Pilot Interaction with Automated Airborne Decision Making Systems

William B. Rouse

Grant NSG-2119
August 1981
Pilot Interaction with Automated Airborne Decision Making Systems

William B. Rouse
Coordinated Science Laboratory
University of Illinois
Urbana, Illinois 61801

Prepared for
Ames Research Center
under Grant NSG-2119

NASA
National Aeronautics and Space Administration
Ames Research Center
Moffett Field, California 94035
TABLE OF CONTENTS

Introduction ................................................. 1
Computer-Aided Multi-Task Decision Making ............... 3
Computer-Aided Problem Solving .......................... 10
Planning and Problem Solving Behavior .................... 14
Conclusions ................................................ 20
References .................................................. 22
INTRODUCTION

Automation is increasingly finding its way into the aircraft cockpit. To a certain extent, the aircraft will soon be able to almost fly itself. This trend leads one to question the role of the pilot in semi-automated and automated aircraft.

If computer technology becomes capable of complete automation of the pilot's task and, if the chance of system failure is absolutely zero, then aircraft pilots can eventually be eliminated. However, it is unlikely that a fail-safe system will be produced, except perhaps in the distant future. Further, even if the system was fail-safe, the public might not be willing to fly on an aircraft without a pilot. Thus, for quite some time, there will be pilots in the cockpit.

What role should the pilot fill in the cockpit of the future? One possibility is to have him perform all those tasks that cannot as yet be automated. Unfortunately, this may lead to his having only an incoherent set of bits and pieces of tasks to perform. Also, the workload level may be so low that the pilot becomes bored and his performance degrades.

Boredom and performance degradation become especially important when pilot workload suddenly becomes very high due to an emergency such as a failure of the computer system. If the pilot has not been involved with flying the aircraft, how can he be expected to suddenly take over the decision making from the computer? Thus, an issue that arises concerns how involved with
the system the pilot has to be to assure acceptable performance on his part during an emergency.

Further, if one of the pilot's main tasks is to maintain himself so as to be able to acceptably respond to unforeseen situations, then it is interesting to consider the pilot's ability to detect such situations, diagnose their causes, and take appropriate actions. Also, the complexity of a highly automated and tightly integrated air traffic system may require that the pilot respond quickly and flexibly to a wide range of situations, the number of which may be so large as to prohibit rote responses.

This report summarizes the results of a six-year program of research which addressed the issues noted above. Each research project within this program is reviewed quite briefly. The interested reader can find complete treatments of these projects in the referenced papers and reports.
COMPUTER-AIDED MULTI-TASK DECISION MAKING

The decision making tasks to be performed in flight management can be divided into three categories:

1. Those decisions which crew members must make,
2. Those decisions which the computer must make,
3. Those decisions that either crew members or the computer could make.

With increasing sophistication of computer technology, the third category of decision making tasks is becoming larger and larger. This is the type of tasks to which we chose to address our research.

In considering this problem, we chose the criterion of trying to minimize the delay in successful completion of all tasks while also maintaining the crew's workload at a level conducive to their responding appropriately to unusual events. We defined workload as the fraction of time that the crew is busy, as opposed to time spent scanning or involved in non-system related tasks. While this definition is rather simplistic, it does conform with classical time-line analysis approaches. Further, fraction of time busy would certainly be an attribute of any more elaborate workload formulations.

In order to be able to predict the impact of any specific allocation of tasks between crew and computer, one needs a model of crew decision making in flight management. This model must allow one to describe both humans and computers in similar terms.
To this end, we decided to view the human as a time-shared computer. Considering the literature on analysis of time-shared computer systems, it is immediately apparent that a queueing theory formulation is appropriate. Thus, we modeled human decision making as a preemptive priority queueing system. It is a priority system in that some tasks are more important than others. It is preemptive in that some tasks (e.g., an autopilot malfunction) require immediate attention and thus, when they occur, are allowed to "go to the head of the line" from a queueing perspective.

With such a model, we used simulation to study alternative approaches to allocating tasks [Rouse, 1977]. It soon became apparent that system performance could only be optimized if one avoided a strict allocation of functions between human and computer. Instead, tasks should be assigned to the decision maker (human or computer) who is, at the moment, most capable of performing the task. These results led us to the conclusion that task allocation should be dynamic and adapt to time-varying aspects of the environment.

But, how should task allocation adapt to the situation? Resorting again to the literature of queueing theory, we were able to extend some available results to obtain the conclusion that a fairly simple scheme was appropriate. Namely, the second decision maker (the computer) should be utilized whenever the sum of the number of tasks to be performed, weighted by the relative importance of each task, exceeded a threshold. The threshold can
be determined analytically for some special cases or, via simulation for more general cases. The optimal value of the threshold was found to depend on the number of tasks, arrival rates of tasks, and service rates of tasks [Chu and Rouse, 1977; Chu, 1978].

With the task allocation problem formulated, the next step was to obtain empirical human decision making data. A flight management scenario was developed [Rouse, Chu, and Walden, 1976]. Using this scenario, our first experiment produced verification that a queueing model of human decision making was appropriate while also providing estimates of service rates and error probabilities [Walden, 1977; Walden and Rouse, 1977, 1978].

With this data, we were able to estimate threshold values as a function of task arrival rates. This led to our second experiment which considered the effects of having the computer as a backup decision maker. Both objective performance measures as well as subjective ratings were measured. It was found that the allocation policy mentioned above produced significant improvements in system performance and was also well-accepted by the subjects in the experiment. Further, the queueing model provided a reasonable description of human performance, even to the extent that the workload predictions of the model and the subjective ratings of subjects were highly correlated. Thus, the model may be useful for predicting levels of workload in a variety of multi-task situations [Chu, 1978; Chu and Rouse, 1978, 1979].
While our queueing model is fairly good at describing how much time the human spends in performing various tasks and predicting the total workload, the model says nothing about how the human detects that tasks must be performed or about how well the tasks are performed. Thus, two other efforts were directed at these issues.

To consider event detection, we developed a process monitoring scenario where subjects had to indicate whether or not they thought a dynamic process had changed characteristics [Rouse and Greenstein, 1976a]. Recognizing that many such tasks would not fit a linear gaussian systems formulation, we avoided an estimation and control theory construct. Instead, we developed a model based on feature extraction approaches of pattern recognition [Rouse and Greenstein, 1976b]. Our first experiment yielded results that compared quite favorably with the model [Greenstein and Rouse, 1978].

Our second experiment focused on the joint problem of event detection and attention allocation. In this experiment, subjects had to trade-off detected probabilities of failures, times to implement actions, and costs of delaying actions in order to reach an allocation decision. Using queueing theory, two models of attention allocation were developed. One model employed a very simple rule to rank order processes for servicing while the other model involved a more global optimization. While both models compared quite favorably with human performance, it was somewhat surprising to find that the simpler model actually
produced the most favorable comparisons [Greenstein, 1979a; Greenstein and Rouse, 1979, 1980].

This modeling effort has quite a few implications [Greenstein, 1979b]. While instrument scanning is one area of application, the feature extraction approach is also amenable to modeling how air traffic controllers detect deviations of aircraft from commanded trajectories. Considering aircraft with cockpit display of traffic information systems, the model appears to be applicable to describing the human's ability to detect changes in the behavior of neighboring aircraft and subsequently allocate increased attention to them.

As noted earlier, our queueing theory model of human decision making in flight management is satisfactory for predicting how much time the human devotes to each task. However, the model only considers performance metrics in terms of probability of task completion. This is not completely satisfactory for control tasks where RMS deviations are also important. Thus, given the queueing model's prediction of the fraction of time (if any) which the human will spend controlling, one would like to predict the control task performance.

One approach to this problem would be to use conventional models of control task performance that include fractions of attention as free parameters. The main difficulty with this approach is that available models are based on the assumption that the human controls continuously while, in our flight management experiments, this is clearly an unreasonable
assumption. Thus, one needs a model that allows for intermittent control actions and also provides a tight link with the overall queueing theory formulation.

To approach this problem, we initially developed a simple heuristic model [Govindaraj and Rouse, 1978]. This model assumes that the human calculates a decision function using a weighted difference between a displayed map and the extrapolated aircraft trajectory. If this decision function exceeds a threshold, the aileron control is held at a maximum value until the maximum bank angle is reached. If the decision function is within the threshold, the aileron is moved so as to return the bank angle to zero.

A simulation experiment was conducted with this model to determine its sensitivity to the aircraft dynamics, characteristics of the map, and the model parameters (weighting function and threshold). Several interesting results were obtained, especially the fact that the model became unstable under conditions similar to those which cause naive subjects to become unstable controllers.

While this model looked promising, we wanted to obtain a more analytical formulation. Thus, we returned to looking at optimal control formulations. First, we solved the optimal preview control problem for deterministic paths (i.e., maps) and then, concentrated on determining how to incorporate discrete events. This effort led to the following formulation.
Within an optimal control formulation, we employed a quadratic cost functional that included weighting on errors (Q) and weighting on control effort (R). Scheduling a discrete event amounts to determining the optimal time to make R/Q very large. This is due to the fact that making R/Q very large will result in very small (effectively zero) control gains and also, will compensate by exerting increased control when the gains are non-zero (i.e., normal R/Q). Thus, our problem was considerably simplified. We developed a procedure for optimally placing large values of R/Q based on a moving window of minimal values of absolute control rate. Comparing this model with human performance resulted in substantially more favorable comparisons than were possible with the earlier heuristic model [Govindaraj, 1979; Govindaraj and Rouse, 1979a,b, 1981].

More recently, our efforts were devoted to integrating the above work with Chu, Govindaraj, Greenstein, and Walden into a coherent framework. This review effort included surveying a wide range of models and empirical results and producing a comprehensive review. As a result of this review, a framework was developed within which human-computer interaction in dynamic systems in general can be viewed [Rouse, 1981].
COMPUTER-AIDED PROBLEM SOLVING

Within our research in flight management, we also have become concerned with the human's role as a problem solver in advanced automated aircraft. While we initially stressed trouble-shooting of failures in aircraft systems (e.g., hydraulic system), we have now come to also emphasize problem solving in terms of emergency and abnormal procedures.

We first developed a trouble-shooting task which abstractly represented what the crew might have to do when diagnosing the failure of one of their systems. After two experiments [Rouse, 1978a], we developed a model of human fault diagnosis abilities based on a few pattern-evoked heuristics as well as concepts from the theory of fuzzy sets [Rouse, 1978b].

The model offered a very succinct description of human performance and motivated some display design notions as well as ideas for how the trouble-shooting task might be extended to match reality more closely. However, among many colleagues there was a consensus that crews will do very little airborne trouble-shooting of their systems. Notable exceptions to this consensus were two individuals in industry who felt that the crew should diagnose failures as well as possible while in the air to avoid excess turnaround time while the aircraft is on the ground. This type of feedback led us to seek and receive support from another agency for the trouble-shooting work which we directed at the training of flight mechanics.
Starting in 1978, we began to look at the human as a problem solver in a more general sense. As a framework for pursuing this topic, we spent some time contrasting flight management with more conventional management domains. We concluded that an essential issue in the design of on-board flight management information systems is an understanding of crew members as information seekers [Rouse and Neubauer, 1978].

Pursuing this issue further, we considered the areas in which the crew can realistically be said to be solving management problems. We concluded that the pilot is a manager in the sense that he manages the aircraft's internal world so as to meet the demands of the external world. In other words, the pilot is a manager (i.e., problem solver and decision maker) who is responsible for what happens inside the aircraft [Rouse, 1978c].

This realization led us to turn our attention to the internal world of the aircraft. In this way, we became interested in aircraft systems (i.e., electrical, hydraulic, etc.) and, in particular, emergency and abnormal procedures. These interests have caused us to focus on information seeking behavior related to emergency and abnormal procedures.

Studying this problem, three issues struck us. First, the huge loose-leaf notebooks in which these procedures are contained seem very difficult to work with. Second, the procedures are mostly in text rather than graphical form. Thus, the crew members have to transform a spatially oriented set of symptoms to a non-spatially oriented text presentation and then, back to a
spatially oriented set of actions. Third, the procedures are highly "proceduralized" and seem to allow little room for innovation should a totally unexpected and unanticipated event occur.

In 1979, we initiated study of two of these issues. One study considered the use of color graphics for representing procedural information. Two experiments considered the effect of various coding schemes on the human's ability to perceive relationships in a schematic representation of a system. We were somewhat surprised to find that color coding did not produce significantly better performance once subjects were fully trained [Neubauer and Rouse, 1979].

A second effort in this area led to the development of a task scenario for studying the effects of alternative approaches for retrieving and displaying procedural information. A first experiment evaluated hardcopy, softcopy, and intelligent softcopy manuals [Rouse and Rouse, 1980]. The manuals were abstractions of 747 emergency procedures. It was found that the additional features of an intelligent softcopy manual were necessary if computer-based manuals were to be clearly superior to hardcopy manuals.

These results demonstrated that an appropriately designed computer-based information system could produce substantial benefits. This led us to design and develop a more complete information system concept for implementation and evaluation in our GAT-II simulator [Rouse, Rouse, and Hammer, 1980].
Preliminary experiments were performed and the results indicated the superiority of the computer-based system in terms of lessening the frequency of serious pilot errors. This study also led to several changes in the computer-based system.

A full-scale experimental evaluation of the system was then conducted utilizing four two-person crews flying normal, emergency, and double-emergency full-mission scenarios with either hardcopy or computer-based information systems. The essential features of the computer aiding included:

1. Automatic cross-referencing among procedures, including returning from cross-references,
2. Automatic "dimming" of procedure steps that the computer could detect to be completed (68% of all steps),
3. Automatic reminders of procedural steps that were intentionally skipped.

The results of this study indicated that a well-designed computer-based system can virtually eliminate certain classes of human errors [Rouse and Rouse, 1981].
PLANNING AND PROBLEM SOLVING BEHAVIOR

In 1978, we reviewed the literature of man-machine systems as it related to modeling man-machine interaction in realistically complex tasks [Johannsen and Rouse, 1978, 1979]. At that time, we concluded that planning constituted a particularly important aspect of human problem solving behavior that had received relatively little attention from those who pursue human factors issues in systems design. For example, while everyone seems to agree that a map display or cockpit display of traffic information will impact the flight crew's planning process, it is difficult to empirically support this hypothesis. The main reason for this difficulty is that we really do not know how to measure planning.

With this background in mind, we set out to study the planning process in flight management. This began with the notion of depth of planning. By depth, we mean level of detail which can range from broad and sketchy to specific and concrete. Our hypothesis was that planning with respect to a particular task need not be very deep if: 1) The amount of time until the task must be performed is large; 2) It appears that the environment will be "hospitable" to successfully completing the task; 3) The task is not critical to mission success. However, if one or more of these conditions ceases to hold, then depth of planning will increase to the extent that the conditions are not satisfied. In other words, the depth of planning associated with a particular task will be very great if the task must be
performed immediately, may be difficult to accomplish, and is critical to mission success.

To study this hypothesis, an HFB 320 Hansa Jet simulator at the Research Institute for Human Engineering in the Federal Republic of Germany was employed. The HFB 320 is a twelve passenger, twin engine jet used for both military and commercial purposes. It normally has a two-man crew. Using this simulator, two experiments were performed using nine professional HFB 320 pilots who flew several 20 to 32 minute missions from cruise to touchdown, in some cases including several cycles of a holding pattern.

Three flight conditions were studied: 1) normal, 2) abnormal involving possible diversion to another airport because of snow or fog and reduced visibility, 3) emergency involving an unexpected engine failure or complete loss of hydraulic pressure or both. Several online questionnaire techniques were used to assess a pilot's depth of planning during each of the flights. Besides this subjective data, numerous objective measurements were also collected.

The results of the first experiment [Johannsen and Rouse, 1980] were somewhat speculative but, nevertheless, indicated that conscious planning was most pronounced in abnormal situations, while normal situations followed the standard scripts and emergency situations relied on the use of highly learned procedures since time did not allow the luxury of planning. Further, there appeared to be substantial differences between the planning processes of different pilots.
The second experiment [Johannsen and Rouse, 1981] involved a factorial study of scenarios (i.e., normal, abnormal, and emergency), flight phases (e.g., initial approach, final approach, and landing), and level of automation (i.e., manual and autopilot). Depth of planning, subjective workload, and flight performance were measured. Although, there were numerous results, only two were somewhat counterintuitive and deserve mention.

First, it was found that the autopilot mode during abnormal scenarios reduced planning while the autopilot mode during emergency scenarios increased planning. Fortunately, this surprising result was explainable. It was noted above that the abnormal scenarios involved events outside of the aircraft (i.e., runway closures) while the emergency scenarios involved events inside the aircraft (i.e., engine and hydraulic system failures). When the runway is closed, the pilot's main task is holding and waiting unless, of course, the delay becomes excessive. When the autopilot is available, the pilot's main task is automated and thus, the need to plan is lessened and planning decreases.

On the other hand, during emergencies the pilot's task involves controlling the aircraft and dealing with the engine and/or hydraulic system failure. In this case, if the autopilot takes over the control task, the pilot is freed to plan with respect to the implications of the failure and hence, planning is increased. Thus, the effects of automation are subtle and dependent on the nature of events.
The second surprising result involves the correlation of subjective assessment of workload with planning activity and flight performance. As might be expected, workload was highly correlated with the level of control activity necessary to maintain aircraft altitude and attitude. However, workload was uncorrelated with level of planning activity. Thus, it appears that pilots do not perceive an increase in workload due to the increased planning necessary to cope with abnormal or emergency events. This may be due to the fact that planning is an internal process rather than an external activity.

While the above HFB 320 experiments have provided interesting insights into planning as it is affected by numerous variables, these types of experiment present two particular problems. First, they are very time-consuming to design, develop, and execute. Second, the amount of data resulting is almost overwhelming even when only a few subjects are utilized.

For these reasons, we decided to create a simplified problem solving environment to be used as a complement to the robust flight tasks possible within the HFB 320. The result was a task called PLANT (Production Levels and Network Troubleshooting) which was designed to include aspects of problem solving that are typical in flight environments but, at the same time, simple enough to allow subjects to learn the basics of the task quickly and thereby enable us to establish a large subject pool.

PLANT is basically a large set of tanks interconnected by valves that may be opened or closed. Fluid, the product of PLANT, flows between two interconnected tanks in proportion to
the difference in the height of the fluid in each tank. The human's goal is to configure PLANT by opening and/or closing valves so as to maximize the production of fluid (i.e., the total flow through the network per unit time) subject to the constraint that no tank become completely empty or full. This constraint leads to a tradeoff between optimization (i.e., maximizing production) and stabilization (i.e., staying safely within the limits of fluid levels). In addition, valves can fail which leads to closure of interconnections and fluid buildup. These failures must be detected, diagnosed, and corrected for the product to have the correct consistency.

PLANT is somewhat analogous to flight management in that the initial portion of a production run involves configuring the network (i.e., takeoff) while the latter portion of the run involves monitoring the network (i.e., cruise). Further, in the event of failures, reconfiguration may be necessary. In such situations, the human must cope with three goals: optimization, stabilization, and detection/diagnosis.

Two experiments were performed using PLANT. The first experiment was a factorial study of network size and failure rate [Rouse and Morris, 1981a]. The second experiment was a factorial study of network size, failure rate, display noise levels, and time constraints [Rouse and Morris, 1981b]. The most important conclusion of these studies was that the subjects appeared to have difficulty in coordinating the goals of optimization, stabilization, and detection/diagnosis. In particular, subjects tended to focus on diagnosis to the extent that production
possibilities were severely compromised. (It is interesting to note that a similar type of focusing was found for some subjects in the GAT-II experiments noted above.) This finding appears to have implications for the design of computer aids for assisting humans to coordinate tasks.
CONCLUSIONS

The six-year program of research summarized in this report included fifteen formal experimental studies and the development of a variety of models of human behavior based on queueing theory, pattern recognition methods, control theory, fuzzy set theory, and artificial intelligence concepts. While these studies and models are important products of this research program, a more important product is the well-tested, automation-oriented design concepts that have emerged. In this final section, we will review these concepts.

Our efforts in computer-aided multi-task decision making have shown that automation decisions need not be static. Instead, performance improvements can be gained if the use of automation is adaptive in that it is used when necessary but avoided when unnecessary. More specifically, automation is invoked when the crew's workload increases to the extent that they will not be able to successfully complete all necessary tasks. Otherwise, the crew performs all of its normal functions. In this way, the crew maintains its skills while also having the advantages of automation when necessary.

Our efforts in computer-aided problem solving have shown that automation can greatly aid the crew by performing the bookkeeping aspects of problem solving. In particular, the computer can help the crew by keeping track of what has been done and the implications of these actions. This type of automation can substantially reduce the frequency of human errors while
still leaving the crew with overall responsibility for the problem solving.

Our studies of planning and problem solving behavior have illustrated some of the subtle effects of automation, particularly in the planning studies, and indicated where automation could be helpful. For example, the problem solving studies showed that performance could be improved if the human was aided in coordinating goals, especially in failure situations. This type of higher-level aid would be quite different from traditional aircraft automation.

To conclude, this program of research has produced several well-tested concepts for using automation to aid crews in decision making, problem solving, and planning. This work has shown the potential benefits to be gained by using automation to assist crews rather than incrementally attempting to replace the crew by automation.
REFERENCES


This report summarizes the results of a six-year program of research in the area of computer-aided multi-task decision making, computer-aided problem solving, and planning and problem solving behavior. The products of this research included 15 formal experimental studies and a variety of models of human behavior based on queuing theory, pattern recognition methods, control theory, fuzzy set theory, and artificial intelligence concepts. Of equal importance, this research produced several well-tested concepts for using automation to aid crews in decision making, problem solving, and planning.
End of Document