Predicting Controller Capacity in Supervisory Control of Multiple UAVs

M.L. Cummings, *Member, IEEE*, and P.J. Mitchell

***Abstract*— In the future vision of allowing a single operator to remotely control multiple unmanned vehicles, it is not well understood what cognitive constraints limit how many vehicles and related tasks a single operator can manage. This paper illustrates that when predicting the number of unmanned aerial vehicles (UAVs) a single operator can control, it is important to model the sources of wait times caused by human-vehicle interaction, especially since these times could potentially lead to system failure. Specifically, these sources of vehicle wait times include cognitive reorientation and interaction wait time, queues for multiple vehicle interactions, and loss of situation awareness wait times. When wait times were included, predictions using a multiple homogeneous and independent UAV simulation dropped by up to 67%, with loss of situation awareness as the primary source of wait time delays. Moreover this study demonstrated that even in a highly automated management-by-exception system, which should alleviate queuing and interaction wait times, operator capacity is still affected by situation awareness wait time, causing a 36% decrease over the capacity model with no wait time included.**

***Index Terms*—multiple unmanned vehicles, supervisory control, Fan-out, operator capacity**

I. INTRODUCTION

With the recognition that intelligent autonomy could allow a single operator to control multiple vehicles (including air, ground, and water), instead of the converse which is true today, there is increasing interest in predicting the maximum numbers of autonomous vehicles an operator can control. Indeed, the Office of the Secretary Defense's Roadmap for unmanned aircraft systems (UASs) specifically calls for such future architectures [1]. Because of the increased number of sensors, volume of information and operational demands that will naturally occur in a multiple vehicle control environment, excessive cognitive demands will likely be placed on operators. As a result, efficiently allocating attention between a set of dynamic tasks will be critical to both human and system performance. To this end, this paper will discuss recent efforts to model human capacity for management of multiple unmanned/uninhabited aerial vehicles (UAVs), outline an extension to a previous operator capacity model, and demonstrate, using experimental evidence, how the upper bound predictions can significantly change with the new model.

II. BACKGROUND

While a number of research efforts have experimentally demonstrated in simulations that under various levels of autonomy, operators can control anywhere from one to twelve unmanned air vehicles [2-4] and one to eight ground vehicles [5, 6], there has been very little research in developing a model for predicting controller capacity. In terms of operator intervention, control can range from manual control (e.g., actually flying the UAV) to supervisory control (higher-level planning and goal state changes). Operators engaged in manual control will necessarily be able to control fewer vehicles than in supervisory control since manual control requires significantly higher levels of dedicated attention to lower level skill-based cognitive activities, with fewer resources available for higher level planning tasks. Indeed, the tasks and their required attentional resources of each vehicle will likely limit human capacity, as well as the number of vehicles.

In one of the few attempts to theoretically predict an upper bound for an operator controlling multiple independent homogeneous unmanned vehicles, it has been proposed that the number of ground robots a single individual could control can be represented by Eqn. 1 [7-10] . In this equation, FO (Fan Out) equals the number of robots a human can effectively control, NT (Neglect Time) is the expected amount of time that a robot can be ignored before its performance drops below some pre-determined threshold, and IT (Interaction Time) is the time it takes for a human to interact with the robot to raise performance to an acceptable level. Thus, the total capacity is the summation of all neglect and interaction times divided by the interaction time. While originally intended for ground-based robots, this work has direct relevance to other unmanned systems in the air or on/under the water, as these systems are becoming more autonomous and will move from the manual control domain to that of multiple vehicle supervisory control. Because these vehicles are assumed to be homogenous, the assumption is the operators are responsible for a homogeneous set of tasks.

$$FO = \frac{NT + IT}{IT} = \frac{NT}{IT} + 1 \qquad \textbf{(1)}$$

The ratio of operator interaction time to an overall mission time like neglect time is similar to another metric known as

Manuscript received November, 2005. This work was supported in part by Boeing Phantom Works, the Office of Naval Research, and Mitre.

M. L. Cummings is an Assistant Professor in the Massachusetts Institute of Technology Aeronautics and Astronautics Department, Cambridge, MA 02139 phone: 617-252-1512; fax: 617-253-4196; e-mail: missyc@ mit.edu.

P.J. Mitchell is a controls development engineer with General Electric Energy in Power Generation Technology (e-mail: paulj.mitchell@ge.com).

utilization (UT) or percent busy time. UT is the ratio of time an operator is actively engaged in a task (including the three major elements of information processing: perception, cognition, and action) to the overall operating time of a system. Several studies have demonstrated that UT is a metric that can be used to evaluate human-automation interaction, and that operators working beyond 70% utilization experience significantly degraded performance [2, 11, 12]. However, UT is a gross measure of human-automation interaction time in that it does not discriminate among different kinds of interaction time (which will be discussed below).

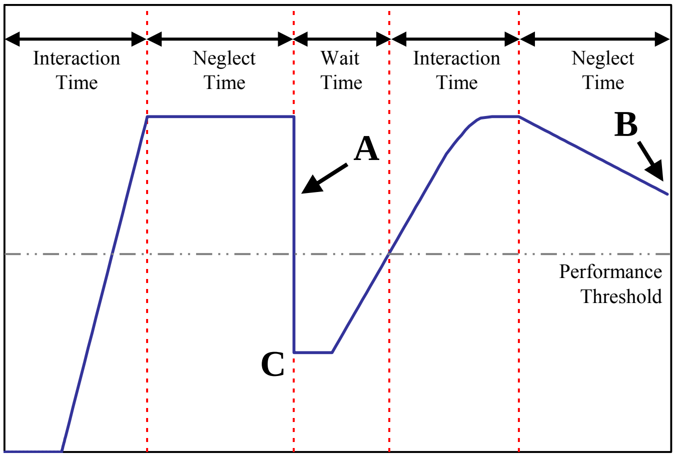

Fig. 1. The relationship between interaction, neglect, and wait times

Equation 1 is an improvement over the more general UT metric in that it compares specifically neglect and interaction times and is a useful measure to establish an ideal upper bound for operators. However, it suffers from the same limitations as UT in that it does not capture the different components of interaction time. Furthermore, we propose there remains an additional critical variable that must be considered when modeling human control of multiple vehicles, regardless of whether they are on the ground or in the air, and that is the concept of Wait Time (WT). In complex problem solving tasks, humans are serial processors in that they can only solve a single complex problem or task at a time [13], and while they can rapidly switch between tasks, any sequence of tasks requiring complex cognition will form a queue, and consequently, wait times for these tasks will build. In the context of a system of multiple vehicles in which two or more vehicles will likely require attention simultaneously from a human operator, wait times are significant in that as they increase, the actual number of vehicles that can be effectively controlled decreases. Moreover, while task and human-computer interface dependently, if humans are cognitively saturated or have low situation awareness, they may not realize that a particular vehicle is in need of attention, thus inducing another wait time until the task is recognized.

Figure 1 illustrates the relationships between interaction, neglect and wait times in terms of overall multiple vehicle system performance. The role of the operator in general is to ensure all vehicles operate at some predetermined minimally acceptable level of performance. Because of the intermittent supervisory control aspect of multiple vehicle control, operators will only attend to those vehicles in need (interaction time) to bring them to this acceptable performance level, hence the increase in performance during IT in Fig. 1. During neglect time periods, the vehicles operate without the need for operator intervention until such time that performance drops below the threshold. Point A in Fig. 1 represents a discrete event that terminates NT, which causes the vehicle to require operator assistance such as a system failure (e.g., an engine loss or loss of a sensor). However, termination of an NT state is not necessarily caused be singular, discrete events. Point B represents performance degradation during NT which causes vehicle performance to eventually drop below the performance threshold, e.g., a slow degradation of an inertial navigation system. In both NT cases, once performance has dropped below an acceptable level, the vehicle must wait until the human recognizes the problem, solves the problem internally, and then communicates that goal to bring the vehicle to an acceptable state so that it can move into a NT state. As can be seen in Fig. 1 at point C, if the problem is not addressed immediately, the vehicle must wait and operate at some non-optimal performance level. As will be discussed in depth in the next section, this delay can be caused by either a loss of operator situation awareness (the operator does not recognize the vehicle is in a degraded state), or the operator is overloaded with more critical tasks from other vehicles. Moreover, since operators cannot generally and instantly bring performance above the threshold, there is an additional interaction wait time while the operator is actively attending to a vehicle (e.g., commanding it back to the correct course) to achieve the desired performance level.

Interaction and neglect times are important in predicting operator capacities for handling multiple vehicles, especially for those domains that are time-critical and high risk like UAVs, as WT could become a critical point for possible system failure. Even for UUVs (unmanned underwater vehicles) that must surface for communications, waiting is not only sub-optimal, it can be extremely hazardous. Moreover, ground-based robots (or unmanned ground vehicles, UGVs) engage in time-critical missions such as search and rescue, which would be negatively impacted if the problem of WT was not addressed. While most robots and vehicles can be preprogrammed to follow some predetermined contingency plan if they do not receive required attention, mission success will likely be significantly degraded if wait times grow unexpectedly.

A. Wait Times

As outlined above, from a robot or vehicle perspective, WT imposed by human interaction (or lack thereof) can be decomposed into three basic components: 1) wait time in the human decision-making queue (WTQ), 2) interaction wait time (WTI), and 3) wait time due to loss of situation awareness (WTSA). For example, suppose an operator is controlling two robots on a semi-autonomous navigation task (much like the Mars Rovers). While typical operations involve human interaction with a single vehicle, there will be times when both vehicles require attention near-simultaneously. When this occurs, once the human operator begins assisting the first robot, the second robot must wait while the operator solves the problem and then issues the commands to it (WTI_1). For the second robot, the time it waits in the queue (WTQ_2) is effectively WTI_1.

IT is the time during which a human's attention is focused on a single vehicle in order to solve a problem or induce some change to improve performance above a specified threshold. From the human perspective, IT includes the time required to determine the nature of the problem, solve the problem, and communicate that solution with some type of feedback. Thus, the vehicle must wait some period of time during the "interaction" due to the human decision-making process. In teleoperation, where the human is directly controlling a robot's movements and positions, interaction wait times might be very small, and occur in rapid succession as the controller gets sensor feedback and adjusts commands accordingly. However, in other supervisory control scenarios that require minimal manual control but significant cognitive input, such as the need to provide a new mission to a UAV, WTI can be quite large depending on the complexity of the problem.

WTSA is perhaps the most difficult wait time component to model because it represents how cognitively engaged an operator is in a task. Situation awareness (SA) is generally defined as having three levels, which are: 1) the perception of the elements in the environment, 2) the comprehension of the current situation, and 3) the projection of future status [14, 15]. While SA can decrease under high workload due to competition for attentional resources [16], it can also decrease under low workload due to boredom and complacency [17]. If an operator does not realize a vehicle needs attention (and thus experiences a loss of SA), the time from the initial onset of the event to actual operator recognition of the problem could range from seconds to minutes.

Equation 2 categorizes general wait time that is not part of human-robot interaction as the summation of wait times that result from queues due to near-simultaneous arrival of problems plus the wait times due to the operator loss of SA.

$$WT = \sum_{i=1}^{X} WTQ_i + \sum_{j=1}^{Y} WTSA_j \qquad \textbf{(2)}$$

$$FO = \frac{NT}{IT\,(WTI\,) + WT} + 1 \qquad \textbf{(3)}$$

WT: Wait time
WTQ: Queuing wait time
WTSA: Wait time caused by a loss of situation awareness
WTI: Wait time due to human decision making, nested in overall interaction time
X = Number of human-automation interaction queues that build
Y = Number of time periods in which a loss of SA causes wait time

Equation 3 demonstrates how the Fan-out equation would change as a result of the inclusion of wait times. For the remaining discussion in this paper, WTI will be considered to be subsumed in IT, which includes time needed for the human to make decisions about a new vehicle problem and then communicate the solution to the vehicle.

While not explicitly linked to interaction time, wait time as a function of WTSA and WTQ occurs in the denominator of Eqn. 3 because WT represents time that should have been spent in the interaction task due to a degraded vehicle state but was not due to human attention inefficiency (either a loss of situation awareness or an inefficient sequencing of tasks in a queue.) Thus, wait times occur only after a vehicle has moved from the NT period to an IT period but must wait until the operator attends to it to achieve a performance increase. Wait times, in effect, increase overall IT.

The original Fan-out equation (1) represents a theoretically perfect system with instantaneous system response, as well as humans who induce no system delays. The modified Fan-out in Eqn. 3 represents a more conservative upper bound which accounts for inefficiencies in human attention allocation (WTSA), limits on the human ability to attend to multiple complex tasks simultaneously (WTQ), as well as the inherent delay that will always accompany the human information processing loop of perception, cognition, and action that occur in every decision making process (WTI). It should be noted that the modified Fan-out (Eqn. 3) is not intended to make an accurate prediction of exactly how many vehicles a person could control, but merely set an upper limit given human and system limitations.

B. Levels of Automation

One of the primary variables that will influence operator capacity in the supervision of multiple vehicles is the level of system autonomy. The challenge in achieving the one-controlling-many goal for management of multiple unmanned vehicles in the future is to determine how automation can be used to reduce human workload. Higher levels of system autonomy will increase NT, which should decrease IT and associated wait times. In terms of aiding the operator in complex decisions, automation decision support can range from fully automatic, where the operator is completely left out of the decision process, to minimal levels where the automation offers basic data filtering or recommendations for the human to consider [18]. For rigid tasks that require no flexibility in decision-making and with a low probability of system failure, higher levels of automation (LOAs) often provide the best solution in terms of operator workload [19]. However, even partially automated systems can result in measurable costs in human performance, such as loss of situational awareness, complacency, skill degradation, and decision biases [20].

In the context of managing multiple vehicles, increasing the levels of automation should reduce workload and wait times by effectively reducing interaction time, but there are several measurable costs for both operators and the system in general. Loss of situation awareness and the propensity for automation bias are significant problems that can result when automation authority increases. For example, at high levels of automation where the system takes over execution of some or all functions, overall wait times are expected to decrease as automation can generally make faster decisions than humans, and the number of opportunities for lower level human errors should be reduced. Superficially, it seems that the system should perform well at higher LOAs, but under abnormal and unexpected conditions, automation could fail, possibly causing a catastrophic event to occur. This is particularly problematic under uncertain or novel conditions because the human operator may not understand the situation well enough to determine if the automation is working correctly and if it requires intervention.

III. MAUVE: The Experimental Test Bed

In order to study how increasing autonomy influences the proposed different components of wait time in the multiple unmanned vehicle control domain, a dual screen simulation test bed named the Multi-Aerial Unmanned Vehicle Experiment (MAUVE) interface was developed (Fig. 2). This interface allows an operator to supervise four homogeneous and independent UAVs simultaneously and intervene as the situation requires. In this simulation, operators are responsible for supervising four UAVs tasked with destroying a set of time-sensitive targets in a suppression of enemy air defenses (SEAD) mission. Because the simulated UAVs are highly autonomous, they only require that operators provide high-level mission planning and execution actions as inputs to the UAVs. The UAVs launch with a pre-determined mission plan, so initial target assignments and routes are already completed. The operator's job is to monitor each UAV's progress, re-plan aspects of the mission in reaction to unexpected events, and in some cases manually execute mission critical actions such as arming and firing of payloads.

MAUVE UAVs are capable of 6 high-level actions: traveling enroute to targets, loitering at specific locations, arming payloads, firing payloads, performing battle damage assessment, and returning to base, generally in this order. Battle damage assessment (BDA) is the post-firing phase of combat where it is determined whether the weapon(s) hit the target and if the desired effect was achieved. In MAUVE, BDA is semi-automated in the sense that operators are responsible for scheduling BDA in advance, but the UAV performs it automatically after firing, if scheduled.

A. *Navigation Display*

The left-hand side of the MAUVE interface in Fig. 2 is known as the navigation display, and it consists of a map display and a mission planning and execution panel. Both elapsed time and time remaining in absolute and relative terms are shown on the top right of the map display. The map display represents a two-dimensional spatial layout of the battle space, updated in real-time. Threat or hazard areas, circular in shape, have a striped yellow pattern, and can be dynamic throughout scenarios, changing size, locations, disappearing entirely, or emerging as time progresses.

The UAVs, always held constant at four, independently change colors according to their current action which include Arm Payload, Fire Payload, Move to Next Target, and Return to Base. Arming and firing are only enabled if the pre-established rules of engagement (RoE) of the simulation are met. For arming, the UAV must be directly on top of a target within pre-determined arming or firing windows. For firing, the UAV should be armed at the correct target. The Move to Next Target button allows operators to bring UAVs out of loiter patterns in case of scheduling problems, and the Return to Base button causes all future targets, waypoints, and loiter points to be deleted from the mission plan. Subsequently, a straight line path is planned directly back to base and is intended for emergency scenarios.

Targets are designated by a diamond-shaped icon and are assigned a priority of high (H), medium (M), or low (L). Active targets are differentiated from inactive targets by their color, which is either red or gray on the display, respectively. Waypoints, shown on the map display with black triangle icons, represent UAV turning points. In addition, UAVs can be loitered at specific points, and typically a UAV loiters for a user-specified amount of time before moving. However, the departure from a loiter pattern must be commanded by the operator. UAV routes can be changed in minor ways by selecting a particular waypoint or loiter point and dragging it to the desired location. More significant routing changes, such as the addition or removal of waypoints, loiter points, or targets, can be accomplished using the mission planning and execution panel. Routing changes are typically only required as a result of unexpected scenarios and represent real-time replanning.

Operators are provided with a "Request TOT Delay" button which allows them limited opportunities to manipulate the Time-on-Targets (TOTs) for those targets assigned. Operators can request a TOT delay for a given target for two reasons: 1)

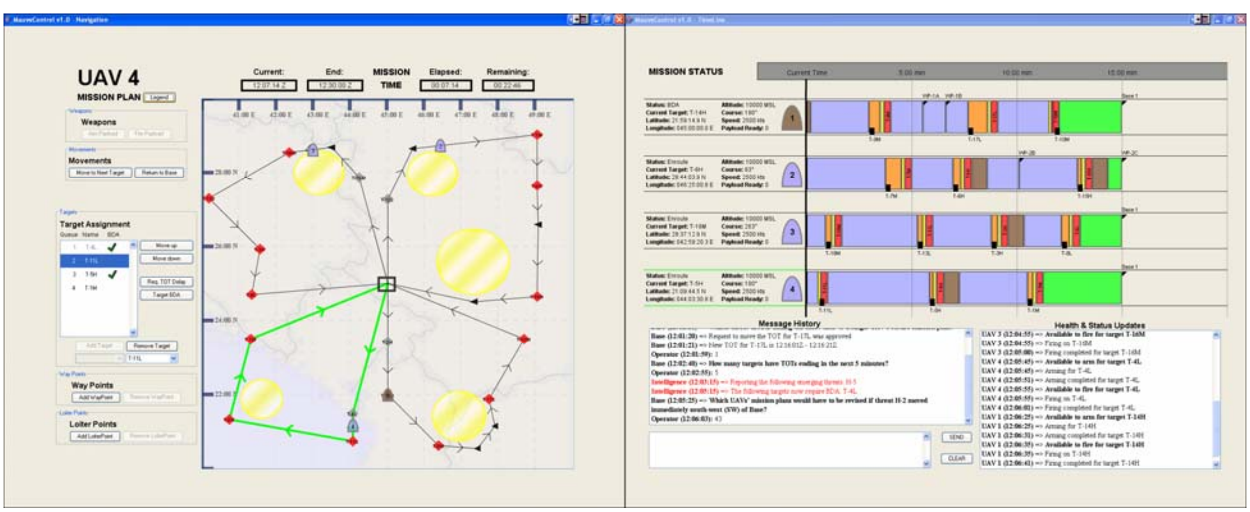

Fig. 2. MAUVE Interface

According to the current mission plan, they are predicted to arrive late to a target, and therefore will miss their deadline, or 2) for workload purposes, i.e., if operators feel they need to spread out a very high workload time block to manage the UAVs more effectively. However, this function should be used with care because moving back one target's deadline likely changes the UAV's arrival time at all subsequent targets. It is important to highlight this change of TOT as a *request*, not a command, and operators' requests can be approved or denied. In MAUVE, the probability of approval is a function of how far in advance of the deadline the request is sent, as would likely be the case in true military situations. When a TOT deadline is immediately approaching, the chance of approval is zero but nearly 1.0 when requested 15 minutes in advance, which is the limit for the decision support, to be discussed in the next section. In MAUVE, a request always takes 5 seconds for response, and during this intervening time, no other TOT requests can be made.

B. *The Decision Support Display*

The right-hand side of the MAUVE simulation in Fig. 2 provides decision support for task management, and it consists of a UAV status window, chat box, UAV health and status updates, and the decision support window. The status window at the top left of the decision support display gives operators low-level, detailed information for each UAV such as current target, current action being performed, latitude and longitude, course, and weapons information. Speed and altitude are also shown, although they are not directly controllable by operators.

The bottom left of the decision support display has a text-based datalink communication tool known as a chat box that contains a time history of all human communication interactions. The chat box window displays various notification messages that appear in response to scenario events or actions taken by users, as well as periodic task-relevant questions for operators to answer from a simulated superior officer. One message that is particularly important to operators is notification that a TOT request is accepted or denied. The bottom right of the decision support display contains a UAV health and status notification window which separates human communications in the simulation from system communications and only contains messages from individual UAVs.

The manipulation of the appearance and functionality of the actual task management decision support (the timeline block) is the primary independent variable for the experiment that will be discussed in the subsequent section. The basic premise of the decision support is to simplify standard air tasking order (ATO) data and combine it in a single interface with up-to-date mission planning information. An ATO provides a schedule of events for a mission as well as required resources. Information contained in an ATO includes which aircraft have been assigned to certain strikes, times on targets, waypoints, and call signs to be used on those missions. Since the focus of this research was to determine how increasing automation would affect wait times, four versions of ATO decision support were created in MAUVE: termed manual, passive, active, and super active, respectively. These will be described in detail below.

The manual LOA level of decision support (Fig. 3a) presents all required ATO and mission planning information

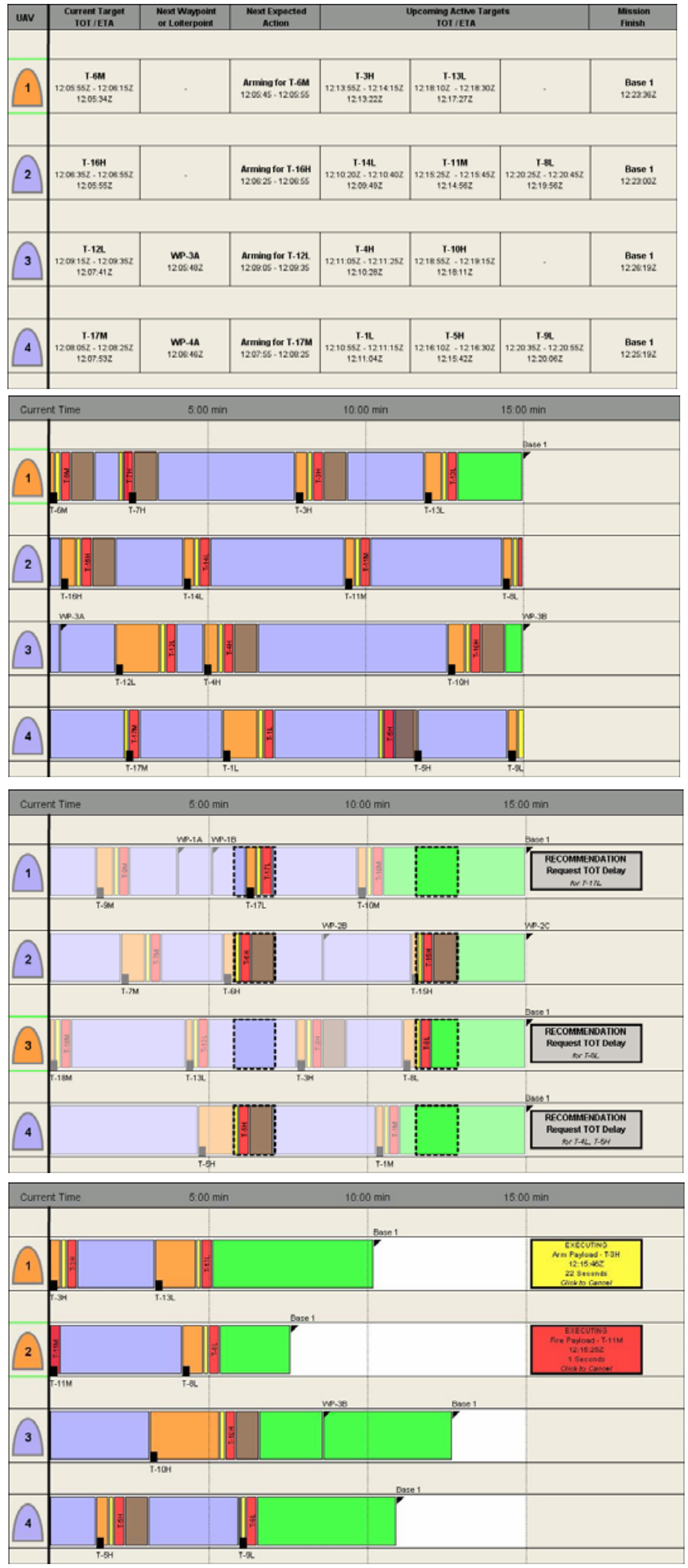

Fig. 3. Increasing Levels of Mission Autonomy
(a) Manual (b) Passive (c) Active (d) Super active

in a text-based table format. Current TOT windows and estimated times of arrivals (ETAs) for up to the next 4 targets in a UAV's timeline are presented. The next waypoint or navigation point on the current route segment is provided, as well as the final ETA for arrival at base in the last column. A low level of automated assistance is provided to the user through the "Next Expected Action" column, which tells the user what they should be doing next and at what time, according to the ATO. This information is updated dynamically to reflect changing ATO requirements and mission planning changes initiated by the user.

The passive LOA (Fig. 3b) assimilates all of the ATO and current mission information contained in the manual level and transforms it into a horizontal timeline format, color coded by action. The major difference between the passive and the manual level is entire schedule integration for users using a graphical format versus text-based. The visual timelines are relative, with the left side representing predicted UAV actions in the near future and the right side up to 15 minutes into the future. Target ETA's are represented by black rectangles on the bottom of each timeline, and waypoint, loiter point and base arrival times are marked by black triangles on the top of each timeline. The TOT windows, arming windows and BDA periods are represented by red, yellow and brown blocks of time.

All of the information in the graphical timeline is available under manual automation, but it cannot be obtained without substantial cross-checking of table entries. Moreover, the manual table format does not facilitate schedule comparisons across different vehicles. With this visual representation, recognizing problems with the current schedule is perceptually-based, allowing users to visually compare the relative location of display elements, instead of specific times to one another. This level of decision support is termed passive because the automation is not performing any tasks except transforming the basic ATO and mission planning information into a graphical format.

The active LOA (Fig. 3c) uses the same horizontal timeline format as the passive automation level but provides additional help from the computer. In the active version, an algorithm searches for periods of time in the schedule that it predicts will cause high workload for the operator, directing the operator's attention toward them. The computer identifies a high workload area, or "bottleneck" as a period of time during which multiple UAVs are scheduled to simultaneously execute mission critical actions which require human interaction, including arming, firing, or performing BDA. The automation draws attention to these areas of concern by a reverse shading technique, in which the "bottlenecks" are highlighted while the rest of the timeline's colors are muted but still visible. As no information is hidden only made less salient, the operator's attention can be directed to the appropriate areas of the schedule while allowing him to maintain SA for the overall mission.

In addition to identifying areas of high workload, the computer also recommends a course of action to alleviate the high workload areas, such as moving a particular TOT. Computer recommendations appear in gray boxes to the right of each relevant UAV's timeline, and subjects have several options: 1) They can acknowledge a high workload area but take no action, 2) They can follow the recommendation to relieve the projected high workload area by shifting a TOT, or 3) They can make other mission planning changes to ensure the high workload area does not occur, such as deleting a target from a UAV's plan. While the automation makes locally optimal recommendations, the algorithm is not globally optimal. Following the computer's recommendation to relieve a high workload area removes that particular schedule conflict but sometimes creates another in the process.

The reverse shading technique, in conjunction with the recommendations, permits operators to make local changes to alleviate workload and immediately see the effect on the global plans of all the UAVs. The purpose of the active level of automation is to help operators identify future time periods of potential high workload farther in advance, so that they can avoid them or at least be better prepared to handle them. This level of decision support is termed active because the automation actively aids operators by narrowing down a set of possible solution alternatives for high workload problems.

The super active LOA (Fig. 3d) also builds upon the passive level visual timeline, but instead of making recommendations to the operator, as in the active LOA, a management-by-exception (MBE) approach is taken. MBE occurs when automation notifies a human that it is going to take some action and gives the operator a limited time to veto the automation. In this experiment, MBE occurs in the super active condition when the computer automatically executes the arming and firing actions for all UAVs at each target, when the rules of engagement are met. Operators are given 30 seconds to intervene with the automation. For example, in order to fire, a UAV has to be located at the particular target it was due to fire on, already armed, and within the TOT window for that target. While the automation handles the actual execution of tasks, the operator is still responsible for determining if the arming and firing actions are appropriate, as well as replanning actions and manipulating routes to ensure the UAVs arrive at the correct targets on time. For the 30 seconds prior to every arming and firing action, exception boxes appear to the right of the timeline that allow the operator to veto these actions. The color of the box indicates which action the UAV is preparing to perform, red for firing and yellow for arming.

IV. Experimental Validation

Figure 4 depicts our hypotheses in terms of how wait times would be influenced by the MAUVE different levels of automation. Because increasing levels of automation should theoretically reduce operators' workload, we proposed that WTI would be progressively lower for the increasing levels. As previously discussed, WTQ depends heavily on WTI, which is expected to decrease with increasing levels of automation in MAUVE. Therefore, WTQ should follow the same decreasing trend as level of automation increases.

For WTSA, the highest level of automation (super active), should theoretically eliminate any wait time due to the loss of SA but only for expected events. When unexpected events occur, operators may have low SA due to complacency, and therefore, they may incur WTSA by not noticing that the mission plan needs adjustment or that a UAV needs attention. In contrast, given the lack of decision support, the manual level should produce the largest total amount of situation awareness wait times of all LOAs. We hypothesized that the active and passive levels should have the lowest accumulated WTSA, because operators are continually interacting with the vehicles and are thus "in-the-loop" as opposed to "on-the-loop" for those operators with super active control. The hypotheses of Fig. 4 are ranked on an ordinal scale both to normalize the variables and because no prior evidence exists to formulate more quantitative hypotheses. In order to validate these hypotheses, an experiment with the MAUVE simulation interface was conducted, which will be described below.

A. Apparatus, Participants, & Procedures

Training and testing was conducted on a four screen system called the multi-modal workstation (MMWS) [21]. A total of 12 subjects took part in this experiment, 10 men and 2 women. Subjects were recruited based on whether they had UAV, military and/or pilot experience. The subject population consisted of undergraduates and graduate students, as well as those from the local reserve officer training corps (ROTC) and active duty military personnel. All were paid $10/hour. In addition, a $50 incentive prize was offered for the best performance. The age range of participants was 20 – 42 years with an average age of 26.3 years. Nine participants were members of the ROTC or active duty officers. While no subjects had operational UAV experience, nine participants had piloting experience. The average number of flight hours among this group was 120.

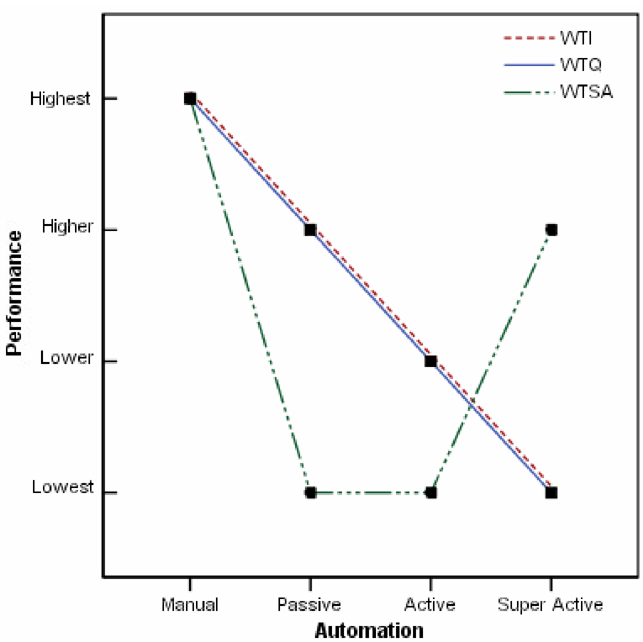

Fig. 4. Experimental Hypotheses

Subjects had two main objectives in this experiment: 1) To guide each UAV so that together, all UAVs properly executed the required missions, which changed over time, and 2) To answer periodic questions about the situation from commanders, which was a secondary task. All subjects received between 90 and 120 minutes of training which culminated in a 15 minute training scenario similar to, but easier in difficulty than the testing sessions. Subjects were required to achieve a predetermined minimum level of performance in this scenario, making no more than two re-planning or targeting mistakes, to move onto the actual test sessions.

Following training, participants tested on two consecutive 30 minute sessions, which represented low and high workload scenarios. These were randomized and counter-balanced to prevent a possible learning effect. The low replanning condition contained 7 replanning events, while the high replanning condition contained 13. These numbers were based on objective and subjective feedback gained through pilot testing. Each simulation was run several times faster than real time so an entire strike could take place over 30 minutes (instead of several hours as is commonplace in actual strikes).

B. Experimental Design

The primary independent variable of interest in this experiment was the level of decision support (Fig. 3). A secondary factor, level of replanning was included to simulate low and high workload scenarios. The level of decision support was between-subjects, and the level of replanning was within-subjects. The statistical model used for the majority of the analyses was a 2x4(3) repeated measures linear mixed model. Three subjects were nested within each automation level, and both independent factors were fixed while subjects represented a random factor. Age was used as a covariate in all analyses and for all reported results $\alpha=0.05$ unless stated otherwise. The dependent variables were the different components of wait time (WTI, WTQ, and WTSA) and as will be described in more detail below, were captured through mouse clicks and data file time stamps of screen objects.

V. Results

A. Interaction Wait Time (WTI)

In this experiment, WTI was recorded from the time an operator selected a degraded UAV to the time a solution was communicated to the UAV (mean = 81.5s, SD (standard deviation) = 57.6s). WTI was significant for level of automation ($F(3,13) = 8.08$, $p = 0.003$) and not significant for level of replanning ($F(1,12.1) = 2.13$, $p = 0.170$). From post-hoc analyses, two homogeneous subsets of automation levels were found, super active/active and passive/manual, meaning that super active and active LOAs produced statistically the same wait times, as did passive and manual. Figure 5 demonstrates how these results compared to predictions, which were generally in line with expectations. It is worth noting for these and the remaining results that the relatively low subject number limits the generalizability of these results.

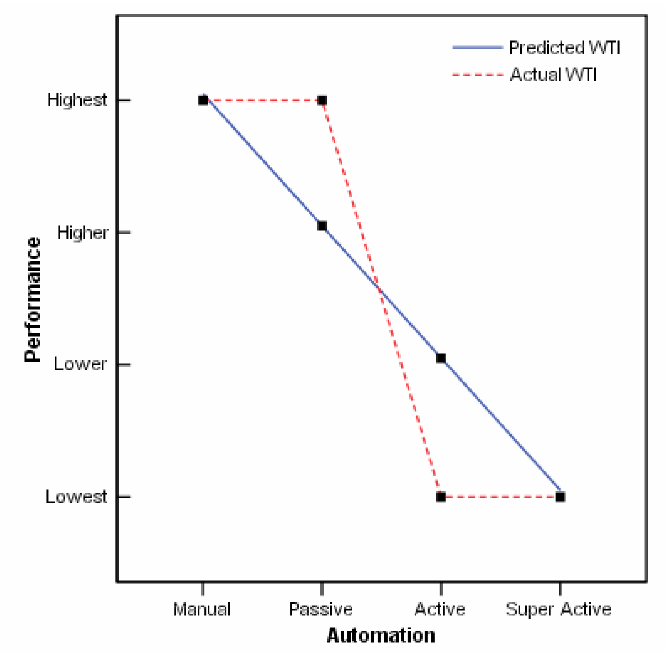

Fig. 5. Interaction Wait Times Results

B. *Wait Time in the Queue (WTQ)*

Wait time spent in the queue was recorded whenever two or more vehicles required operator attention simultaneously, and the operator moved immediately to the vehicle(s) waiting in the queue after finishing a task with another vehicle (mean = 35.0s, SD = 43.4s). Figure 6 illustrates the results as compared with predictions. The amount of wait time spent in the queue was significant for level of replanning (F(1,12.3) = 18.08, p = 0.001) and marginally significant for level of automation (F(3,13.2) = 3.10, p = 0.063). The unusual spike in WTQ for the active level of automation under high replanning in Fig. 6 warranted further investigation which showed that for the high replanning condition, the active level had significantly more WTQ than both the super active and passive levels (A (active) vs. SA (super active), p = 0.009; A vs. P (passive), p = 0.074). In addition, those subjects with active decision support experienced an equal amount of WTQ as those with the manual level.

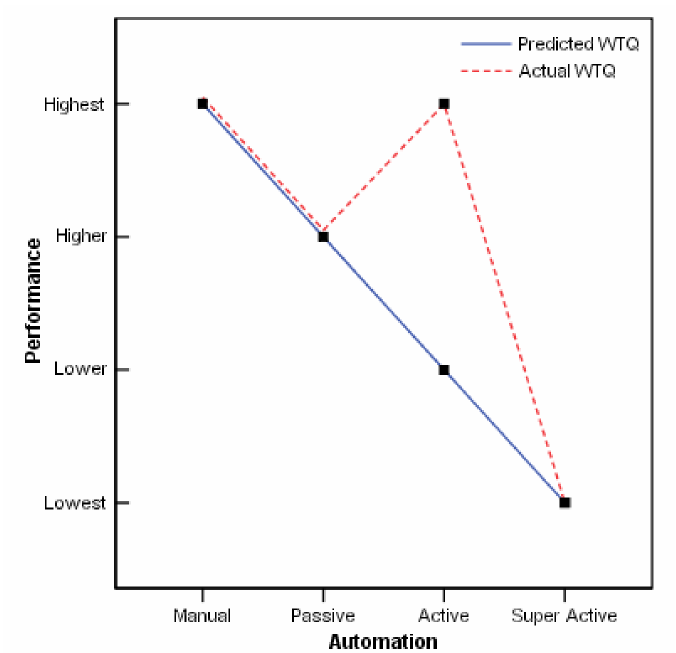

Fig. 6. Wait Time in the Queue Results

C. *Situation Awareness Wait Time (WTSA)*

Situation awareness wait time was recorded when at least one vehicle required attention, but the operator did not realize it (mean = 263.7s, SD = 239.7s). This time interval was marked by operator inactivity, i.e., if a problem was present, and the operator did not attend to it and was not engaged in any other task. Common situations where WTSA was incurred included when subjects forgot to arm and fire on a target and left a UAV loitering unnecessarily, or when a subject flew a UAV into a threat area without attempting to redirect it to a safe area. WTSA was significant for level of replanning (F(1,12.3) = 18.70, p = 0.001) but not for level of automation (F(3,13.2) = 2.14, p = 0.144). However, as depicted in Fig. 7, there was a significant difference in cell means under the high replanning condition between the active and super active level (p = 0.046). Interestingly, under manual automation, there was no significant difference in WTSA between the high and low replanning conditions.

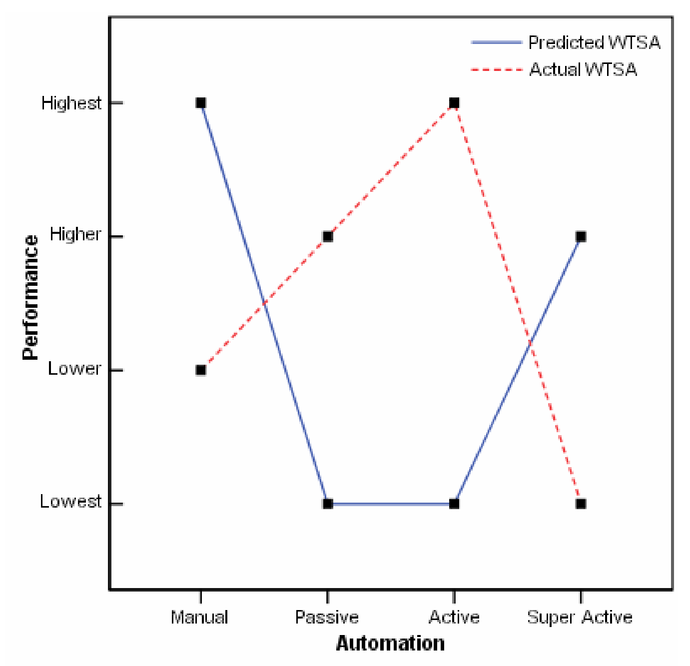

Fig. 7. Situation Awareness Wait Time Results

VI. Discussion

In general, there was a decreasing trend of WTI with increasing levels of automation, which was reasonably consistent with expectations (Fig. 5). Similarly the WTQ results approximately followed the predicted trend of decreasing WTQ with increasing automation level, except for the relatively high average WTQ for active automation (Fig. 6). Quantitative analysis of WTQ accumulation in the active, high replanning test sessions showed that the majority were accumulated in several large queues that formed late in the scenarios when multiple, difficult re-plans were required of the operator.

Further investigation of the anomalous WTQ active condition demonstrated that this problem can be attributed to subjects' inappropriate use of the "Request TOT change" function. It appeared that for those operators in the active condition, when not engaged in a specific task, they fixated on resolving possible areas of predicted high workload [22]. Because of this fixation, the WTSA predictions did not match the actual results (Fig. 7). Subjects using the active level of automation spent significantly more time attempting to achieve an optimal schedule based on computer recommendations, which led to a loss of SA and subsequent significantly higher queuing times.

As can be seen in Fig. 8, of all of the proposed wait time components, WTSA contributes significantly more than the other elements under both low and high workload conditions. There was no significant difference in the relative proportions of WTI, WTQ, or WTSA across the different levels of automation. The proportion of WTSA made up 63% of all wait time in the low replanning condition and 72% in the high replanning condition, showing a trend of increasing proportion of WTSA with increasing workload.

Figure 9 demonstrates how the predictions from Eqn. 1 would change for the high workload condition using the data from this experiment if the proposed wait times are included, as in Eqn. 3. By including wait times in the prediction model, the upper bound of the number of homogeneous independent unmanned vehicles a single controller can manage drops 36-67% based on the effectiveness of the decision support. More importantly, even under a high level of automation represented by the super active management-by-exception decision support, subjects still experienced wait times due to a loss of situation awareness that then led to a reduction in predicted capacity (36%). This significant result demonstrates the importance of including human decision making limitations in an upper bound prediction model, even for a highly automated system. In addition, the increasing WTSA for the lowest three LOAs indicate that operators were approaching their workload limit, culminating in the cognitive saturation of operators under the active condition who were not effectively controlling the number they were assigned,

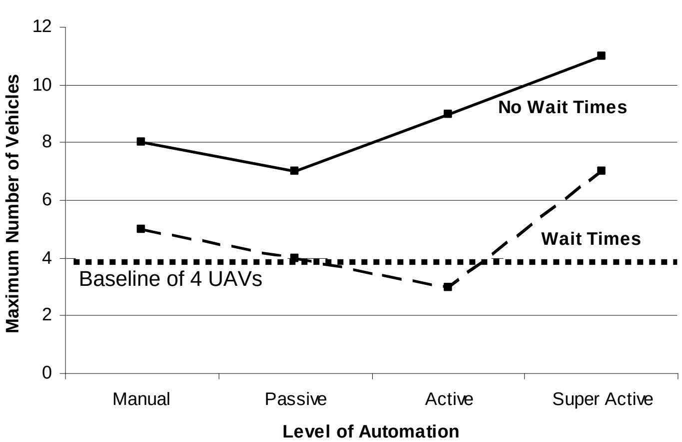

Fig. 9. Operator Capacity Predictions: Original Fan-out Eqn (1) v. Revised Fan-out Eqn (3) Including Wait Times.

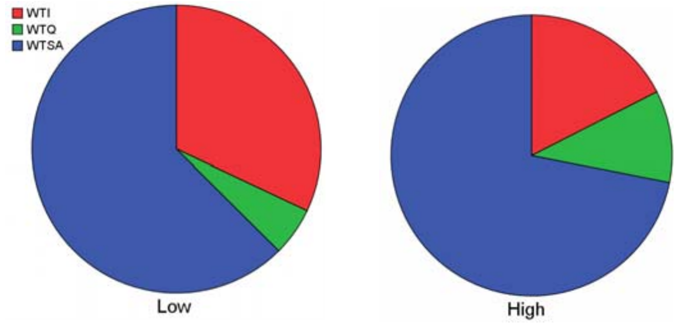

Fig. 8. Wait Time Proportions

(Fig. 9). In the active condition of the experiment, operators clearly struggled with the multiple vehicle management task, driven primarily by WTSA (Figs 7 & 8). This relationship was captured in the upper limit predictions in Fig. 9, which was not captured in the original Fan-out (Eqn. 1) because it does not capture inherent inefficiencies in human decision-making and action.

These results illustrate how critical it is to ensure that when predicting an upper limit for operator capacity based on temporal attributes, the impact of specific automation decision support designs on decision times and situation awareness is critical. These results also highlight the fact that any model that attempts to capture temporal measurements of human-automation interaction is both subject to the task and the context. The significance of the level of replanning across the wait times demonstrates that one external environmental variable can dramatically influence the results, so caution should be exercised in generalizing specific results to other domains.

One additional source of wait time that deserves further investigation not addressed in this study is wait time due to cognitive reorientation (WTCR), which is a component of interaction wait time. WTCR is the time it takes an operator to regain the correct mental model, and situation awareness needed to solve a problem once recognized. Thus, WTCR is primarily a component of WTI but could be related to WTSA in that operators with lower WTSAs will likely have lower WTCR. WTCR represents a switching cost that occurs when an operator is cognitively engaged in one task but then must spend some period of time reorienting to a new problem. This environment of switching attention between multiple vehicles means that operators must not only concentrate attention on a primary task, such as higher level mission planning, but also be prepared for alerts for other events, such as a vehicle system failure. This need to concentrate on one task, yet maintain a level of attention for alerts/information from other potential tasks causes operators to have a conflict in mental information processing.

WTCR and the associated switching costs are very difficult to capture as performance-based simulations such as the one reported in this UAV study cannot accurately and consistently categorize WTCR. These difficulties have also been shown in multiple control of ground robots [9]. Using software that tracked users' cursor movements and activation of control devices, we were able to determine when a subject was engaged with a particular UAV, but we were unable to determine at what point the cognitive reorientation occurred, a subtle transition. This inability to partition WTCR from the overall category of WTI could possibly be addressed through some psychophysiologic measure or perhaps a more carefully crafted scenario. This area deserves more attention because it is unclear how much WTCR contributes to overall wait times, and how technological interventions such as intelligent decision support could possibly mitigate this and other wait times.

VII. Conclusions

This research extends previous work attempting to predict the number of homogeneous and independent unmanned vehicles a single operator can control. We propose that any predictive model of operator capacity that includes human-in-the-loop remote interaction should include various sources of wait time which include wait time due to human-computer interactions (including cognitive reorientation), queuing wait time, and wait time due to a loss of situation awareness. Using data from a simulation examining control of multiple homogeneous and independent UAVs, capacity predictions that included these sources of delay dropped by up to 67%, with loss of situation awareness as the primary source of wait time delays. Furthermore, this study demonstrates that both interface design components, as well as the level of decision support automation, can be a significant contributor to overall wait times. Moreover, even in a highly automated management-by-exception system which should alleviate queuing and interaction times, this study demonstrated that operator capacity is still affected by situation awareness wait time, causing a 36% decrease over the capacity model with no wait time. Further work is needed to more accurately capture cognitive reorientation times which represent a switching cost as well as loss of situation awareness wait times.

Acknowledgments

We would like to acknowledge the critical support of our superb undergraduate research team, Dave Pitman and Albert Leung. We would also like to thank Dr. Tom Sheridan and the authors of refs [3-5, 9], as well as the reviewers, whose comments were critical in more accurately framing this research.

References

[1] Office of the Secretary of Defense, "Unmanned Aircraft Systems (UAS) Roadmap, 2005-2030," DoD, Washington DC 2005.
[2] M. L. Cummings and S. Guerlain, "Developing Operator Capacity Estimates for Supervisory Control of Autonomous Vehicles," *Human Factors*, vol. 49, pp. 1-15, 2007.
[3] S. R. Dixon, C. D. Wickens, and D. Chang, "Comparing quantitative model predictions to experimental data in multiple-UAV flight control," presented at Human Factors and Ergonomics Society 47th Annual Meeting, Denver, 2003.
[4] H. A. Ruff, S. Narayanan, and M. H. Draper, "Human Interaction with Levels of Automation and Decision-Aid Fidelity in the Supervisory Control of Multiple Simulated Unmanned Air Vehicles," *Presence*, vol. 11, pp. 335-351, 2002.
[5] R. Parasuraman, S. Galster, P. Squire, H. Furukawa, and C. Miller, "A Flexible Delegation-Type Interface Enhances System Performance in Human Supervision of Multiple Robots: Empirical Studies With RoboFlag," *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, vol. 35, pp. 481-493, 2005.
[6] S. A. Rehfeld, F. G. Jentsch, M. Curtis, and T. Fincannon, "Collaborative Teamwork with Unmanned Ground Vehicles in Military Missions," presented at Human Computer Interaction International, Las Vegas, 2005.
[7] D. R. Olsen and M. A. Goodrich, "Metrics for Evaluating Human-Robot Interactions," presented at Performance Metrics for Intelligent Systems, Gaithersburg, MD, 2003.
[8] D. R. Olsen and S. B. Wood, "Fan-out: Measuring Human Control of Multiple Robots," presented at CHI2004, Vienna, Austria, 2004.
[9] M. A. Goodrich, M. Quigley, and K. Cosenzo, "Task Switching and Multi-Robot Teams," presented at 3rd International Multi-Robot Systems Workshop, Washington DC, 2005.
[10] J. W. Crandall, M. A. Goodrich, J. Dan R. Olsen, and C. W. Nielsen, "Validating Human-Robot Interaction Schemes in Multitasking Environments," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 35, 2005.
[11] W. B. Rouse, *Systems Engineering Models of Human-Machine Interaction*. New York: North Holland, 1983.
[12] D. K. Schmidt, "A Queuing Analysis of the Air Traffic Controller's Workload," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 8, pp. 492-498, 1978.
[13] A. Chapanis, F. C. Frick, W. R. Garner, J. W. Gebhard, W. F. Grether, R. H. Henneman, W. E. Kappaif, E. B. Newman, and A. C. Williams, "Human Engineering for an effective air navigation and traffic control system," P. M. Fitts, Ed. Washington DC: National Research Council, 1951.
[14] M. R. Endsley, "Design and Evaluation for Situation Awareness Enhancement," presented at Proceedings of the Human Factors Society 32nd Annual Meeting, Santa Monica, CA, 1988.
[15] M. Endsley, "Measurement of situation awareness in dynamic systems," *Human Factors*, vol. 37, pp. 65-84, 1995.
[16] A. Andre and C. Wickens, "When Users Want What's Not Best for Them," *Ergonomics in Design*, vol. October, 1995.
[17] M. D. Rodgers, R. H. Mogford, and B. Strauch, "Post Hoc Assessment of Situation Awareness in Air Traffic Control Incidents and Major Aircraft Accidents," in *Situation Awareness Analysis and Measurement*, M. Endsley and D. J. Garland, Eds. Mahwah, NJ: Lawrence Erlbaum Associates, 2000, pp. 73-112.
[18] T. B. Sheridan and W. L. Verplank, "Human and Computer Control of Undersea Teleoperators," MIT, Cambridge, Man-Machine Systems Laboratory Report 1978.
[19] M. R. Endsley, "Situation Awareness in Aviation Systems," in *Handbook of Aviation Human Factors*, D. J. Garland, J. A. Wise, and V. D. Hopkin, Eds. Mahwah, NJ: Lawrence Erlbaum, 1999.
[20] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, "A Model for Types and Levels of Human Interaction with Automation," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 30, pp. 286-297, 2000.
[21] G. Osga, K. Van Orden, N. Campbell, D. Kellmeyer, and D. Lulue, "Design and Evaluation of WarfighterTask Support Methods in a Multi-Modal Watchstation," SPAWAR, San Diego 1874, 2002.
[22] M. L. Cummings and P. J. Mitchell, "Automated Scheduling Decision Support for Supervisory Control of Multiple UAVs," *Journal of Aerospace Computing, Information, and Communication*, in press.

Mary L. Cummings (M'03) received her B.S. in Mathematics from the United States Naval Academy in 1988, her M.S. in Space Systems Engineering from the Naval Postgraduate School in 1994, and her Ph.D. in Systems Engineering from the University of Virginia in 2003. A naval officer and military pilot from 1988-1999, she was one of the Navy's first female fighter pilots. She is currently an assistant professor in the Aeronautics & Astronautics Department at the Massachusetts Institute of Technology. Her previous teaching experience includes instructing for the U.S. Navy at Pennsylvania State University and as an assistant professor for the Virginia Tech Engineering Fundamentals Division. Her research interests include human supervisory control, human-uninhabited vehicle interaction, bounded collaborative human-computer decision making, decision support, information complexity in displays, and the ethical and social impact of technology.

Paul J. Mitchell received his B.S. in Mathematics and Engineering from Queen's University, Canada in 2002, and his M.S. in Aeronautics and Astronautics from the Massachusetts Institute of Technology in 2005. He is currently a controls development engineer for General Electric Energy in Power Generation Technology. His research and professional interests include human supervisory control, human-computer interfaces and decision support, particularly for uninhabited vehicles, and the application of advanced control methods to aerospace and power generation systems.