Managing Multiple Uninhabited Aerial Vehicles: Changes in Number of Vehicles and Type of Target Symbology

Jeremy T. Nelson, Austen T. Lefebvre, and Terence S. Andre
United States Air Force Academy
Colorado Springs, Colorado
jeremy.nelson@wpafb.af.mil, austen.lefebvre@usafa.af.mil, terence.andre@usafa.af.mil

ABSTRACT

This study attempted to more fully develop the theoretical framework for a single operator monitoring several uninhabited aerial vehicles (UAVs). Twenty-four subjects from the United States Air Force Academy participated in a study to determine the performance effects of managing multiple UAVs. Subjects monitored three and five UAVs simultaneously using the Multi-Modal Immersive Intelligent Interface for Remote Operation (MIIIRO) synthetic task environment. In addition to workload and performance measures, we examined the effectiveness of two types of target symbology (a stylized symbol set and the MIL-STD-2525B symbols) while controlling the UAVs in a simulated task. Results showed five UAVs created a significant degradation in performance as well as significant increase in workload (subjective and objective). Also, MIL-STD-2525B symbols were shown to have significantly better recall than stylized icons as implemented in the MIIIRO synthetic task environment.

ABOUT THE AUTHORS

2Lt Jeremy T. Nelson recently graduated from the United States Air Force Academy with a major in Human Factors. He has worked on numerous research projects in many different areas ranging from mentoring programs to Predator UAV landing performance. He also worked with the Air Force Research Laboratory to help develop a photopic imaging system during a summer research program. While at the Air Force Academy, 2Lt Nelson served on the Institutional Review Board and as Cadet-in-Charge of the USAFA radio station and the Psychology Club. 2Lt Nelson is currently assigned to Wright-Patterson Air Force Base in Dayton, Ohio as a crew systems engineer.

2Lt Austen T. Lefebvre recently graduated from the United States Air Force Academy with a major in Human Factors. During his time at the Air Force Academy he was a cadet squadron commander and performed research at the Air Force Research Laboratory in Mesa during a summer research program. 2Lt Lefebvre is currently assigned to the Air Force Academy before he begins his official assignment at Sheppard AFB, Texas where he will be attending undergraduate pilot training.

Lt Col Terence Andre is the Deputy Department Head for Cadet Operations in the Department of Behavioral Sciences and Leadership at the United States Air Force Academy. He directs the department's $1.2M laboratory infrastructure and research program for faculty and cadets. Lt Col Andre is an assistant professor in the department, teaching courses in human factors, system design, and human-computer interaction. Lt Col Andre graduated from the Air Force Academy in 1987 and received his Masters in Industrial Engineering from Cal Poly and his PhD in Industrial and Systems Engineering from Virginia Tech.

Managing Multiple Uninhabited Aerial Vehicles: Changes in Number of Vehicles and Type of Target Symbology

Jeremy T. Nelson, Austen T. Lefebvre, and Terence S. Andre
United States Air Force Academy
Colorado Springs, Colorado
jeremy.nelson@wpafb.af.mil, austen.lefebvre@usafa.af.mil, terence.andre@usafa.af.mil

INTRODUCTION

As technology increases, Uninhabited Aerial Vehicles (UAVs) are able to operate more autonomously, requiring little direct operator control. Consequently, operators will have a new role of monitoring and maintaining supervisory control over multiple semi-autonomous UAVs (Draper, Calhoun, & Ruff, 2003). In order to keep the operator "in the loop" for optimal situation awareness, workload, and decision-making, certain aspects of this management task must be examined. This can be thoroughly examined using the Multi-Modal Immersive Intelligent Interface for Remote Operation (MIIIRO), which is a synthetic task environment for studying display formats and interface design when controlling multiple UAVs.

In previous research, Hitt, Brinton, and Walton (2003) discuss the usefulness of the surface management system used by air traffic controllers and traffic managers citing the system's success in reaching the goal of providing information and decision support for tactical and strategic air traffic control tasks to increase situational awareness of operations through knowledge of current and future demands. This is similar to the goal of the MIIIRO, as the operator acts much like an air traffic controller and traffic manager when managing the UAVs. Thus, the MIIIRO system is a good platform to properly assess this probable future scenario.

Addressing the number of UAVs a single operator can monitor was originally tackled by Draper et al. (2003) in their testbed simulation study using the MIIIRO. However, using two or four UAVs showed no significant differences, except when combined with other variables. Their study suggested an expanded experiment to include monitoring of three and five UAVs. This would allow for a gamut of information and a better conceptual understanding of the human operator limits.

Past research also shows that different target symbologies produce different levels of performance (Havig, Jenkins, & Geiselman, 2002). As reported by Havig and colleagues (2002), the military standard symbology often performs better than many other types of symbology, especially in straightforward tasks (e.g., looking at a computer monitor directly in front of you). As the current MIIIRO symbology has not been evaluated as to its effectiveness, manipulation of the MIIIRO default and military standard symbologies may prove beneficial. Other studies have shown that more stylized icons that more closely resemble the objects they represent have exhibited greater effectiveness (Sanders & McCormick 1993).

Our expectation in this experiment is that an operator's ability to monitor and respond to UAV mission events will differ as a function of number of UAVs. Specifically, by increasing the number of UAVs there will be a significant degradation in performance as workload increases. Also, it is expected that the operator's performance will improve with the use of stylized icons over that of the MIL-STD-2525B. This is because the stylized icon set should naturally map to the objects they represent, thus allowing for increased recall performance later.

METHOD

Participants

Participants for this study were twenty-four United States Air Force Academy (USAFA) cadets enrolled in the Introduction to Behavioral Sciences for Leaders class. All participants were between the ages of 18 – 22 (22 male, 2 female). Participation was voluntary and participants could receive extra credit for their class by completing a short worksheet after the experiment. Participants came from all classes (freshmen, sophomores, juniors, and seniors).

Materials

Testing was conducted using a Dell 3.2 GHz, Pentium 4 processor with two 19" liquid crystal display (LCD) monitors (see Figure 1). For the synthetic task, the experiment used the MIIIRO workstation from the Synthetic Interface Research for UAV Systems (SIRUS) Laboratory at Wright-Patterson Air Force Base (AFB), Ohio. The two LCDs were placed side-by-side. The left monitor displayed what is called the Tactical Situation Display (TSD) (see Figure 2) and the right monitor displayed an image queue (see Figure 3). The TSD provides the participant with a top-down view of the different UAVs and their routes over a map of the terrain area involved. The image queue displayed the images taken by the UAVs allowing the participant to complete the task of identifying and selecting enemy tanks in the images. Participants completed the required tasks using a keyboard and mouse.

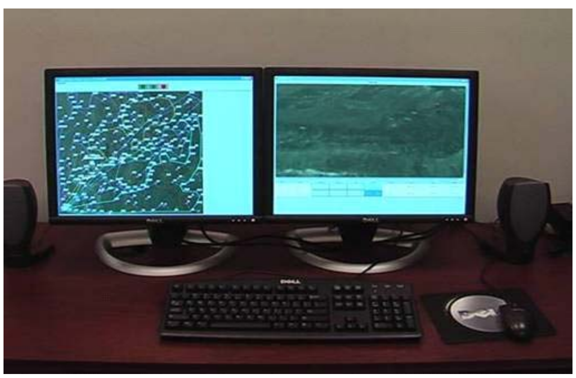

Figure 1. MIIIRO Workstation Arrangement

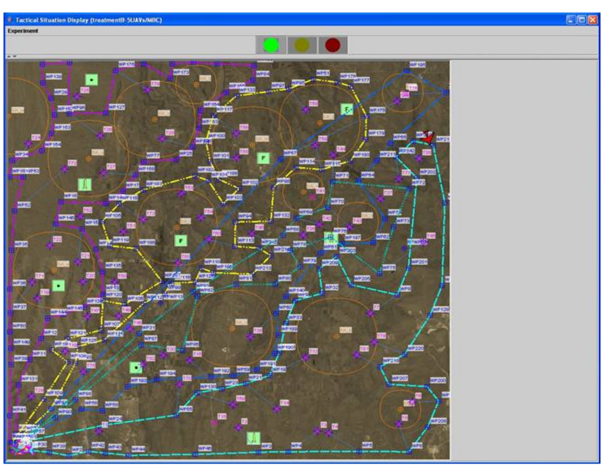

Figure 2. Tactical Situation Display

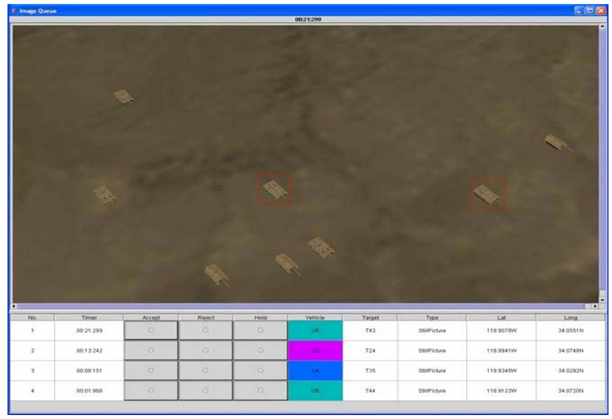

Figure 3. Image Queue View of MIIIRO Workstation

Design and Procedure

A mixed subjects design was utilized for this experiment. The twenty-four participants were randomly assigned to two groups. Each group was exposed to a different type of symbol set, MIL-STD-2525B or Stylized Icons, for the experiment (see Figure 4). Participants read an explanation of the experiment and they were trained extensively in the use of the MIIIRO program. They were first taught how to successfully complete each task and then they completed practice trials on the computer. They were required to achieve perfect performance in all tasks before they were allowed to move on to completing the actual experiment.

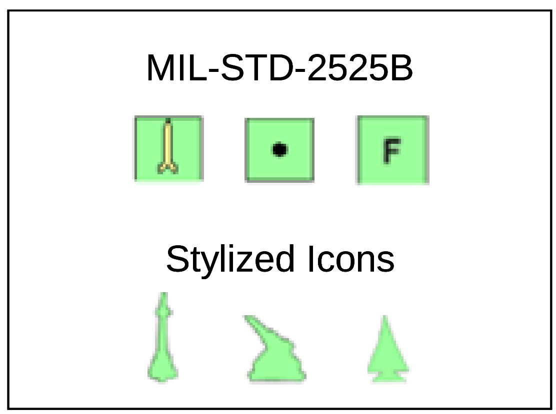

Figure 4. MIL-STD-2525B and Stylized Icons

All participants completed one trial with 3 UAVs and one trial with 5 UAVs, and each group was counterbalanced to decrease order effects. Each trial was 8 minutes long. Participants were required to manage five different tasks while completing each trial. These tasks, in order of priority for the participant to complete them were, identifying unidentified aircraft, overseeing the replan of UAV routes, identifying and selecting target tanks on the image queue, completing task on a Mission Mode Indicator, and counting symbols on the TSD for recall later. The task of identifying unidentified aircraft (UA) required the participant to click on a red icon that represented a UA, and this would prompt a task box to appear where the participant would type "TIGER1" and then hit the "enter" key or click on an "OK" tab on the dialogue box. The replan task required participants to accept or reject a suggested replan for a particular UAV route by the computer. These replans occurred for multiple reasons, and a participant would always accept a replan unless the new route passed through a threat ring or crossed another UAV's route. The image queue task required participants to evaluate an image that one of the UAVs had taken that contained tanks. The participants determined which tanks were enemy tanks and selected them. After they had identified what they thought were all of the enemy tanks, they would accept the image. If there were no enemy tanks in an image, the participants were required to reject the image. The fourth task required participants to attend to the mission mode indicator (MMI), three lights (green, yellow, and red light) across the top of the TSD. If the light on the MMI was green, participants did not need to take any action. However, the light would occasionally change from green to yellow. Once this occurred, the participant had 10 seconds to click on the light and type a number sequence identical to one displayed above it in a box that appears. The symbol recall task required participants to count the number of symbols (MIL-STD-2525B or Stylized Icons) they saw on the TSD and what they were. There were three different types of possibilities for symbols to represent: fighter aircraft, artillery pieces, or missiles. After the participant completed the trial, they would then recall the number of each type of symbols they saw. All of these tasks were designed to simulate the level of workload expected of an operator in a multiple UAV management environment.

At the end of each trial, participants were asked to recall the number and type of symbols they saw, answer a three question questionnaire that asked them about their perceived situational awareness, perceived task difficulty, and perceived task performance, and they then filled out a modified Cooper-Harper Rating Scale. Once participants had completed both trials and post-trial questions, they answered four questions about the symbol set used in the experiment.

The MIIIRO program has multiple measures of a participant's performance in the tasks described above with the exception of symbol recall. Because of the multiple measures, there were four measures that were identified as providing the best overall indicators of performance. These measures were the percent of enemy tanks targeted, the number of MMI events completed by the participant, the average time it took a participant to reject or accept a suggested replan by the computer, and the average time it took a participant to identify an unidentified aircraft.

RESULTS

The MIIIRO program recorded over 70 different performance measures. Four of the measures provided adequate understanding of the workload and overall performance by participants. The measures that were used included the average action time for replan events and unidentified aircraft events, targets that were prosecuted, and the number of mission mode indicator tasks the participant completed. The subjective measures evaluated were the participant's perceived task difficulty, performance, situational awareness, and the workload.

Using a multivariate analysis of variance (MANOVA), there was a significant difference among the UAV conditions for the performance measures. Based on this multivariate test, we conducted individual ANOVAs of the main effects.

Enemy Targets Prosecuted

Mean values for the percent of enemy targets prosecuted for 3 and 5 UAVs (n = 24) are shown in Table 1. Performance on enemy targets prosecuted was not significantly different for 3 versus 5 UAVs [$F(1,23) = 0.21$, $p > .10$].

Table 1. Objective Performance Data from MIIIRO Workstation

Performance Measure	3 UAV Condition	5 UAV Condition
Enemy Targets Prosecuted (% identified correctly)	92	92
Mission Mode Indicator (% accomplished correctly)	91	71
Replan Average Action Time (seconds)	6.10	7.70
Unidentified Aircraft Identification - Time to Respond (seconds)	8.48	11.87

Mission Mode Indicator

Mean values for the number of mission mode indicator tasks completed for 3 and 5 UAVs (n = 24) are shown in Table 1. Performance on the mission mode indicator tasks was significantly better for 3 versus 5 UAVs [$F(1,23) = 37.50$, $p < .001$].

Replan Average Action Time

Mean values for the replan average action time for the 3 and 5 UAVs (n = 24) are shown in Table 1. Performance on the replan average action time tasks was significantly better for 3 versus 5 UAVs [$F(1,23) = 15.38$, $p < .001$].

Unidentified Aircraft Identification

Mean values for the percentage of unidentified aircraft identified for 3 and 5 UAVs (n = 24) are shown in Table 1. Performance on the unidentified aircraft task was significantly better for 3 UAVs [$F(1,23) = 49.23$, $p < .001$].

Subjective Data

Mean values for the subjective data results for the 3 and 5 UAVs (n = 24) can be found in Table 2. There was a significant difference between 3 and 5 UAVs for all subjective measures. The participants perceived significantly better situational awareness when managing 3 UAVs [$F(1,23) = 33.64$, $p < .001$]. Participants perceived task difficulty for 5 UAVs to be significantly greater than 3 UAVs [$F(1,23) = 33.45$, $p < .001$]. Participants perceived their performance to be significantly better for 3 UAVs [$F(1,23) = 44.51$, $p < .001$]. Participants also perceived the workload experienced for 5 UAVs to be significantly greater [$F(1,23) = 49.72$, $p < .001$].

Table 2. Participant Ratings of Performance and Workload

Question Item	3 UAV Condition	5 UAV Condition
Perceived Situational Awareness	5.67	4.46
Perceived Task Difficulty	4.04	5.79
Perceived Task Performance	5.08	3.63
Perceived Workload	4.17	6.96

Symbol Data

The recall data for symbols were analyzed by adding up the total number of errors per trial per participant. This allowed for comparisons between the two symbol groups. One participant's recall data was found to be an extreme outlier compared to the rest of the data (7 and 20 standard deviations from the sample mean). For this reason, the data was dropped.

The final data was a comparison of the two symbol sets used in the experiment. The mean number of recall errors and standard deviation for the MIL-STD-2525B and stylized icon groups are shown in Table 3. Our analysis showed a significant difference between the two groups with MIL-STD-2525B demonstrating a lower recall error rate [$F(1,23) = 8.57$, $p < .01$].

Table 3. Recall Performance based on Type of Symbol Set

Symbol Set	Mean (# of errors)	Std dev.
MIL-STD-2525B		
3 UAVs	0.80	0.28
5 UAVs	2.40	0.72
Stylized Icons		
3 UAVs	1.18	0.27
5 UAVs	2.91	0.68

DISCUSSION

Our results lend support for the first hypothesis that increasing the number of UAVs will have a significant impact on the performance and workload of a participant. However, the results give contrary support to the second hypothesis of stylized icons having improved recall over MIL-STD-2525B.

Except for the percent of enemy targets prosecuted, both objective and subjective performance were shown to significantly decrease as participants supervised five UAVs versus three. Also, workload as measured objectively by reaction times and subjectively by the post-trial questionnaires (including a modified Cooper-Harper Scale), showed a significant increase with a greater number of UAVs under a single operator's control. This was all in congruence with the first stated hypothesis.

There may be several reasons for the lack of support for the second hypothesis. First, the contrast of the icons may have played a major role. The MIL-STD-

2525B icons appeared as green blocks surrounding the inner symbol on the screen. This allowed the icons to be easily distinguished from the often cluttered background. Further, the shape was very noticeable compared to most other objects displayed on the TSD. Even though the stylized icons were the same green color as the MIL-STD-2525B boxes, they did not stand out from the background as well. Their contours may have blended into the background more easily and been lost by the clutter. It was apparent after running this experiment that an area to be looked at is that of display clutter. While the placement of each icon set was attempted to be the same, slight variations may have existed. Additionally, if a re-route or other change in displays occurred that occluded an icon, the MIL-STD 2525B still maintained a possibility of being seen, while the stylized icon may have been completely hidden.

This study is not without some limitations. The use of Air Force Academy cadets as participants brings up the question of generalizability. Do their results generalize to possible UAV supervisors? This question can be answered two ways. First, the purpose of this study is to help build the theoretical framework by which this possible (and probable) future scenario can be further studied. As with any study, future experiments should include real sensor operators (and future UAV supervisors). Also, the position of the supervisor was viewed as not one of actual piloting of the UAV. As such, participants lacking flight experience are not considered any less valid than those with such experience.

Also, this study should be revised to include more realistic tasks that would be experienced by a supervisor in this position. The purpose of the mission mode indicator tasks was to simulate a possible short checklist a supervisor may have to perform. In addition, the participant's testing environment was one of a quiet room with no auditory disturbance or tasks. Furthermore, this experiment included no anomalies during flight. For instance, what would happen to performance and workload if one of the UAVs experienced a problem (such as a fire, getting hit by a missile, etc.)? These authors suspect performance and workload would significantly change to the point of perhaps justifying only two or three UAVs per supervisor. These are issues that can be sorted out once the framework is in place to further this area of research.

One benefit of this experiment is examining the effectiveness of managing multiple UAVs with the potential of making future recommendations regarding the viability of saving money for military (and possible civilian) operations. If fewer humans are required to actively monitor a single UAV, to the point of one operator monitoring several UAVs, then money is saved. This is an area that should be looked into more thoroughly with more externally valid tasks and environments, as well as larger sample sizes and trials.

REFERENCES

De Matos, P. A. & Powell, P. L. (2003). Decision support for flight re-routing in Europe. *Decision Support* Systems, *34, 397-412*.

Draper, M., Calhoun, G., & Ruff, H. (2003). Operator interface research testbed for supervisory control of multiple UAVs: capabilities and initial experimental results. Unpublished manuscript, Air Force Research Laboratory, AFRL/HEC, WP AFB, OH.

Draper, M., Calhoun, G., Ruff, H., Harrison, C., Vidulich, M., & Fontejon, J. (2003). MIIIRO: baseline workload assessment. Unpublished manuscript, Air Force Research Laboratory, AFRL/HEC, WP AFB, OH.

Endsley, M. R. & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness, and workload in a dynamic control task. Ergonomics, 42, 462-493.

Havig, P. R., Jenkins, J. C., & Geiselman, E. E. (2002). A comparison of HMD ownship status symbology and frame of reference orientation during two aircraft control tasks. Unpublished manuscript, Air Force Research Laboratory, AFRL/HEC, WP AFB, OH.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1994). Applied Statistics for the Behavioral Sciences (3rd ed.). Boston: Houghton Mifflin Company.

Hitt, J. M., Brinton, C. J., & Walton, D. H. (2003). Demonstration of the surface management system: an air traffic control decision support tool. Proceedings of the Human Factors and Ergonomics Society, Denver, CO.

Persson, R., Garde, A. H., Hansen, A. M., Orbaek, P., & Ohlsson, K. (2003). The influence of production systems on self-reported arousal, sleepiness, physical exertion, and fatigue-consequences of increasing mechanization. Journal of the International Society for the Investigation of Stress, 19, 163-171.

Sanders, M. S., & McCormick, E. J. (1993). Human Factors in Engineering and Design (7th ed.). New York: McGraw-Hill.