Pilot performance and preference for short cycles of automation in adaptive function allocation

S. F. Scallen, P. A. Hancock and J. A. Duley

Human Factors Research Laboratory, University of Minnesota, 1901 Fourth St S.E., Minneapolis, MN 55455, USA

The present experiment examined pilot response to the rapid cycling of automation. The experiment was conducted using a multi-task simulation environment consisting of tracking, fuel management, and system monitoring sub-tasks. 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 either 15, 30 or 60 sec. These cycle times were completely crossed with three levels of tracking difficulty giving nine within-subject conditions which lasted 5 min each. Performance was measured on each of the sub-tasks, as was pilot fatigue level and subjective workload for the respective conditions. Results indicated that both difficulty and cycle duration significantly affected tracking performance which was degraded with task difficulty and longer cycle times. Fuel management and system monitoring performance were unaffected by tracking difficulty and automation duration. However, a subsequent analysis was conducted using the 15 sec period immediately following each automation episode as a 'window' of performance. A different pattern of results was observed. Tracking performance was similarly affected by difficulty, but was no longer affected by cycle duration. Furthermore, fuel management error indicated a trend toward better performance in low difficulty conditions. Results illustrate micro trade-offs within sub-tasks and macro trade-offs between sub-tasks. Overall, the results support the contention that excessively short cycles of automation prove disruptive to performance in multi-task conditions.

Keywords: automation, cycles of automation, workload, adaptive task allocation

In ergonomics, the traditional approach to task allocation describes the capabilities of the human versus the machine, and attempts to divide tasks upon the basis of this comparison. In the context of its early formulations (Fitts, 1951) this was a reasonable strategy given that many systems under consideration presented relatively consistent demands (see also Craik, 1947). However, this traditional approach is less effective when unexpected conditions are encountered, especially where a shift in the allocation of tasks is desired during actual performance. This desire has prompted the development of a dynamic adaptive approach, in which the task allocation profile between the human and the machine changes as a function of performance. This new approach is described as adaptive because the control of the onset and the offset of specific tasks, depending on load, is shared between the human and the machine. It is dynamic because changes in task allocation are proposed to occur in real-time (Hancock & Chignell, 1987; 1989; Morrison *et al*, 1993; Parasuraman *et al*, 1990). The present work focuses on this adaptive approach in the realm of aviation operations. It is proposed that dynamic allocation will enhance pilot

performance by better management of information flow and task demands, allocating pilot's resources appropriately and continuously over time. However, enacting adaptive allocation represents a considerable practical challenge since this approach requires an input or trigger to initiate the shift in allocation. There is a number of potential contexts in which a trigger can be specified.

In the first context, some threshold or component of the operational environment may be specified as a trigger. Identified objects, unidentified objects, environmental perturbations or threshold levels of some property are examples of environmental triggers. However, the use of the environmental context to define adaptive triggers is practically limited. The effectiveness of such an approach would necessarily rely on the ability to predict most, if not all, environmental characteristics. Unfortunately, the assessment of all possible properties of the environment is unlikely simply because of the proliferation of potential conditions relating to operator performance. Further, such assessment is impractical since the information necessary would rapidly overwhelm the processing abilities

of both human and machine. Alternately, a second context may be specified in relation to the aircraft itself, where the trigger for task allocation is a function of the aircraft's aerodynamic limitations. At present, this approach is perhaps the most advanced because of the plethora of information readily available about the performance of the aircraft. This strategy is currently employed in some safety features of 'fly by wire' aircraft which prevent the pilot from generating dangerous conditions such as a stall. While this approach may work for the relatively stable conditions of commercial transportation, it is a much more problematic question in high-performance tactical aircraft. Attempts to define 'safe' limits of tactical operations are paradoxical in light of the fact that these operations by their very nature are 'unsafe'.

A third way in which context may be specified is through pilot competence. In this strategy, changes in allocation are predicated upon the momentary assessment of pilot behavior. Pilot performance is monitored by the system where the violation of some performance criteria triggers a shift in task allocation. This procedure has high face validity since the overall goal is the efficient performance of the human–machine system in toto. However, there is a flaw in this reasoning. The purpose of the pilot in high performance aircraft is to perform those functions not easily replicated by the machine. Many of these functions are intimately linked to the reaction and decision-based response to unexpected or unusual conditions. The inability to foresee all conditions and interactions of conditions which bound pilot behavior obviates the formalization of human performance goals. As we are unable to specify deterministically all the goals and reactions of the pilot, we cannot tell what is 'efficient' and what is 'inefficient' flight performance in different circumstances (see Branton, 1987). We can provide flight envelope protection, e.g. terrain avoidance, but the momentary interchange of tasks within this envelope cannot be founded on performance alone. Hence, we need a further assessment of pilot state beyond performance capability alone. We have previously advocated the use of perceived workload to fulfill this function (Hancock & Chignell, 1987; Hancock et al, 1993; Hancock and Meshkati, 1988).

We propose that the most logical and practical context for the identification of an adaptive trigger is the fourth approach which lies in the interaction between the human and the machine (see also Morrison et al, 1993). Thus, knowledge about the pilot's momentary performance and energetic state is combined with information about aircraft status (and potentially mission status) to initiate allocation change (see also Hilburn et al, 1993; Raeth et al, 1994). When the combination of information from these different sources exceeds some threshold in the algorithm, tasks are allocated to the system. However, the interaction context is not without its drawbacks. We can imagine conditions in which the inputs to that algorithm reach a threshold value and trigger automation; but systems like those found in tactical aircraft have exceptional output capabilities and after only a brief instant, performance conditions may stabilize to such an extent that manual control is returned to the pilot. However, the workload associated with re-capturing manual control might be sufficient to trigger another episode of automation. Consequently, there exists a potential for the human–machine system to border on the threshold for re-allocation. This could produce a potentially uncontrollable oscillation of manual and automated control. We refer to these conditions as automation cycling which is defined as the frequency of automation change over a specific time period. If uncontrolled, the oscillation between manual and automated control could prove particularly destructive to overall performance. Practically, short episodes of automation may also prove so distractive that the pilot may simply shut the system off. Hence, the failure to understand pilot response to short episodes of automation might obviate a fundamental purpose of dynamic adaptive allocation, i.e. the regulation of pilot workload and the optimization of pilot and aircraft performance.

There is little research which has examined cycling of automation in any systems. Those researchers who have examined short cycles of manual and automated control have typically not examined periods of less than 10 min (see Glenn et al, 1994; Hilburn et al, 1993; Parasuraman et al, 1992). Generally, these studies have failed to demonstrate evidence of short cycle deficits but did confirm the existence of automation benefits for tasks concurrent with automated tasks. While we believe that this research marks an important first step in identifying the effects of short cycle automation, existing research contains two features which may limit its ability to identify many salient performance characteristics.

First, short episodes of automation have seldom been examined with blocks of automation less than 10 min. We believe that the definition of 'short' automation should practically be extended to include durations much shorter than 10 min. The Gulf War provided a pertinent example of an extremely short cycle of automation. It has been reported that the total time over target was frequently under 30 sec. During this period, pilots off-loaded other tasks to focus on the delivery of the weapon payload, providing an episode of automation of 30 sec or less. We believe that extending experimental analysis of short episodes of automation to include durations of less than 1 min has practical significance for tactical fighter aircraft and for many other systems which are subject to brief unstable changes (e.g. nuclear power plants, stock markets, and industrial production). The second strong limiting feature of previous research is the employment of methodologies in which change in automation only occurs between discrete trial blocks. This single episode approach fails to capture the significance of multiple cycles and the dynamics of change as they actually occur.

In sum, dynamic adaptive automation necessarily requires a trigger for the shift in task allocation and three contexts have been considered in an attempt to identify an automation trigger; the environment, the aircraft, and pilot competence. In each case significant practical limitations have been identified. A fourth context, the interaction of pilot, aircraft and environment represents the most viable context for adaptive automation. However, like the others, this hybrid context may produce circumstances in which automation is susceptible to frequent oscillation. Therefore,

the purpose of the present experiment was to examine pilot performance effects associated with extremely short, multiple cycle automation. The practical aim of the research was to produce design recommendations for human–machine systems which are susceptible to oscillations between manual and automated control.

Method

Experimental participants

Six rated pilots (five male, one female) were solicited for participation in the study. The pilots had a 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 (e.g. Cessna), two pilots were primarily military aviators (e.g. F-16), and one pilot was employed by a major Mid-West passenger airline and flew commercial aircraft (e.g. Boeing 747, 727). All participants were in professed good health at the time of testing.

Experimental tasks

We developed a general multi-task simulation platform: MINSTAR (see Hancock *et al*, 1993 for an extended description). This provided tracking (psycho-motor), fuel management (cognitive), and system monitoring (perceptual-motor) tasks, representing three flight relevant domains (Parasuraman *et al*, 1992). The three sub-tasks were displayed on two VGA monitors mounted on the forward cockpit of a fixed-base aircraft shell. Simulation based experimentation was employed because it promoted manipulation of flight tasks without hazard, provided an extended time-frame for such exploration beyond that which is available in real systems, facilitated measurement accuracy and was essential for creating consistent and replicable conditions between pilots (for discussion of the advantages of simulation see Flexman and Stark, 1987; Moray, 1993). The two-dimensional compensatory tracking sub-task moves a crosshair on the sum of seven sine waves throughout a target area. The goal for the pilot is to make corrective movements with a flightstick in order to bring the moving cursor in alignment to a fixed target cursor. The difficulty of the task is manipulated by modifying the amplitude and frequency of the sine waves. The fuel management sub-task displays five rectangular shaped fuel tanks connected by six fuel pumps. The two outermost tanks are targeted as the goal tanks and lose fuel 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 is manipulated by initiating failure(s) of the fuel pumps. The monitoring sub-task displays five lights and four graduated gauges. The goal for the pilot is to reset the lights or gauges whenever they deviate from their normal status. The spatial orientation of the three experimental tasks is displayed in *Figure 1*.

Experimental measures

Pilot performance on the tracking sub-task was quantified as root mean square error (RMSe). Performance on the fuel management sub-task was quantified as the absolute deviation from the goal level (in gallons),
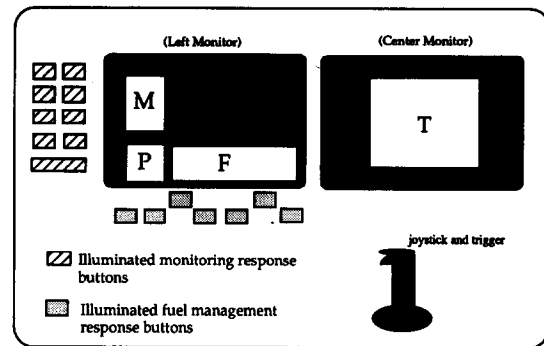

Figure 1 The MINSTAR Display. T (Tracking), F (Fuel management), M (System monitoring) and P (Psychological assessment). The figure shows the left and center monitors and the approximate location of the sub-tasks and response buttons. The arrangement of the fuel management response buttons accurately maps pump arrangement on the VDT. Monitoring response buttons for lights are color coded and gauges are numbered 1–4 to encourage accurate response mapping

averaged for the two goal tanks. Performance on the system monitoring sub-task was quantified as response time (in tenths of seconds) for each light or gauge deviation. Subjective measures of mood state, fatigue and workload measures were also collected in order to obtain a fuller description of pilot behavior. A measure of pilot affect was obtained by The Profile of Mood States (POMS) inventory (McNair *et al*, 1971) which measures six affective states: tension, depression, anger, vigor, fatigue and confusion. Mental fatigue was measured by a portable critical flicker frequency (CFF) device (Jimbo Engineering Corporation Fatigmeter™, model number JM-101). The device presents a single visual stimulus which systematically decreases in frequency (55–20 pulses/sec) until a 'flicker' is perceivable, at which time the subject responds. CFF has been used in the past as an indicator of fatigue in vigilance-type tasks (Baschera and Grandjean, 1979). Workload was assessed via the subjective workload assessment technique [SWAT] (Reig and Nygren, 1988) by having subjects respond on a three point scale to the questions: How much spare time do you have? (time load), What is your stress level? and What is your mental effort? These dimensions are adaptations of factors proposed as major contributors to subjective workload (Jahns, 1973; Johanssen *et al*, 1979; Kahneman, 1973; Moray, 1982; Sheridan and Simpson, 1979). Procedures for the administration of the SWAT were adapted according to the observations of Biers and MacInerney (1988). The SWAT was used in preference to other workload scales because it presents a minimal load in itself and, hence, can be used in real time and with minimal task disturbance.

Experimental conditions and design

Three automation durations were selected for the tracking sub-task. A 15 sec duration was conceived as a realistic lower boundary for short automation, a 60 sec duration as an upper boundary, with a 30 sec duration being intermediate. It is critical to note that the length of the period of manual control matched that for the period of automation. Thus, a 15 sec duration automa-

tion period was cycled with a 15 sec duration manual performance period. This relationship also held for the 30 and 60 sec duration. Thus, results concerning tracking performance when identified for having influence for automation duration represent performance during the interpolated manual control interval. Three levels of difficulty (low, medium, high) were selected for the manual portions of the tracking sub-task. Levels were chosen based on a preliminary study of all tracking difficulty levels available. Thus, the present study was conducted as a within-subjects 3 × 3 (duration vs difficulty) repeated measures design. To ensure that the subject monitored the tracking task during automation a 10% automation failure rate was introduced. During a failure the automated tracing task returned to manual control shortly after being automated. To ensure pilots were cognizant of the control mode, automation was cued by a change in the tracking display and configuration.

Experimental procedure

All pilots began the experimental session by signing an informed consent and responding to the POMS inventory. Afer a practice session each pilot completed the nine 5-min conditions. Intertrial periods were approximately 30 sec and order of conditions was counterbalanced. SWAT measures of workload were obtained 15 sec before the end of each trial. The CFF measure was obtained before the first trial and after the third, fourth, sixth, and last trials. The CFF instrument was placed in a standardized head-on position on the dashboard (62° angle, 18″ viewing distance). Following the nine trials, a post-trial POMS questionnaire and a subject debrief survey were administered.

Results

Tracking

Mean RMS error for cach subject in each of the nine conditions was entered into a 3 × 3 (difficulty × automation duration) repeated measures analysis of variance (ANOVA) procedure. Results indicated main effects for difficulty (F(2, 10) = 48.33, $p < 0.01$] and automation duration [F(2, 10) = 6.54, $p < 0.05$]. Post-hoc tests, using Tukey's procedure, revealed significant differences between all difficulty levels and between the 15 sec duration and the 60 sec duration. Data for tracking RMS error by automation duration, are displayed in *Figure 2*.
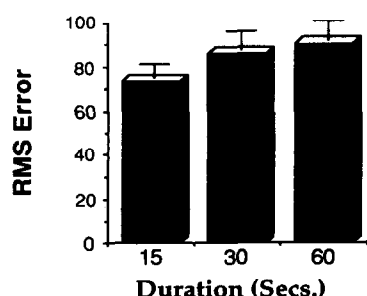
System monitoring

Each 5 min condition contained 10 monitoring deviations (two per min). Mean response time in sec, was obtained for each condition and entered into repeated measures ANOVA. Results indicated no significance for either main effect or the interaction. A similar analysis was conducted for the number of missed monitoring deviations with no significant results.

Fuel management

Fuel management error for each condition was calculated as the absolute deviation from 2500 averaged for all samples in a condition. The obtained fuel management error for each subject, for each condition, was entered into 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 for each condition. Data were analyzed via a 3 × 3 (difficulty × automation duration) repeated measures ANOVA for each of the questions. The analyses for time load and mental effort revealed marginal significance for automation duration [F(2, 10) = 3.75, $p = 0.06$ and F(2, 10) = 3.29, $p = 0.08$, 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 3*. Pre- and post-trial POMS questionnaires were scored for the six scales according to the 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 indicated significant decreases for Anger and Depression scales in the post trial testing session [t(5) = 2.31, $p = 0.06$ and $t(5) = 2.69, p < 0.05$, respectively]. Critical flicker frequency data were subjected to a repeated measures ANOVA. No significant differences were observed.

'Critical window' analysis

In order to examine time dependent effects of automation episodes, performance on each sub-task was

Figure 2 Tracking performance, as represented by RMS Error, by duration of the tracking automatic cycle. Standard error bars are shown
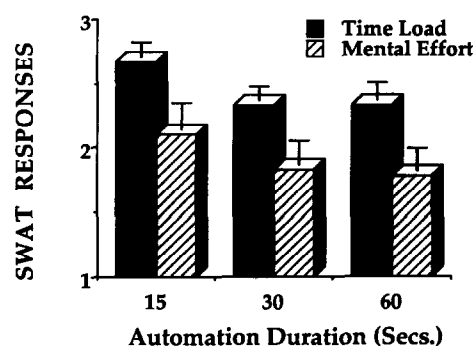
Figure 3 Pilot's responses to the SWAT questions: 'What is your time load?' and 'What is your mental effort?' A high response indicates the perception of increased time load or mental effort. SWAT responses were collected on-line 15 sec before the end of each conditons

examined for a period immediately following each tracking automation episode. A 15 sec window was selected since this interval facilitated comparison across manual tracking durations. Thus, performance on each of the three sub-tasks was examined for a 15 sec period following each episode of tracking automation, regardless of the automation duration. Because automation duration was a manipulated variable, but total trial time was held constant there was an unequal number of complete manual to automation to manual cycles among the different durations. In a 5 min trial with 15 sec automation durations there were 10 complete cycles, with 30 sec automation durations there were five complete cycles, and with 60 sec automation durations there were two cycles.

RMS error data for tracking performance were obtained for the first 15 sec following each automation episode. An overall mean was then calculated for each trial. The mean data were entered into a repeated measures ANOVA procedure. The analysis indicated a main effect for difficulty [$F(2, 10) = 16.71, p < 0.01$]. Follow-up tests indicated significant differences between each of the three levels of tracking difficulty. In this analysis, there were no significant effects for automation cycle duration ($F < 1$).

Fuel management error was obtained for the first 15 sec following each automation episode. Data were treated similarly to those of the tracking analysis, except that each 15 sec 'critical window' yielded only three data points as fuel error was calculated every 5-sec. Mean fuel errors for each trial were entered into a repeated measures ANOVA procedure. The analysis indicated a marginally significant effect for tracking difficulty [$F(2, 10) = 3.809, p = 0.059$]. Data for fuel error, by level of tracking difficulty, are presented in *Figure 4*.

No attempt was made to examine a critical window for the system monitoring responses as the 15 sec window did not contain enough system monitoring deviations per trial to calculate a meaningful representation of response time.
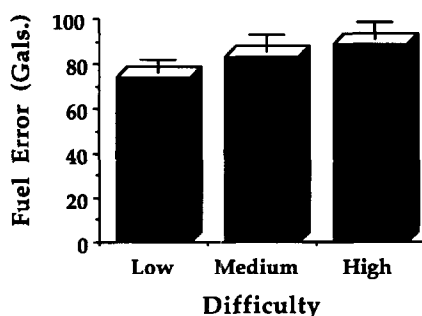

Figure 4 Fuel Management Performance in the 'Critical Window.' Fuel Management Performance by level of tracking difficulty in the 'critical window' following re-acquisition of the tracking sub-task. This figure demonstrates that aviators' fuel management performance were negatively affected by level of tracking difficulty in the first 15 sec following a mode change from automated tracking to manual tracking. Note that aviators' performance was better during the low tracking difficulty than in the high tracking difficulty.

Aviator debriefing responses

Five of the six pilots responded in full to the debrief questionnaire, the majority agreed that the automation of the tracking sub-task decreased their workload. However, pilots reported the cyclic changes from manual to automated control of tracking were moderately distracting. Three of the five respondents reported they were not able to immediately perform the tracking sub-task to the best of their ability after automation episodes. Two the these pilots reported it was a relatively short period of time before they gained satisfactory performance and that this was moderately related to cycle duration. Surprisingly, pilots reported only a moderate ability to discriminate between the three different cycle durations.

Discussion

There are several important observations which come from the present experiment. First, there appears to be a trade between performance and workload. Performance was best at the 15 sec cycle duration, but mental effort and temporal demand scores were significantly greater here. As we have experienced in companion experiments (Duley *et al*, 1994) this implies a strategic trade-off by pilots. The fact that the shortest cycle duration was most beneficial for performance is counter-intuitive. It was hypothesized that the longer cycle times would prove most beneficial, while short cycle times would prove distractive. In a preliminary study, aviators reported the shorter cycles of automation were associated with significant performance distraction. In this preliminary study subjective comments indicated that very short cycles especially hindered their ability to re-acquire manual control. However, the data from the present experiment did not support this assertion.

The relationship between tracking performance and workload can be classified as a micro trade-off, within the tracking sub-task. In effect, pilots did better at tracking only at the cost of working harder. Importantly, this micro trade-off seemed insulated from performance on the other sub-tasks. This insulation may be related to the perception of tracking as the principle task in the environment. Consistent with our previous research (Hancock *et al*, 1993; Hancock *et al*, 1993) pilots reported tracking as the principal sub-task. Thus, tracking was the most sensitive to experimental maniplations in spite of evidence that some primary task performance measures can be insensitive to shifts in workload (see Wierwille *et al*, 1985). In arriving at their conclusion, Wierwille *et al* did not discuss the operator's tendency to assign primacy to a task, a strategy clearly reported by the aviators in this experiment. As Hart and Bortolussi (1984) suggested, workload may be a function of two components, load imposed by the operation of the system, and load imposed by the individual. We believe the difficulty manipulation represented the load imposed by the operation of the system and the pilots tendency to assign primacy to one of the tasks, in this case tracking, represented the load imposed by the individual. Therefore, it was the interaction of these two sources of workload that resulted in the sensitivity of the tracking task to experimental manipulations and the insulation

of these effects from other tasks. This dual-load phenomenon resulted in the micro take-off between performance and workload.

The principal findings of this study indicate that the effects of task-load and automation duration were confined to the specific sub-task under automation. While we have elaborated on the 'dual-workload' explanation for this pattern of results, other explanations should be addressed. One explanation is that the system monitoring and fuel management sub-tasks were not sensitive to the experimental manipulations. The main effect for tracking difficulty produced the expected differences for each level, indicating that the difficulty manipulation affected performance as predicted. 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 anticipated. The insulation of the fuel management task from increases in tracking workload appears to conflict with previous research demonstrating that mediational tasks are particularly sensitive to task loading (Wierwille et al, 1985). However, our present experiment differs in that our design held task-load on the mediational task constant while manipulating task-load on the principal task. It would appear that task loading on a principal task can produce performance decrements, but task-load effects can also be insulated within a principal task. This is related to our concept of micro-tradeoffs.

Influence of the 'window' of analysis

The analyses of data for the 'critical window' indicated an alternative and informative view of results. When examining a 'window' immediately following reacquisition of manual tracking control, aviator performance on the tracking task was influenced by task-load, where performance was significantly better in the lowest difficulty level and worsened toward the most difficult. These results, in conjunction with the original tracking analyses, further validate the successful manipulation of task-load in the present experiment. However, contrary to the original tracking analysis, automation duration no longer exhibited an effect on tracking performance. Therefore, it may be concluded task-load consistently affected tracking performance throughout manual performance, exhibiting consistent affects throughout the full period of manual control. Automation effects on tracking error accumulated in the interval of manual performance following the automation period. The fact that automation frequency had no affect in the initial 15 sec window, but did have a cumulative effect up to 60 sec, implies a progressive degradation of capability within the manual portion of automation cycles. Although automation duration had no effect on tracking during the first 5 sec, it did affect other sub-task performance. Automation duration exhibited influences on fuel management early in the re-acquisition phase, but its affect was dissipated over time and eventually counterbalanced transient effects associated with automation change. The interpretation of the effects of automation duration on the tracking and fuel management sub-tasks therefore center on the 'window' of performance examined.

Similar to tracking performance, fuel management performance exhibited a different pattern of results,

dependent on the window of analysis. While the original results indicated that fuel management was insulated from both task load and automation duration manipulations, the 'critical window' analysis revealed a different pattern. In this case, fuel management performance was affected by the task load of the tracking task, where performance was better for low tracking task-loads and worsened toward high tracking loads. This trade-off between sub-tasks we interpret as a macro-trade-off where manipulations of the principal tracking task were related to performance on a secondary sub-task.

The 'critical window' analyses, in conjunction with the original analysis, raise important questions concerning time dependent effects for factors such as task load and automation duration. The window of analysis is pivotal to the interpretation one draws from the results. If the goal of the operator is to maintain consistency on all sub-tasks, at all times, particular concern may need to be directed toward the period of performance immediately following episodes of short automation. This interpretation is consistent with a dynamic adaptive strategy. In our case a dynamic adaptive strategy would result in the automation of the fuel management sub-task during the first moments of the re-acquisition of tracking control. As a result, the time-dependent detriment in fuel management performance would be mitigated by automated intervention. The employment of adaptive allocation strategies in an attempt to negate time-dependent performance decrements in decision-making sub-tasks would thus be consistent with the proposed goals of adaptive allocation (Morrison et al, 1993).

Implications for design

A fundamental goal of the present research is to identify implications for the design of human–machine systems which are susceptible to oscillations between manual and automated control. The foremost aim is the reconception of hybrid human–machine systems to include periods of automation significantly less than the 5 or 10 min periods previously studied. The present experiment demonstrated that episodes of automation even down to 15 sec duration have an impact on operator performance. Given these effects we do not yet know if even shorter episodes of automation affect performance, however we are investigating these possibilities. In any system it is important to control oscillations which threaten to produce destructive instability. This can be achieved by introducing a damping factor. In the case of adaptive allocation systems we propose a moratorium strategy in which there is a minimum frequency with which the system can either assume or relinquish task control. For design purposes this represents a minimum time threshold in which particular sub-tasks are prohibited from change in control status. With Jordan (1963) and others, we accept that there may be switching between tasks and our recommendation at present is only for suppression within a particular task. Clearly further experimentation is necessary to distinguish whether this principle should be applied generally across all sub-tasks or components of any complex operation environment.

Another design implication is the identification of micro trade-offs within tasks and macro trade-offs

between tasks. It is important to consider which components are subject to automation and which are not since the capability for change status alone appears to have direct effects on performance. Complicating this automation issue are the measurement questions associated with analytical evaluation. Thus, system evaluation is not just a matter of looking at performance, but knowing when to look and where to look. Adaptive systems which cycle between manual and automated control must therefore be evaluated in light of an historic profile of performance. Both continuous (time-based) performance and specific (event-based) actions are paramount to obtaining an accurate picture of performance (Poulton, 1965). Finally, the convergence or divergence of data is an important issue. Pilots in the present study subjectively reported an increase in workload during shorter cycle durations, yet objective performance data indicated better performance at shorter rather than longer cycles. Determining the effects of automation cycles by either of these methods alone would provide an incomplete picture of system performance. We contend that the evaluation of all the man–machine systems must include both obective performance data and subjective participant data.

Acknowledgements

References

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

Biers, D. W. and MacInerney, P. 1988 'An alternate approach to analyzing SWAT data' *Proceedings of the Human Factors Society* 31, 63–66

Branton, P. 1987 'In praise of ergonomics – a personal perspective' *International Review of Ergonomics* 1, 1–20

Craik, K. J. W. 1947 'Theory of human operator in control systems. I: The operator as an engineering system' *British Journal of Psychology* 38, 56–61

Duley, J. A., Scallen, S. F. and Hancock, P. A. 1994 'The response of experienced pilots to interface configuration changes for adaptive allocation' *Proceedings of the First Automation Technology and Human Performance Conference* (pp. 134–141) Washington, D.C.

Fitts, P. M. (ed.) 1951 *Human Engineering for an Effective Air Navigation and Traffic Control System* National Research Council Washington, D.C.

Flexman, R. E. and Stark, E. A. 1987 'Training simulators' in Salvendy G. (ed.) *Handbook of Human Factors* (pp. 1012–1038) John Wiley, New York

Glenn, F., Barba, C., Wherry, R. J., Morrison, J., Hitchcock, E. and Gluckman, J. P. 1994 'Adaptive automation effects on flight management task performance' *Proceedings of the First Automation Technology and Human Performance Conference* (pp. 33–39) Washington, D.C.

Hancock, P. A. and Chignell, M. H. 1987 'Adaptive control in human–machine systems' in Hancock, P. A. (ed.) *Human Factors Psychology* (pp. 305–345) North Holland, Amsterdam

Hancock, P. A. and Chignell, M. H. (eds). 1989 *Intelligent Interfaces: Theory, Research, and Design* North Holland Press, Amsterdam

Hancock, P. A., Duley, J. A. and 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. and Meshkati, N. 1988 *Human Mental Workload* North Holland, Amsterdam

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

Haart, S. G. and Bortolussi, M. R. 1984 'Pilot errors as a source of workload' *Human Factors* 26, 545–556

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

Jahns, D. W. 1973 'A concept of operator workload in manual vehicle operations' *Forschungsinstitut Anthropotechnik*, Meckenheim, Bericht Nr. 14

Johanssen, G., Moray, N., Pew, R., Rasmussen, J., Sanders, A. and Wickens, C. 1979 'Final report of experimental psychology group' in Moray, N. (ed.) *Mental Workload* Plenum, New York

Jordan, N. 1963 'Allocation of functions between man and machines in automated systems' *Journal of Applied Psychology* 47, 161–165

Kahneman, D. 1973 *Attention and Effort* Prentice-Hall, Englewood Cliffs, N.J.

McNair, D. M., Lorr, M. and Droppleman, L. F. 1971 *Profile of Mood States Manual* Educational and Industrial Testing Services, San Diego, CA

Moray, N. 1982 'Subjective mental workload' *Human Factors* 24, 25–40

Moray, N. 1993 'Reducing the risk in automation – is there an optimal solution?' *Proceedings of the 2nd International Congress on System Safety* Budapest, Hungary

Morrison, J. G., Cohen, J. P. and 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. and Barnes, M. 1990 *Theory and design of adaptive automation in aviation systems* (Technical Report) Naval Air Development Center, Warminster, PA

Parasuraman, P., Bahri, T. and Molloy, R. 1992 *Adaptive automation and human performance: I. Multi-task performance characteristics* (Technical Report No. NAWCADWAR 92035-60) Warminster, PA: Naval Air Warfare Center

Parasuraman, P., Bahri, T., Molloy, R. and Singh, I. L. 1992 *Adaptive automation and human performance: II. Effects of shifts in the level of automation on operator performance* (Technical Report No. NAWCADWAR-92036-60) Warminster, PA: Naval Air Warfare Center

Poulton, E. C. 1965 'On increasing the sensitivity of performance measures' *Ergonomics* 8, 69–76

Raeth, P. G., Noyes, J. L. and Montecalvo, A. J. 1994 'Trust-enhancing sensor and information fusion for knowledge-based cockpit decision aids' *Proceedings of the 3rd International Workshop on Human–Computer Teamwork, The Human–Electronic Crew: Can we Trust the Team?* Cambridge, UK

Reid, G. B. and Nygren, T. E. 1988 'The subjective workload assessment technique: A scaling procedure for measuring mental workload' in Hancock, P. A. and Meshkati, N. (eds) *Human Mental Workload* (pp. 185–218) North Holland Amsterdam

Sheridan, T. B. and Simpson, R. W. 1979 *Towards the definition and measurement of the mental workload of transport pilots* (FTL Report R79-4) Cambridge, MA: Massachusetts Institute of Technology, Flight Transportation Laboratory, Department of Aeronautics and Astronautics

Wierwille, W. W., Rahimi, M and Casali, J. G. 1985 'Evaluation of 16 measures of mental workload using a simulated flight task emphasizing mediational activity' *Human Factors* 27, 489–502