On the Process of Automation Transition in Multitask Human–Machine Systems

Peter A. Hancock, *Member, IEEE*

***Abstract*—This paper examined the effects of different forms of automation invocation, the reconfiguration of the task display when automated, and the influence of the specific modality that warned of the manual/automation transition on operator performance. Thirty-two experienced pilots engaged in a multiple-task situation consisting of tracking, monitoring, and fuel management subtasks, representative of typical aviation demands. Automation of the tracking task could be invoked in four different ways: 1) system-initiated automation (SIA); 2) pilot command by negation (PCN); 3) pilot command by initiation (PCI); and 4) pilot-initiated automation (PIA). Pilots were warned of the mode change between manual and automated controls by either a visual, an aural, or a combined visual and aural cue. The display of the subtask while in automation was reduced in size and placed either in a central or peripheral location. Results indicated that SIA had a differential effect on tracking performance as compared to all other forms of automation invocation. The respective location of the automated display had it effects in the fuel management subtask, whereas monitoring capability remained stable across manipulations. A significant three-way interaction between invocation procedure, display location, and warning modality illustrated the selective disadvantage of the visual warning combined with the central location under the PCI procedure on tracking response. Measures of subjective response suggested that visual warning modality was slightly more taxing than either an auditory or a combined auditory and visual warning. Pilots also experienced elevated fatigue when the system initiated the automation. These results confirm that both performance and subjective perception of multitask demand are greater when the system controls the option to automate. A qualitative model is presented, which provides an approach for the integrated assessment of human performance with adaptive systems.**

***Index Terms*—Adaptive task allocation, automation, dynamic interface display, multitasking, perceived workload.**

I. Introduction

THE NATURE of automation and its advantages and disadvantages with respect to operations of modern complex systems continues to be the topic of significant contention and ongoing debate [1]–[9]. Proponents of automation cite the progressive increase in information load as their basic rationale for the implementation of automation. In contrast, critics assert that automation reduces the status of the human operator to a mere "monitor" or sporadic "button pusher," even referring to operators as system "baby sitters" [10]. Two developments have the potential to resolve this putative conflict. One is the penetration of evermore-reliable digital technology, creating automated systems that have diminishingly low failure rates. The second is the evolution of systems that are adaptive, in which the capabilities and limitations of both machine and human are considered as linked elements in a dynamic sharing of obligatory task demands [11], [12]. As part of a more general recognition of the advantage of human-centered design [1], adaptive systems seek to enact a flexible operational strategy by implementing control changes dynamically as opposed to the one-time and static assignment of tasks to either operator or system. Given the growing reality of this latter evolution, the primary purpose of this paper was to examine the effect of different automation invocation procedures on performance in a realistic aviation-representative multitask environment. In this context, aviation is a particularly relevant application realm since it has often been the pioneer arena for a variety of innovative human–machine developments, including adaptive automation [13], [14].

Adaptive function allocation in complex systems seeks to achieve a major goal of advanced automation, that being the regulation of operator workload. However, adaptive function allocation as a design strategy does not identify exactly how the shift from operator to system control should occur. There are many ways through which such adaptive allocation implementation could take place. For example, the system could initiate automation based on the fracture of some preagreed performance threshold. Such a threshold might be a reflection of aircraft performance or some psychophysiological trigger, which indicated an incipient pilot overload state. Conversely, the pilots themselves could choose to change to an automated mode as they perceived themselves to be reaching demand saturation. The latter approach would be contingent largely on personal subjective assessment rather than on objective system state [15]. Thus, there is a wide spectrum of possible methods to choose from for the automation trigger, and indeed, several of these candidates have already begun to be examined [16]–[18].

In adaptive automation, one might select from among these various invocation triggering mechanisms. However, where does the ultimate authority for invoking automation lie? This

Manuscript received September 20, 2005; revised October 2, 2006. This work was supported by the Naval Air Warfare Center under Contract N/N62269-92-0211, with Dr. J. Morrison as the Contract Monitor. The completion of this work was facilitated by the Department of Defense Multidisciplinary University Research Initiative, which was administered by the Army Research Office under Grant DAAD19-01-1-0621. This paper was recommended by Associate Editor J. Adams.

The author is with the Multidisciplinary University Research Initiative/Operator Performance Under Stress Research Laboratory, Department of Psychology, and Institute for Simulation and Training, University of Central Florida, Orlando, FL 32816 USA, and also with the Department of Civil and Environmental Engineering, University of Central Florida, Orlando, FL 32816 USA (e-mail: phancock@pegasus.cc.ucf.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMCA.2007.897610

question has been addressed theoretically by several commentators under the title of control/management [19] and control logic [20]. Sheridan [21], for example, conceptualized this general authority structure under "levels of automation" and identified ten divisions on a continuum from 100% automated control to 100% manual control. Sheridan's division refers to specific operational configurations. However, for the present purposes, his levels of control can also be conceived as the locus of authority for the invocation of automation. At Sheridan's now derived level 1, the operator is solely responsible for the change from manual control to automated control. This process would involve both the assessment of self-performance and the subsequent execution of the automation invocation procedure itself. At the other end of the continuum, the system would be solely responsible for the execution of the change in control authority depending on a transition across some preset threshold. This paper examines various forms of these invocation procedures, some of which are close to the respective extremes as previously described. However, it also examines additional procedures in which the system proceeds with automation except under the negative command of the pilot and another condition in which the automation warns the pilot of the current poor level of performance but takes no explicit action beyond this, relying on the pilot to make the decision on change of control authority. Variations of each of these procedures have been recommended as possible invocation strategies in aviation as well as in other complex systems [22]–[24]. At the time that the present experiment was conducted [25], it represented one of the first empirical examinations of this range of possible strategies that have subsequently been evaluated in a number of more recent research studies [26]–[29].

While the primary purpose of the present investigation was to determine how these differing degrees of automation invocation authority influenced performance, there were two additional purposes. The first of these was to evaluate the ways in which the pilot could be warned of the change that had been made in control status. Thus, the present experiment embedded a comparison of a visual versus an auditory versus a combined visual and auditory warning to indicate automation status change. These warnings represented formats that have been used frequently in electronic cockpits. Second, an explicit effort was made to evaluate the veracity of a previous finding concerning the independent effects of reducing the size and/or relocating the display of the component subtask after automation had occurred [26] and to see how such different interface changes affected performance on the remaining subtasks. This previous finding had suggested that placing the reduced size display of the automated task in a central location had actually been less effective than placing it in a peripheral location, which is a result that appeared to be counter to the functional proximity model for display arrangement [27]. The present procedure provided an opportunity to test the reliability of that finding with a much expanded sample of experienced pilots in association with the new set of invocation procedures. It was hypothesized that, first, SIA would prove more effective but more stressful than all other conditions, whereas performance with the recommended automation procedures would prove more effective than performance under the discretionary invocation procedure. Second, the multimodal warning would prove the most effective, which is an assertion that is based on the notion of sensory cue redundancy. Third, it was expected that workload would vary according to the competition for resources that are expressed by the various warning modality combinations, such that those involving visual warnings either singly or in combination would prove more demanding than auditory warnings alone. Finally, it was anticipated that peripheral relocation of the automated display would prove more effective than central relocation, which is founded on the outcome of a prior experimental evaluation [26].

II. Experimental Method

A. Experimental Participants

To evaluate these respective propositions, 32 experienced pilots (30 males and 2 females) were recruited from the general area through local advertising. The majority of the commercial pilots were recruited from the professional participants who were undergoing advanced training at the Northwest Air Training Company, which was located in the local area. Pilot experience ranged from General Aviation with 100 h flying time under Visual Flight Rules (VFR) to Commercial Aviation with 16 000 h including professional flight time under Instrument Flight Rules (see Table I). The majority of the pilots flew under Part 135 of the Federal Aviation Regulations. The pilots, whose mean age was 32.5 years, were divided into four separate groups of eight each for the between-participant factor of automation invocation procedure. Groups were generally balanced according to their flying experience, specifically accounting for the number of hours in General Aviation versus Commercial and Corporate Aviation. Experience itself was not directly considered as a factor in the present experiment since there was no way to account for the nature of such experience on the present multitask environment beyond the notation of the number of hours flown, the specific aircraft flown, and the flight qualification status.

B. Experimental Tasks

1) MINSTAR Test Facility: The MINSTAR multitask environment served as the experimental platform for the present experiment. This middle-level fidelity flight-task simulation facility has been previously described in detail elsewhere [25], [30]. Briefly, MINSTAR presents a multitask display in which 2-D compensatory tracking, fuel management, and display monitoring are presented as individual component subtasks. These subtasks can be performed individually, in combination with each of the others, or under varying modes of automation as required by the experimental procedure. A schematic of the MINSTAR test facility is shown in Fig. 1.

2) Tracking Subtask: The goal for the participant in the tracking subtask was to engage in corrective movements via the flight stick to bring the moving cursor (shown as the cross on the right screen in Fig. 1) in alignment with a fixed target at the center of the display. The difficulty of the task was manipulated by modifying the amplitude and frequency of the sine waves,

TABLE I
PILOT EXPERIENCE BY INVOCATION GROUPING

Subject	Age	Civil Ratings	Total Flight Hours	Aircraft Experience
SYSTEM INITIATED AUTOMATION				
1	42	Com, SEL, IFR	300	Cherokee 140, 180
3	45	SEL, MEL, IFR	500	T-2A&B, TA-45
10	22	MEL, IFR	190	Cessna 152, 172, PA180
11	47	SEL, IFR	240	Cessna 172, 182, Piper
12	20	SEL, IFR	230	-
18	42	ATP	550	DC-9/MD80, DC-10, C-13, C-9
19	22	VFR	120	Cessna 172
21	28	ATP, CFII, MEI	4000+	DC-6, SF-340, BAe3100
PILOT COMMAND BY NEGATION				
2	26	ATP	5000	F-27, DH-8
5	30	SEL, ME, ATP	4500	Com a/c & aerobatic
9	40	ATP, CFII	16000	B727, DC-9, SW-4
13	23	ME, SE, IFR	800	King Air 200, Navajo
14	23	Com, ME, IFR	225	Cessna 172, 310
15	30	Com, SEL	416	Cessna 150, 180
24	21	IFR	200	-
25	24	Com, IFR	230	Piper Cadet, Piper Arrow
PILOT COMMAND BY INITIATION				
4	36	ATP	12000	BD03100, SF-340
6	34	SEL, IFR	250	Cessna 150, 180, Beech 19
7	45	ATP	12000+	A320, B727, DC-9, KC-135
8	47	Com, IFR, ME, CFI	3250	250 hours
16	28	ATP, SE, ME	4150	F-27, DC-9, B727
17	46	ATP, CFII, MEI	3000	BE58, MU-2
29	21	SEL/VFR	125	Cessna 152, 172
30	38	SEL/VFR	105	Cessna 152, 172
PILOT INITIATION OF AUTOMATION				
20	-	CFI	425	Cessna's
22	24	Com, CFI	345	Cessna, Beech
23	23	ASMEL, CFII	260	-
26	37	Com, ME, CFI, IFR	700	UH-1H
27	47	SEL, SES	286	Cessna 150, 180, PA18-90
28	-	ATP, Com, CFI,..	8000	PA23, Cessna 401, 172, P2-Y
31	37	ATP	5900	DC10, B737, B727, A320, T37, T38, KC135, T37, T38, KC235
32	28	Com, CFII, MEI	1900	BE58, BE55, PA44, PA34

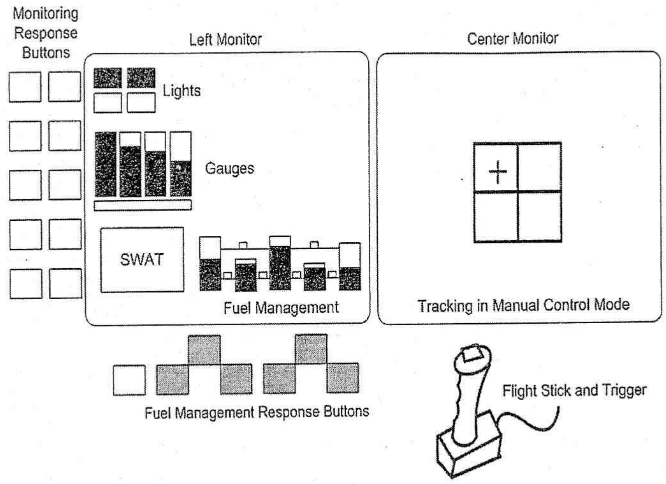


Fig. 1. Overall MINSTAR task configuration. On the right CRT is the 2-D tracking task that is controlled by the flight stick, with the automation button shown on top. On the left CRT is the fuel management task, which is shown at the bottom right, and the monitoring task, which is composed of light and gauge monitoring and shown at the upper left. The appropriate response buttons to the latter two tasks are directly mapped to each display component. Adjacent to the fuel management display and under the monitoring display is the workload assessment (SWAT) window, which is shown here in operation mode. SWAT responses were requested periodically, as detailed in the text.

which control the position and velocity of the moving cursor [26]. The specific characteristics that drove the present tracking task have been described in detail previously [31] and represent the medium level of demand that was detailed in that work. Performance was measured by a root-mean-square (rms) error score that was taken each second of the 5 min of each respective trial's duration.

3) Fuel Management Subtask: The goal in the fuel management subtask was to manually control the on/off status of the respective fuel pumps to maintain a target level of 2500 gal in the two outer tanks (shown at the bottom right of the left display screen in Fig. 1). Difficulty in this subtask was manipulated by initiating failures in the fuel pumps that were positioned between tanks. Performance on the fuel management subtask was assessed using two derived but complementary error measures. The first form of measurement, which reflected central tendency, was constant error. This was calculated as the participant's mean deviation from the target value of 2500 gal in each of the far left and far right tanks. The second measure, which reflects response stability, was variable error (VE). This was calculated as the standard deviation of each participant's response. VE does not depend on whether or not the participant was close to the target level itself since it is calculated about the participant's own mean and therefore provides a measure of dispersion around the self-generated measure of central tendency. In a general sense, the fuel management task was itself one that required a generalized tracking-type response. However, this form of tracking is contingent on much less frequent and individually discrete responses in terms of pump activations, as compared to the continuous input of the central tracking task. In terms of a timeline of performance over each respective 5-min trial, the tracking subtask and the fuel management subtask were continuously present (except when tracking automation was initiated). However, the following monitoring subtask presented only periodic deviations from normal status, and the timeline of monitoring events is described in Section II-B4.

4) Monitoring Subtask: The goal in the last of the three subtasks was to monitor the collective display of lights and gauges, as shown in Fig. 1, and to press the appropriate buttons to reset them whenever they deviated from their normal state. The frequency of these respective deviations, of which there were a grand total of 23 per 5-min trial, was under experimental control via a predetermined programmed script. The lights component of the monitoring subtask had two green lights that were normally on, two red lights that were normally off, and a yellow light that was also normally off. When any of these lights changed status, the participant had to press the appropriate function key to restore the light to its normal condition. The second element of the monitoring subtask required attention to four graduated gauges. The participant was required to monitor each gauge for deviations above criterion yellow (first hash mark) or red (second hash mark) levels. Participants were instructed to interpret deviations above the yellow hash mark as a warning. However, when the indicator rose above the red hash mark, participants were required to press the appropriate key to reset the indicator as quickly as possible. These respective monitoring displays approximate those seen in modern-day fly-by-wire aircraft. In each condition, there were three deviations of the red and three deviations of the green lights, as well as four deviations of the yellow light. In respect of the gauges, there were five deviations of the gauge when the warning reached yellow level and eight deviations where the gauge reached the red level. As noted, the combination of these various deviations gave a grand total of 23 script-driven deviations, which were randomly distributed across each condition. However, such scripts were consistent across participants; thus, each individual saw the same respective signals in their same temporal location. The timeline of these deviations distributed them roughly equally across the 5-min trial duration. Performance on the overall monitoring subtask was quantified in terms of response time (in seconds) to each light or gauge deviation and the number of missed signals and false alarms made in each condition.

5) Profile of Mood States (POMS) Questionnaire: Since invocation procedures can be triggered by both objective performance level and pilot subjective workload state, it is important to understand each of these types of response. Thus, two reflections of subjective response were recorded in addition to the objective measures of response capacity. Mood state has been found to be a predictor of change in performance capacity [32]. Thus, the POMS [33] questionnaire was administered in its standard version. The POMS asks respondents to rate each of 65 items in terms of how well those items describe his/her feelings "during the past week including today." While the standard instructions were appropriate for pretrial assessment, the time frame that was specified in the instructions was inappropriate for the posttrial assessments, which were aimed at evaluating immediate mood state. The POMS has previously been adapted so that respondents report their mood state "right now." Therefore, these modified instructions were used in the posttrial sessions of the present experiment. The 65 items of the POMS were scored to make up scales measuring six identifiable moods or affective states, namely: 1) Tension–Anxiety; 2) Depression–Dejection; 3) Anger–Hostility; 4) Vigor–Activity; 5) Fatigue–Inertia; and 6) Confusion–Bewilderment. It is the scores on these respective scales that are reported here. It was anticipated that the lack of control over automation, as evident in the SIA condition, would be reflected in elevated scores on component elements of the overall POMS scale.

6) Subjective Workload Assessment Technique (SWAT): The second form of subjective assessment that was used in the present experiment was the SWAT [34], [35]. This reflection of subjective response was particularly appropriate to assess the reaction to the pilot-initiated invocation procedure, which is the most common current form of adaptive automation interaction mode. The standard SWAT procedure was modified here so that at the 4:45 mark of each 5-min trial, a visual display of the SWAT scale was posted on the left cathode ray tube (CRT), as shown in Fig. 1. At that time, the experimenter then asked the pilot to respond to the following three questions in reference to the current trial: 1) What is your time load—Do you have much spare time, some spare time, or little spare time? 2) What is your stress level—low, medium, or high? 3) What is your mental effort—low, medium, or high? The experimenter recorded the numerical responses (1, 2, or 3 for low, medium, and high levels, respectively) for each of the three questions for each participant in each trial. It was these responses that were entered into subsequent analysis. Since the SIA condition is the one in which the decision control is removed from the pilot and since such absence of control is know to be a critical factor in stress elevation, it was expected that the SIA condition would result in higher levels of workload as reflected in the SWAT scores. It was further anticipated that as the level of pilot control over automation increased, the associated level of workload, as reflected in the SWAT scores, would be reduced.

C. Experimental Manipulations

In the present experiment, the automation invocation procedure was the between-participant factor. A between-participant design was used to avoid the potentiality for negative transfer between conditions that may be anticipated if automation invocation had been a within-participant factor. However, all participants experienced each of the modality configurations that warned of automation invocation, namely an auditory warning on its own, a visual warning on its own, and a multimodal combined visual and auditory warning. Each of these respective warning signals occurred at exactly the same time as the automation commenced. As well as these manipulations, all participants experienced the two different locations of the automated tracking display, which thus represented the second and final within-participant factors. When automated, the tracking display was always reduced in size. However, as well as this size reduction, which explicitly indicated automation status, the size-reduced display could appear in one of two different locations. Either it appeared in the center of the screen, or it was placed in the lower right-hand corner of the respective CRT. The overall experimental design was therefore a mixed one with both within-participant (warning modality and automated display location) and between-participant (automation invocation procedure) factors.

1) Automation Invocation Procedures: The four methods for the invocation of automation for the tracking subtask were: 1) SIA; 2) pilot command by negation (PCN); 3) pilot command by initiation (PCI); and 4) pilot-initiated automation (PIA). The procedure for each method is described later in the text. With the exception of the PIA condition, all other invocation methods used a constant rms error performance criterion to initiate the processes of tracking automation.

SIA: In each instance that the pilot did not maintain an rms error of 70 units or less for a period of 250 ms, the system automated the tracking subtask. The duration of the automation was 15 s, after which manual control was returned to the pilot.

The rms values for automation onset in terms of both the 70-unit level and timing in terms of the 250-ms duration, as well as the absolute duration of that epoch of automation. i.e., 15 s, were each derived from previous experimentation [36], which had shown them to be appropriate values. However, it is important to emphasize that both the spatial criterion for error and the length of automation duration in the present experiment were empirically derived thresholds. While it is likely that these thresholds will vary according to the specific operational system in question and it may well be that different operators will themselves choose different values according to their preference, there is presently a significant research opportunity to show how such thresholds could be derived from modeling, simulation, or theory. This paper retains the empirically driven strategy and ensures that the threshold is constant across the different invocation procedures.

PCN: In each instance that the pilot did not maintain an rms error of 70 units or less for a period of 250 ms, the system warned the pilot that automation was to be initiated unless the pilot specified that it did *not* occur. The pilot could allow the system to automate or negate the occurrence of automation by pushing the appropriate button within 5 s. If the pilot wished to negate the automation, he/she should first bring the cursor within the rms error threshold before actuating the response button marked with "NA." If automation occurred, the duration was for 15 s, after which manual control was again returned to the pilot.

PCI: In each instance that the pilot did not maintain an rms error of 70 units or less for a period of 250 ms, the system recommended that the pilot automate the tracking subtask. The pilot could choose to accept this recommendation by pulling the trigger on the flight stick. If the pilot ignored the recommendation but continued to perform "poorly" and beyond the rms error threshold, the system continued to recommend automation until tracking returned to within the 70-unit rms error limit. The pilot also had the option of choosing when to exit automation, which was accomplished again via the trigger on the flight stick.

PIA: In this condition, there was no imposed threshold for performance, and the pilot could choose to automate whenever he/she wished. However, in initial instructions to the pilot before each trial, it was recommended that he/she use automation whenever they felt that their performance was deteriorating on the three subtasks, i.e., when 2500 was not consistently being achieved on the fuel management, they were not responding quickly and accurately to monitoring deviations, and tracking performance was not constantly maintaining the cursor near to the central cross-hair target. The pilot's response to this recommendation was evaluated in a postperformance debriefing survey. In general, these differing forms of invocation procedure can be located on a continuum of operator control that is analogous to that represented for operational status in Sheridan's ten levels of automation [21] and Billings and Woods [19] control/management spectrum (see Fig. 2).

2) Warning Modalities: The first within-participant factor was the warning modality, which was used to indicate the change in automation status. These warnings were composed of either a visual, an auditory, or a combined multimodal visual and auditory warning. Each of these forms of warning was crossed completely with the automated tracking display relocation. The auditory warning consisted of a single computer-generated "beep" when automation was invoked. A double "beep" signified that automation had been turned off. The comparable visual warning consisted of a message below the tracking display that indicated "automation on" in the SIA condition, "automation imminent unless negated" in the PCN condition, or "automation recommended" in the PCI condition. When automation occurred, the message read "automation on" for all procedures. When manual control was returned to the pilot, the message then read "automation off" in all methods. In the multimodal condition, these respective warnings were combined. On the basis of cue redundancy, it was anticipated that the combined cue condition would prove to be the most effective. Furthermore, on the basis of attentional distribution, it was expected that the aural cue would prove the next most effective and, on the principle of visual–visual competition, demand that the visual cue would be the least effective of the three conditions that were tested.

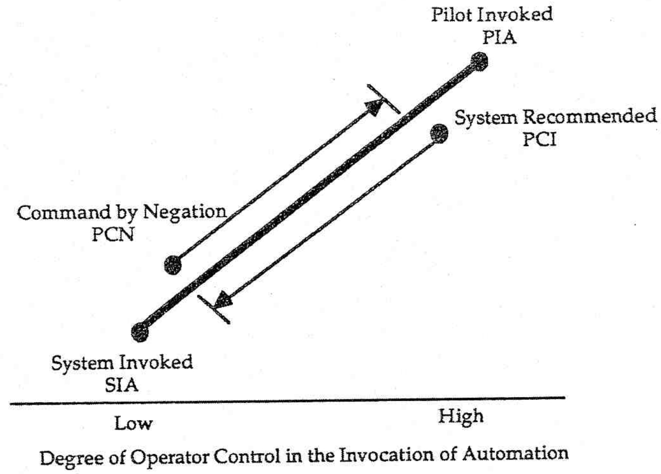


Fig. 2. Level of automation on the vertical axis from high to low versus the degree of operator control. Individual invocation techniques are fitted within this operational space.

3) Display Relocation: When the tracking subtask was changed in status to automated control, the tracking display changed in one of two ways. Either the display could shrink but remain in the same central location, or the display would shrink and be relocated at the lower right of the display screen in a peripheral location [31]. This manipulation was done to examine the effect of a single indicator of automation status (i.e., a shrunken display size) versus two indicators of automation status (i.e., a shrunken and relocated display) on performance. As a within-participant factor, each individual experienced each of these changes in the course of testing.

D. Experimental Procedure

Each participant read and signed the required voluntary consent form for participation. A baseline POMS questionnaire was then administered. Following the completion of this questionnaire and familiarization with the SWAT technique, each pilot was given instruction as to the nature of the three individual subtasks. Practice sessions were then provided for each of the three subtasks individually, as well as an additional practice session on the overall multitask environment until each individual pilot felt comfortable with his/her level of mastery. While there may have been a degree of learning specific to this experimental environment, the stratified random distribution of experimental participants was used to balance such effects across conditions. All relevant experimental conditions were then explained to each participant with instructions for and explicit demonstrations of all three forms of warning as to automation status change as well as demonstrations of each of the automated display locations. Following this instruction, practice again was provided. Each pilot then completed six 5-min trials. At the 4 : 45 mark of each trial, the pilot was asked to provide his/her SWAT response. Following the completion of all experimental trials, the participant was asked again to complete the POMS questionnaire. Finally, the pilot completed a postperformance questionnaire pertaining to his/her flight experience and his/her reactions to the experiment, the testing facility, with an added opportunity for open-ended comments. This paper focused explicitly on automation of the tracking subtask only for two specific reasons. First, previous work [26] demonstrated that the pilots ubiquitously treated the tracking as the subtask of primary importance, and second, automating the tracking was expected to achieve a major goal of automated systems in general, i.e., a reduction in operator workload.

III. Experimental Results

The performance results for the effect of invocation procedure partially confirmed the initial hypothesis. The only performance effects of the between-participant invocation variable were evident exclusively in response on the tracking subtask. In contrast, the only performance effects for either of the within-participant variables, either warning modality or display relocation effects, were seen in the fuel management subtask response. Results also indicated that performance on the monitoring subtask remained essentially stable under manipulations of each of the independent variables. These effects are first considered separately. In each case, data were analyzed using a repeated-measures analysis of variance (ANOVA). When the ANOVA results proved to pass the preset level ($p < 0.05$) for the test of significance, the Tukey pairwise comparison procedure was employed to differentiate specific effects.

A. Overall Tracking Error

The first measure that was considered was the total rms tracking error in each of the respective automation invocation conditions. Thus, rms error for the manual control portion of such tracking performance was calculated for each participant. Results of the repeated-measures ANOVA on these data revealed a main effect for automation invocation procedure ($\mathbf{F}[3, 28] = 10.813$, $p = 0.0001$). *Post hoc* evaluation revealed a significant difference between the SIA and all other procedures, and this pattern is illustrated graphically in Fig. 3. Such an outcome is closely related to the overall time spent in automation, as subsequently indicated.

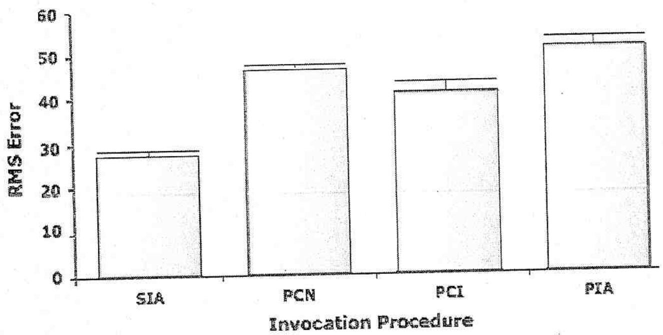


Fig. 3. Tracking rms error by the invocation procedure.

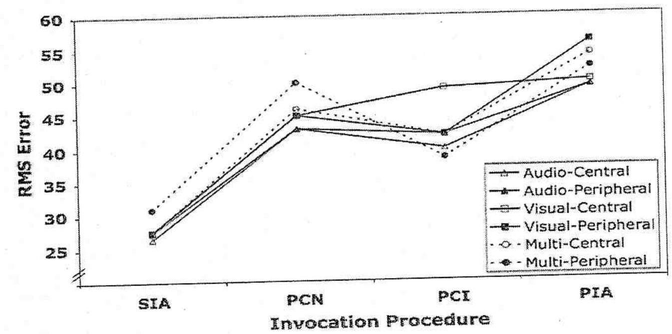


Fig. 4. Invocation procedure by automated display location and warning modality interaction effect. Note that the one exceptional point is the visual/central combination in the PCI condition.

There were no main effects or two-way interactions involving warning format or display location on tracking rms error. However, there was a significant three-way interaction ($\mathbf{F}[6, 56] = 2.462$, $p = 0.0348$), and as shown in Fig. 4, the one obvious exceptional condition was the PCI procedure when the warning was presented visually with the automated display that was placed in the central location. In this display condition, error was elevated as compared to all of the peer combinations that were experienced under the PCI invocation procedure. Such an interaction supersedes the main effects that are associated with this particular manipulation. This pattern accords with that which has been found by others concerning manipulations of adaptive automation [37] and is discussed further in Section III-B.

B. Time to First Automation

One of the facets of tracking response that can be easily assessed is the time that it takes for the first episode of automation to occur. To measure this value, each individual trial was examined, and the time to first automation (in seconds) was recorded. This measure was conceived as an uncontaminated reflection of the between-group invocation procedure effect since the within-subject manipulations (i.e., the warning modality differences and the automated display location differences) were not presented until after any episode of automation began. Data for each trial were entered into an ANOVA with no within-subject factors because of the contingency that was noted. Analysis here indicated a significant effect for invocation procedure ($\mathbf{F}[3, 181] = 6.502$, $p = 0.0003$), and Tukey's procedure distinguished that the SIA group was

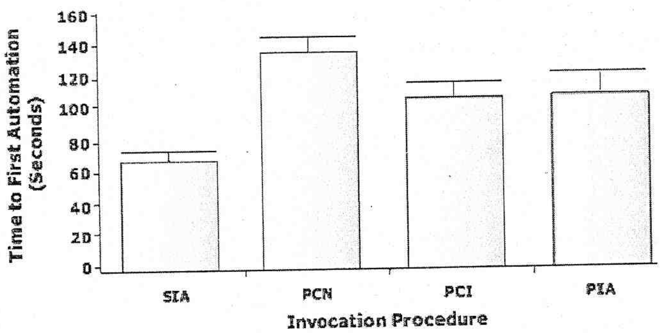


Fig. 5. Time to first automation episode parsed by the automation invocation procedure that was employed.

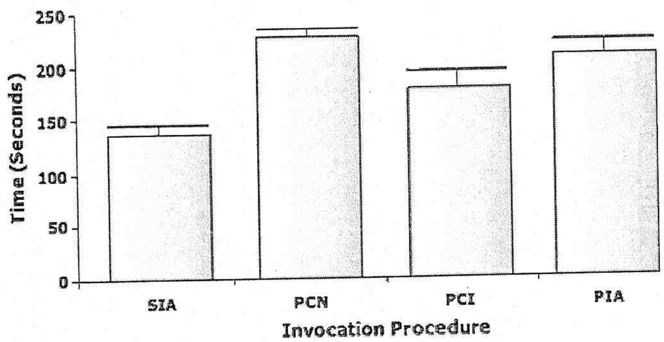


Fig. 6. Time spent in manual tracking control mode parsed by the automation invocation procedure that was employed.

significantly different from all other groups. As shown in Fig. 5, these latter three groups did not differ significantly between themselves.

C. *Time Spent in Automation*

As the procedures for tracking automation were manipulated as a between-participant variable, the different groups spent a different total time in manual control. The interval spent in manual control was calculated for each trial, and these data were entered into a repeated-measures ANOVA. This analysis also revealed a main effect for invocation procedure ($\mathbf{F}[3, 28] = 3.364$, $p = 0.0326$). *Post hoc* tests showed a significant difference between the SIA and PCN groups. Pilots in the SIA group spent significantly less time in the manual tracking mode than the PCN group. The mean time spent in tracking manual control for each automation invocation procedure is shown in Fig. 6. These data cannot be attributed to individual differences in tracking skill across the groups [31]. This possibility was tested by first standardizing the rms error according to the actual time that each pilot spent in manual control. This process yielded a measure of manual tracking performance per unit of time for each trial. These standardized data were then subjected to a repeated-measures ANOVA, which revealed no significant group effects. We can thus conclude that the significant tracking effects were not due to the inherent differences between individuals but were due to the different invocation procedures that were employed.

These findings can be interpreted in a number of different ways. Since intrinsic skill difference across the groups does not appear to be responsible, it is some facet of the invocation procedure that appears to be critical. Such results could be interpreted in terms of a simple timing model in which the automation of the SIA condition differs from the others in terms of the immediacy of its onset, i.e., the automated system was turned on immediately when the threshold was crossed and thus involved no human response activation lag time. However, this pattern of results may be interpreted also in terms of pilot preferences. For example, there is no barrier to the pilot in the PIA condition immediately turning on automation and sustaining it there throughout the whole trial. Evaluation of the current pattern implicates the timing effect and argues that any performance threshold value for control transfer must be most carefully set. Interpretation of these invocation effects has also to be considered in light of the subjective response of the pilots involved, and this overall portraiture is discussed in the following section.

D. *Fuel Management Effects*

The effects for warning modality and display location were assessed, and the only significant main effect occurred for automated display location on fuel management error. The repeated-measures ANOVA revealed a main effect on constant error for the automated tracking display location ($\mathbf{F}[1, 28] = 5.001$, $p = 0.0335$), in which the fuel error for the automated tracking display in the central location exceeded the fuel error when the automated tracking display was in the peripheral location (i.e., 48 versus 10 gal, respectively). In addition, there was a companion main effect for the variability in fuel management error, again for the change in automated tracking display location ($\mathbf{F}[1, 28] = 5.788$, $p = 0.023$). On this occasion, however, the variable error in the central display location was significantly lower than the variable error in the peripheral location (i.e., 94 versus 106 gal, respectively). Thus, participants were closer to the target value on average when the display was shown peripherally, but at the same time, they were significantly more variable in their response as compared to the central display location. No other performance effects for either of the within-participant factors reached the threshold level of significance.

E. *Signal Monitoring Effects*

The response times (in seconds) were obtained for each monitoring deviation, and mean response times were then calculated for each of the experimental conditions. These times were entered into a repeated-measures ANOVA. Results revealed no significant interactions or main effects whatsoever. Similarly, the number of missed monitoring deviations and false responses were recorded and entered in a repeated-measures ANOVA. This analysis also showed no significant interactions or main effects.

F. *Subjective Measures*

Subjective measures of affective demand were collected via the POMS questionnaire and SWAT procedure. In terms of

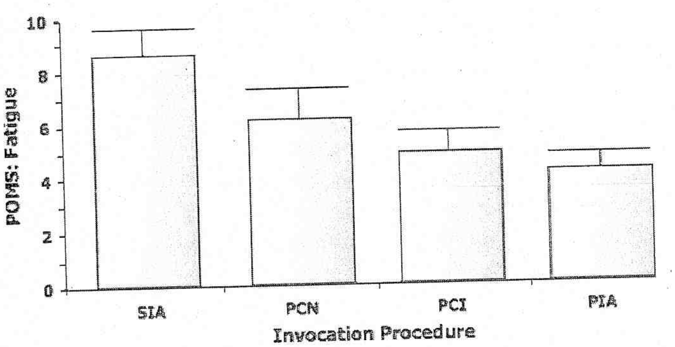


Fig. 7. Values for the Fatigue scale of the POMS assessment technique differentiation by the invocation procedure that was employed.

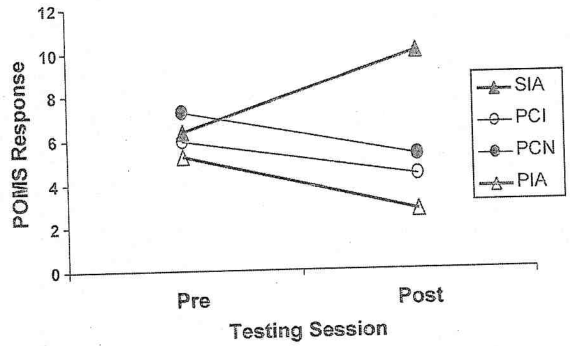


Fig. 8. Preevaluation versus postevaluation of the POMS Fatigue responses showing the evident increase with the SIA as compared to all other invocation procedures.

SWAT results, the only significant result was that for "time" load, which revealed a main effect for the within-participant variable of warning modality ($\mathbf{F}[2, 56] = 3.924$, $p = 0.0254$). *Post hoc* evaluation using Tukey's procedure revealed that the visual warning modality resulted in a significantly higher time load than the multimodal warning, whereas the auditory warning exhibited no difference from either of these time load values. Scores from the pretrial and posttrial POMS questionnaires were recorded for the six scales. A seventh score, reflecting the total mood disturbance, was also obtained by summation across all scales (where the Vigor scale was scored negatively). Pretrial and posttrial scores for the seven scales were analyzed using a repeated-measures ANOVA, and results indicated one significant interaction and three main effects. First, there were significant effects for the Anger ($t[28] = 13.561$, $p < 0.0001$) and the Depression ($t[28] = 7.913$, $p < 0.0001$) subscales, each of which declined across the experimental session. The POMS also showed a significant main effect for invocation procedure ($\mathbf{F}[3, 28] = 3.033$, $p = 0.0457$), which showed a progressive increase in fatigue as the degree of system control increased, and this trend is illustrated in Fig. 7. Since the SWAT values did not vary with invocation procedure, the degree of change of fatigue state is most probably related to the issue of personal situational control or, more generally, the idea of autonomy [38]. This propensity to increase affective levels of fatigue when automation status is under the control of the machine is a further evidence of the importance of user-centered control and user-centered design in the creation of adaptive automation. Finally, there was a significant test session by invocation procedure interaction ($\mathbf{F}[3, 28] = 3.04$, $p = 0.0454$), in which the level of fatigue increased across sessions for the system-initiated invocation, whereas fatigue declined for all other forms of automation invocation, and this pattern of results is shown in Fig. 8.

G. Questionnaire Results

At the end of the experiment, pilots responded to a debrief questionnaire, which posed a number of queries about the testing procedure and also provided the opportunity to express any open-ended comments [39]. Not all participants chose to provide such comments. One specific question concerned the choice of preferred form of tracking automation display. In respect of this, an equal number of pilots chose the change in display size as compared with the change in display size with a change in its location. One pilot expressed a preference for neither a change in display size nor a change in location, whereas another commented that the change in automated subtask display to the peripheral location could "*decrease situation awareness and hamper scan habits.*" This specific pilot was highly experienced in "glass-cockpit" aircraft operations having more than 4000 h in aircraft such as the SAAB 340 and the Jetstream. In respect of issue of display proximity, this pilot makes an important point, and future experimentation needs to address systematically the issue of proximity in respect to automated displays. Such proximity and compatibility tests can be guided by the postulations of Wickens and Carswell [40] on general display configuration. Most pilots preferred the multimodal warning method, which combined the visual message with the auditory warnings. However, one private (VFR-rated) pilot commented that it would have been useful to have had discretionary control over the activation of the warning mode, i.e., to choose the preferred warning mode by varying phase of flight.

Pilots were also asked if they prioritized the three subtasks and, if so, in what order. The overwhelming majority identified the tracking subtask as of primary importance followed by the fuel management and systems monitoring as of equal but secondary priority. Pilots were also asked: "Did tracking automation alter your workload?" They could respond on a scale of 1 (decreased) to 7 (increased), with the response of "no change" corresponding to 4 on the scale. Most pilots indicated that automation decreased his/her workload, with the predominant number of responses in the range of 1–2. The following open-ended comments were also made with respect to the question of workload by the more experienced pilots [his/her flying experience is described briefly in the brackets]:

[16 000 h, B727, DC9] "*. . .automation allowed me to concentrate more on fuel and warning light monitoring.*"

[4500 h] "*. . .learned to rely on* George (the pilot's term for the autopilot)—*blew off hand-flying when aircraft* (tracking system) *became unstable.*"

[12 000 h, BD3100, SF340] "*. . .as the exercise went forward, the tracking became so intense that it was impossible to keep the cursor centered. Automation had to be used.*"

IV. Discussion of Results

If adaptive task allocation and reallocation is posited as one vital form of solution to the problems of interaction with complex automated systems, a critical question must concern *how* the transition of control between manual status and the various conditions of automation is to be accomplished. Furthermore, it is not simply a question of who initiates this transition, but how that transition event is to be communicated to the human operator, during both the actual process of transfer itself and for the duration of the time when the automation is active. The present investigation is part of a program of research [26], [36], [41], [42] that systematically examined the effects of different invocation procedures, the way in which the warning as to state transition was communicated, and the effect that altered display representation of automation had on pilot's response on a representative real-world task. In this context, it is crucial to use individuals with significant professional expertise since task naive individuals exhibit a radically different pattern of response [42].

The principal effects for invocation procedure in this experiment were seen in the results for rms error in the tracking task. Given the primacy that pilots attribute to this task, this effect for the major independent variable is not unexpected. The first outcome of note is the time to the first episode of automation. Here, we can see a strong effect for SIA invocation. Given the absolute difference in time to first automation, this effect cannot simply be due to the inherent lag in automation onset in the recommended conditions (i.e., PCN and PCI). Unlike the results for the summed level of rms error, the difference here is not averaged across the whole trial but represents one unique time for each respective condition. The difference that was exhibited well exceeds any lag in operator response time. Neither can this lag be the causal difference from the PIA condition since, in the latter condition, pilots were free to begin automation whenever they wished. This result, however, tells us that setting the "threshold" for automation takeover is a fundamental design issue. With respect to this time to first automation, the difference that was noted between the SIA condition and all other conditions may well be an expected one. After all, the purpose of having a system-initiated condition is to emphasize the accuracy and precision with which such a configuration can be engineered. However, what is also of interest is the lack of any significant difference between each of the other conditions. This implies that regardless of the different form of invocation procedure, the collective response of the pilots in each of these groups provided very similar times at which automation was first invoked. One possible conclusion here is that the subjective apperception of when automation should be initiated is driven by a common mechanism across the groups in which pilots had a degree of control, and this includes the PCN. Again, this emphasizes the important difference between a machine-driven architecture and an operator-driven architecture as to when automation is necessary.

Although the time to the first automation event has an effect on total time spent in manual control, the overall pattern of response is not the same for the two measures. The SIA and PCI conditions exhibit the lowest value of overall rms error and also the lowest time in manual control. Descriptively, the PCN condition shows the longest time to the first automation episode as well as a slightly higher overall time in manual control. Nevertheless, PCN shows slightly lower overall rms error than the PIA condition. While each of these respective differences individually did not reach the preset level of acceptable significance, in an experiment of the present sort, and particularly in combination, they represent the most interesting trends, which indicate a need for further exploration in relation to the choice of invocation procedure.

In terms of the overall tracking error, the ordering of conditions shows evidently that the SIA condition produced the lowest error, each of the recommendation conditions (i.e., PCI and PCN) produced the next lowest level of error, whereas the PIA condition produced the highest level of error. As previously noted, this pattern of findings is not due to an inherent difference in tracking skill level between the respective groups. An explicit test of this proposition using a measure of error per unit time in manual control excludes this possibility. The tracking results are thus a direct result of differences in invocation procedure, which led to differential intervals in the automated state. Although the SIA condition showed the best level of performance, this superiority was not directly evident in the performance levels of the other associated tasks. This could be due to the higher level of fatigue and workload associated with this condition and would represent a trade of effort for performance. This pattern represents a direct association of performance level and subjective workload response but an evident insensitivity between primary task performance and subsidiary task performance, given that pilots identified the tracking as of the greatest priority [43].

There was one single interaction that involved invocation procedure with the warning modality and display location. This interaction was directly due to the elevated error rate of the visual warning with the automated display at the central location for the pilot command by initiation procedure. This one exception is considered in a later discussion of overall performance. Given the general pattern of findings for tracking, the first experimental hypothesis, i.e., that performance would be differentiated by invocation procedure, is supported here. A subsidiary aspect of this hypothesis, i.e., that SIA would prove significantly different from all other forms of automation initiation, also received support. However, another corollary that the recommended automation invocation would differ from discretionary invocation of automation was not supported. The idea that retaining at least some degree of overall pilot control would be superior to the system-invoked automation was not consistent with the present tracking findings. Although as was evident in the data for the subjective response, gains in tracking performance level are somewhat offset by greater disturbance to operator state.

The model of operator control and invocation procedure that is shown in Fig. 2 displays a linear continuum. However, the tracking data do not confirm this linear interpretation or the implication of an interval-like scale of control. Clearly, SIA distinguishes itself on a number of aspects of tracking response. However, the pilot-invoked procedures involving shared authority were between themselves far less differentiable. Theoretically, the model displays a continuum, but behaviorally, the data indicate much more of a dichotomy between system control and each of the other invocation procedures involving operator control. Thus, describing invocation control alternatives in terms of a linear continuum is a potentially misleading representation. This linear representation may be reasonable as a first-pass theoretical heuristic, although there is no explicit protestation by Sheridan [21] that these levels be interpreted in a strict linear manner. The present pattern of performance outcome indicates that there is a clear separation between system-initiated invocation and all other procedures. Whether this division is specific only to the present multitask assemblage directly applicable to aviation or applies over a wider range of operations of real-world complex systems awaits further evaluation.

The presently reported effects for fuel management confirmed and extended previous systematic findings [26]. In a highly consistent fashion with this previous outcome, the between-participant invocation procedure exerted no significant effect on fuel management error. However, as with the previous findings, there was a systematic effect on fuel management error for display location. With this present expanded sample of professional pilots, it was again observed that the mean level of the fuel management error was significantly higher when the automated display appeared in the central location than in the peripheral location. This confirmatory pattern indicates that these are robust trends, which are not simply the result of the nuances of a single investigational procedure. As with the explanation of the previous findings, the expectation here is that the reason for this pattern of mean fuel management error is due to the differences in the sequence of visual scanning under these differing display conditions. However, the present experiment extends this understanding by including a further analysis of the variability of error as well as the mean level of that error. These additional findings indicated that there was a greater variability in response when the automated display was in the peripheral location. This added facet of the performance outcome further supports the contention of a differential frequency of visual sampling between different display formats. It implies less frequent fixations in the peripheral display condition but more extended dwell times when such fixations are made. This outcome deserves considerable further study using sophisticated eye-tracking methods to evaluate fixation locations and dwell times to fully explore the pilot's response to change in display characteristics with adaptive automation. The null effects for monitoring also partly replicate a component of a previous procedure [26]. In the latter study, there was no effect for display location on gauge monitoring, which is an outcome pattern that is repeated here. However, in the previous study, there were significant differences for monitoring of the light displays, which diverge from the finding in this paper. It is possible to provide any number of potential explanations for such a difference, but resolution clearly awaits further experimental evaluation in this realm.

When these objective outcomes are all we have, then advocacy for pure SIA would appear to be evidently justified. However, in addition to the quantitative measures, we have representations of pilot's qualitative reactions, and these results also have to be factored into the overall appreciation of the multitask experience. As is clear from the POMS fatigue data, the performance advantages of the SIA condition came with a price. In the present situation, we can see that SIA induced greater absolute level of fatigue when compared with all other invocation procedures. More tellingly, this fatigue level was increasing across the testing interval, while all other forms of invocation procedure saw a small decrease in fatigue. Thus, we have a tradeoff between the acute level of momentary performance efficiency and the chronic level of experienced fatigue. It is also probable that the SIA condition could also have reduced situation awareness, as reported by previous investigators [7]. As a general overview of the multitask environment, the present results, in both performance and affective findings, tend to reinforce the idea of the importance of choice. It has been shown that the tasks that remove choice prove most stressful and, in the long term, destructive [44]–[46]. A degree of evidence for this principle is seen in the data reported here.

With respect to warning modalities, it was anticipated that the multimodal warning would be the most effective, simply on the well-known and established principle of sensory signal redundancy [47]. However, the warnings in this experiment were designed to be ecologically valid representations of the sorts of warnings that pilot's actually experience in the cockpit [48]. Therefore, it was not possible to generate visual and aural warnings of exact psychophysical equivalence, and indeed, whether this is even theoretically possible is still subject to dispute. While the present findings show a slight advantage for multimodal expression on one aspect of one of the subjective response scales, there was no evident consistent advantage for any particular form of warning in the objective performance data. Given this set of outcomes, it is probably reasonable to refrain from making any definitive statements about the design of possible warning modality effects from the present procedure and to encourage future work on methods to warn the operator concerning transfer of control between manual and automated modes [19].

One obvious theoretical and practical concern that is raised by the present procedure is the criterion setting for automation initiation. The 15-s interval for which automation persisted here was derived from empirical evaluation of the length of automation that suppressed any unstable oscillations of control switching between operator and system [36]. For example, consider the situation in which the automation suddenly takes control, as in the case of the SIA condition. Immediately, the pilot is offloaded, and the system then, sensing this alleviated state of the pilot, quickly returns manual control. This sudden action of returning control, in turn, instantly overloads the pilot; thus, the system, which is aware of this change, again takes back control. Such continual momentary switching of control represents a form of unstable oscillation that, if it persists, would cause significant decrement in overall performance and, most probably, a radical loss of situation awareness. We have previously suggested [36] that to damp out this unwanted oscillation, there should be an imposed latency during which the automation, having once switched, now cannot switch again. Our prior experimental evaluation [36] identified an interval of 15 s as the standard minimum duration that should be imposed for this moratorium, and hence, this value was used in the present experiment. Of course, there may be exceptions to this interval in practical circumstances, such as when an emergency event like a Traffic Collision Avoidance System (TCAS) or Ground Proximity Warning System (GPWS) occurs in the said interval but under most operational circumstances, this nonswitching moratorium appears justified. If 15 s is the temporal constraint on control switching, what represents the spatial constraint? In the present experiment, the 70-unit rms error threshold was again derived empirically, prior to the beginning of experimentation. In part, this threshold was chosen so that as the tracking became more difficult, there would eventually be the necessity to automate, which is an aim that was achieved given the observations of pilots here. After all, in one sense, too liberal or too conservative, a threshold negates the very idea of adaptation in the first place. However, it is important to understand that this threshold or criterion setting is a design decision. True, it too can be dynamic and set and reset in relation to local conditions, in the fashion of a meta-adaptive strategy. However, given the penetration of adaptive automation into a spectrum of different systems, it is probable that such thresholds will, at first, be largely empirically derived. Just where such a threshold level is set promises to affect which form of invocation procedure is preferred and which is the most effective in terms of objective performance.

V. General Discussion

To the present, discussion has focused on the effects of specific manipulations on specific dependent variables, and this partitioned evaluation has proved informative. However, as a phenomenological experience, the multitask environment is generally appreciated as a unified whole, not as a distinct set of three highly disparate subtasks that are totally separated in space and time [49]. In the laboratory, in simulation, and in real world, it is possible to continue to derive an ever-increasing number of dependent measures from these components of a multitask set. Inevitability, a certain number of these will produce significant effects. The difficulty then is to distinguish meaningful effects from an obligatory probabilistic background of "significant" effects. One guideline in this search concerns the consistency of a pattern of findings as reflected in both the objective and subjective responses of the participants. The present procedure has demonstrated this property, particularly when considered in conjunction with a previous comparable investigation [26].

In multitask environments, we have still yet to develop a satisfactory procedure by which to derive a measure of "overall effectiveness." This latter measure would need to take in to account issues such as the importance and the urgency of each respective subtask component as well as the operator's perception of their momentary relative importance. Optimized measures such as reaction time are only relevant when operators have to respond as quickly as possible. Quite often, such responses can be "satisficed" rather than optimized [50]. Essentially, this means that in aviation, for example, pilots can respond within a window of opportunity, and anywhere within that window, faster response does not necessarily connote better

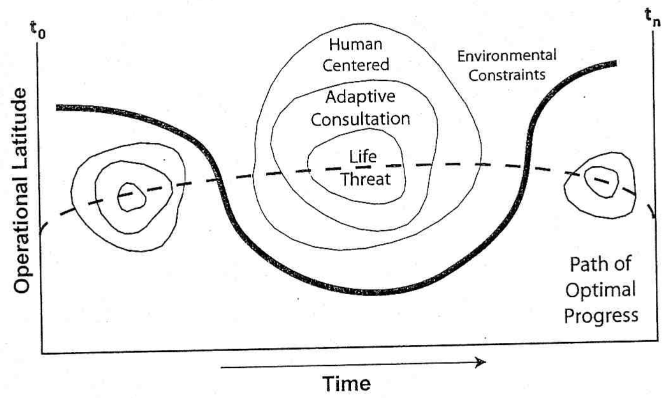


Fig. 9. Descriptive context model for the implementation of differing levels of adaptation. The circular contours represent zones of operation that differ in the level of operator control. The environmental context (the thick line) drives possible levels of adaptation. A path of optimal progress (the dotted line) is sought by the control-sharing algorithm, but occasionally, not all levels of adaptation are available if the mission goal is to be achieved.

response. In constructing any metric of "overall effectiveness," how one trades poor performance on a critical task with good performance on a less important task is still to be effectively determined. Also, how does one integrate subjective response and match changes in such values against objective task performance? It is this form of collective measure that will be most critical in developing functional adaptive systems, which must operate on this overall value and not on any selected piecemeal component. What is clear is that the simple proliferation of the number of response dependent variables does not answer this central question of adaptive automation in complex multitask systems [7], [51]–[56].

Although there have been some initial formulations [57], a quantitative model to answer such questions in adaptive automation is still some way off. In the absence of an effective quantitative model, one viable choice is the presentation of a qualitative alternative, and this conception is presented here in light of the results that were obtained. It is now clear from a programmatic series of work [11], [25], [26] and the systematic observation of other researchers, theorists, and experimenters [7], [16], [17], [24], [37], [51], [54]–[61] that adaptive allocation can be viewed in terms of system of nested envelopes of protection. What drives the system are the needs to achieve preset mission goals. What bounds the system are the constraints on the operator and on the technical capacities to hand and the environmental conditions that are encountered. These latter three factors collectively compose the "environmental constraints." The illustration in Fig. 9 shows a descriptive but dynamic view of the process of adaptive allocation. Initially, the environmental constraints (shown by the thick line) are considerable, but the range of available operational latitude allows any of the three general allocation strategies to be engaged. As time progresses, the environmental constraints are loosened (as, for example, in straight and level flight). Operational latitude increases, and the ability to use more human-centered strategies becomes evident. Finally, as environmental constraints are once again tightened, at the right of the illustration (as, for example, in pilot incapacitation, engine failure, or nap-of-the-earth flight), the possible range of strategies is radically reduced. Note that in each circumstance, preservation of life predominates, and adaptation and subsequently full human-centered approaches are subservient.

The inner region of operational latitude represents a severe system's threat envelope. Conditions that threaten to violate these bounds are largely dealt with by the system level of automation initiation. This form of automation initiation is particularly appropriate in emergency circumstances such as the incapacitation of either pilot or some vital aircraft component, when options to share response between human and machine are highly curtailed. Within these outer boundaries lies a region of consultation. This is typified by the interactive strategies of command by negation and advisories as explored here. Finally, the inner region represents a totally human-centered strategy in which all automation decisions are initiated by the pilot. We would like to maintain the pilot in this region of optimal control by generating a path of progress that avoids the greatest degrees of threat, but of course, this is not always possible as numerous environmental contingencies act to perturb the path of progress. These perturbations are communicated in terms of tasks to be performed, and hence, reconfiguration of the task matrix is periodically needed. The present descriptive model has the advantage of combining all forms of automation initiation strategy that are dependent on context and preserving safety of operation in the outer envelope. As we are better able to specify context and also better able to characterize the momentary "overall effectiveness" of the operator and the system, we will be able to create ever more sophisticated adaptive opportunities. It is the specification of this optimal constraint and capacity-driven pathway that looks to represent the best possibility for the successful future of advance human–machine systems [62], [63].

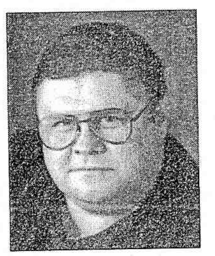
Acknowledgment

The author would like to thank the participants and particularly J. Duley and S. Scallen, as well as T. Oron-Gilad, A. Duley, and J. Ross, for their help and the unknown reviewers for their insightful comments. The views expressed are those of the author and do not necessarily represent the position or policy of any of the named individuals or agencies.

References

[1] C. E. Billings, *Aviation Automation: The Search for a Human-Centered Approach*. Mahwah, NJ: Erlbaum, 1992.
[2] R. M. Fitts, Ed., *Human Engineering for an Effective Air Navigation and Traffic Control System*. Washington, DC: Nat. Res. Council, 1951.
[3] N. Jordan, "Allocation of functions between man and machines in automated systems," *J. Appl. Psychol.*, vol. 47, no. 3, pp. 161–165, 1963.
[4] W. B. Rouse, "Adaptive aiding for human/computer control," *Hum. Factors*, vol. 30, no. 4, pp. 431–443, Aug. 1988.
[5] W. B. Rouse, "Human–computer interaction in multi-task situations," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-7, no. 5, pp. 293–300, May 1977.
[6] W. B. Rouse, "Twenty years of adaptive aiding: Origins of the concept and lessons learned," in *Human Performance in Automated Systems: Current Research and Trends*, M. Mouloua and R. Parasuraman, Eds. Hillsdale, NJ: Erlbaum, 1994, pp. 249–255.
[7] R. Parasuraman, M. Mouloua, and R. Molloy, "Effects of adaptive task allocation on monitoring of automated systems," *Hum. Factors*, vol. 38, no. 4, pp. 665–679, Dec. 1996.
[8] P. A. Hancock and S. F. Scallen, "The future of function allocation," *Ergon. Des.*, vol. 4, no. 4, pp. 24–29, Oct. 1996.
[9] E. L. Weiner and R. E. Curry, "Flight deck automation: Problems and promises," *Ergonomics*, vol. 23, no. 10, pp. 995–1011, 1980.
[10] L. Bainbridge, "Ironies of automation," *Automatica*, vol. 19, no. 6, pp. 775–779, 1983.
[11] P. A. Hancock, M. H. Chignell, and A. Loewenthal, "KBAM: A prototype knowledge-based adaptive man–machine system," in *Proc. 9th Congr. Int. Ergonom. Assoc.*, I. D. Brown, R. Goldsmith, K. Coombes, and M. A. Sinclair, Eds., Bournemouth, U.K., 1985, pp. 103–105.
[12] P. A. Hancock and M. H. Chignell, "Mental workload dynamics in adaptive interface design," *IEEE Trans. Syst., Man, Cybern.*, vol. 18, no. 4, pp. 647–658, Jul./Aug. 1988.
[13] D. O'Hare and S. N. Roscoe, *Flightdeck Performance: The Human Factor*. Ames, IA: Iowa State Univ. Press, 1990.
[14] E. L. Wiener and D. C. Nagel, Eds., *Human Factors in Aviation*. New York: Academic, 1989.
[15] C. Bubb-Lewis and M. W. Scerbo, "Does desire for control affect interactions in an adaptive automation environment?" in *Automation Technology and Human Performance*, M. W. Scerbo and M. Mouloua, Eds. Mahwah, NJ: Erlbaum, 1999, pp. 124–128.
[16] L. J. Prinzel, F. G. Freeman, M. W. Scerbo, P. J. Mikulka, and A. T. Pope, "A closed-loop system for examining psycho-physiological measures for adaptive task allocation," *Int. J. Aviat. Psychol.*, vol. 10, no. 4, pp. 393–410, 2001.
[17] A. T. Pope, E. H. Bogart, and D. S. Bartolome, "Biocybernetic system validates index of operator engagement in automated task," *Biol. Psychol.*, vol. 40, no. 1/2, pp. 187–195, 1995.
[18] R. Parasuraman, "Effects of adaptive function allocation on human performance," in *Proc. FAA/NASA Conf. Artif. Intell. and Human Factors Air Traffic Control and Aviation Maintenance*, Daytona Beach, FL, 1993, pp. 147–157.
[19] C. E. Billings and D. D. Woods, "Concerns about adaptive automation in aviation systems," in *Human Performance in Automated Systems: Research and Trends*, M. Mouloua and R. Parasuraman, Eds. Hillsdale, NJ: Erlbaum, 1994, pp. 264–269.
[20] R. Parasuraman, M. Mouloua, R. Molloy, and B. Hilburn, "The adaptive function allocation for intelligent cockpits (AFAIC) program: Interim research and guidelines for the application of adaptive automation," Naval Air Warfare Center, Warminster, PA, Tech. Rep., 1993.
[21] T. B. Sheridan, *Humans and Automation: System Design and Research Issues*. Santa Monica, CA: Wiley, 2000.
[22] J. G. Morrison and J. P. Gluckman, "Preliminary program plan for the adaptive function allocation for intelligent cockpits (AFAIC) program," Naval Air Develop. Center, Westminster, PA, Tech. Rep., 1991.
[23] M. W. Scerbo, "Theoretical perspectives on automation," in *Automation and Human Performance*, R. Parasuraman and M. Mouloua, Eds. Mahwah, NJ: Erlbaum, 1996, pp. 37–63.
[24] D. B. Kaber, J. M Riley, K. W. Tan, and M. R. Endsley, "On the design of adaptive automation for complex systems," *Int. J. Cogn. Ergon.*, vol. 5, no. 1, pp. 37–57, 2001.
[25] P. A. Hancock, J. A. Duley, and S. F. Scallen, "The control of adaptive function allocation," Univ. Minnesota, Minneapolis, MN, Tech. Rep., HFRL, No. 6, 1994.
[26] P. A. Hancock, "The effects of automation invocation procedure and dynamic display relocation on performance in a multi-task environment," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 37, no. 1, pp. 47–57, 2007.
[27] N. Sarter and B. Schroeder, "Supporting decision-making and action selection under time pressure and uncertainty: The case of inflight icing," *Hum. Factors*, vol. 43, no. 4, pp. 573–583, 2002.
[28] K. Johnson, L. Ren, J. Kuchar, and C. Oman, "Interaction of automation and time pressure in a route replanning task," in *Proc. Int. Conf. Human Comput. Interaction Aeronaut.*, Cambridge, MA, 2002, pp. 132–137.
[29] T. L. Chen and A. R. Pritchett, "Development and evaluation of a cockpit decision aid for emergency trajectory generation," *J. Aircr.*, vol. 38, no. 5, pp. 935–943, 2001.
[30] P. A. Hancock, M. S. Coyle, X. Zheng, and V. Kelker, "The development of the Minnesota systems task allocation research (MINSTAR) facility," Univ. Minnesota, Minneapolis, MN, Tech. Rep. HFRL 93-07, 1993.
[31] S. Miyake, P. Loslever, and P. A. Hancock, "Individual differences in tracking," *Ergonomics*, vol. 44, no. 12, pp. 1056–1068, Oct. 2001.
[32] C. Beedie, P. C. Terry, and A. M. Lane, "The Profile of Mood States and athletic performance: Two meta-analyses," *J. Appl. Sport Psychol.*, vol. 12, no. 1, pp. 49–68, 2001.
[33] D. M. McNair, M. Lorr, and L. F. Droppleman, *Manual for the Profile of Mood States*. San Diego, CA: EdITS, 1992. (Rev. ed.).

[34] G. B. Reid and T. E. Nygren, "The subjective workload assessment technique: A scaling procedure for measuring mental workload," in *Human Mental Workload*, P. A. Hancock and N. Meshkati, Eds. Amsterdam, The Netherlands: North Holland, 1988, pp. 185–218.
[35] P. A. Hancock and N. Meshkati, Eds., *Human Mental Workload.* Amsterdam, The Netherlands: North Holland, 1988.
[36] S. F. Scallen, P. A. Hancock, and J. A. Duley, "Pilot performance and preference for short cycles of automation in adaptive function allocation," *Appl. Ergon.*, vol. 26, no. 6, pp. 397–403, Dec. 1995.
[37] M. P. Clamann, M. C. Wright, and D. B. Kaber, "Comparison of performance effects of adaptive automation applied to various stages of human–machine system information processing," in *Proc. Human Factors Ergonom. Soc.*, 2002, vol. 46, pp. 342–346.
[38] R. A. Karasek, R. S. Russell, and T. Theorell, "Physiology of stress and regeneration in job related cardiovascular illness," *J. Stress*, vol. 8, no. 1, pp. 24–42, 1982.
[39] Y. J. Tenney, W. H. Rogers, and R. W. Pew, "Pilot opinions on cockpit automation issues," *Int. J. Aviat. Psychol.*, vol. 8, no. 2, pp. 103–120, 1998.
[40] C. D. Wickens and M. Carswell, "Proximity compatibility and principle: Its psychological foundation and relevance to display design," *Hum. Factors*, vol. 37, no. 3, pp. 473–494, Sep. 1995.
[41] P. A. Hancock, G. Williams, S. Miyake, and C. M. Manning, "The influence of task demand characteristics on workload and performance," *Int. J. Aviat. Psychol.*, vol. 5, no. 1, pp. 63–85, 1995.
[42] P. A. Hancock, S. F. Scallen, and J. A. Duley, "Initiation and cessation of automation: Location versus configuration change," Univ. Minnesota, Minneapolis, MN, HFRL, Navy Rep. 3, 1993.
[43] P. A. Hancock, "Effect of control order, augmented feedback, input device and practice on tracking performance and perceived workload," *Ergonomics*, vol. 39, no. 9, pp. 1146–1162, Sep. 1996.
[44] R. A. Karasek, "Job demands, job decision latitude, and mental strain: Implications for job redesign," *Administ. Sci. Q.*, vol. 24, no. 2, pp. 285–308, Jan. 1979.
[45] P. A. Hancock and J. S. Warm, "A model of stress and sustained attention," *Hum. Factors*, vol. 31, no. 5, pp. 519–537, Oct. 1989.
[46] P. A. Hancock and P. A. Desmond, Eds., *Stress, Workload and Fatigue.* Mahwah, NJ: Erlbaum, 2001.
[47] M. S. Sanders and E. J. McCormick, Eds., *Human Factors in Engineering and Design*, 7th ed. New York: McGraw-Hill, 1993.
[48] A. Stokes, C. Wickens, and K. Kite, *Display Technology: Human Factors Concepts.* Warrendale, PA: SAE, 1990.
[49] P. A. Hancock, "Hours of boredom, moments of terror,—Or months of monotony, milliseconds of mayhem," presented at the 9th International Symposium on Aviation Psychology, Columbus, OH, 1997.
[50] H. A. Simon, *The Sciences of the Artificial.* Cambridge, MA: MIT Press, 1969.
[51] S. W. A. Dekker and D. D. Woods, "MABA-MABA or abracadabra? Progress on human-automation coordination," *Cogn. Technol. Work*, vol. 4, no. 4, pp. 240–244, 2002.
[52] J. Gluckman, M. Carmody, J. Morrison, E. Hitchcock, and J. S. Warm, "Effects of allocation and partitioning strategies of adaptive automation on task performance and perceived workload in aviation relevant tasks," in *Proc. Int. Symp. Aviat. Psychol.*, R. S. Jensen, Ed, 1993, vol. 7, pp. 150–155.
[53] D. D. Woods, "Cognitive demands and activities in dynamic fault management: Abduction and disturbance management," in *Human Factors of Alarm Design*, N. Stanton, Ed. London, U.K.: Taylor & Francis, 1994, pp. 63–92.
[54] D. B. Kaber and M. C. Wright, "Automation-state changes and sensory cueing in telerobot control," presented at the International Ergonomics Association Society, 2003.
[55] N. Moray and T. Inagaki, "Adaptive automation, trust, and self-confidence in fault management of time critical tasks," *J. Exp. Psychol.*, vol. 6, no. 1, pp. 44–58, 2000.
[56] D. B. Kaber and J. M. Riley, "Adaptive automation of a dynamic control task based on secondary task workload measurement," *Int. J. Cogn. Ergon.*, vol. 3, no. 3, pp. 169–187, 1999.
[57] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, "A model for types and levels of human interaction with automation," *IEEE Trans. Syst., Man, Cybern.*, vol. 30, no. 3, pp. 286–297, May 2000.
[58] B. Hilburn, R. Molloy, D. Wong, and R. Parasuraman, "Operator versus computer control of adaptive automation," in "The Adaptive Function Allocation for Intelligent Cockpits (AFAIC) Program: Interim Research and Guidelines for the Application of Adaptive Automation," Naval Air Warfare Center, Warminster, PA, pp. 19–24, Tech. Rep. NAWCADWAR-93031-60, 1993.
[59] C. A. Miller and M. Hannen, "The rotorcraft pilot's associate: Design and evaluation of an intelligent user interface for a cockpit information," *Knowl.-Based Syst.*, vol. 12, no. 8, pp. 443–456, Dec. 1999.
[60] C. A. Miller, "Human–computer etiquette: Managing expectations with intentional agents," *Commun. ACM*, vol. 47, no. 4, pp. 31–34, Apr. 2004.
[61] N. Sarter and D. D. Woods, "Pilot interaction with cockpit automation II: An experimental study of pilots' model and awareness of the flight management system," *Int. J. Aviat. Psychol.*, vol. 4, no. 1, pp. 1–28, 1994.
[62] P. A. Hancock and J. L. Szalma, "The future of neuroergonomics," *Theor. Issues Ergon. Sci.*, vol. 4, no. 1/2, pp. 238–249, Jan. 2003.
[63] R. Parasuraman, "Neuroergonomics: Research and practice," *Theor. Issues Ergon. Sci.*, vol. 4, no. 1/2, pp. 5–20, Jan. 2003.



Peter A. Hancock (A'05–M'06) received the Ph.D. degree from the University of Illinois, Urbana, in 1983.

He is a Provost Distinguished Research Professor in the Multidisciplinary University Research Initiative/Operator Performance Under Stress Research Laboratory, Department of Psychology, and Institute for Simulation and Training, and the Department of Civil and Environmental Engineering, University of Central Florida, Orlando. He has authored more than 400 refereed publications. His current work concerns the evaluation of behavioral response in high-stress situations. His theoretical works concern humans and technology and the possible future for this symbiosis.

Prof. Hancock is a Fellow and past President of the Human Factors and Ergonomics Society. He received the Sir Frederic Bartlett Medal of the Ergonomics Society of Great Britain for lifetime achievement in 2000. He received the Franklin Taylor Award of the American Psychological Association and the Jerome Ely Award of the Human Factors and Ergonomics Society in 2001. He received the Jastrzebowski Medal of the Polish Ergonomics Society for contributions to world ergonomics in 2002. He received the Liberty Mutual Medal of the International Ergonomics Association in 2003. He received the Norbert Wiener Award of the IEEE Systems, Man, and Cybernetics Society in 2006.