EFFECTS OF AUTOMATION RELIABILITY ON HUMAN MONITORING PERFORMANCE

Brian Oakley, Mustapha Mouloua, & Peter Hancock
University of Central Florida
Orlando, FL

Human decisions to rely on automation of a simulated flight task were monitored in the present study. Participants performed the Multi Attribute Task (MAT) battery previously developed by Comstock and Arneguard (1990). Twenty-eight participants were tested for 7 different reliability levels. Decision accuracy, false alarm rates, and reaction time to simulated automation failures were collected for each participant over nine 10-minute sessions. The results indicated that the rate of automation failures detected varied inversely with automation reliability level. As automation level increased, detection rate decreased thereby indicating automation-induced monitoring inefficiency. However, the mean reaction time of correct detection rate decreased as automation reliability level increased. No significant effect of reliability was obtained for the false alarm rate data. The implications for training and systems design are also outlined.

INTRODUCTION

The implementation of advanced automation technology in aviation systems has resulted in several benefits (e.g., all-weather flying, increased safety, fuel efficiency, decrease in physical workload, etc.). However, there are several automation costs that have also been attributed to increased automation use. These behavioral problems include automation-induced monitoring inefficiency, loss of situation awareness, degradation of manual flight skills, inability to revert to manual control in case of automation failures, and increased mental workload. According to the Aviation Safety Reporting System (ASRS), there are many instances in which monitoring failures were linked to excessive trust in or over reliance on automation such as the autopilot or FMS (Mosier, Skitka, and Korte, 1994).

In general, human monitoring of automated systems for possible malfunctions or failures has resulted in increased demands on human operators (Chambers and Nagel, 1985; Parasuraman, 1987; Wiener, 1988). An extensive body of literature has well documented several aspects of these behavioral problems across different domains of applications (Mouloua and Parasuraman, 1994; Parasuraman and Mouloua, 1996; Mouloua and Koonce, 1997, Scerbo and Mouloua, 1999). Research on vigilance has also shown that human detection of low-probability events is poor during sustained watch keeping operations (Davies and Parasuraman, 1982). Monitoring automated subsystems for automation failures has also been poor when automation is static (Parasuraman, Molloy, and Singh, 1993; Singh, Molloy, and Parasuraman, 1993). These results have been attributed to the performance costs associated with static automation. In contrast, adaptive automated tasks in which the allocation of function between the machine and operator is not static have been touted to be superior (Rouse, 1988, Parasuraman, Mouloua, and Molloy, 1996). Because most of these studies have used different reliability levels, it is not clear whether some of these effects can also be attributed to automation reliability.

Operator’s trust in or over-reliance on automation is an important factor in determining when a system can be trusted to function appropriately and when a system is unreliable and trust is unwarranted (Muir, 1988). Subsequently, May, Molloy, and Parasuraman (1993) examined the effects of automation reliability on performance of an automated monitoring system in a pair of
experiments. While the first experiment reported that the detection rate of automation failures decreased as the level of automation reliability increased; experiment 2 failed to report any significant reliability effect with only two reliability groups. This could be because the range between the two groups was not wide enough and could have caused high individual differences (Singh, Molloy, & Parasuraman, 1993); thereby resulting in substantial error variance in experiment 2.

The present study was designed to further investigate factors affecting the establishment of automation trust in the context of monitoring performance. Automation reliability was varied over 7 different levels ranging from 5% to 95%. It was hypothesized that increased levels of automation reliability would result in degraded performance of the system-monitoring task.

METHOD

Twenty-eight volunteers participated in this study. A 7 (reliability) by 3 (session) mixed factorial design was used. The reliability of the automation system in the monitoring task was manipulated as a between-subjects (5%, 20%, 35%, 50%, 65%, 80%, and 95% reliability) factor. The within-subjects factor in the factorial design was the three 30-minute sessions. Automation reliability was defined as the percentage of the 16 system malfunctions correctly detected by the automation routine in each of the 10-minute blocks. A modified version of the Multi-Attribute Task Battery (MAT) developed by Comstock and Arnegard (1990) was used for each session. The revised MAT is comprised of three tasks: tracking, system monitoring, and resource management.

The system-monitoring task was automated to detect any offsets in which pointers on one of four dials went out of range. Upon detection, the malfunction was reset automatically. Periodically the automation system would fail to detect the malfunction. Upon this occurrence, the subject was instructed to manually correct the offset by pressing the corresponding function key. If 10 seconds elapsed without detection, the offset was automatically corrected. A first-order, two-dimensional compensatory tracking task was used as the second task. The task of the subject involved keeping an aircraft icon in the center of the tracking display. The fuel management task was designed to simulate the tasks performed in managing the fuel system of an aircraft. The subjects were instructed to compensate for fuel depletion by pumping fuel from the supply tanks to the main tanks. Root mean square (RMS) error was computed for both the tracking and fuel management tasks.

Subjects were instructed to perform all three tasks manually in a 10-minute training period and to give equal attention to each of the tasks. The subjects then began three 30-minute sessions, which were comprised of three 10-minute blocks. The subjects were informed that the system-monitoring task had been automated and their primary tasks were tracking and fuel management. In addition, they were also advised that the automation routine may not be 100% reliable and therefore they needed to monitor it for automation failures.

RESULTS AND DISCUSSION

Figure 1 shows the mean percentage of the automation failures detected for each of the seven groups. As the level of reliability increased, the detection rate decreased. An analysis of variance (ANOVA) of the detection rate showed a significant effect of automation reliability, $F(6,21) = 5.88$; $p < .005$.
An initial analysis of the data revealed no difference between the groups due to a resource trade-off effect between the three tasks of the MAT. In other words, a high percentage of detection of the automation failures could be obtained at the expense of the performance on the other two tasks (tracking and fuel management). There were no significant differences in performance on the resource management task or the tracking task between the reliability groups. Therefore, it can be concluded that the decrease in monitoring, as a function of automation reliability, was not due to the trading of resources to accommodate higher performance on the manual tasks.

These results partially support previous findings by May et al., (1993) indicating that the more reliable an automated system, the poorer the operator’s monitoring performance will be but only to a certain level. The decrease in the detection rate of automation failures with increasing reliability was not completely linear. These findings indicate that poor automation monitoring reflects over-reliance on automation as reliability increases.

REFERENCES


