callen, S.F., Duley, J.A., & Hancock, P.A. (1994). Pilot performance and preference for cycles of automation in adaptive function allocation. In: M. Mouloua and R. Parasuraman (Eds.).Human performance in automated systems: Current research and trends. (pp. 154-160), Erlbaum: Hillsdale, N.J.

Pilot Performance and Preference for Cycles of Automation in Adaptive Function Allocation

Stephen F. Scallen, Jacqueline A. Duley, & Peter A. Hancock

Human Factors Research Laboratory
University of Minnesota

ABSTRACT

The present experiment examined pilot response to short duration cycles of automation in an adaptive task allocation context. The purpose of the work was to seek acceptable minimal cycle times for automation take-over of sub-tasks in a multi-task environment. We used MINSTAR, a purpose built multi-task environment, which consisted of tracking, fuel management, and monitoring sub-tasks. Six experienced pilots performed the three-part task for nine 5-minute trials. Monitoring and fuel management sub-tasks were performed manually in all conditions. The tracking sub-task cycled between manual and automated control at fixed intervals of 15, 30, and 60 seconds. These cycle times were completely crossed with three levels of tracking difficulty giving the nine within-subject conditions. Performance was measured on each of the three component sub-tasks, while subjective workload, and mental fatigue were also evaluated. Both tracking difficulty and cycle duration had an effect on tracking performance. Tracking efficiency decreased with difficulty level and increasing cycle duration. Fuel management and monitoring sub-tasks appeared to be unaffected by tracking difficulty and automation cycle length. These results are examined in light of the determination of optimal cycle times in complex systems employing adaptive automation and task allocation.

INTRODUCTION

The traditional approach to task allocation in multi-task environments is a static strategy which prescribes the specific task and time appropriate for human or machine control (Fitts, 1951). This strategy limits the sharing of tasks since the allocation decisions are made *a priori*, but is effective if task demands are mostly predictable. What is problematic is the occurrence of unforeseen or unanticipated conditions in which a change in the agreed policy is desirable during actual performance. Thus, the real-time allocation of functions between the human operator and automated sub-systems forms the basis for the adaptive allocation strategy (Hancock & Chignell, 1989; Parasuraman, Bahri, Deaton, Morrison, & Barnes, 1990). However, as might be expected, adaptive allocation provides a considerable theoretical and practical challenge since context contingent action is necessary. Thus, an important element is the identification of meanings and derivations of context which must be specified. Here we examine four ways in which 'context' can be specified. Essentially the discussion of context seeks to identify a trigger or criterion that mediates task allocation between the human and machine.

First, 'context' may be specified by the environment. However, the assessment of all properties of the environment related to the performance of the pilot and the aircraft is unlikely simply because of the proliferation of possible conditions. Such assessment would surely be impractical since the information necessary would overwhelm any processing ability. Second, 'context' may be specified by the aircraft, where the trigger for task allocation is a function of the performance limitations of the aircraft itself. At the present time this approach is perhaps the most advanced because of the plethora of information readily available about the performance of the aircraft. This strategy is enacted in some current 'fly by wire' aircraft such as the A320 Airbus. While this approach may work for the mild and placid conditions of commercial transportation, it is a more problematic question in high performance

Scallen, S.F., Duley, J.A., & Hancock, P.A. (1994). Pilot performance and preference for cycles of automation in adaptive function allocation. In: M. Mouloua and R. Parasuraman (Eds.).Human performance in automated systems: Current research and trends. (pp. 154-160), Erlbaum: Hillsdale, N.J.

tactical operations. Currently, there are few task allocation strategies mediated by the context of the aircraft.

Third, 'context' may be specified completely by pilot competence. That is, changes in allocation are predicated by momentary assessment of pilot behavior. In this sense, the context is the momentary capability of the pilot. We might also take flight performance directly as the critical metric of instantaneous pilot ability. This has extensive face validity since the overall goal is the efficient performance of the system. However, there is a fatal flaw in this simple reasoning. The purpose of the pilot in the high performance aircraft is to perform those functions not easily replicated by the machine. Also, the adaptive allocation strategy is purposely designed to ease the impact of unusual conditions. Therefore, we cannot specify a single goal of the system in the very circumstances for which we are advocating adaptive task allocation and interchange. As we will not be able to specify the goal, we cannot tell what is 'efficient' and what is 'inefficient' flight performance. We can provide flight envelope protection, e.g., terrain avoidance, but the momentary interchange of tasks cannot be founded on external performance alone. Hence, we have to have a further assessment of pilot state. In the past and in present work, we have advocated the use of perceived workload to fulfill this function (Hancock, Duley, Scallen, 1993).

Eventually, we, like others (e.g., Morrison, Cohen, & Gluckman, 1993) propose that the information upon which dynamic allocation of task demands is predicated will depend upon an interaction between human and machine status. Thus knowledge about the pilot's momentary performance and energetic state will be combined with information about aircraft (and potentially mission) status to initiate change. This statement represents the fourth, and we believe most viable, way in which 'context' can be specified. That is, a property of the interaction of pilot, aircraft and environment. In principle, this view has been advocated also by Hilburn, Molloy, Wong, and Parasuraman (1993). Thus, task allocation changes are envisaged to occur based on an algorithm triggered by inputs from pilot, aircraft, and environmental sources. We can well imagine conditions in which the inputs to that algorithm reach a threshold value for triggering allocation change. After only a brief instant, performance conditions stabilize to such an extent that manual control is re-initiated. However, the workload associated with re-capturing manual control is sufficient to trigger re-allocation again, and so on. There exists, therefore, the potential that the complete system will constantly border on the threshold for change, producing an uncontrolled and uncontrollable oscillation of allocation changes. These changes are referred to as cycles of automation and are defined by the frequency with which automation is turned on/off over a period to time. If uncontrolled, oscillation of adaptive allocation would prove particularly destructive to overall performance. It has been previously determined that excessively short automation cycles can limit the effectiveness of adaptive automation in enhancing operator performance of both primary flight and monitoring tasks (Hilburn, Molloy, Wong, & Parasuraman, 1993). Furthermore, the possible distraction and disturbance to the pilot may inspire him/her to simply shut the system off. Hence, the failure to understand pilot preferred rate of task re-allocation (cycles) might obviate the whole purpose of adaptive allocation. Therefore, the purpose of the present experiment is to examine pilot response to differing frequency oscillations in the automation of a tracking sub-task. We seek guidelines as to acceptable oscillation rates and approaches to suppression of unacceptable rates.

METHOD

Experiment Participants

Six rated pilots (5 male, 1 female) were solicited for participation in the study. The pilots had mean age of 43.3 years and a mean of 3708 total flying hours. Three pilots indicated primary experience with small single or double engine planes (i.e., Cessna), two pilots were primarily Navy aviators (i.e., F-16), and one pilot was employed by a major Mid-west passenger airline (747, 727). All participants were in professed good health at the time of testing.

MINSTAR Test Facility

The MINSTAR multi-task environment served as the experimental platform. We have described this facility, in detail, previously (Hancock, Scallen, & Duley, 1993). Briefly, it presents a multi-task environment in which two dimensional compensatory tracking, fuel management, and monitoring are presented as individual sub-tasks. These sub-tasks can be performed individually, in combination, or under varying modes of automation as required by the experimental procedure (Wiener & Curry, 1980). The two dimensional compensatory tracking sub-task utilizes a crosshair which moves on the sum of seven preselected sine waves throughout the target area. The goal for the pilot is to make corrective movements via a joystick in order to bring the moving cursor in alignment to a fixed target cursor at the center of the display. The difficulty of the task can be manipulated by modifying the amplitude and frequency of the sine waves. The difficulty manipulation is accomplished via predetermined programming scripts. The fuel management sub-task displays 5 rectangular shaped fuel tanks connected by 6 fuel pumps. The two outermost tanks are targeted as the goal tanks and deplenish at a constant rate. The goal for the pilot is to manually control the 'on' or 'off' status of the pumps in order to maintain a target level of fuel in the two goal tanks. The difficulty of the task can be manipulated by initiating failure(s) of the fuel pumps. The difficulty manipulation is accomplished via predetermined programming scripts. The monitoring sub-task displays 5 lights and 4 graduated gauges. The goal for the pilot is to reset the lights or gauges whenever they deviate from their normal status. Monitoring deviations are controlled via predetermined programming scripts.

Experimental Conditions and Design

Three automation durations were selected for the tracking sub-task (15 sec., 30 sec., 60 sec.). The 15 sec. cycle duration was conceived as a realistic lower boundary for automation cycles. The 60 sec. cycle duration was conceived as the upper boundary for the examination of short cycle automation as intended in this study. The 30 sec. cycle was conceived as intermediary to the other two cycles. Three levels of difficulty were selected for the manual portions of the tracking sub-task. Levels were chosen based on a test study of all tracking difficulty levels available for the MINSTAR tracking sub-task. The three significantly different levels were labeled Low, Medium, and High.. Thus, the present study was conducted as a within subjects 3 X 3 (duration vs. difficulty) repeated measures design. To ensure that the subject monitored the tracking task during its automated cycle a 10% automation failure rate was introduced. Pilots were cued to the automated tracking mode by a change in the tracking display and configuration.

Experiment Procedure

All pilots began the experimental session by signing an informed consent and responding to the Profile of Mood States inventory (McNair, Lorr, & Droppleman, 1971) The 65 items of the POMS are scored to make up scales measuring six identifiable mood or affective states: Tension, Depression, Anger, Vigor, Fatigue, and Confusion. Pilots practiced each sub-task separately and then practiced the combined sub-tasks. After the practice session subjects completed the 9 five-minute trials. Inter-trial periods were approximately 1 minute. Order of conditions was counterbalanced across subjects. On line subjective measures of time load, stress level, and mental effort were obtained via the Subjective Workload Assessment Technique (SWAT) at the 4:45 mark or each 5 minute trial. The details of this procedure have been given by Reid and Nygren (1988). Prior to the first, forth, seventh, and at the end of the ninth trial, data for

Scallen, S.F., Duley, J.A., & Hancock, P.A. (1994). Pilot performance and preference for cycles of automation in adaptive function allocation. In: M. Mouloua and R. Parasuraman (Eds.).Human performance in automated systems: Current research and trends. (pp. 154-160), Erlbaum: Hillsdale, N.J.

critical flicker frequency (CFF) was obtained as a measure of mental fatigue. CFF has been used in the past as an indicator of fatigue across levels of difficulty of repetitive or vigilance-type tasks (Baschera & Grandjean, 1979). The CFF instrument was placed inside the cockpit at an angle of approximately 62 degrees to the horizon. All pilots viewed the instrument head-on. Following the nine trials, a post-trial POMS questionnaire and a subject debrief survey were administered. Monitoring task performance was quantified as the response time, in seconds, to each light or gauge deviation. The number of missed signals and false responses were also recorded. Fuel management task performance was quantified as the absolute mean deviation of fuel level for the two goal tanks, sampled every 5 seconds. Error was averaged for the two measured tanks. Tracking performance was quantified as the deviation of the crosshair from the center of the display, as a root mean square (RMS) error, sampled every second.

RESULTS

Tracking

A mean RMS error was calculated for each subject, for each of the nine conditions. The mean RMS error for each subject were entered into a 3 X 3 (difficulty X automation duration) repeated measures ANOVA. Results of the ANOVA for RMS error data indicated a main effect for difficulty ($F(2,10)=48.33$, $p<.01$) and a main effect for automation duration ($F(2,10)=6.54$, $p<.05$). Post-hoc follow up tests, using Tukey's procedure, were conducted for main effect difficulty and revealed significant differences between all difficulty levels (low, medium and high). Post-hoc follow up tests, using Tukey's procedure, were conducted for main effect automation duration and revealed significant differences between the 15 second duration and the 60 second duration. Data for tracking RMS error by automation duration, are displayed in Figure 1.

Monitoring

Each five minute trial contained ten monitoring deviations (2 per minute). Response times, in seconds, were obtained for each monitoring deviation and mean response times were calculated for each of the nine conditions. Mean response times for each subject were entered into a 3 X 3 (difficulty X automation duration) repeated measures ANOVA. Results of the ANOVA for response times indicated no significance for either main effect or the interaction. A similar analysis was conducted for the number of missed monitoring deviations for each condition. The repeated measures ANOVA revealed no main effects or interaction.

Fuel Management

The level of fuel was averaged for the two goal tanks. Fuel management error for each condition was calculated as the absolute deviation from 2500, summed and averaged for all samples in a condition. The obtained fuel management error for each subject, for each condition, was entered into a 3 X 3 (difficulty X automation deviation) repeated measures ANOVA. Results of the ANOVA revealed no trends for either main effects or the interaction.

Affective Measures

A response value was obtained for the SWAT time load, stress level, and mental effort questions. Data were analyzed via a 3 X 3 (difficulty X automation duration) repeated measures ANOVA, for each of the questions. The analyses for time load and mental effort revealed marginal significance for main effect automation duration ($F(2,10)=3.75$, $p=.06$ and $F(2,10)=3.29$, $p=.08$, respectively). No effects were demonstrated for stress level. Data for subjective measures of time load and mental effort, by automation duration are presented in Figure 2. Pre and post trial POMS questionnaires were scored for the six scales according to the

Scallen, S.F., Duley, J.A., & Hancock, P.A. (1994). Pilot performance and preference for cycles of automation in adaptive function allocation. In: M. Mouloua and R. Parasuraman (Eds.).Human performance in automated systems: Current research and trends. (pp. 154-160), Erlbaum: Hillsdale, N.J.

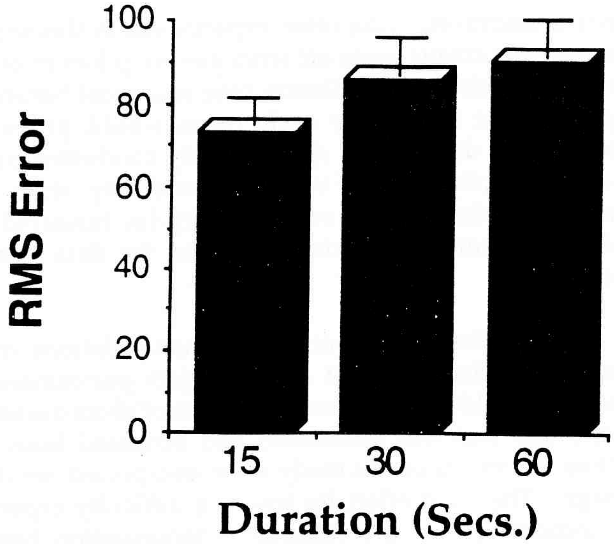

Figure 1. Tracking performance, as represented by RMS Error, by duration of the automation cycle. Standard error bars are shown.

instruction manual. A seventh score, reflecting total mood disturbance, was obtained by summing across all scales (scoring Vigor negatively). Pre trial and post trial scores for the seven scales were subjected to matched pairs t-tests. Results indicted significant differences for Anger and Depression scales ($t(5)=2.31, p=.06$ and $t(5)=2.69$, $p<.05$, respectively). Critical flicker frequency data obtained for the four testing periods was subjected to a repeated measures ANOVA. No significant differences were demonstrated.

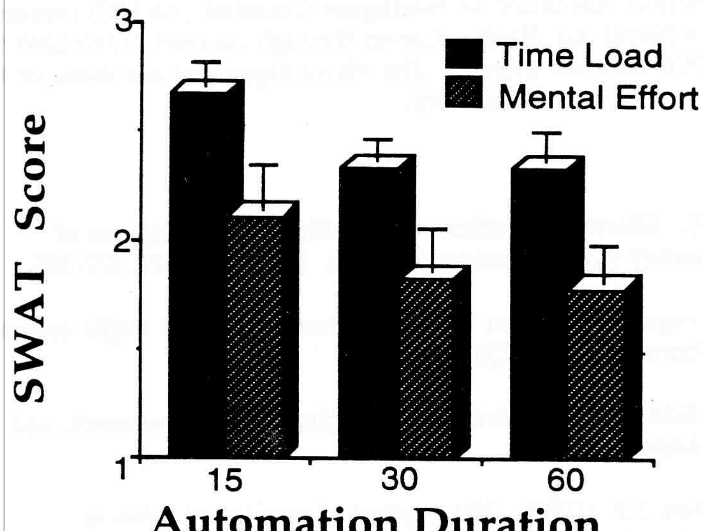

Figure 2. Pilot's responses to the questions: "What is your time load?" and "What is your mental effort?'. A high response indicates the perception of increased time load or mental effort.

DISCUSSION and CONCLUSIONS

There are a number of important observations which come from this experiment. First, there was a trade of performance for workload. That is, performance was best at the 15 sec. cycle time, but workload was also increased at this cycle time. Thus, both mental effort and temporal demand increased as performance improved. Like other experiments in this sequence (Duley, Scallen, & Hancock, 1993), this suggests covert trade-off strategies by pilots in order to sustain efficient performance. The fact that the shortest oscillation time was most beneficial is also counter-intuitive. We had suspected that the longer cycle times would prove most efficacious while short cycle times would prove distractive. A pilot study conducted prior to the onset of this study demonstrated that pilots were very distracted by short cycle automation. Most pilot subjects complained bitterly that very short cycles hindered their ability to re-acquire manual control after an automation cycle. However, the data from this experiment did not support this assertion.

A second important observation was that the two experimental manipulations for the tracking sub-task (difficulty, automation duration) did not effect subject performance on monitoring or fuel management sub-tasks. It would appear, then, the effects of short duration of automation were confined to the specific task that was automated and insulated from other tasks in the multi-task environment. While the results of this study were unexpected, we do not believe they were an artifact of the design. The main effect for tracking difficulty expectedly produced differences for each level, indicating that the difficulty manipulation behaved exactly as we had planned. Concerning the automation duration manipulation, subjective data indicated that pilots perceived increased levels of time load and mental effort for the shortest duration, again, as we had planned. The authors of this study strongly recommend increased research efforts in the area of cycle duration, especially in extremely short cycles (15 secs. or shorter) and cycles between 60 sec and five minutes. These suggested areas of research are currently underrepresented in research concerning the adaptive automation context.

ACKNOWLEDGMENTS

We would like to thank the pilots who participated in this study and other studies involving the Navy Adaptive Function Allocation in Intelligent Cockpits (AFAIC) research. This research was supported by the Naval Air Warfare Center through contract N/N62269-92-0211. Dr. J. Morrison was the NAWC contract monitor. The views expressed are those of the authors and do not necessarily represent those of the Navy.

REFERENCES

Baschera, P., & Grandjean, E. (1979). Effects of repetitive tasks with different degrees of difficulty on critical fusion frequency (CF) and subjective state. *Ergonomics, 22,* 377-385.

Fitts, P.M. (1951). (Ed.). *Human engineering for an effective air navigation and traffic control system.* Washington, D.C., National Research Council.

Hancock, P.A. & Chignell, M.H. (Eds.). (1989). *Intelligent interfaces: Theory, research, and design.* Amsterdam: North Holland Press.

Hancock, P.A., Duley, J.A., & Scallen, S.F. (1993). *The response of experienced pilots to interface configuration changes for adaptive allocation.* (Technical Report). Naval Air Warfare Center, Warminster, PA.

Hancock, P.A., Scallen, S.F., & Duley, J.A. (1993). *Initiation and cessation of automation: Location versus configuration change.* (Technical Report). Naval Air Warfare Center, Warminster, PA.

Hilburn, B., Molloy, R., Wong, D., & Parasuraman, R. (1993). *Operator versus computer control of adaptive automation.* (Technical Report NAWACADWAR-93031-60). Naval Air Warfare Center, Warminster, PA.

Scallen, S.F., Duley, J.A., & Hancock, P.A. (1994). Pilot performance and preference for cycles of automation in adaptive function allocation. In: M. Mouloua and R. Parasuraman (Eds.).Human performance in automated systems: Current research and trends. (pp. 154-160), Erlbaum: Hillsdale, N.J.

McNair, D.M., Lorr, M., & Droppleman, L.F. (1971). *Profile of mood states manual.* San Diego, CA: Educational and Industrial Testing Services.

Morrison, J.G., Cohen, J.P., & Gluckman, J.P. (1993). *Prospective principles and guidelines for the design of adaptively automated crewstations.* (Technical Report NAWACADWAR-93031-60). Naval Air Warfare Center. Warminster, PA

Parasuraman, R., Bahri, T., Deaton, J.E., Morrison, J.G., & Barnes, M. (1990). *Theory and design of adaptive automation in aviation systems.* (Technical Report). Naval Air Development Center, Warminster, PA.

Parasuraman, P., Molloy, R., & Singh, I.L. (1993). Performance consequences of automation-induced "complacency". *The International Journal of Aviation Pychology, 3*(1), 1-23.

Reid, G.B., & Nygren, T.E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P.A. Hancock & N. Meshkati (Eds.). *Human mental workload.* (pp. 185-218). Amsterdam: North Holland.

Wiener, E.L., & Curry, R.E. (1980). Flight-deck automation: Promises and problems. *Ergonomics, 23,* 995-1011.

Scallen, S.F., Duley, J.A., & Hancock, P.A. (1994). Pilot performance and preference for cycles of automation in adaptive function allocation. In: M. Mouloua and R. Parasuraman (Eds.).Human performance in automated systems: Current research and trends. (pp. 154-160), Erlbaum: Hillsdale, N.J.