAHONY, M., & Hautus, M. (2008). The signal detection theory ROC curve: Some applications in food sensory science. *Journal of Sensory Studies*, 23(2), 186-204.
- Palmer, J., & McLean, J. (1995). Imperfect, unlimited-capacity, parallel search yields large set-size effects. *Society for Mathematical Psychology, Irvine, CA*,
- Palmer, J., Ames, C. T., & Lindsey, D. T. (1993). Measuring the effect of attention on simple visual search. *Journal of Experimental Psychology Human Perception and Performance*, 19, 108-108.
- Palmer, S. E. (1992). Common region: A new principle of perceptual grouping. *Cognitive Psychology*, 24(3), 436-447.
- Paul, F. (2013). Application of the augmented operator function model for developing cognitive metrics in persistent surveillance ., 1-273. doi:2013
- Petty, R. E., Wegener, D. T., & Fabrigar, L. R. (1997). Attitudes and attitude change. *Annual Review of Psychology*, 48(1), 609-647.
- Pham, T., Cirincione, G. H., Verma, D., & Pearson, G. (2008). Intelligence, surveillance, and reconnaissance fusion for coalition operations. Paper presented at the *Information Fusion, 2008 11th International Conference on*, 1-8.

- Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review*, 117(3), 864.
- Privitera, C. M., Renninger, L. W., Carney, T., Klein, S., & Aguilar, M. (2010). Pupil dilation during visual target detection. *Journal of Vision*, 10(10) doi:10.1167/10.10.3
- Ramos-Álvarez, M. M., Moreno-Fernández, M. M., Paredes-Olay, C., & Rosas, J. M. (2013). A methodological proposal based on signal detection theory for the study of dissociation between sensory and decision processes in the context of olive oil tasting. *Food Quality and Preference*, 28(1), 71-76.
- Reid, G. B., & Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. *Human Mental Workload*, 185, 218.
- Roper, Z. J., Cosman, J. D., Mordkoff, J. T., & Vecera, S. P. (2010). Perceptual load corresponds to known factors influencing visual search. *Journal of Vision*, 10(7), 1277-1277.
- Ruff, H. A., Calhoun, G. L., Draper, M. H., Fontejon, J. V., & Guilfoos, B. J. (2004). *Exploring automation issues in supervisory control of multiple UAVs*. SYTRONICS INC DAYTON OH.
- Shafiullah, G., Gyasi-Agyei, A., & Wolfs, P. (2007). Survey of wireless communications applications in the railway industry. Paper presented at the *Wireless Broadband and Ultra Wideband Communications, 2007. AusWireless 2007. the 2nd International Conference on*, 65-65.
- Srinivasan, S., Latchman, H., Shea, J., Wong, T., & McNair, J. (2004). Airborne traffic surveillance systems: Video surveillance of highway traffic. Paper presented at

- the *Proceedings of the ACM 2nd International Workshop on Video Surveillance & Sensor Networks*, 131-135.
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31(1), 137-149.
- Sulman, N., Sanocki, T., Goldgof, D., & Kasturi, R. (2008). How effective is human video surveillance performance? Paper presented at the *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, 1-3.
- Tabachnick, B. G., & Fidell, L. S. (2001). Using multivariate statistics. Published by California State University, Northridge.
- Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance*, 8(2), 194.
- Tremblay, S., Vachon, F., Lafond, D., & Kramer, C. (2012). Dealing with task interruptions in complex dynamic environments are two heads better than one? *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 54(1), 70-83.
- Trent, S. A., Patterson, E. S., & Woods, D. D. (2007). Challenges for cognition in intelligence analysis. *Journal of Cognitive Engineering and Decision Making*, 1(1), 75-97.
- Trinh, T. T., & Kuchar, J. K. (1999). Study of visual cues for unmanned aerial vehicle waypoint allocation. Paper presented at the *Digital Avionics Systems Conference, 1999. Proceedings. 18th, , 1* 4. D. 5-1-4. D. 5-8 vol. 1.

- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5(2), 207-232.
- Vachon, F., Vallières, B. R., Jones, D. M., & Tremblay, S. (2012). Nonexplicit change detection in complex dynamic settings what eye movements reveal. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 54(6), 996-1007.
- Verghese, P. (2001). Visual search and attention: A signal detection theory approach. *Neuron*, 31(4), 523-535.
- Verghese, P., & Nakayama, K. (1994). Stimulus discriminability in visual search. *Vision Research*, 34(18), 2453-2467.
- Wegener, D. T., & Petty, R. E. (1997). The flexible correction model: The role of naive theories of bias in bias correction. *Advances in Experimental Social Psychology*, 29, 141-208.
- Weigle, C., & Banks, D. C. (2008). Analysis of eye-tracking experiments performed on a Tobii T60. *Electronic Imaging 2008*, 680903-680903-12.
- Wichchukit, S., & O'Mahony, M. (2010). A transfer of technology from engineering: Use of ROC curves from signal detection theory to investigate information processing in the brain during sensory difference testing. *Journal of Food Science*, 75(9), R183-R193.
- Wickens, C. D., Dixon, S., & Chang, D. (2003). *Using interference models to predict performance in a multiple-task UAV environment-2 UAVs* (No. AHFD-03-9/MAAD-03-1). ILLINOIS UNIV AT URBANA-CHAMPAIGN SAVOY AVIATION HUMAN FACTORS DIVISION.

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, 114(1), 152-176.

Young, A. H., & Hulleman, J. (2013). Eye movements reveal how task difficulty moulds visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 39(1), 168-190. doi:10.1037/a0028679