3-2020

The Impact of Changing the Size of Aircraft Radar Displays on Visual Search in the Cockpit

Justin R. Marsh

THE IMPACT OF CHANGING THE SIZE OF AIRCRAFT RADAR DISPLAYS ON VISUAL SEARCH IN THE COCKPIT

THESIS

Justin R. Marsh, Captain, USAF

AFIT-ENV-MS-20-M-226

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENV-MS-20-M-226

THE IMPACT OF CHANGING THE SIZE OF AIRCRAFT RADAR DISPLAYS ON VISUAL SEARCH IN THE COCKPIT

THESIS

Presented to the Faculty

Department of Systems Engineering and Management

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Systems Engineering

Justin R. Marsh, MBA

Captain, USAF

March 2020

DISTRIBUTION STATEMENT A.
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AFIT-ENV-MS-20-M-226

THE IMPACT OF CHANGING THE SIZE OF AIRCRAFT RADAR DISPLAYS ON VISUAL SEARCH IN THE COCKPIT

Justin R. Marsh, MBA

Captain, USAF

Committee Membership:

Dr. Michael E. Miller
Chair

Dr. Eric E. Geiselman
Member

Dr. Paul R. Havig
Member

AFIT-ENV-MS-20-M-226

Abstract

Advances in sensor technology have enabled our fighter aircraft to find, fix, track, target, engage (F2T2E) at greater distances, providing the operator with more data within the battlefield. Modern aircraft are designed with larger displays while our legacy aircraft are being retrofitted with larger cockpit displays to enable display of the increased data. While this modification has been shown to enable improvements in human performance of many cockpit tasks, this effect is often not measured nor fully understood at a more generalizable level. This research outlines an approach to comparing human performance across two display sizes in future F-16 cockpits. The results show that increases in display size can increase search times under some circumstances even when the displays include a large number of tracks, actually reducing human performance.

Acknowledgments

I would like to express my sincere appreciation to my faculty advisor, Dr. Michael Miller, who gave me the freedom to explore various topics too numerous to list. Thank you for guiding me along my research adventure and for your profound expertise (and patience) as I did so.

Thank you to my future home, AFLCMC/WWM, who was always a willing sponsor and welcomed me to many meetings and trips without hesitation. I am excited to become an official member of the Viper team in March.

To my fellow 20M GSE family, who impress me on a daily basis with their intelligence, character and passion, thank you for supporting my academic pursuits. Especially, Lt David King, for your MATLAB and Python expertise that was invaluable in helping me overcome numerous coding challenges that I inadvertently created.

Finally, thank you to my beautiful wife, Zyrene, for the support, encouragement and patience over the last 18 months of this journey, when I had to trade husband time for student time.

Justin R. Marsh

Table of Contents

Page

List of Figures

Page

List of Tables

Page

THE IMPACT OF CHANGING THE SIZE OF AIRCRAFT RADAR DISPLAYS ON VISUAL SEARCH IN THE COCKPIT

I. Introduction

General Issue

Our legacy aircraft, such as the F-16 Fighting Falcon, are subject to sensor and performance upgrades to permit us to maintain a competitive advantage against our enemies and expand, as well as evolve our mission sets. Both of these situations result in the need to present more data to the operator in the cockpit. Presenting modern sensor information to operators in the cockpit often requires more addressable pixels than legacy displays can provide. Feedback such as "clutter" and "data overload" are commonly heard frustrations of pilots while using legacy displays to view information collected by more modern sensors. Such is the case on the F-16, where operators are now forced to zoom and pan to view the information provided by their sensors, or worse, completely disregard information by hiding layers of information from the display. Consequently, there is an increased potential for an operator to miss critical information or incorrectly interpret that information, especially during high task load situations.

So how did the F-16 pilot-vehicle interface (PVI) reach its current state where the capabilities of its displays are insufficient to provide easy access to the available sensor information? Program office engineers attest that it was not sudden, but a gradual increase in information presented to the operator over the last decade for the two reasons discussed above. Since it was fielded in mid-1970's, the F-16 has been modernized with numerous performance and sensor upgrades. At its time of conception, it was designed as a "simple air-superiority day fighter" but has since evolved into a multi-role, multi-

mission fighter with lethal capabilities in any condition. Remarkably, the F-16 pilots have taken the increase in the platform's responsibilities in stride, enabling the platform to remain the workhorse fighter for the United States and dozens of nations world-wide for 40 years. Avionics and display upgrades have been minimal, with the most recent completed in the mid-1980's. The Air Force has stated that another avionics display upgrade is needed to facilitate the latest sensor upgrades to detect, track and identify a greater number of targets quicker and at longer ranges, signaling another shift towards longer-range standoff engagements.

Program offices have several options for increasing the number of addressable pixels in the cockpit. If the area of the display is constrained to a maximum, the program office can increase the resolution of display (i.e. apply smaller pixels, thus providing more pixels in the same area). However, smaller symbol sizes may become too difficult to read without increasing overall symbol size. If there is space in the cockpit, display area can be increased by increasing the physical size of the screen or adding a separate display. For its currently planned avionics upgrade, the F-16 is adding a third, larger cockpit display by removing primary flight gauges and allowing primary flight information to be displayed electronically. The USAF is currently in the process of retrofitting the fleet of F-16's with this new display.

In addition to the ability to display more data, new displays often have advantages in terms of improved luminance, color, and contrast. Research has even suggested that the performance of some pilot tasks, such as stick flying and performing navigation, is better with larger displays (Chen, Liao, & Yeh, 2011; Stelzer & Wickens, 2006; Tan, Czerwinski, & Robertson, 2006). However, very little research has been done to address

the effect of display size on more fundamental mediating factors such as attention, perception, and motor components (Chen et al., 2011). In research by Chen et al., it was found that in certain conditions, the visual search performance of humans decreased with increases in display size (Chen et al., 2011). Their conclusion presents a challenging counterpoint to the benefits of larger display sizes that is worth investigating. F-16 pilots rely on finding information as quickly and accurately as possible to apply it to their evolving game plan. If Chen and colleagues results can be replicated with conditions that are more consistent with F-16 display conditions, then there is some trade-space with the display upgrade after all, i.e. performance isn't always better with the bigger display. To a system engineer in a program office, this would be valuable data to inform decisions made in the acquisition and employment of the weapon system.

Problem Statement

The F-16 program office is currently in the midst of a sensor and avionics upgrade that includes the addition of a new 6 by 8-inch display, called the Center Display Unit (CDU), shown in Figure 1. The larger display has more addressable pixels for displaying sensor data, which results in a lower chance of symbol overlap than is likely in the traditional Multi-Function Display (MFD) at 4.24 by 4.24 inches. Therefore, it is presumed that CDU will enable the pilot to respond to a greater number of sensor detections. However, research suggests that a pilot's visual search performance may actually be worse in larger displays in certain search scenarios (Chen et al., 2011). Quantifying the human performance differences between using the MFD and CDU would allow for a better understanding of the human-machine performance in the F-16. Further,

this research has implications for the F-35, T-7, F-15E(X) and other USAF tactical aircraft which are being designed with or upgraded to include larger displays to enable the presentation of sensor information.

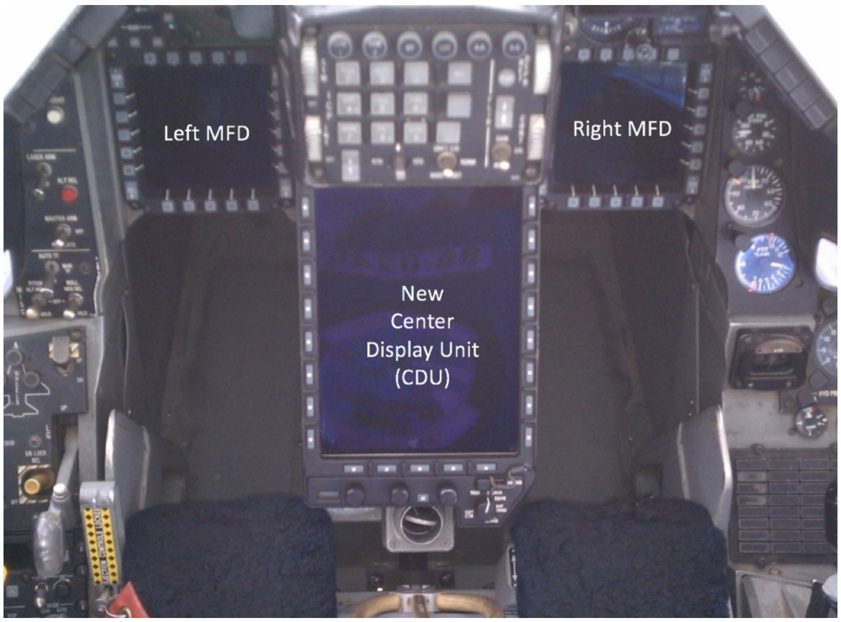


Figure 1 – F-16 Cockpit Showing Multi-Function Displays

Research Focus

This research focused on the differences in human performance while using two different size displays in the F-16 cockpit. The two displays are the legacy MFD and the new CDU shown in Figure 1 of the previous section. Previously, the main display of sensor information was accomplished through the left and right MFD's. The information from radar, targeting video or other sources is displayed on various pages or views selectable by the pilot. Both MFDs are identical and have the same menu options selectable through the option selection buttons located around the displays or through buttons on the stick. The new CDU (15k pixels/in^2) has a similar pixel size to the MFD

(16k pixels/in^2) but is 114% larger, enabling it to display the same sensor information as the MFD's but with more addressable pixels. Pilots interact with the CDU in the same manner as the MFDs.

The view of interest in this research is a 5th generation-like view, which fuses information from a variety of sensors within the same spatial area. Information from these sensors are fused through a central processor and displayed on the same view, instead of separate page views. As a result, fused or integrated views such as these generally have both a larger number and variety of symbols displayed to the pilot compared to the non-fused, 4th generation type views. Research by Kroft and Wickens (2002), suggests these integrated sensor displays result in faster reaction times and greater accuracy for answering questions about airspace awareness compared to accessing separate displays. Although these authors note that the increased clutter requires the use of time-consuming decluttering techniques on integrated displays. However, their results suggest that the combined benefits of reduced scanning and larger display size outweigh the costs of clutter (Kroft & Wickens, 2002). The F-16 is transitioning to 5th generation avionics capabilities with its latest modernization efforts. Figure 2 illustrates the differences between the 4th generation and 5th avionics architectures. For this study, "radar displays" will refer to 5th generation, integrated avionics.

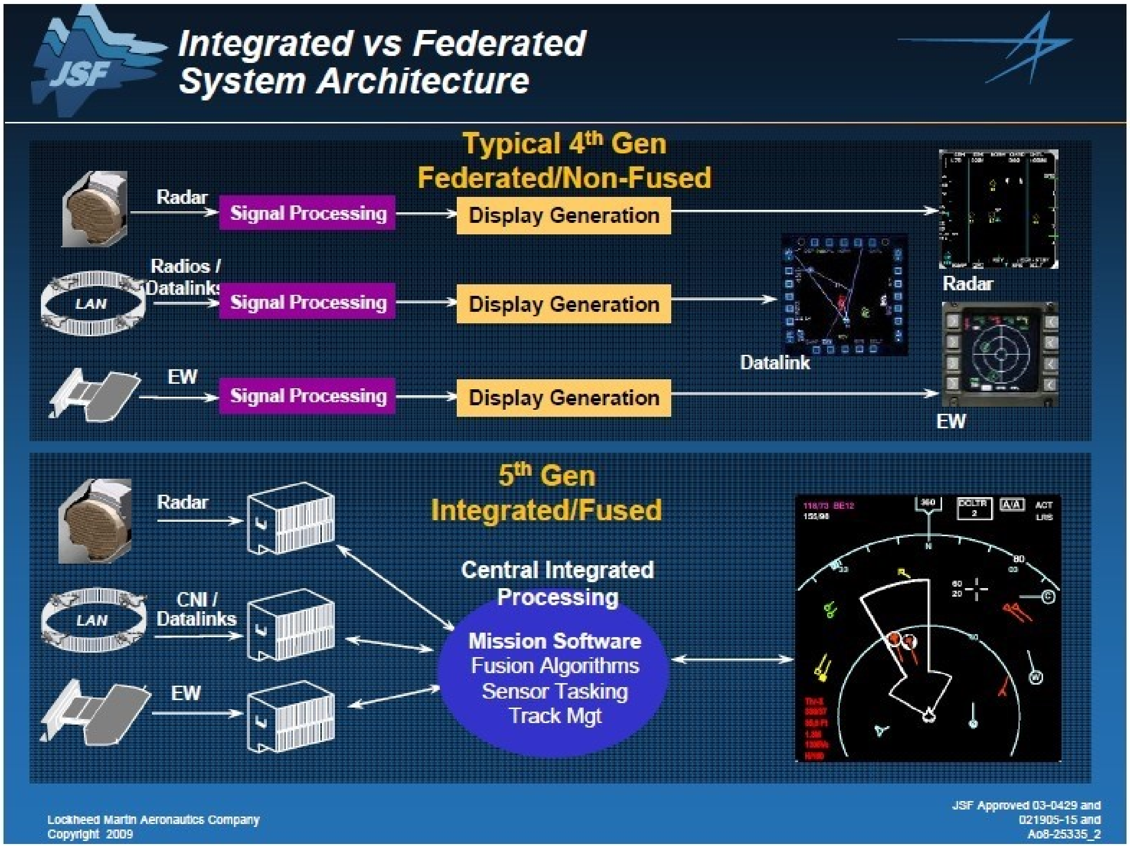


Figure 2 – Fourth and Fifth Generation Avionics Architectures (Hermelin, 2013)

The attribute of human performance investigated was visual attention through the task of a visual search. Attention is an abstract concept that is sometimes described as a searchlight, with both a breadth and direction. The breadth is divided into what we want to process (focused attention) and that which we must process but do not want to (divided attention) (Wickens & Hollands, 2000). The direction of the search light is guided by how our selective attention knows when, what, and where in the environment to illuminate (Wickens & Hollands, 2000). Visual attention is usually assessed in terms of reaction time and/or response accuracy because human performance improves as more attention is directed to a task (Chen et al., 2011). The nature of visual attention, prior

research, and the application of this information to this study are explained in detail in Chapter 2.

Investigative Questions

This thesis attempts to answer the research questions listed below:

1. How can we quantify changes in visual search performance in the cockpit of the F-16?
2. How does a pilot's visual search performance while using the MFD and CDU differ?
3. How can this information be used create predictive models of human-machine performance in the F-16 cockpit?

Methodology

After a review of relevant research, a human subjects experiment was designed and performed. A representative F-16 test apparatus was designed in SolidWorks to simulate the viewing conditions of an F-16 cockpit. The display scenes were created using MATLAB. The experiment was designed and executed using the PsychoPy 3.0 software package. Experimental data on reaction times and accuracy was collected to compare the differences between both displays in the F-16.

Assumptions and Limitations

The following assumptions and limitations apply to the current research:

1. It was assumed that the sensor views on the 2 displays were identical and displayed the same information, with the same sized symbols, permitting the

additional pixels on the CDU to be used to increase separation between symbols.

2. It was assumed that the pixel size, color, contrast, and brightness of both displays was identical. In reality, these displays are different in all these areas but the differences, were not believed to be significant. Assuming both displays to be identical simplified the experiment, the interpretation of the results and allowed them both displays to be represented using the same monitor.
3. It was assumed that the Primary Flight Display (PFD) would be displayed at the bottom of the CDU, providing a 6 by 6-inch area for on the CDU for display of the experimental stimuli.
4. F-16 pilots could not be used for the human subjects experiment because of time and test personnel constraints. Instead, AFIT (Air Force Institute of Technology) student officers were used as the participants. For the same reason, the number of available participants was limited, resulting in a relatively small sample size.
5. The experiment was conducted using a test rig that was designed to replicate the viewing conditions of the F-16 cockpit. A more accurate F-16 simulator was not available to conduct the study.
6. The experiment was conducted in controlled laboratory environment. The participants were only exposed to displays of sensor detections on the MFD and CDU. Other information, which can be provided in the real world like flying sensations, flying visuals and sounds, weather and environmental

conditions, radio communications, and other mission information, was not provided to the participants in this experiment. Although this experiment did not reflect real world conditions, a standardized synthetic environment could make it possible to compare the participant's reaction time and accuracy under a controlled environment.

7. The experiment assumed participants had no memory of the previous sensor display. In other words, each view of sensor information was completely unique and not a slight variation of the previous view. A participant couldn't use their memory to more quickly identify what was different about the new scene. Each trial was therefore a new task with a comparable reaction time.

Implications

The study presents potential impacts across the acquisition and employment of fighter aircraft new and old by using F-16 as an example. Although the USAF is transitioning towards the ability to engage our enemy at longer distances, resulting in more time to make decisions, the ability to make decisions as quickly and accurately as possible in fighter aircraft is still important and relevant. This is particularly true in the F-16, which maintains an air-to-air combat role in certain situations. If the ability to find information quickly is hindered on the CDU in certain situations, then designers may need to put measures in place to assist in those situations. This research will touch on a few of those measures and provide suggestions to interface designers. The impacts of displaying 5th generation sensor information should be fully understood by designers, engineers, and operators. This author expects the amount of sensor information and sizes

of cockpit displays will continue to increase. Therefore we should be cognizant of the impacts suggested in this research.

Preview

The first chapter stated the purpose and objective of this research, an overview of the method, assumptions and limitations, and this study's significance. Chapter 2, Literature Review, contains the theoretical framework for this study. This chapter presents a review of the literature relevant to human performance in the cockpit and the effect of increasing display size. Chapter 3, Methodology, describes and justifies the data collection method used for this research. This chapter also outlines how the data will be analyzed. Chapter 4, Results, addresses the results from data analysis. This chapter contains results from the human factors experiment, including participants' reaction times, and accuracy scores. Finally, Chapter 5, Discussion, Recommendation, and Conclusion, addresses the meaning of the study's findings and contains the overall conclusion and suggested areas for future research.

II. Literature Review

Overview

This chapter provides the theoretical framework for the current research through a review of the relevant literature. Specifically, Chapter 2 provides answers the first research question and provides evidence towards predicting the second research question listed in Chapter 1. The chapter begins with a look at two studies that show how human performance can benefit from increasing display size. In search of a more generalizable, non-specific measure of human performance in the cockpit, visual attention is explored in-depth. The section provides a general discussion on the nature of visual attention, its models, the effect of display size, and the issue of clutter. Finally, the section concludes with how visual attention can be applied to tasks a pilot performs in the cockpit.

Display Size Effects in Aviation Applications

Over the past decade, research has indicated that display size does affect human perception and performance (Stelzer & Wickens, 2006; Tan et al., 2006). In fact, results have suggested that human performance is better with large displays than with conventional, small displays (Chen et al., 2011). Stelzer and Wickens (2006) performed a study comparing human performance in a flight control task with 3 displays, sized 10º x 7º, 20º x 15º, and 36º x 27º visual angle respectively. Visual angle is widely used in the visual science community to describe display sizes because it incorporates the viewing distance and size of object or display into the measurement. For flight control, pilots exhibited less path error and greater stick activity with the 36º x 27º display, which was attributed both to greater enhanced resolution and to the fact that larger depictions of

error lead to greater urgency in correcting deviations (Stelzer & Wickens, 2006). Similarly, research by Tan et al. (2006) found that in virtual navigation, increasing field of view from 47º to 120º increased navigation performance of all users on average. They attributed the results to better optical flow cues offered by the large displays (Tan et al., 2006).

The aviation tasks explored in these studies are relevant and helpful to unmanned aerial vehicles or helmet-mounted displays where virtual information is used to guide a plane. However, they do not characterize human performance on a more generalizable level, such as human attention. Human attention deserves consideration because it is a component of our daily lives and the tasks we perform. Absent technology that predicts the information we need before we need it, humans will always need to search for a target piece of information among distracting information. Like looking for a particular pen in a drawer full of pens and pencils, a pilot performs a similar visual task with the information displayed in the cockpit. But unlike a non-hazardous office environment, a pilot flying in enemy territory needs to be able to perform this task in the shortest amount of time possible. They are required to process the information from 2 to 3 displays as quickly as possible to permit them to react faster than an adversary. Needless to say, the speed and accuracy with which a pilot performs a visual search is absolutely critical. The future of integrated avionics will include more information that competes for our attention, so a better understanding of the impact of this information to visual attention will become increasingly important.

Visual Attention

Decades of visual search experiments have resulted in a paradigm for characterizing human performance in visual search tasks. In this paradigm, efficiency of the search is defined as the slope of the function relating reaction time (RT) to the set size (Wolfe, 2001). Reaction time is defined as the amount of time required to make a correct "target-present" or "target-absent" response (Wolfe, 1998). Set size is defined as the total number of items present on the display (Wolfe, 1998). Therefore, this slope represents the rate at which items can be processed in a search (Wolfe, 2010). Parallel search tasks, such as finding a red spot among green spots, are perfectly efficient with a slope of zero or an infinitely unlimited rate. In other words, a person will find the red spot among any number of green spots in the same amount of time. Tasks that are sensitive to the set-size are called serial search. For example, in finding a "T" among "L's", reaction times will increase as a linear function of set-size. In serial search, a person much search through an average of half the items in order to find the target (Neisser, 2014). To understand the visual search tasks a pilot may encounter on the F-16, we first have to turn our attention to the nuances of visual attention.

Using the searchlight analogy for focused or selective visual attention, our search for information resembles a visual sampling process (Wickens & Hollands, 2000). Our visual sampling system is an artifact of the biology of our eye and brain. The processing power of our visual system is limited. As such, the resolution of our eye is greatest within the 2 degree center of our eye, also known as the fovea (Miller, 2019). Therefore, we have the best ability to resolve detail within this area of our vision. Our peripheral vision is the area outside of the fovea, in which resolution and color sensing decreases

eccentrically as distance from the fovea increases (Miller, 2019). Sensitivity to motion declines at a far less rate, allowing us to use our periphery to cue off motion that we later look at with our fovea (Wickens, Lee, Liu, & Becker, 2013). We can only confirm the presence of a specific target by directing our attention to that target (Wolfe, 2010). While we can attend to information outside our fovea, our eyes typically, although not always, move to locate the fovea to be coincident with the area of attention.

Having established our visual sampling system, we can describe the visual sampling process with two components: the fixation and saccade (Wickens & Hollands, 2000). A fixation is characterized by a location (the center of the fixation), useful field of view (diameter around the central location from which information is extracted), and dwell time (how long the eye remains at that location) (Wickens & Hollands, 2000). Saccadic movements are discrete, jerky movements that jump from one fixation in the visual field to the next and direct the fovea to an object or region of interest (Findlay & Walker, 2012; Wickens & Hollands, 2000). In normal viewing, several saccades are made each second and their destinations are selected by cognitive brain process without any awareness being involved (Findlay & Walker, 2012). Figure 3 shows the patterns of saccades and fixations made when viewing a picture of a face for 3 minutes (Findlay & Walker, 2012).

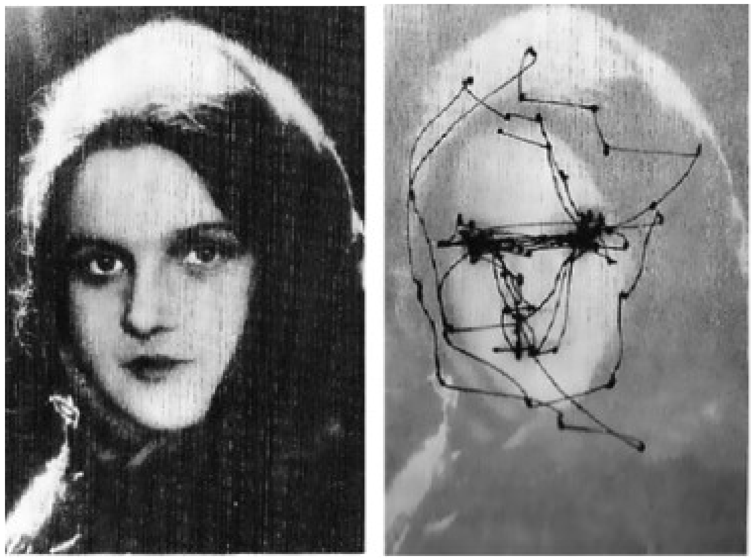


Figure 3 – Scan Path Showing Saccades and Fixations Made

Since it is not possible to fully process all of the stimuli in our full visual field at the same time, we need a source of guidance which relies on information in our periphery to drive changes in fixation location and attention (Wolfe, 2010). Sources of information for guidance or pre-attentive attributes include color, motion, size, and many others. The saliency of differentiable cues depends the type of cues and where they fall within our retinal eccentricity (Loschky, McConkie, Yang, & Miller, 2005). In other words, certain spatial cues may be salient when registered close to the fovea but be less salient with increased eccentricity as the resolution of the eye decreases with increasing eccentricity. Since these can provide differences that can be distinguished by our peripheral visual system in order to guide where to look with our fovea in one single step, this type of search is commonly called parallel search or feature search (Wolfe, 2010). Parallel search can be perfectly efficient, such as the earlier example of finding a red spot among green spots.

Without sources of guidance, we are forced into a serial or conjunction search process, attending one-by-one to every piece of information. If pre-attentively equivalent items are large and well-separated, the unguided search would range between 20 to 40 milliseconds per item in target-present conditions (Wolfe, 2010). However, if targets are not well-separated then longer eye fixation may be required to discern target details (Wolfe, 2010). For example, if T's and L's were slightly overlapping or touching, longer fixations are required because the shapes are not immediately clear. In that case, the search rate will be limited by the rate of eye movements, which is typically 3 or 4 per second (Wolfe, 2010). Serial, unguided search with poor target separation is considered the least efficient type of search.

The above examples describe situations where the target is present among distractors. The same performance cannot be expected for all target-absent conditions because of an asymmetry in visual search. RTs for target-absent conditions are about twice those of target-present conditions (Wolfe, 2001). While at first this may seem illogical, the systematic behavior can be understood when you consider basic features and our pre-attentive processing. If you have a situation in which a search for an A among Bs is highly efficient whereas a search for a B among As is less efficient, this is a hint that stimulus A has an added basic feature that B does not (Wolfe, 2001). In the target-absent condition, no basic features differ from those of the distractors, reducing the efficiency of the search. Thus, the presence of a search asymmetry can be one mark of a basic, pre-attentive feature. Research of asymmetries in visual search has been a valuable source of insight into pre-attentive visual processing and has led researchers to draw conclusions about how we use attention for many decades (Wolfe, 2001).

Models of Attention

In 1967, Ulric Neisser proposed a division of visual processing into pre-attentive processes that operate in parallel across the entire visual field and limited capacity processes that are restricted or deployed to loci or by objects of attention (Neisser, 2014). Anne Treisman and Garry Gelade further built on Neisser's theory by concluding that different features are registered early, automatically and in parallel across the visual field, while objects are identified separately and only at a later stage, which requires focused attention (Treisman & Gelade, 1980). This idea was incorporated in their Feature Integration Theory (FIT) model, which is one of the best known and most accepted theories in the field of visual attention today (Frintrop, 2006; Treisman & Gelade, 1980). FIT proposed a two-stage process. The pre-attentive stage captures the basic features of color, orientation, and intensity in parallel subconsciously. The individual features of an object are then combined in the focused-attention stage, permitting the perception of the whole object.

In 1989, an evolution of the FIT model was introduced by Jeremy Wolfe, called the Guided Search (GS) model (Wolfe, 1993). GS is now recognized among the most important theories in the field of psychophysical models of visual attention (Frintrop, 2006). Wolfe modified the basic FIT idea, by adding the concept of guidance to the model. Wolfe saw that search efficiency is governed by the ability to guide attention towards likely targets and the speed with which the distractors can be rejected (Wolfe, 2018). Guidance in GS is achieved through a combination of top-down (i.e., knowledge of the objects and features) and bottom-up (i.e., detection of features) processes to create a ranking of items by attentional priority, facilitating efficient search.

Like FIT, the GS model has undergone significant revisions and is now on its fourth iteration, appropriately named GS 4.0. Figure 4 below shows the architecture of this model. It is described as a parallel–serial hybrid with a parallel front end, followed by an attentional bottleneck with a serial selection rule that feeds parallel object recognition processes (Wolfe, 2007). Bottlenecks are a new addition to this visual search model, that are governed by visual selective attention since attention covers a very wide range of processes in the nervous system. Additionally, Wolfe has taken inventory of all the “diverse empirical phenomena” found in visual search experiments and attempted to address these in GS 4.0. Specifically, he states the model should account for the phenomena of set-size, presence/absence, features and target distractor similarity, distractor heterogeneity, flanking/linear separability, search asymmetry, categorical processing, and guidance (Wolfe, 2007). A few of these phenomena have been discussed earlier and a full discussion of each would be entirely too long and complex to explore here. Regardless, continued development of visual search models highlights how complex the phenomena is. Wolfe’s latest update on GS is appropriately titled “Current Progress with GS 4.0”, 30 years from its inception (Wolfe, 2007).

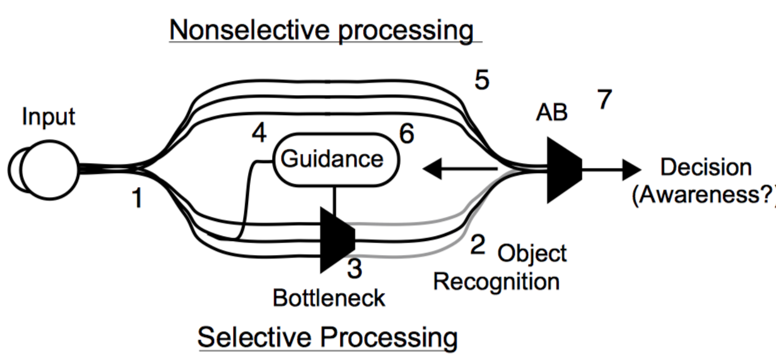


Figure 4 – Wolfe's Guided Search 4.0 Architecture

Neisser's division of visual processing, Treisman's FIT, and Wolfe's GS show us the mix of mechanisms a person uses in visual search depends on the task i.e. the specific target and its relationship to its distractors (Neisser, 2014; Treisman & Gelade, 1980; Wolfe, 1993). Therefore, symbology characteristics are obviously important in the design of cockpit displays. Symbol shape, color, features, and size all effect the mechanisms we use to execute a visual search. While it may be difficult to distinguish parallel from serial in visual search tasks, one can confidently say that a majority of tasks are a combination of both (Wolfe, 2007). A model such as GS 4.0 could be used to predict the performance of the user when performing a specific task for a given set of target and distractor symbols, but what about the other attributes of the cockpit, such as display size? Display size is another design decision that has to be made by system designers. Like symbology, display size likely impacts the pilot's visual scan performance. The larger the display, the more addressable pixels are generally available to present information to the pilot. However, the larger the display, the more useful pre-attentive features will be in guiding a pilot's search across the larger area. Otherwise a pilot may need to utilize a serial search process, which may require many eye fixations on a large display, making the search process too inefficient for successful completion of their mission.

Display Size Effects

Fortunately, the effect of display size on visual attention has been explored in recent research. Chen et al. attempted to invoke pure serial and parallel visual search tasks on 3 different size displays at 16°, 32°, and 60° horizontal fields of view (Chen et

al., 2011). For the parallel search task, participants were asked to find the letter "O" or the absence of an "O" among letter "V" distractors. Here, the letter "O" can be differentiated from "V" using one basic feature, which should have induced a pre-attentive or guided search process. For the serial task, participants were asked to find the letter "L" or the absence of the letter "L" among letter "T" distractors. Since the letters "L" and "T" share more than one basic feature, the authors believed this would induce a serial search process. All letters were well-separated and presented at the same size. Eye tracking was not used, so the presence of a true serial or parallel search process was only inferred by the slope of the function relating RT to the set size function.

The results indicated that all three main effects, three two way interactions, and one three way interaction were statistically significant (Chen et al., 2011). Overall, they found that set size and display size had a very small effect on search time in the parallel search condition (Chen et al., 2011). In contrast, search time increased with set size and display size in the serial search condition (Chen et al., 2011). The RT growth rate in the serial condition, averaged cross target-present and absent, was found to be around 4 msec/item/degree of visual angle compared to nearly 0 for the parallel search condition (Chen et al., 2011). The authors attribute the increase in search time to a search cost raised by the increase in display size. It is likely that more eye fixations are required to cover the increased area, however no data on fixations was collected in their research.

The search cost associated can also be seen in a technique for predicting serial search performance developed by Neisser in 1976 (Neisser, 2014). It states that if each inspection takes a relatively constant time, I, and the expected location of the target is

unknown beforehand, then it is possible to predict the average time it will take to find the target as the following,

$$T = \frac{N \times I}{2}$$

where *I* is the average inspection time for each item, and *N* is the total number of items in the search field (Wickens et al., 2013). Neisser found that, on average, the target will be encountered after half of the targets have been inspected (sometimes earlier, sometimes later) (Wickens et al., 2013). If you apply this model to the research by Chen et al. (2011), you see that if the number of items stays constant, the only way to increase the average time, *T*, to find a target is if the average item inspection time, *I*, increases. These findings indicate more about the composition of *I*, which includes components of visual sampling discussed earlier. The average inspection time, *I*, can be decomposed further into the size and duration of both saccades and fixations. Therefore, one can conclude that in a larger display with all other variables the same, if the time to inspect each item is greater, then an increase in the saccadic and fixation characteristics was the likely cause.

Issue of Clutter

A threat to the external validity of many visual search experiments revolves around the spacing of objects in the visual search scene. Targets and distractors within these experiments are often "well-separated" and evenly distributed throughout the scene. The astute engineer would realize that in reality this is rarely the case. In sensor displays, such as those in the F-16, objects detected are likely to occur in groupings or clusters for a variety of reasons. For example, air threats such as enemy aircraft, usually fly in formations of 2 or more. Ground threats, such as surface to air missiles, have multiple

pieces of equipment arranged in a particular pattern or cluster. Enemy populated areas on the ground may contain clusters of targets due to their higher density. The worst case of target overlap occurs if correlation of signals from multiple sensors for the same target is not being performed correctly. The result of this phenomena are multiple symbols, all essentially representing the same object, displayed in a closely spaced cluster and often overlapping with each other. While information separation on the display needs to be great enough for pilots to read salient features, this separation is often not present in realistic aircraft displays.

It becomes evident that these models need to be able to handle the structures and hierarchies that clusters form in a visual search scene. Wolfe hints at this ability in GS4 by stating that "scene statistics can guide deployments of attention is a new feature of GS4" (Wolfe, 2007) but doesn't explicitly describe how. Recent research by Yaoda Xu has provided some insight into how the presence of item clustering might impact visual search performance (Xu, 2010). In her research, Xu successfully predicts that it depends on the nature of the visual search. The experiment consisted of two tasks. The first was a simple feature search in which observers searched for the target letter "T" among distractor letters "O" and judged whether the target bottom of the letter "T" was pointing to the right or to the left when oriented horizontally (Xu, 2010). The second was a more challenging spatial configuration in which observers searched for the target letter "T" among distractor letter "L"s and judged the letter's orientation (Xu, 2010). The total number of items was fixed, but 6 cluster sizes were tested. The results provided evidence that item clustering impaired a simple feature search but facilitated a difficult spatial configuration search (Xu, 2010).

Fortunately, research has shown that asymmetry phenomena also holds true in cluttered displays (Yamani & McCarley, 2011). A study performed by Yamani and McCarley, demonstrated the robustness of the canonically oriented and mirror-reversed letter asymmetry in the presence of heavy display clutter (Yamani & McCarley, 2011). They asked participants to find canonical and reversed N's in geospatial images with various levels of clutter. The asymmetry proved to be true, in that the "reserved N" was easier to find than the "canonical N." Their findings are important because it shows that the design of display symbology to produce visual search asymmetries can offset the cost of visual clutter, maximizing detectability of task-critical information in complex displays (Yamani & McCarley, 2011).

Application to Research

The visual search task a pilot performs when looking at the display of sensor data is not unlike the phenomena explored in visual search research. For example, a pilot in the cockpit of an F-16 looks at their displays either seeking a particular piece of information (top-down) or by scanning for unexpected information (bottom-up). As research has shown, whether parallel serial mechanisms are employed by the pilot depends on the task, specifically the target and its distractors (Neisser, 2014; Treisman & Gelade, 1980; Wolfe, 1993). Assuming that they have already moved their eyes to the display of interest, they could first be guided by any pre-attentive features, such as color, size, or motion. They would then direct their attention in these areas first by making an eye movement. If no pre-attentive features are present, the pilot will make an eye movement to focus their attention at a particular aspect of the screen dictated by training.

For instance, if it is a radar map, a pilot would start at the nearest range and work their way out, increasing in range. If the objects they are attending to are not large, nor well-separated, they will then fixate on each object until they can differentiate salient information. If salient information is present, then their target has been found. If no salient features are present, they may restart the process by making another eye movement and continuing their scan. It is also possible that pre-attentive features may become more apparent during the scan, so they could be guided by these in the middle of the search. Visual search appears to have excellent validity with tasks a pilot performs in a cockpit and is a worthy factor of consideration.

Summary

Predicting the human performance expected for a pilot to attend to information from a sensor display seems to be an insurmountable task. Depending on their intended piece of information and its location within the scene, their search could be parallel, serial, or a combination of both. Guiding features are only salient if they appear within certain areas of the field of view. Therefore, removing artificial target eccentricity limitations almost guarantee that the search will always be a mix of serial and parallel processes. A serial search is the most concerning because it appears to have a cost associated with the size of the display, in addition to the number of distractors. A larger display increases the chances of information will be placed further into the periphery of the viewer, which potentially degrades the saliency and overall effectiveness of guiding attributes. The end result is more serial search. Chen et al. (2011) performed their research on displays with horizontal viewing angles of 16°, 32°, and 60°. Using a CAD

cockpit model, the MFD and CDU on the F-16 are estimated to have visual angles within the range tested by Chen et al. Therefore, this author expects that the CDU will have a higher average search cost than the MFD with all other variables constant. In our future of integrated avionics and the large data sets they provide, this search cost will be important to understand to maintain maximum effectiveness against our enemies. Based on this framework, the next chapter describes the methodology to be employed in the design of a human factors experiment to quantify the search cost in terms of human performance.

III. Methodology

Chapter Overview

This chapter contains data collection methods used for this research and outlines how the data will be analyzed. For this research, an experiment was designed to measure the anticipated effect of increased search cost associated with performing a visual search on a larger display. Then the experiment was conducted using human participants and the results were analyzed to address the hypothesis.

Hypothesis

The physical size of the display is positively related to the visual search time.

Participants

Ten participants (1 female and 9 males) with ages between 21 and 40 voluntarily participated in the study. Six participants were between the ages of 21 and 30, while the remaining four had ages between 31 and 40. All of the participants reported normal or corrected-to-normal vision. Four of the ten participants reported near-sighted vision and were wearing corrective glasses or contacts. Full participant details are shown in Appendices E and G. Participants were recruited through e-mail across the Air Force Institute of Technology Department of Systems Engineering and Management.

Experimental Design

The experiment performed was a classic visual search experiment. The experiment included four independent variables which were manipulated as within participants variables in a full-factorial experimental design. The variables and their associated specifications are as follows.

Independent Variables

The first independent variable was cockpit display, which was either the MFD or CDU. The MFD is 8° x 8° (4.25" x 4.25"), 524 x 524 pixels, located 18.3° down & 11° left, 34.8" look distance. The CDU is 17° x 17° (6" x 6"), 768 x 768 pixels, located 18.3° down & on-center, 33.8" look distance. The second independent variable was distractor levels 1 through 7, which were designed to increase scene complexity. More details are shown in Table 1 of the Scene Design section. The third independent variable was target condition, which was either present or absent. The fourth independent variable was set-size, which was either 20, 40, or 60 items per scene.

Dependent Variables

The dependent variables included reaction time and error rate. Reaction time was defined as the amount of time that is required to make a correct "target-present" or "target-absent" response. This time was determined from the difference between the time a scene was presented and the first mouse button pressed by the participant. Error was determined by a participant's incorrect key response, either indicating a target as present when it was not or vice-versa.

Scene Design

Traditionally, visual search experiments are performed using common symbols, such as letters and common shapes. However, in the case of F-16 sensor displays, there are plenty of symbols to draw inspiration from. Figure 5 below shows a simple radar view on the F-16, which is composed of squares and circles for basic shapes. Those shapes have a few modifiers such as fill, vector lines, and number labels. In reality, there

are significantly more types of symbol variations and options, many of which are defined in MIL-STD 2525D (United States of America Department of Defence, 2014). The experiment was designed around the basic circle and square, which are common shapes. Symbol modifiers, such as those discussed above, were made to see the impact of changes in distractor symbology. A slightly smaller filled-white square was used as the target because it represents a new radar detection that a pilot would have to recognize among objects already being tracked by the radar. The smaller size of the square allows for a guided search by the human participant when searching among squares. Size is one of the undoubted attributes, also called pre-attentive attributes, that has been demonstrated to guide search in a large number of studies (Wolfe & Horowitz, 2008).

Figure 5 – Example Radar View

Symbol pixel sizes were approximated from actual F-16 MFD views and were held constant on both displays. Each scene was composed of an equal number of circle

and square basic shapes. The final experimental design included seven levels of distractors or unique task situations in target-present and absent conditions. Examples of the target-present Levels 1 and 7 are shown in Figure 6. A complete list of distractor level examples are shown in Appendix A.

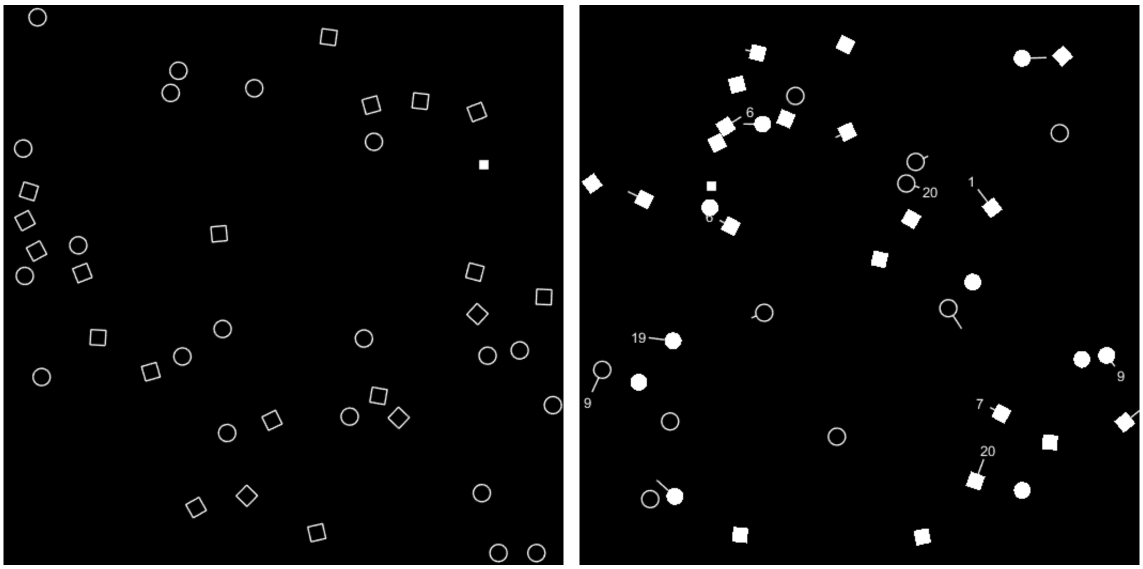


Figure 6 – Target-Present Distractor Level 1 (left) & Level 7 (right) Scene Examples

The levels were designed with the intention of increasing level of difficulty, starting from a parallel-like search in Level 1 and increasing the amount of serial search up to Level 7. This was accomplished by using symbol modifiers to create more unique symbol types contained within the scene. Color was not used to simplify the experimental design. The resulting scenes were not operationally realistic or feasible because they were artificially generated for the purposes of comparing search performance between the 2 displays. The incremental approach allows for more clarity in how distractors affect the search and better covers the types of visual search conditions that may actually occur on the F-16.

Table 1 below describes the physical differences between each of the levels, including number of unique symbols that occur as result of the fill, vectors, and numbers. As the modifiers are applied the number of unique symbols increases within the scene.

Table 1 – Scene Distractor Level Descriptions

Level	Description	Number of Unique Symbols
1	Unfilled squares & circles	2
2	Unfilled squares, filled circles	2
3	Unfilled circles, filled squares	2
4	50/50 unfilled/filled squares, filled circles	3
5	50/50 unfilled/filled circles, filled squares	3
6	Same as 5, but add random vectors on half the squares & circles	5
7	Same as 6, but add random numbers on half the vectors	7*

*Level 7 is challenging to define because it contains a number label located near the end of the vector line, shown in Figure 7. Therefore, the text label could either be considered part of the symbol or another grouping of symbols in itself.

Figure 7 – Distractor Level 7 Ambiguity

Set size, as traditionally defined, is the total number of distractor items that the human participant must search through in the scene. The set sizes used in the experiment were 20, 40, and 60. The smallest set size used was 20 items, 10 squares and 10 circles, and was determined to be within the range of current AESA radar capabilities (*Lockheed Martin F-16 Block 70 India Brochure*, 2016). Accounting for the additional sources of

sensor data on integrated radar displays, the set sizes of 40 and 60 are estimated to be feasible, but unlikely to be encountered in the real-world. A custom MATLAB program was used to draw scenes randomly with all the characteristics required for the experiment. The shapes were drawn using binary image functions and then placed in random locations in the scene, using a normal random distribution. Overlap was detected by dilating the shapes and detecting any intersections with previously placed shapes. If a shape intersected another, the program would replace the shape in another random location until overlap was not detected. Scenes were generated ahead of time and not during the experiment.

Experimental Apparatus

The experimental apparatus consists of one LCD monitor, monitor stand, desktop computer, and peripherals. The monitor was 24" ViewsonicVG2455-2K IPS liquid crystal display monitor with 2560 x 1440 pixels. The monitor matches the F-16 MFD and CDU pixel size within 6% and was the nearest match commercially available. The extruded aluminum monitor stand was custom-built to place the monitor at 27° from vertical, which creates a visual angle that approximately replicates the F-16 cockpit environment when participants are placed according to the dimensions in Figure 8. The monitor stand dimensions are shown in Appendix B. The viewing angles were derived from a representative CAD model of the F-16 cockpit provided by the 711th Human Performance Wing, Wright Patterson AFB, OH. Since an anthropometric human model was not available, visual angles were estimated using the F-16 minimum sitting eye height of 30.2" (Zehner, 2002). A standard desktop computer with a dedicated graphics

card running PsychoPy 3.0, standard computer mouse, USB number pad with blue mechanical switches, and non-adjustable chair were also used.

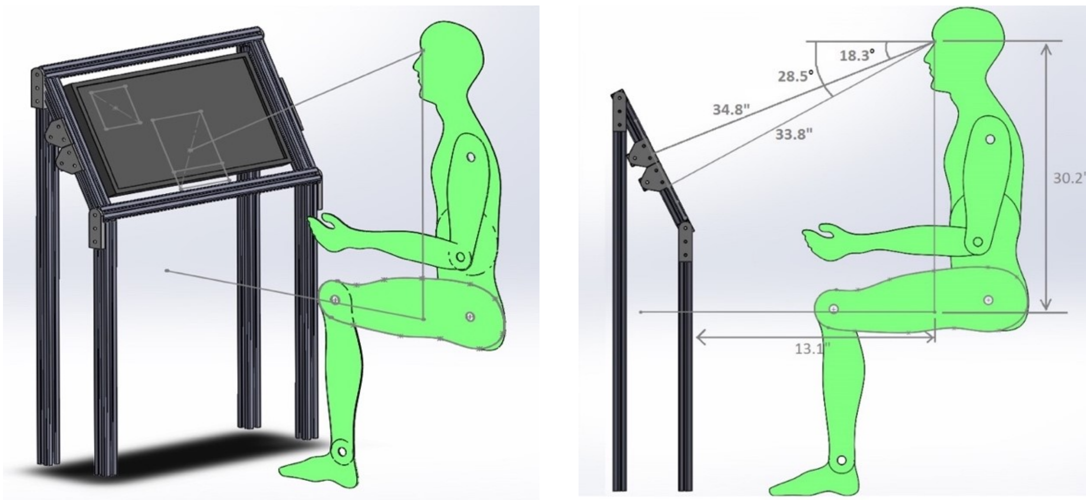


Figure 8 – Isometric View (left) and Side view (right) of Experimental Set-up

Figure 9 – Test Apparatus & Input Pad

Procedure

The experiment was designed in PsychoPy 3.0, which is an open source platform for designing and running experiments in Python (Peirce et al., 2019). The PsychoPy

Builder interface was used to define the overall structure before customizing the other unique Python code functions. Figure 10 below shows that the experiment structure consisted of two small loops for each display nested in one large loop. Custom python scripts were written to allow for full randomization within the structure and other required functions. For instance, the first scene presented has an equal chance of being either CDU or MFD. Since each display had a different size and position on the screen, each required separate modules on the builder. Each module randomly draws a scene from the specific display database without replacement. The end result was an experiment that was unique for every execution.

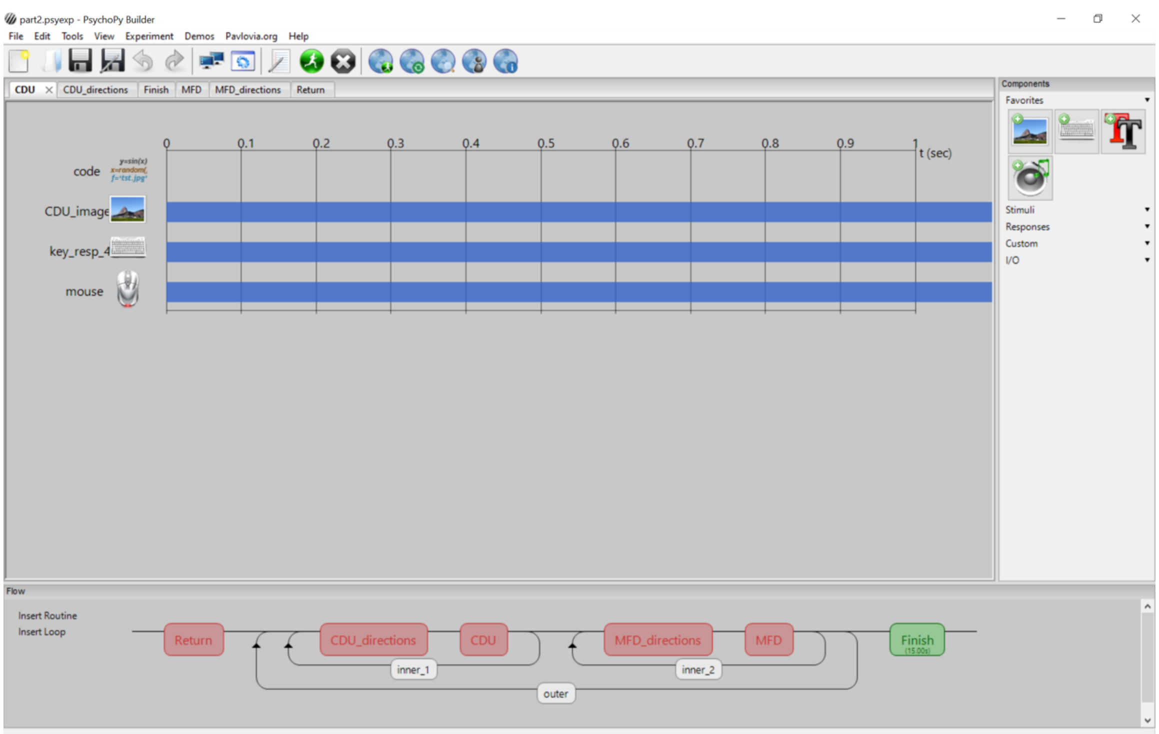


Figure 10 – Experiment Design in PsychoPy Builder

The experiment was designed for human participants to be presented with several hundred search tasks or scenes, each one estimated to take under 2 seconds. As in

traditional visual search experiments, the scenes were split 50/50 for target-present/absent conditions (Wolfe & Horowitz, 2008). Due to time scheduling convenience and user fatigue concerns, the experiment was designed be executed within one hour of time. A pilot study was used to estimate timing. Using 1 hour as the time constraint, each scene combination could be replicated 5 times. A scene replicate looks unique but contains the same number and types of symbols. The locations of each are different for each replicate. A total of 410 scenes were presented to each participant in a fully randomized order across all scenes with no blocking. Table 2 below shows the test matrix used in the experiment.

Table 2 - Test Matrix Repeated for Each Display

Distractor Level	**Scene ID**	**Set Size**	**Target Present Replications**		**Target Absent Replications**
1	1	20	5		5
	2	30	5		5
	3	60	5		5
2	4	20	5		5
	5	30	5		5
	6	60	5		5
3	7	20	5		5
	8	30	5		5
	9	60	5		5
4	10	20	5		5
	11	30	5		5
	12	60	5		5
5	13	20	5		5
	14	30	5		5
	15	60	5		5
6	16	20	5		5
	17	30	5		5
	18	60	5		5
7	19	20	5		5
	20	30	5		5
	21	60	5		5
		Subtotal	105		105
			Total scenes	210	

On test day, participants were asked a series of pre-experiment screening questions to determine if they qualified to participate before consenting to begin. The survey used is shown in Appendix D. Then the participants were positioned to have a visual sightline distance of 34” to the center of the screen. Each participant was required to keep their left index and middle fingers on the “target-present” and “target-absent” indicator keys respectively, as shown in Figure 11 below.

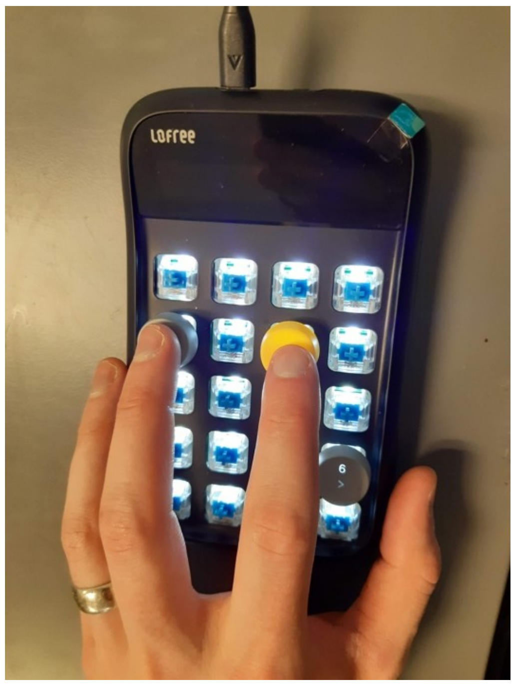


Figure 11 – Number Pad Finger Placement

Once positioned, the participants were presented a series of instructions on the screen. They were asked to work quickly and accurately, balancing both equally, as if they were flying in an F-16. Then a series of 8 practice trials, 4 target-present and 4 target-absent, were provided to practice the keystrokes. Prior to each scene, the participant is provided with the location of the next scene on the screen so that they may move their eyes to that area of the display. In this way, variable eye movement to the

area of the display is not captured in the reaction time measurement. When a scene was presented, the participant indicated that the target was present or absent by pressing the respective keys. For target present conditions, they would move their mouse to the target location and left click on the target symbol. After the participant completed 105 random scenes, a 10-minute break was provided. The experiment was completed after the participant completed the remaining 105 scenes. A post-experiment survey was administered to collect their feedback on the tasks they were asked to perform and collect demographic data. The survey used is shown in Appendix F.

Output Data Analysis

Data on the dependent variables of reaction time and error were collected through the PsychoPy 3.0 program. The program allows for robust logging of keystrokes and simultaneous timing of experimental events limited to the refresh rate of the monitor being used. Errors were determined by comparing what scene was presented and what response button was pressed by the participant. For a trial to be considered a correct target identification, participants had to first indicate correctly on the number pad and then right click with the mouse on the target symbol within a specified radius. The program captured a screenshot of each mouse click with a green circle, drawn at that radius. An example is how in Figure 12 shown below. The screenshots were analyzed posttest using a simple Python routine utilizing OpenCV to determine if the target square was contained within the green circle.

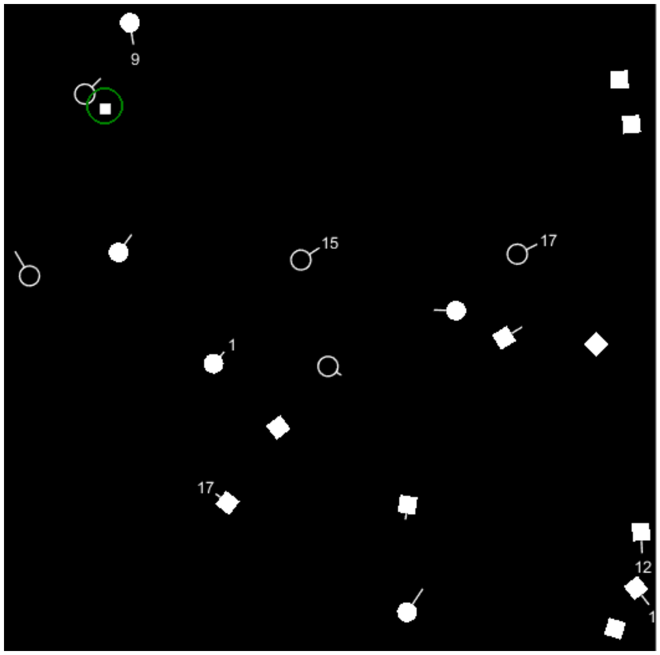


Figure 12 – Target Click Radius Screenshot Example

IV. Analysis and Results

Chapter Overview

This chapter details the analysis and results of the human subjects experiment performed in the study. The analysis was divided by the two types of tasks performed, target-present and target-absent. SPSS was used to perform a repeated measures ANOVA on each of the main effects and their interactions. The results are presented below. Results for the participant pre-and post-experiment surveys are shown in Appendix E and G, respectively.

Target-Present Analysis

Error

As discussed earlier, error was determined by a participant's incorrect key response, indicating a target as not present when it was in a target-present scene. No errors were determined from the screenshot analysis using the Python OpenCV routine. Therefore, all correct target-present responses were determined to be valid and no data points were rejected. Overall, the total errors for both displays were less than 6% for all participants in the target-present condition. Results by participant shown in Table 3 below.

Table 3 – Target-Present Error by Participant

Participant	Total Errors	%	Participant	Total Errors	%
1	8	2%	6	4	1%
2	11	3%	7	9	2%
3	7	2%	8	20	5%
4	21	5%	9	9	2%
5	18	4%	10	2	0%

An interesting result occurred in the comparison of error by display type, shown in the Graph below. The errors were on average 87% greater on the larger CDU. Since error rates were so low overall, this was not considered a significant result.

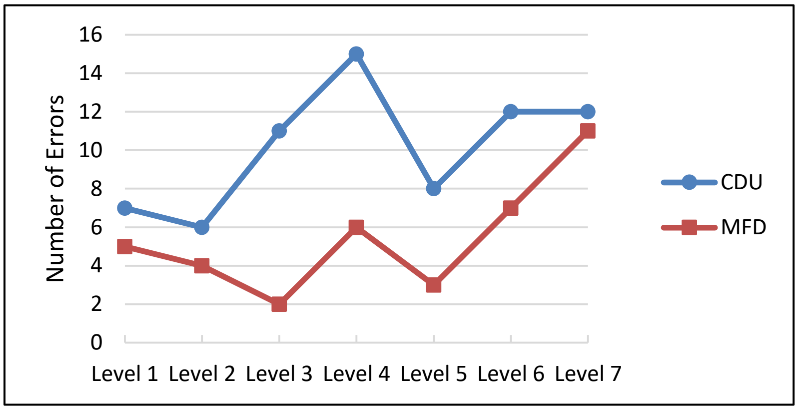


Figure 13 – Target-Present Errors by Display

Reaction Time

Reaction times for target-present trials were analyzed using a three-way, within-subjects analysis of variance with display (MFD or CDU), set size (20, 40, or 60), and distractor levels (1 through 7). The data violated Mauchly's test of sphericity and therefore a Greenhouse-Geisser correction was applied to adjust for the lack of sphericity. The three main effects were found to be significant with display [F(1, 9) = 32.3, MSE = 6.1, p = 0.000, $\eta_p{}^2 = 0.78$), set size (F(1.7, 15.0) = 25.9, MSE = 15.6, p = 0.000, $\eta_p{}^2 = 0.85$] and distractor level [F(2.0, 18.2) = 34.1, MSE = 11.3, p = 0.000, $\eta_p{}^2 = 0.79$]. Display*distractor level [F(3.8, 34.5) = 3.8, MSE = 0.49, p = 0.012, $\eta_p{}^2 = 0.30$] and set size*distractor level [F(2.9, 26.1) = 4.6, MSE = 2.38, p = 0.011, $\eta_p{}^2 = 0.34$] interactions were also found to be significant. The Greenhouse-Geisser adjusted ANOVA table is shown below.

Table 4 – Adjusted Analysis of Variance Results Target-Present

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Display	Sphericity Assumed	6.110	1	6.110	32.336	0.000	0.782
	Greenhouse-Geisser	6.110	1.000	6.110	32.336	0.000	0.782
Error(Display)	Sphericity Assumed	1.701	9	0.189			
	Greenhouse-Geisser	1.701	9.000	0.189			
SetSize	Sphericity Assumed	25.920	2	12.96	49.249	0.000	0.845
	Greenhouse-Geisser	25.920	1.663	15.591	49.249	0.000	0.845
Error(SetSize)	Sphericity Assumed	4.737	18	0.263			
	Greenhouse-Geisser	4.737	14.96	0.317			
DisLvl	Sphericity Assumed	22.778	6	3.796	34.085	0.000	0.791
	Greenhouse-Geisser	22.778	2.019	11.279	34.085	0.000	0.791
Error(DisLvl)	Sphericity Assumed	6.014	54	0.111			
	Greenhouse-Geisser	6.014	18.18	0.331			
Display * SetSize	Sphericity Assumed	0.545	2	0.273	2.916	0.080	0.245
	Greenhouse-Geisser	0.545	1.869	0.292	2.916	0.085	0.245
Error(Display*SetSize)	Sphericity Assumed	1.683	18	0.093			
	Greenhouse-Geisser	1.683	16.82	0.100			
Display * DisLvl	Sphericity Assumed	1.888	6	0.315	3.836	0.003	0.299
	Greenhouse-Geisser	1.888	3.828	0.493	3.836	0.012	0.299
Error(Display*DisLvl)	Sphericity Assumed	4.430	54	0.082			
	Greenhouse-Geisser	4.430	34.46	0.129			
SetSize * DisLvl	Sphericity Assumed	6.901	12	0.575	4.571	0.000	0.337
	Greenhouse-Geisser	6.901	2.898	2.381	4.571	0.011	0.337
Error(SetSize*DisLvl)	Sphericity Assumed	13.588	108	0.126			
	Greenhouse-Geisser	13.588	26.09	0.521			
Display * SetSize * DisLvl	Sphericity Assumed	2.248	12	0.187	1.842	0.050	0.170
	Greenhouse-Geisser	2.248	3.933	0.572	1.842	0.143	0.170
Error(Display*SetSize*DisLvl)	Sphericity Assumed	10.984	108	0.102			
	Greenhouse-Geisser	10.984	35.4	0.310			

The main effects plots are shown in Figure 14. In Figure 14a, we see expected results with RT increasing as set size increases with a strong correlation (r=0.985). Due to the few number of data points, a T-test was not calculated to test the significance of the relationship. However, as discussed in Ch II, the set size and reaction time correlation is widely researched and accepted as significant in the community. In Figure 14b, CDU resulted in approximately 0.24 sec (18%) slower RT than MFD, which supports the premise of a search cost associated with larger displays. Figures 14c and d show distractor level is correlated with reaction time. After converting both variables to a

ranked scale, a Spearman correlation coefficient was calculated to be (ρ=0.929), indicating strong correlation. A T-test was then performed to test significance resulting in a p-value of 0.003, confirming a significant correlation.

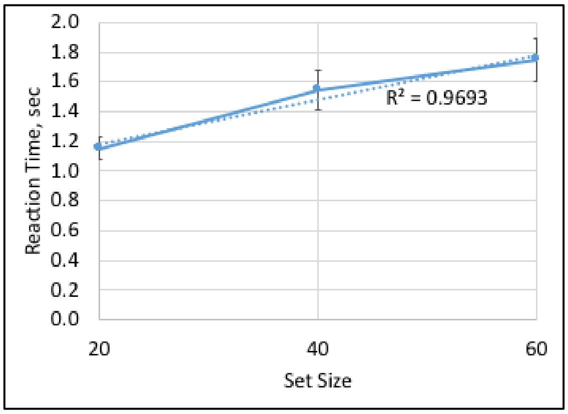


Figure 14a – TP Set Size Main Effect

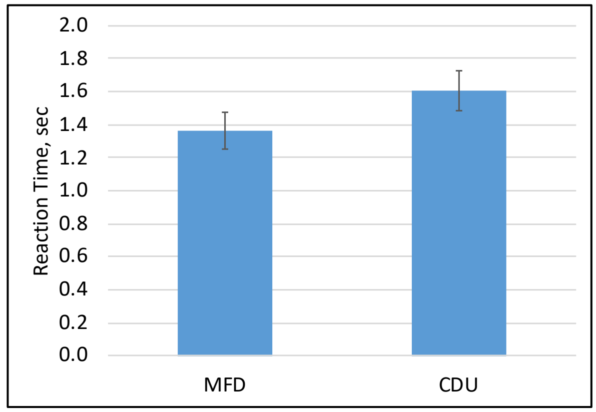


Figure 14b – TP Display Type Main Effect

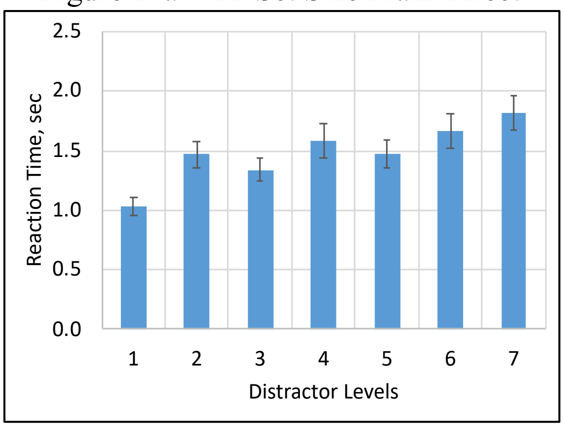


Figure 14c – TP Distractor Level Main Effect

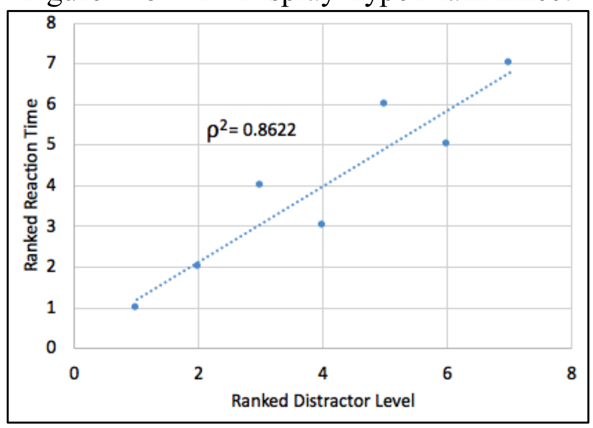


Figure 14d - TP Ranked Distractor Level Main Effect

Plotting the reaction times versus set size allows for a comparison of the search efficiencies through examination of the slope. Figure 15 below shows the moderate slopes for both displays are nearly the same. Since symbol size and distractor levels were identical between the two displays, the search tasks were the same. Therefore, we should expect that both displays have nearly the same slope. The search cost of about 0.24 seconds from scanning a larger display is shown in the offset between the two. The time

to search on the MFD was on average 0.039 seconds per item compared to 0.045 seconds per item on the CDU. The search cost is likely due to the increased number of fixations required to search the larger area.

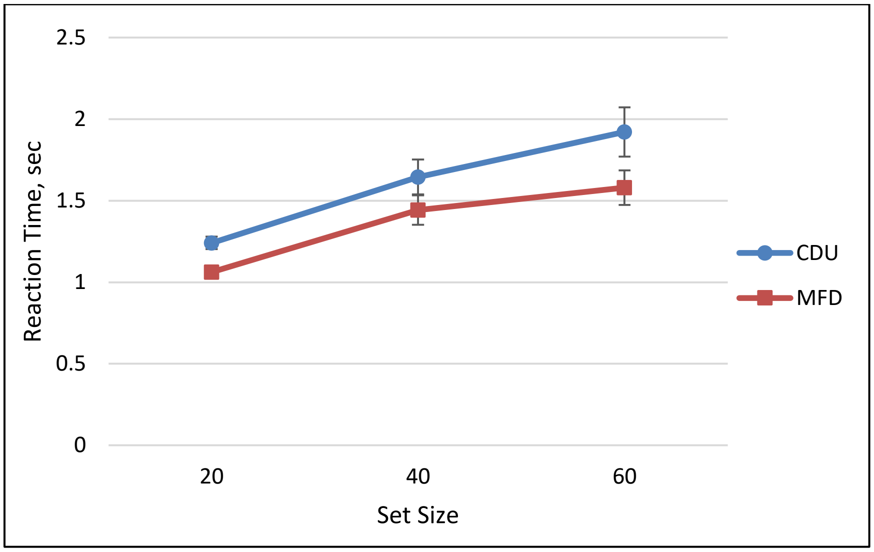


Figure 15 – Target-Present Efficiency by Display

A Bonferroni Pairwise comparison was performed to compare the levels within the main effects. Levels within display and set size were all found to be statistically different than each other with at least ($p <= 0.005$). Distractor levels were found to have more mixed results. Table 5 summarizes those findings and their associated p-values. The green cells indicate the distractor levels which are statistically different from each other. The red cells are not statistically different. Levels 6 and 7, while not statistically different from each other, were found to be statistically different than level 5. In terms of scene design, reaction time for Level 6 was not statistically different than reaction time for level 5 with the addition of the vector lines on 50% of the symbols. The reaction time for Level 7 was not statistically different from the reaction time for level 6 with the addition of the number labels on 50% of the vector lines.

Table 5 – Bonferroni Pairwise Distractor Level Comparison Main Effects TP

Level	1	2	3	4	5	6	7
1		0.000	0.013	0.008	0.004	0.002	0.000
2			0.248	0.904	1.000	0.028	0.000
3				0.09	0.685	0.015	0.000
4					0.705	1.000	0.005
5						0.025	0.000
6							0.291
7							

Plotting the same efficiency plot by distractor level provides some insights into the how the distractor combinations affect search efficiency. Figures 16 and 17 show the results for CDU and MFD, respectively. Level 1 for both displays had the flattest slope, indicating this was a nearly parallel, highly efficient search. As the distractor levels increase, there appears to be an increase in slope as more serial search is mixed into the search task. Level 6 or 7 are the most inefficient search tasks for both displays. These levels contained the largest and second largest number of unique symbols, at 5 and 7 respectively.

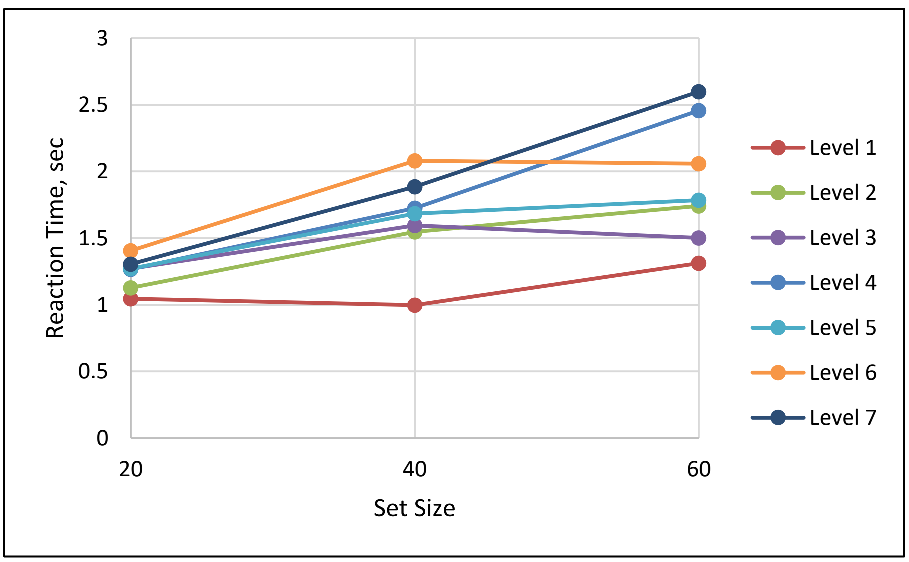


Figure 16 – CDU Target-Present by Distractor Level

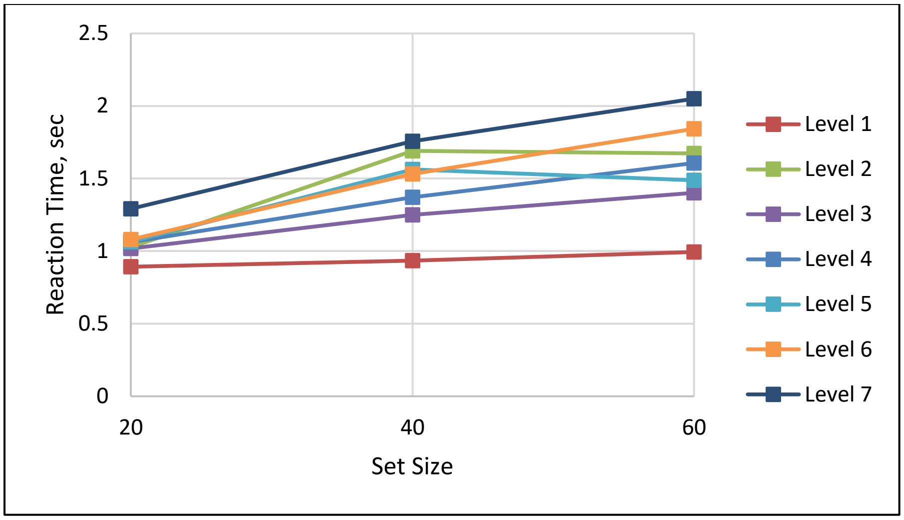


Figure 17 – MFD Target-Present by Distractor Level

Symbol complexity appears to have a greater effect for the large set size. Figure 18 shows a comparison between reaction times for distractor level 5 and level 7. Level 7 distractors were the same as level 5 but with vector lines and labels added on 50% of the symbols. Results in Figure 18 show there is a larger difference in reaction time between the two displays at the set size of 60, compared to the smaller set sizes.

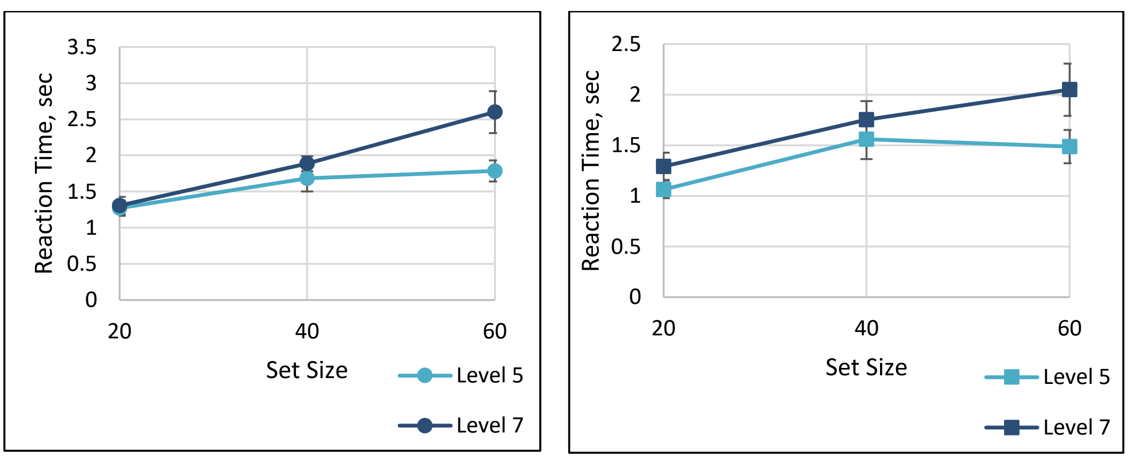


Figure 18 – Distractor Level 5 & 7 on CDU (left) & MFD (right) Target-Present

A single ANOVA was performed on the two factor interactions to determine the effect of factors at their lower levels. The adjusted ANOVA tables for both display types are shown in Table 6 below. Display*distractor level was analyzed by holding display constant and averaging across the set sizes. Distractor levels were found to be significant within each display. Both displays had significant positive correlations with reaction time. Correlations and significance values are shown in Table 7.

Table 6 – Display*Distractor Level Adjusted ANOVA Target-Present

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
CDU							
DisLvl	Sphericity Assumed	4.859	6	0.81	19.575	0.000	0.685
	Greenhouse-Geisser	4.859	2.695	1.803	19.575	0.000	0.685
Error(DisLvl)	Sphericity Assumed	2.234	54	0.041			
	Greenhouse-Geisser	2.234	24.259	0.092			
MFD							
DisLvl	Sphericity Assumed	3.363	6	0.56	24.262	0.000	0.729
	Greenhouse-Geisser	3.363	2.901	1.159	24.262	0.000	0.729
Error(DisLvl)	Sphericity Assumed	1.247	54	0.023			
	Greenhouse-Geisser	1.247	26.106	0.048			

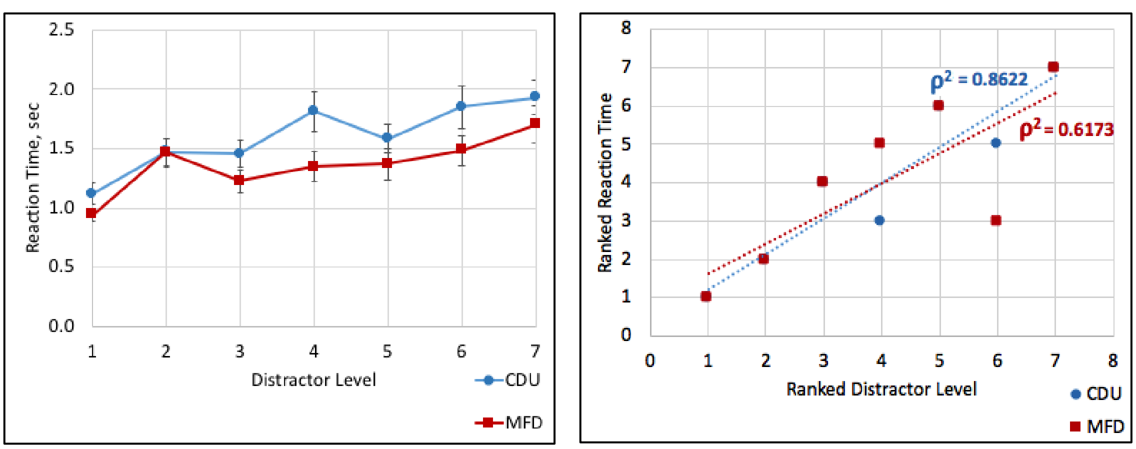


Figure 19 – Display*Distractor Interaction (left) & Ranked (right) Target-Present

Table 7 – Display*Distractor Interaction Correlation & Significance

Display	**Spearman Correlation, ρ**	**T-test (p-value)**
CDU	0.929	0.003
MFD	0.786	0.036

A Bonferroni Pairwise comparison was performed to compare the distractor levels within the CDU and MFD. Table 8 and 9 summarizes those findings and their associated p-values. The green cells indicate the distractor levels which are statistically different from each other. The red cells are not statistically different. Distractor levels within the displays were found to 6 pairs with different statistical conclusions, which was more than expected. Levels 6 and 7 were not statistically different from each other in both displays. Level 5 and 7 were statistically different from each other in both displays, which was also true for the main effects.

Table 8 – Bonferroni Pairwise Distractor Level Comparison CDU Target-Present

Level	1	2	3	4	5	6	7
1		0.001	0.071	0.012	0.011	0.011	0.002
2			1.000	0.085	0.917	0.059	0.006
3				0.145	1.000	0.163	0.004
4					0.287	1.000	1.000
5						0.358	0.009
6							1.000
7							

Table 9 – Bonferroni Pairwise Distractor Level Comparison MFD Target-Present

Level	1	2	3	4	5	6	7
1		0.002	0.030	0.019	0.038	0.002	0.002
2			0.022	1.000	1.000	1.000	0.155
3				0.445	0.937	0.022	0.002
4					1.000	0.243	0.001
5						1.000	0.006
6							0.202
7							

Set size*distractor level was analyzed by holding set size constant and averaging across the display types. The adjusted ANOVA tables for the set size 20, 40, 60*distractor levels are shown in Table 10 below. Distractor levels were found to be significant within each set size. All three set sizes had significant positive correlations with reaction time. Correlations and significance values are shown in Table 11.

Table 10 – Set Size*Distractor Level Adjusted ANOVA Target-Present

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Set Size 20							
DisLvl	Sphericity Assumed	0.69	6	0.115	5.706	0.000	0.388
	Greenhouse-Geisser	0.69	2.164	0.319	5.706	0.010	0.388
Error(DisLvl)	Sphericity Assumed	1.089	54	0.02			
	Greenhouse-Geisser	1.089	19.475	0.056			
Set Size 40							
DisLvl	Sphericity Assumed	5.054	6	0.842	9.912	0.000	0.524
	Greenhouse-Geisser	5.054	2.564	1.971	9.912	0.000	0.524
Error(DisLvl)	Sphericity Assumed	4.59	54	0.085			
	Greenhouse-Geisser	4.59	23.079	0.199			
Set Size 60							
DisLvl	Sphericity Assumed	9.095	6	1.516	19.853	0.000	0.688
	Greenhouse-Geisser	9.095	2.628	3.46	19.853	0.000	0.688
Error(DisLvl)	Sphericity Assumed	4.123	54	0.076			
	Greenhouse-Geisser	4.123	23.655	0.174			

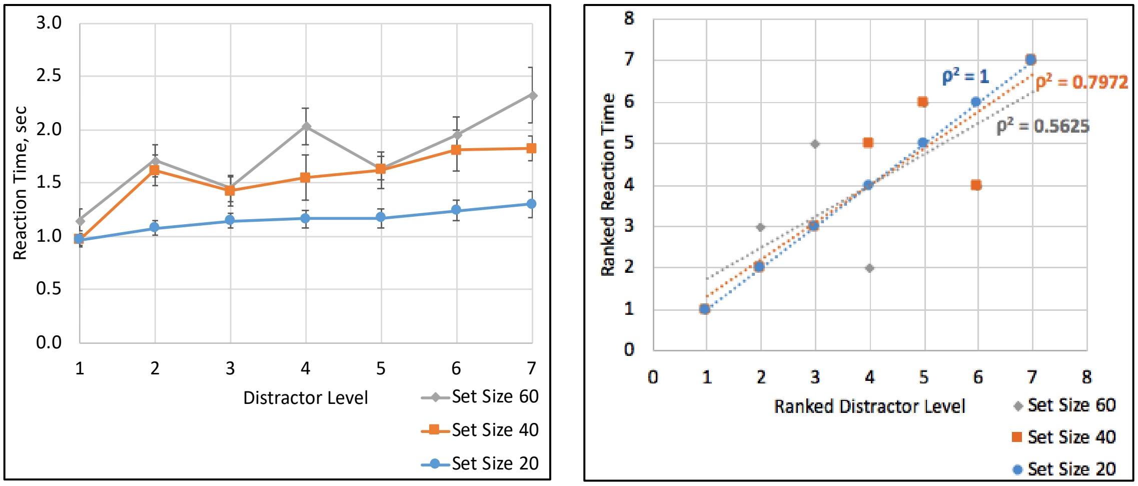


Figure 20 – Set Size*Distractor Interaction (left) & Ranked (right) Target Present

Table 11 – Set Size*Distractor Interaction Correlation & Significance

Set Size	**Spearman Correlation, ρ**	**T-test (p-value)**
20	1.000	0.000
40	0.893	0.007
60	0.750	0.052

A Bonferroni Pairwise comparison was performed to compare the distractor levels within the three Set Sizes. Tables 12, 13, & 14 summarize those findings and their associated p-values. The green cells indicate the distractor levels which are statistically different from each other. The red cells are not statistically different. Distractor levels within the displays were found be mostly not different from each other in set size 20 and 40. However, set size 60 has double the amount of significant distractor levels from the smaller sizes. It appears that the differences between the distractor levels are more pronounced at the largest set size.

Table 12 – Bonferroni Pairwise Distractor Level Comparison Set Size 20 TP

Level	1	2	3	4	5	6	7
1		0.493	0.021	0.111	0.114	0.031	0.276
2			1.000	1.000	1.000	0.08	0.734
3				1.000	1.000	0.908	1.000
4					1.000	1.000	1.000
5						1.000	1.000
6							1.000
7							

Table 13 – Bonferroni Pairwise Distractor Level Comparison Set Size 40 TP

Level	1	2	3	4	5	6	7
1		0.002	0.109	0.175	0.016	0.008	0.000
2			1.000	1.000	1.000	1.000	1.000
3				1.000	1.000	0.554	0.441
4					1.000	0.040	1.000
5						0.161	1.000
6							1.000
7							

Table 14 – Bonferroni Pairwise Distractor Level Comparison Set Size 60 TP

Level	1	2	3	4	5	6	7
1		0.002	0.365	0.002	0.002	0.002	0.007
2			0.800	0.338	1.000	0.578	0.085
3				0.001	0.462	0.010	0.007
4					0.056	1.000	1.000
5						0.248	0.099
6							1.000
7							

Target-Absent Analysis

Error

As discussed earlier, error was determined by a participant's incorrect key response, indicating a target as present when it was not in a target-absent scene. No errors were determined from the screenshot analysis using the Python OpenCV routine, therefore all correct target-absent responses were determined to be valid no data points were rejected. Overall, the total errors for target-absent condition were much less than target-present at 1% or less for all participants. Results by participant shown in Table 15 below.

Table 15 – Target-Absent Error by Participant

Participant	Total Errors	%		Participant	Total Errors	%
1	0	0%		6	1	0.2%
2	0	0%		7	0	0%
3	0	0%		8	1	0.2%
4	4	1%		9	1	0.2%
5	1	0.2%		10	0	0%

Reaction Time

Like the target-present condition, reaction times for target-absent trials were analyzed in a three-way, within-subjects analysis of variance with display (MFD or CDU), set size (20, 40, or 60), and distractor levels (1 through 7). Once again, the data violated Mauchly's test of sphericity and the Greenhouse-Geisser correction was applied to adjust for this assumption violation. The three main effects were found to be significant with display [$F(1, 9) = 75.4$, $MSE = 42.3$, $p = 0.000$, ${\eta_p}^2 = 0.89$], set size

[F(1.0, 9.3) = 37.9, MSE = 157.9, p = 0.000, $\eta_p{}^2$ = 0.81] and distractor level [F(1.7, 15.4) = 46.5, MSE = 123.4, p = 0.000, $\eta_p{}^2$ = 0.84]. Display*set size [F(1.8, 16.5) = 10.6, MSE = 2.2, p = 0.001, $\eta_p{}^2$ = 0.54] and set size*distractor level [F(2.8, 24.9) = 7.8, MSE = 10.4, p = 0.001, $\eta_p{}^2$ = 0.46] interactions were also found to be significant. The Greenhouse-Geisser adjusted ANOVA Table 16 is shown below.

Table 16 – Adjusted Analysis of Variance Results Target-Absent

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Display	Sphericity Assumed	42.322	1	42.322	75.423	0.000	0.893
	Greenhouse-Geisser	42.322	1.000	42.322	75.423	0.000	0.893
Error(Display)	Sphericity Assumed	5.05	9	0.561			
	Greenhouse-Geisser	5.05	9.000	0.561			
SetSize	Sphericity Assumed	163.202	2	81.601	37.878	0.000	0.808
	Greenhouse-Geisser	163.202	1.034	157.88	37.878	0.000	0.808
Error(SetSize)	Sphericity Assumed	38.778	18	2.154			
	Greenhouse-Geisser	38.778	9.303	4.168			
DisLvl	Sphericity Assumed	211.09	6	35.182	46.523	0.000	0.838
	Greenhouse-Geisser	211.09	1.711	123.404	46.523	0.000	0.838
Error(DisLvl)	Sphericity Assumed	40.836	54	0.756			
	Greenhouse-Geisser	40.836	15.395	2.653			
Display * SetSize	Sphericity Assumed	3.987	2	1.993	10.601	0.001	0.541
	Greenhouse-Geisser	3.987	1.838	2.17	10.601	0.001	0.541
Error(Display*SetSize)	Sphericity Assumed	3.385	18	0.188			
	Greenhouse-Geisser	3.385	16.539	0.205			
Display * DisLvl	Sphericity Assumed	2.61	6	0.435	2.257	0.051	0.2
	Greenhouse-Geisser	2.61	3.306	0.789	2.257	0.097	0.2
Error(Display*DisLvl)	Sphericity Assumed	10.408	54	0.193			
	Greenhouse-Geisser	10.408	29.757	0.35			
SetSize * DisLvl	Sphericity Assumed	28.791	12	2.399	7.782	0.000	0.464
	Greenhouse-Geisser	28.791	2.763	10.422	7.782	0.001	0.464
Error(SetSize*DisLvl)	Sphericity Assumed	33.297	108	0.308			
	Greenhouse-Geisser	33.297	24.863	1.339			
Display * SetSize * DisLvl	Sphericity Assumed	2.819	12	0.235	1.157	0.324	0.114
	Greenhouse-Geisser	2.819	4.152	0.679	1.157	0.346	0.114
Error(Display*SetSize*DisLvl)	Sphericity Assumed	21.937	108	0.203			
	Greenhouse-Geisser	21.937	37.367	0.587			

The main effects plots are shown in Figure 21. In Figure 21a, we see expected results with RT increasing as set size increases with a strong correlation (r=0.996). Due

to the few number of data points, a T-test was not calculated to test the significance of the relationship. However, as discussed in Ch II, the set size and reaction time correlation is widely researched and accepted as significant in the community. In Figure 21b, CDU resulted in approximately 0.63 sec (24%) slower RT than MFD, which supports the premise of a search cost associated with larger displays. Figures 21c and d show distractor level is correlated with reaction time. After converting both variables to a ranked scale, a Spearman correlation coefficient was calculated to be (ρ=1.000), indicating perfect correlation. A T-test was then performed to test significance resulting in a p-value of 0.000, confirming a significant correlation.

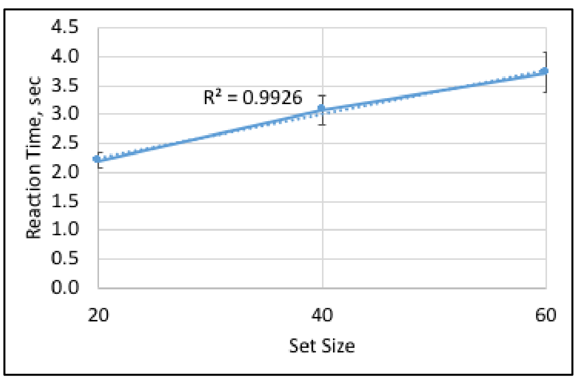


Figure 21a – TA Set Size Main Effect

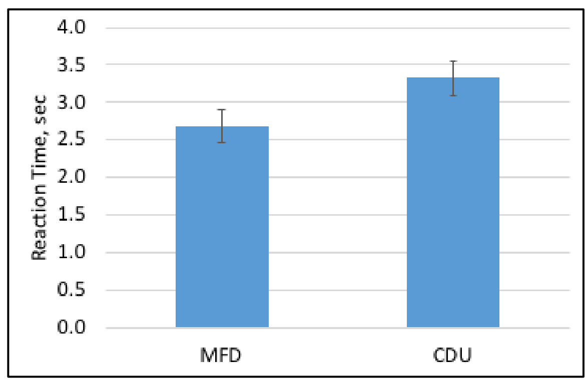


Figure 21b – TA Display Main Effect

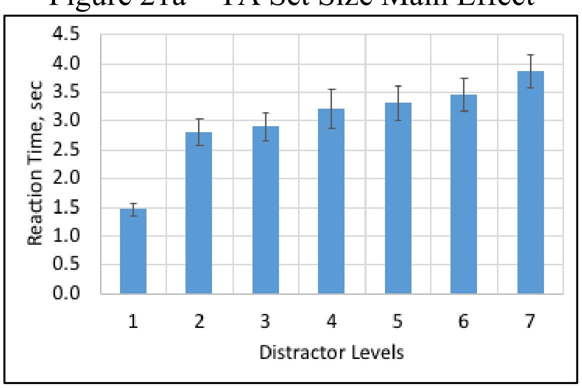


Figure 21c – TA Distractor Level Main Effect

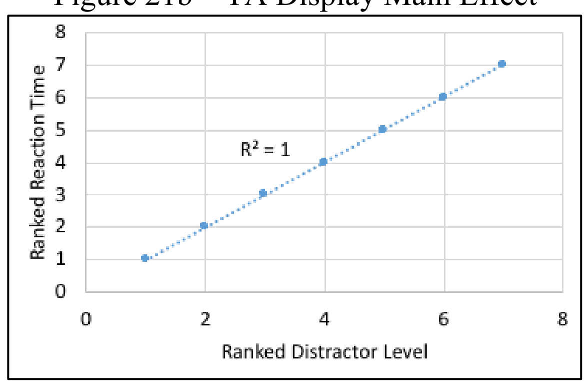


Figure 21d – TP Ranked Distractor Level Main Effect

Plotting the reaction times versus set size allows for a comparison of the search efficiencies through examination of the slope. Figure 22 below shows the moderate slopes of both displays are nearly the same. Since symbol size and distractor levels were identical between the two displays, the search tasks were the same. Therefore, we should expect that both displays should have nearly the same slope. The search cost of about 0.63 seconds from scanning a larger display is shown in the offset between the two. The time to search on the MFD was on average 0.075 seconds per item compared to 0.092 seconds per item on the CDU. Like the target-present condition, the search cost is likely due to the increased number of fixations required to search the larger area.

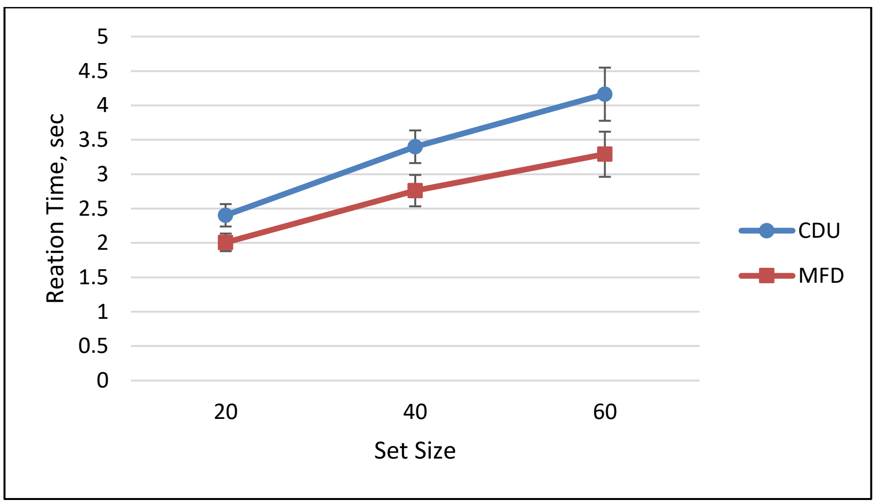


Figure 22 – Target-Absent Efficiency by Display

A Bonferroni Pairwise comparison was performed to compare the levels within the main effects. Levels within display and set size were all found to be statistically different than each other with at least ($p <= 0.005$). Distractor levels were found to have more mixed results. Table 15 summarizes those findings and their associated p-values. The green cells indicate the distractor levels which have statistically different reaction time from each other. The red cells indicate reaction times that are not statistically

different. Unlike target-present, levels 6 and 7 are statistically different from each other. In target-absent, reaction times for levels 5, 6, and 7 are statistically different from each other. In terms of scene design, Level 6 was the same as level 5 with the addition of the vector lines on 50% of the symbols. Level 7 was the same as level 6 with the addition of the number labels on 50% of the vector lines.

Table 17 – Bonferroni Pairwise Distractor Level Comparison TA

Level	1	2	3	4	5	6	7
1		0.003	0.002	0.006	0.001	0.000	0.000
2			1.000	0.256	0.050	0.002	0.000
3				0.409	0.039*	0.001	0.000
4					1.000	0.536	0.004
5						1.000*	0.007
6							0.004*
7							

*Different than target-present

Plotting the efficiency plot by distractor level provides some insights into the how the distractor combinations affect search efficiency. Figures 23 and 24 show the results for CDU and MFD, respectively. Level 1 for both displays had the flattest slope, indicating this was a nearly parallel, highly efficient search. As the distractor levels increase, there appears to be an increase in slope as more serial search is mixed into the search task. Levels 6 or 7 are the most inefficient search tasks for both displays. These levels contained the largest and second largest number of unique symbols, at 5 and 7 respectively.

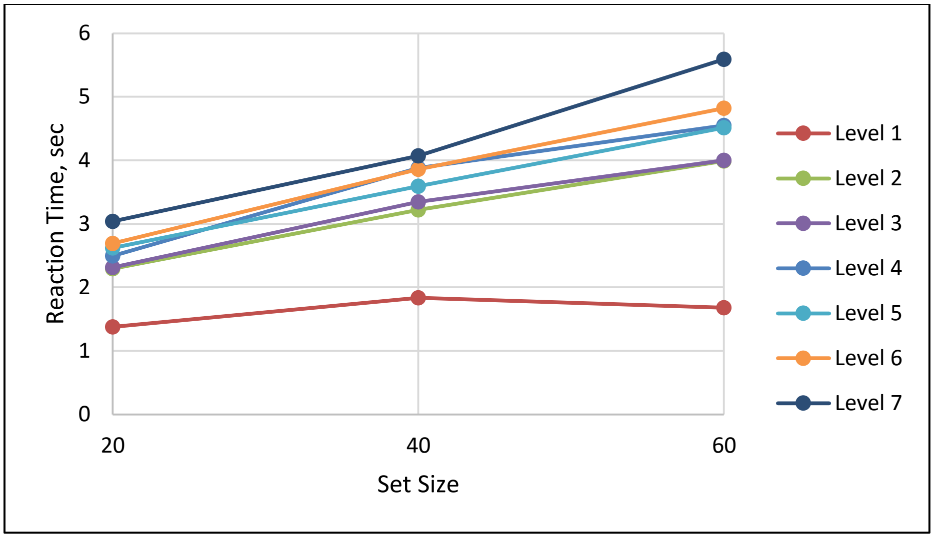


Figure 23 – CDU Target-Absent by Distractor Level

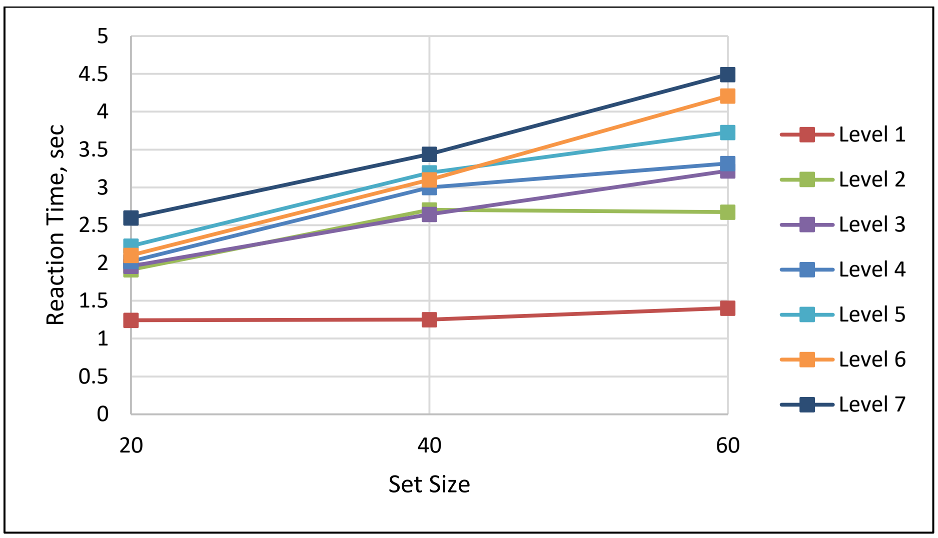


Figure 24 – MFD Target-Absent by Distractor Level

Symbol complexity appears to have a greater effect for the large set size. Figure 25 shows a comparison between distractor level 5 and level 7. Level 7 distractors were the same as level 5 but with vector lines and labels added on 50% of the symbols. Results in Figure 25 show there is a larger difference in reaction time between the two displays at the set size of 60, compared to the smaller set sizes.

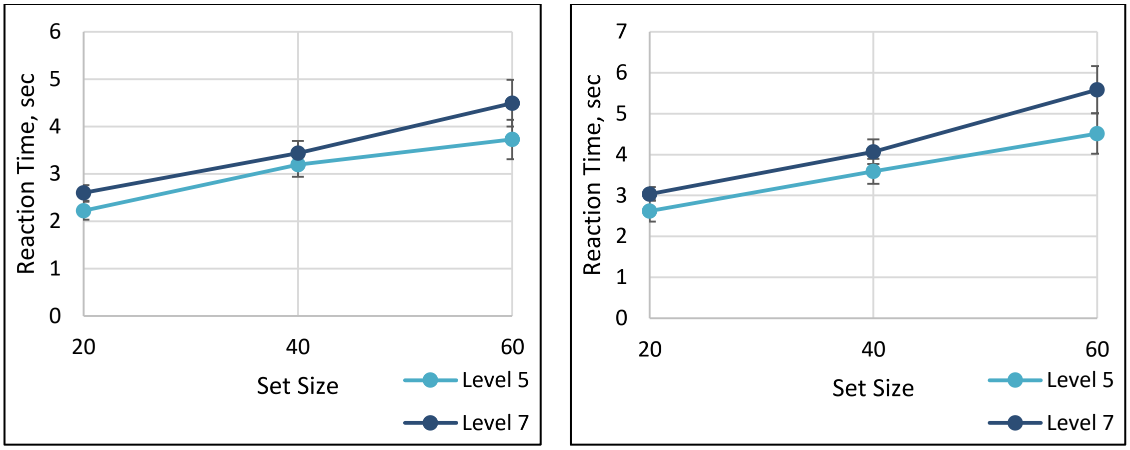


Figure 25 – Distractor Level 5 & 7 on CDU (left) & MFD (right) Target-Absent

A single ANOVA was performed on the two main factor interactions to determine the interaction of factors at their lower levels. The adjusted ANOVA tables for the CDU*Set size and MFD*Set size level are shown in Table 18 below. The interaction plot is shown below in Figure 26. The display*set size interaction was analyzed by holding display constant and averaging across the distractor levels. Set size levels were found to be significant and positively correlated with reaction time for both displays.

Table 18 – Display*Set size Adjusted ANOVA Target-Absent

CDU							
Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
SetSize	Sphericity Assumed	15.575	2	7.788	35.786	0.000	0.799
	Greenhouse-Geisser	15.575	1.087	14.326	35.786	0.000	0.799
Error(SetSize)	Sphericity Assumed	3.917	18	0.218			
	Greenhouse-Geisser	3.917	9.785	0.400			
MFD							
SetSize	Sphericity Assumed	8.309	2	4.155	35.505	0.000	0.798
	Greenhouse-Geisser	8.309	1.066	7.797	35.505	0.000	0.798
Error(SetSize)	Sphericity Assumed	2.106	18	0.117			
	Greenhouse-Geisser	2.106	9.592	0.220			

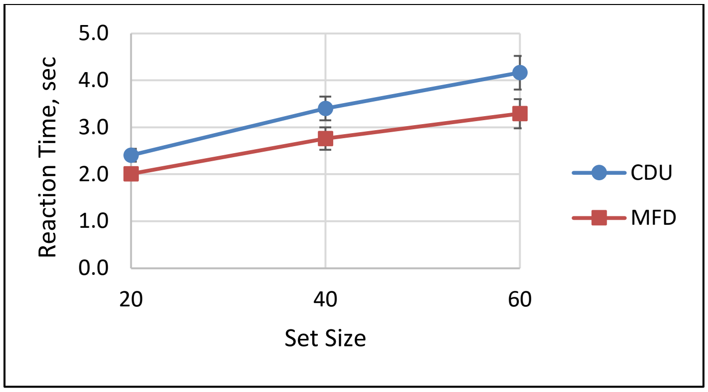


Figure 26 – Display*Set size Interaction Plot Target-Absent

A Bonferroni Pairwise comparison was performed to compare the set size levels within the displays. Table 19 shows the results of the comparison. The green cells indicate the distractor levels which have statistically different reaction time from each other. Set size within CDU and MFD size were all found to be statistically different than each other with at least (p <= 0.005).

Table 19 – Bonferroni Pairwise Set Size Comparison CDU (left) and MFD (right)

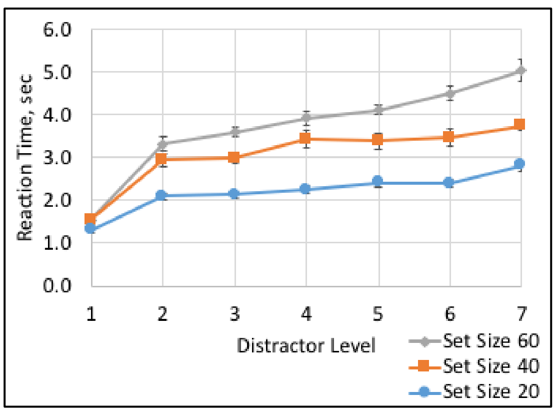
Set Size	20	40	60
20		0.000	0.001
40			0.003
60			

Set Size	20	40	60
20		0.001	0.001
40			0.000
60			

The interaction of set size*distractor level was analyzed by holding set size constant and averaging across the display types. The adjusted ANOVA tables for the set size 20, 40, 60*distractor levels are shown in Table 20 below. Distractor levels were found to be significant within each set size. All three set sizes had significant positive correlations with reaction time. Correlations and significance values are shown in Table 21.

Table 20 – Set Size*Distractor Level Adjusted ANOVA Target-Absent

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Set Size 20							
DisLvl	Sphericity Assumed	12.74	6	2.123	35.551	0.000	0.798
	Greenhouse-Geisser	12.74	2.672	4.767	35.551	0.000	0.798
Error(DisLvl)	Sphericity Assumed	3.225	54	0.06			
	Greenhouse-Geisser	3.225	24.052	0.134			
Set Size 40							
DisLvl	Sphericity Assumed	32.275	6	5.379	23.356	0.000	0.722
	Greenhouse-Geisser	32.275	2.379	13.564	23.356	0.000	0.722
Error(DisLvl)	Sphericity Assumed	12.437	54	0.23			
	Greenhouse-Geisser	12.437	21.415	0.581			
Set Size 60							
DisLvl	Sphericity Assumed	74.926	6	12.488	31.504	0.000	0.778
	Greenhouse-Geisser	74.926	1.777	42.16	31.504	0.000	0.778
Error(DisLvl)	Sphericity Assumed	21.405	54	0.396			
	Greenhouse-Geisser	21.405	15.995	1.338			



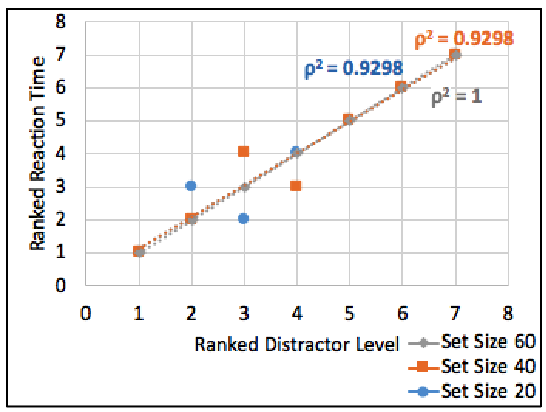


Figure 27 – Set Size*Distractor Interaction Plot (left) & Ranked (right) Target-Absent

Table 21 – Set Size*Distractor Interaction Correlation & Significance

Set Size	Spearman Correlation, ρ	T-test (p-value)
20	0.964	0.000
40	0.964	0.000
60	1.000	0.000

A Bonferroni Pairwise comparison was performed to compare the distractor levels within the three Set Sizes. Tables 19, 20, & 21 summarizes those findings and their associated p-values. The green cells indicate the distractor levels which are statistically different from each other. The red cells are not statistically different. Distractor levels within the displays were found to be mostly not different from each other in set size 20 and 40. Like the interaction in the target-present condition, the differences between the distractor levels are more pronounced at the largest set size.

Table 22 – Bonferroni Pairwise Distractor Level Comparison Set Size 20 TA

Level	1	2	3	4	5	6	7
1		0.004	0.001	0.000	0.002	0.000	0.000
2			1.000	1.000	0.016	0.303	0.021
3				0.561	0.574	0.279	0.003
4					1.000	1.000	0.028
5						1.000	0.794
6							0.097
7							

Table 23 – Bonferroni Pairwise Distractor Level Comparison Set Size 40 TA

Level	1	2	3	4	5	6	7
1		0.017	0.015	0.016	0.001	0.004	0.000
2			1.000	0.646	0.778	0.232	0.142
3				0.494	0.207	0.297	0.028
4					1.000	1.000	1.000
5						1.000	0.205
6							1.000
7							

Table 24 – Bonferroni Pairwise Distractor Level Comparison Set Size 60 TA

Level	1	2	3	4	5	6	7
1		0.002	0.003	0.024	0.003	0.001	0.001
2			0.637	1.000	0.145	0.009	0.005
3				1.000	0.160	0.007	0.001
4					1.000	0.662	0.004
5						1.000	0.025
6							0.389
7							

V. Conclusions and Recommendations

Discussion

The ability to engage our enemy at longer distances has many benefits such as more time and space to make decisions in conflict. The increased range and sensitivity of our sensors enabling this type of engagement brings more data on-board to potentially be displayed to the operator. Fifth generation avionics, with fusion algorithms and large displays, were subsequently developed to handle the increased data load. However, the ability to make decisions as quickly as possible in fighter aircraft is still important and relevant. Ample time and space to make decisions is not a forgone conclusion. Therefore, we should be aware of any impacts to human performance associated with large displays of data in the cockpit. Due to the time-constrained environment of flying fighter aircraft, any impact would be non-trivial and worthy of consideration in acquisition and employment decisions.

The search cost associated with performing a mixed parallel and serial visual search task with a larger display was successfully demonstrated with significance in this study. Using the CDU (17° x 17°) resulted in a 0.24 second (18%) slower search time in the target-present condition and 0.63 second (24%) slower search time in the target-absent condition compared to using the MFD (8° x 8°). The increase in search time is likely due to an increase in the number of fixations and saccades in the visual scan to cover the larger area. These findings support the conclusion that pilots can attend to the same amount of information faster on the MFD versus the CDU. However, this is one aspect of the human performance trade-space that should be considered in acquisition and employment of our aircraft. For example, in the F-16, upgrading to the CDU provides

114% more-pixel area than the MFD. In situations where there is ample time and space to make decisions in the cockpit, a pilot may prefer the extra pixel space and decreased probability of target overlap. However, in an Air-to-Air engagement situation, a pilot may be better suited using the MFDs because they can attend to the same sensor area 18-24% faster than using the CDU. Recommendations on how to use this information in acquisition and employment is discussed next.

Recommendations for Acquisitions & Operators

In most instances, this author believes that the CDU would be the display of preference for the F-16 pilot community. However, program offices and operators should be cognizant of the search cost in terms of human performance associated with using larger displays. In situations when larger display area does not provide an advantage to the pilot, then smaller display sizes should be used to facilitate faster decision-making. Programming default displays in certain aircraft modes would be a feasible solution to prevent pilots from using a display size that puts them at a disadvantage. Logic could also be used to automatically switch to a particular display if targets appear within a certain distance or density in the battlespace. If the larger display is still a preference for other reasons, then a more focused view on the large display using a smaller display area is also a feasible solution.

Recommendations for Future Research

The recommendations for future research revolve around developing better models for human performance in the cockpit. This study identified a search cost associated with larger displays and proposed a theoretical explanation for that cost. It is

likely that the increased search cost is due to an increase in fixation and saccadic phenomena required to cover the larger area. However, gaze or eye tracking data was not collected to test this hypothesis. Therefore, the next step in this study would be to integrate the experiment with an eye-tracking system to collect visual scan data of the participants performing the visual search tasks. Assuming the hypothesis is correct, the comparison of fixation and saccadic behavior would provide insight in the phenomena behind the search cost associated with display size. With this data, more accurate conclusions could be made with regards to human performance using two different display sizes. Without a more realistic simulator and accurate radar imagery, conclusions from this data would still be restricted to comparisons between display sizes.

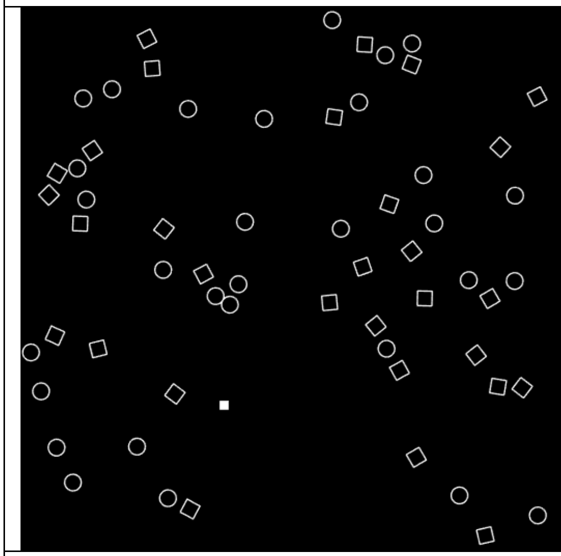
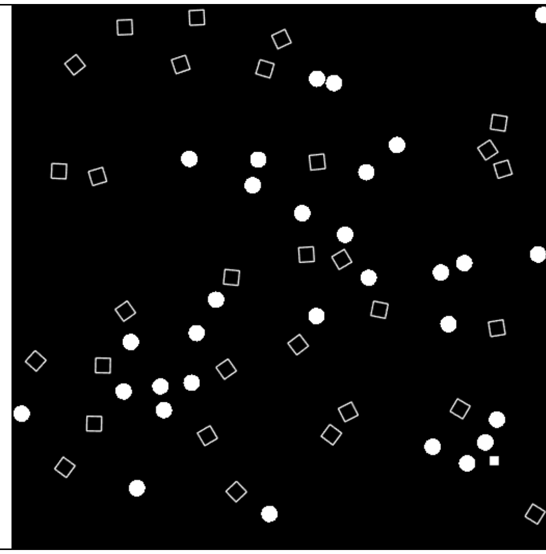
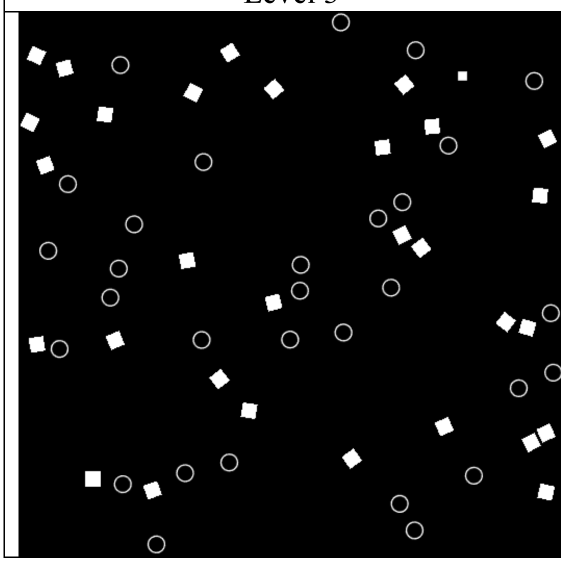
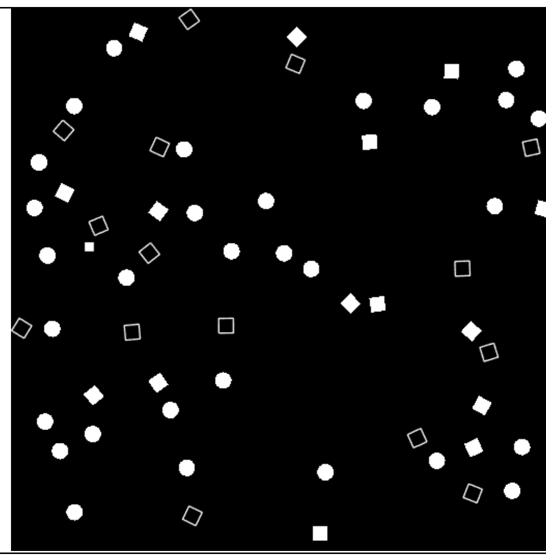
A more accurate measurement of human performance in the cockpit would require using a representative F-16 simulator and realistic radar simulations. Scenes could be developed to represent a variety of operational scenarios in which a key piece of information is sought by the pilot. Unlike in this study, the target information could be varied in addition to the distractors in operational scenarios. The resulting experiment would measure the full-spectrum of human performance across a wide range of operationally representative visual search tasks.

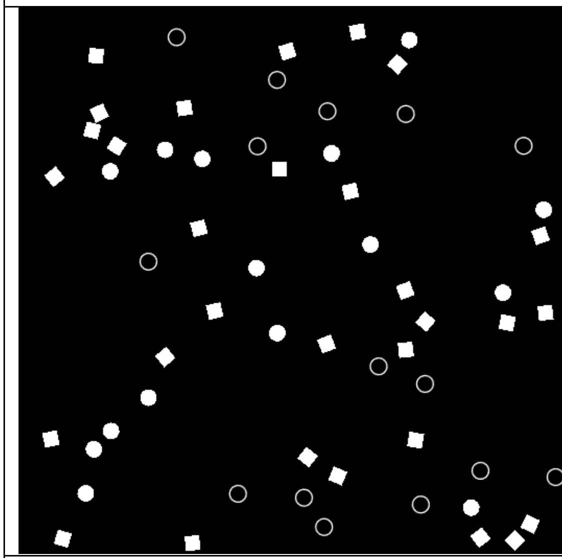
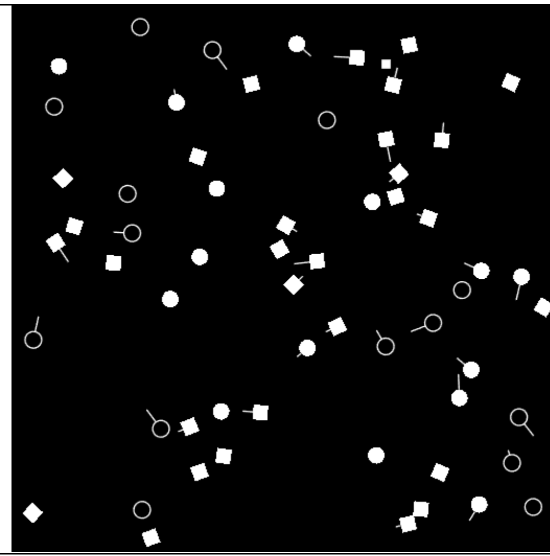
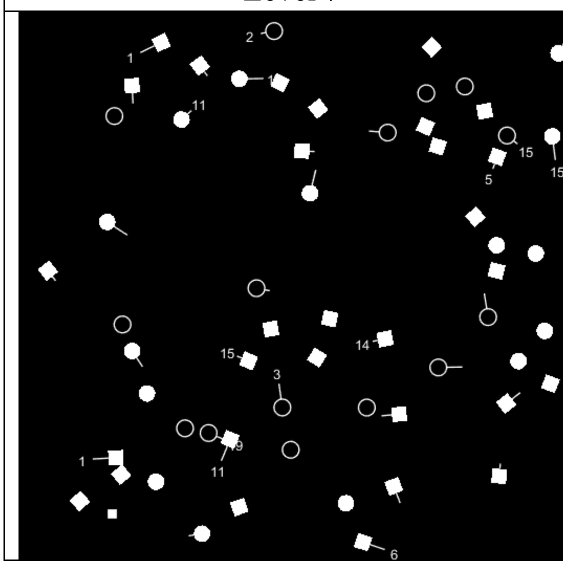
A similar type of experiment could also be performed to determine how the modification of radar symbology effects visual search performance. Symbol modifications are often made without understanding the impact to human performance. Results from this study suggested that modifications of symbology to enable them to carry more information create more complex scenes, which can increase search times. Overall, there was a general trend of increasing search time as distractor levels increased.

A more deliberate test of this phenomena would be helpful to designers who are looking for more ways to communicate information to pilots in the cockpit.

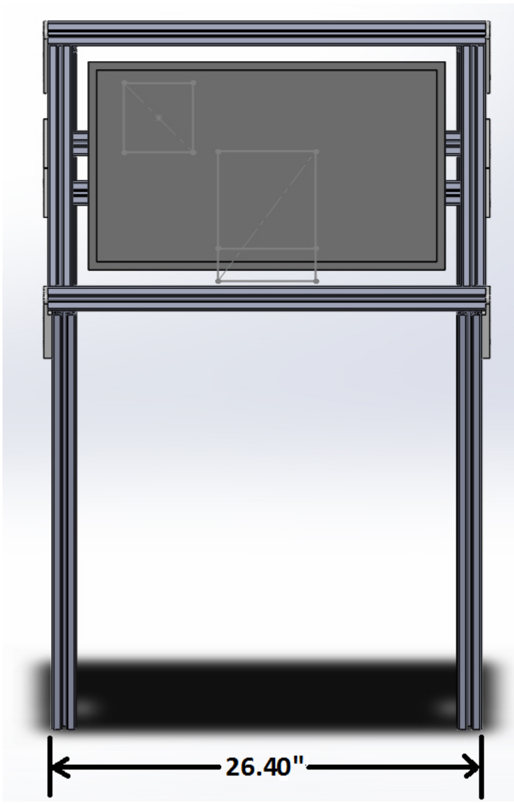
Research has also shown that target eccentricity is a factor in whether certain cues are salient in our periphery or parafovea (Loschky et al., 2005). Determining what types of cues are salient at various locations in the display would be helpful for designers to avoid costly serial search in large displays. Using larger displays increases the probability that targets will be located further in our periphery, increasing the chances of serial search. Therefore, the saliency of attributes in our periphery is more critical in larger displays. This data could help interface designers decide which symbol characteristics facilitate better human-performance. Ultimately, incorporating this type of data into a human performance model would provide a better prediction of human-performance by scene characteristics. It may be possible to accurately predict human performance in terms of search time knowing the size and content of the radar scene together.

Appendix A. Distractor Level Examples (Target-Present Condition)

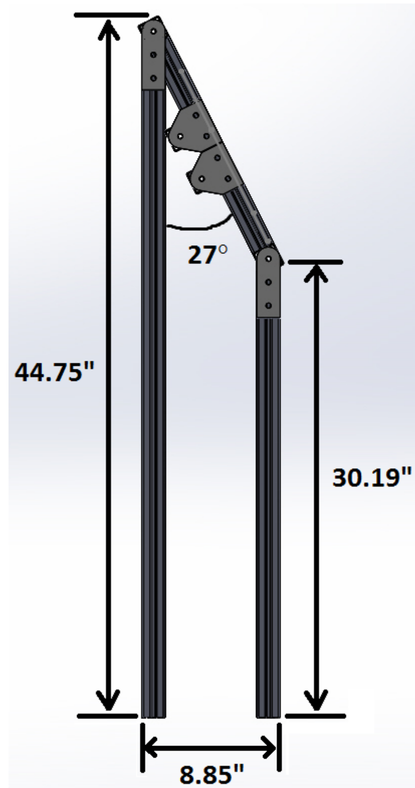
Level 1	Level 2
	
Level 3	Level 4
	

Level 5	Level 6
	
Level 7	
	

Appendix B. Monitor Stand Dimensions

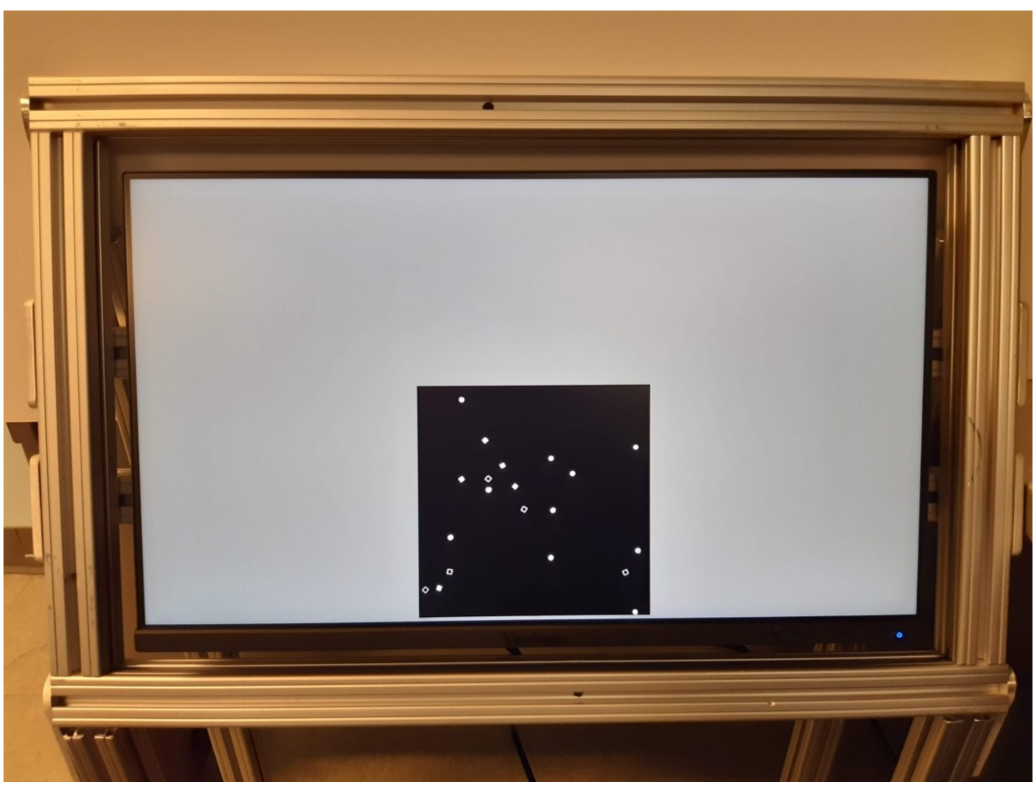


Front View



Side View

Appendix C. CDU and MFD Queuing Sequence




1. Next scene
will appear here
2. Press the
arrow to continue

Appendix D. Visual Search Pre-Experiment Questionnaire

How many hours of sleep do you get on average? ________________________________

How many hours of sleep did you have last night?________________________________

How would you characterize your sleep last night?
Circle one choice: Very Poor, Poor, Fair, Good, Very Good

Did you consume any products with caffeine today?
Circle one choice: yes or no
If yes:
What product(s) did you consume?
____
When did last consume this product?
____
Approximately how much (mg / ounces / cups) of this product have you consumed today?
____
Approximately how much (mg / ounces / cups) of this product do you typically consume?
____

Do you have eye-fatigue or eye-strain problems? Yes or No

Do you use moisturizing eye-drops regularly? Yes or No

Do you have corrected vision? Yes (circle glasses/contacts) or No
If yes: Are you near or far-sighted? Yes (circle near/far) or No
If yes: Are you wearing them now? Yes or No

Do you have any reason(s) to believe that your ability to accomplish visual search tasks during this study today would be abnormal (distracted, overly tired, hungry, stressed, injured, jittery)?
____
____
____

If yes:
Do you still want to participate in the study today? Circle one choice: Yes / No
If no:
Would you like to reschedule participation for another day?
____

Appendix E. Visual Search Pre-Experiment Questionnaire Results

Subject	Hours of sleep on avg	Hours of sleep last night	Quality of sleep	Caffeine Today	Eye-fatigue or strain history	Moisturizing drops use	Corrected Vision	Near/far-sighted	Wearing
1	7	7.5	Good	Yes/normal	No	No	Yes	Near	Contacts
2	7	6.5	Good	Yes/normal	No	No	No	N/A	N/A
3	8	8	Good	Yes/normal	No	No	Yes	Near	Glasses
4	8	8	Good	No/normal	No	No	Yes	Near	Glasses
5	8	7	Fair	No/normal	No	No	Yes	Near	Contacts
6	8	8	Fair	Yes/normal	No	No	No	N/A	N/A
7	8	6	Very Good	No/normal	No	No	No	N/A	N/A
8	7	8	Good	Yes/normal	No	No	No	N/A	N/A
9	7	8	Good	No/normal	No	No	No	N/A	N/A
10	8	6	Blank	Yes/normal	No	Yes	No	N/A	N/A
			Average hours of sleep previous night			7.3			
			Percent w/ Corrected Vision			40%			

Appendix F. Visual Search Post-Experiment Questionnaire

Computer experience:

What sort of electronic devices do you use?

Circle all choices that apply:

Personal computer/Desktop/Laptop
TV/Game Console
Smartphone/Tablet
Enterprise Server
Other, ____
Prefer not to answer

How often do you use electronic devices?

Circle one choice: Daily, A few times a week, Once a week, Never, Prefer not to answer

How often do you play video games?

Circle one choice: Daily, A few times a week, Once a week, Never, Prefer not to answer

If yes: Type of video game? Candy Crush-like, First Person, Other:_______________

Do you use electronic devices in your job?

Circle one choice: Yes, No, Prefer not to answer

Age: 20 and under 21-30 31-40 41-50 51-60+ Prefer not to answer

Are you male or female? Male___ Female____ Prefer not to answer ____

What's your highest education level?

A. Lower than high school
B. Graduated from high school
C. Some college, no degree
D. Associate's Degree
E. Bachelor's Degree
F. Master's degree
G. Ph.D. degree
Prefer not to answer

Have you had pilot training or been trained in the scanning of instruments? Yes or No

If yes: What training? (PPL, FlightSim, etc)

____

____

Performance:

Did you have trouble seeing the display or target? Yes or No Prefer not to answer

If yes: Why do you believe you had trouble?

____

____

Did you feel that you lost focus at all during the experiment? Yes or No Prefer not to answer

If yes: When did you lose focus? First half or Second half

Any more information on when?

____

____

What did you find difficult or easy about this task?

____

____

____

Do you feel like you changed your search or response strategy during the experiment? Yes or No

If yes: When? First half or Second half

Any more information on when?

____

____

Appendix G. Visual Search Post-Experiment Questionnaire Results

Subject	# of Electronic Devices Regularly Used	Electronic device use frequency	Video games		Use in job	Age	Gender	Highest Edu	Recent Flying Training		Trouble seeing the display or target?		Lost focus during experiment		What did you find difficult or easy?	Search strategy change	
1	5	Daily	Never	N/A	Y	31-40	M	BS/BA	N	N/A	Y	Small square would sometimes blend with other shapes, eye fatigue	N	N/A	Difficult to distinguish when many shapes, especially extra lines and numerics, larger screen had impact on eye fatigue having to search larger area	N	N/A
2	7	Daily	Weekly	First-person	Y	21-30	M	BS/BA	Y	Fews hours towards PPL	N	N/A	N	N/A	After extended period of time of staring at screen became difficult to distinguish between circles & squares	Y	Started ruling out empty shapes and using new search patterns to jump to clusters of solid shapes
3	7	Daily	Never	N/A	Y	21-30	M	BS/BA	N	N/A	N	N/A	N	N/A	Easy to remember & manipulate controls. Number labels & target were similar sized so could be confused	Y	I experimented with different scan types: circle, horizontal pass but only did it conciously for a few images
4	7	Daily	Never	N/A	Y	21-30	F	BS/BA	N	N/A	Y	N/A	Y	When there were multiple N targets in a row	When there are multiple where I don't see a target, I lose focus and miss some	Y	2nd half of experiment
5	5	Daily	Monthly	Sudok u	Y	21-30	M	BS/BA	N	N/A	N	N/A	Y	Second half, it was harder to focus my eyes. No loss of concentration	Persistence was the most challenging. Top left view felt easier. It was easier to spot target with peripheral vision than direct	Y	I tried searching with peripheral vision about mid-way through each part
6	5	Daily	Daily	Flight Sim	Y	21-30	M	BS/BA	Y	Glider & aircraft flying at USAFA	N	N/A	Y	I got fatigued towards the end	The target was easy to see when it was by iteself and when there were lots of circles	N	N/A
7	7	Daily	Weekly	All types	Y	21-30	M	BS/BA	Y	USAFA Glider	N	N/A	Y	Whenever there was a lack of targets for a few slides in a row I would start to lose focus and make mistakes/ take too long	Difficult to dtermine if the target was absent. It took me much longer	N	N/A
8	7	Daily	Weekly	Soccer	Y	31-40	M	MS/MA	N	N/A	N	N/A	Y	N info	It will tell me where to focus before I start the task	Y	I took my time so I don't choice the wrong target
9	5	Daily	Weekly	All types	Y	31-40	M	BS/BA	N	N/A	N	N/A	Y	N info	Difficult to stay engaged	N	N/A
10	7	Daily	Daily	MMO/ FPS	Y	31-40	M	MS/MA	N	N/A	N	N/A	N	N/A	Most difficult were absent trials with lots of clutter. Task is also a little fatiguing (didn't greatly affect me though)	N	N/A

Bibliography

Chen, I.-P., Liao, C.-N., & Yeh, S.-H. (2011). Effect of Display Size on Visual Attention. *Perceptual and Motor Skills*, *112*(3), 959–974. https://doi.org/10.2466/22.24.26.pms.112.3.959-974

Findlay, J., & Walker, R. (2012). Human saccadic eye movements. Retrieved from http://www.scholarpedia.org/article/Human_saccadic_eye_movements

Frintrop, S. (2006). *VOCUS A Visual Attention System for Object Detection*. Springer-Verlag Berlin Heidelberg. https://doi.org/10.1007/11682110

Hermelin, S. (2013). Fighter Avionics Part 1. Retrieved from https://www.solohermelin.com/aircraft-systems.html

Kroft, P., & Wickens, C. (2002). Displaying multi-domain graphical database information: An evaluation of scanning, clutter, display size, and user activity. *Information Design Journal*, *11*(1), 44–52. https://doi.org/10.1075/idj.11.1.06kro

Lockheed Martin F-16 Block 70 India Brochure. (2016). Retrieved from https://lockheedmartin.com/content/dam/lockheed/data/aero/documents/F-16_Block_70_Executive_Summary_2017.pdf

Loschky, L., McConkie, G., Yang, J., & Miller, M. (2005). The limits of visual resolution in natural scene viewing. *Visual Cognition*, *12*(6), 1057–1092. https://doi.org/10.1080/13506280444000652

Miller, M. (2019). Human Perception of Color. In *Color in Electronic Display Systems: Advantages of Multi-primary Displays* (1st ed., pp. 19–44). Springer.

Neisser, U. (2014). *Cognitive Psychology: Classic Edition*. *Cognitive Psychology* (Classic). New York: Psychology Press.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., … Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Stelzer, E. M., & Wickens, C. (2006). Pilots strategically compensate for display enlargements in surveillance and flight control tasks. *Human Factors*, *48*(1), 166–181. https://doi.org/10.1518/001872006776412225

Tan, D. S., Czerwinski, M. P., & Robertson, G. G. (2006). Large displays enhance optical flow cues and narrow the gender gap in 3-D virtual navigation. *Human Factors*, *48*(2), 318–333. https://doi.org/10.1518/001872006777724381

Treisman, A., & Gelade, G. (1980). A Feature-Integration Theory of Attention. *Cognitive Psychology*, (12), 97–136. Retrieved from papers3://publication/uuid/69FA9B61-2CE4-4D63-BFCD-F6243419B321

United States of America Department of Defence. (2014). Mil-Std-2525D. *Joint Military Symbology*, (June).

Wickens, C., & Hollands, J. (2000). Engineering Psychology and Human Performance. In *Engineering Psychology and Human Performance* (Third, pp. 69–118). New Jersey: Prentice Hall. https://doi.org/10.4102/sajip.v13i1.457

Wickens, C., Lee, J., Liu, Y., & Becker, S. (2013). *Visual Sensory Systems*. *A Brief Introduction to Human Factors Engineering* (Vol. 2). https://doi.org/10.1111/trf.12177

Wolfe, J. (1993). Guided Search 2.0: The Upgrade. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 1295–1299).

Wolfe, J. (1998). What can 1,000,000 trials tell us about visual search? *Psychological*

Science, *9*(1), 33–39.

Wolfe, J. (2001). Asymmetries in visual search: an introduction. *Perception & Psychophysics*, *63*(3), 381–389. https://doi.org/10.3758/BF03194406

Wolfe, J. (2007). Guided Search 4.0: Current Progress with a Model of Visual Search. *Integrated Models of Cognitive Systems*, 99–120. https://doi.org/10.1093/acprof:oso/9780195189193.003.0008

Wolfe, J. (2010). Visual Search. *Current Biology*, *20*(8), 346–349. https://doi.org/10.1016/j.cub.2010.02.016

Wolfe, J. (2018). Visual Search. In J. Wixted (Ed.), *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* (4th ed., pp. 1–55). https://doi.org/10.1007/978-3-319-47829-6_1514-1

Wolfe, J., & Horowitz, T. (2008). Visual Search. Retrieved from http://www.scholarpedia.org/article/Visual_search

Xu, Y. (2010). The impact of item clustering on visual search: It all depends on the nature of the visual search. *Journal of Vision*, *10*(14), 24–24. https://doi.org/10.1167/10.14.24

Yamani, Y., & McCarley, J. (2011). Visual search asymmetries in heavy clutter: Implications for display design. *Human Factors*, *53*(3), 299–307. https://doi.org/10.1177/0018720811410241

Zehner, G. (2002). Body Size Accommodation in USAF Aircraft. *United States Airforce Research Laboratory*, (January 1998).

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 074-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From – To)*
12-03-2020	Master's Thesis	October 2018 – March 2020

TITLE AND SUBTITLE	
The Impact of Increasing the Size of Aircraft of Radar Displays on Visual Search in the Cockpit	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	**5d. PROJECT NUMBER**
Marsh, Justin R., Captain, USAF	**5e. TASK NUMBER**
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)	**8. PERFORMING ORGANIZATION REPORT NUMBER**
Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 Hobson Way, Building 640 WPAFB OH 45433-8865	AFIT-ENV-MS-20-M-226
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)** AFLCMC/WWM
AFLCMC/WWMA (F-16 Division) 2725 C Street, Bldg 553 Wright Patterson AFB, OH 45433-7424 937-713-6547 and john.silance.3@us.af.mil ATTN: Mr. John Silance	**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRUBTION STATEMENT A. APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Modern aircraft are designed with larger displays while our legacy aircraft are being retrofit with larger cockpit displays to enable display of the increased data. While this modification has been shown to enable improvements in human performance of many cockpit tasks, this effect is often not measured nor fully understood at a more generalizable level. This research outlines an approach to comparing human performance across two display sizes in future F-16 cockpits. The results show that increases in display size can increase search times under some circumstances even when the displays include a large number of tracks, actually reducing human performance.

15. SUBJECT TERMS
Visual search, cockpit displays, human performance, salience, display of sensor data, display size

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Michael E Miller, AFIT/ENV
U	U	U	UU	88	**19b. TELEPHONE NUMBER** *(Include area code)* (937) 255-3636, x4651 Michael.Miller@afit.edu

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18