

5.3 20174

Techniques For Improving Pilot Recovery From System Failures

**Summary of Research
Grant NAG 1-2014**

submitted to
NASA Langley Research Center
Hampton VA 23681-2199

Attention: Anna Trujillo

**Amy R. Pritchett, Sci.D., Principal Investigator
Schools of Industrial and Systems Engineering and Aerospace Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0205
(Tel) 404-894-0199
(Fax) 404-894-2301
Amy.Pritchett@isye.gatech.edu**

December 2001

ABSTRACT

This project examined the application of intelligent cockpit systems to aid air transport pilots at the tasks of reacting to in-flight system failures and of planning and then following a safe four dimensional trajectory to the runway threshold during emergencies. Two studies were conducted. The first examined pilot performance with a prototype awareness/alerting system in reacting to on-board system failures. In a full-motion, high-fidelity simulator, Army helicopter pilots were asked to fly a mission during which, without warning or briefing, 14 different failures were triggered at random times. Results suggest that the amount of information pilots require from such diagnostic systems is strongly dependent on their training; for failures they are commonly trained to react to with a procedural response, they needed only an indication of which failure to follow, while for 'un-trained' failures, they benefited from more intelligent and informative systems. Pilots were also found to over-rely on the system in conditions where it provided false or mis-leading information.

In the second study, a proof-of-concept system was designed suitable for helping pilots replan their flights in emergency situations for quick, safe trajectory generation. This system is described in this report, including: the use of embedded fast-time simulation to predict the trajectory defined by a series of discrete actions; the models of aircraft and pilot dynamics required by the system; and the pilot interface. Then, results of a flight simulator evaluation with airline pilots are detailed. In 6 of 72 simulator runs, pilots were not able to establish a stable flight path on localizer and glideslope, suggesting a need for cockpit aids. However, results also suggest that, to be operationally feasible, such an aid must be capable of suggesting safe trajectories to the pilot; an aid that only verified plans entered by the pilot was found to have significantly detrimental effects on performance and pilot workload. Results also highlight that the trajectories suggested by the aid must capture the context of the emergency; for example, in some emergencies pilots were willing to violate flight envelope limits to reduce time in flight – in other emergencies the opposite was found.

INTRODUCTION

In-flight emergencies often require the pilot to perform two (largely sequential) tasks: first, to diagnosis and remedy the immediate cause of the emergency, be it on-board system failure, weather related, etc., and second, to re-assess and re-plan his or her flight path to execute a safe trajectory that allows for a landing as soon as possible while also meeting numerous safety-related constraints. Technologies to aid pilots with both these tasks are conceivable given recent advances in computer science and available computing power. However, understanding by the aviation community of what functions these systems should perform to truly assist the pilot in the context of a cockpit in an emergency has been limited, for these technologies bring hitherto-unseen capabilities, benefits and potential problems to the cockpit. Likewise, designing these systems can be difficult, for they must mimic and/or support pilot behavior and strategies at these tasks, behavior which is highly context sensitive, complex, and safety-critical.

Therefore, this project examined the application of intelligent cockpit systems to aid air transport pilots at the tasks of reacting to in-flight system failures and of planning and then following a safe four dimensional trajectory to the runway threshold during emergencies. Two studies were conducted, one on each task. The remainder of this report details these studies. Each study was also documented for the general research and design community in numerous conference proceedings and journal papers, including:

- S. Davis and A.R. Pritchett, Alerting System Assertiveness, Knowledge, and Over-Reliance, *Journal on Information Technology Impact (Aerospace Special Edition)*, Vol. 1, No. 3, pp. 119-144, 2000.
- T.L. Chen and A.R. Pritchett, Cockpit Decision Aids for Emergency Flight Planning, *Journal of Aircraft*, Vol. 38, No. 5, pp. 935-943, 2001.
- T.L. Chen and A.R. Pritchett, On-the-Fly Procedure Development for Flight Re-Planning Following System Failures, *Proceedings of the The 38th AIAA Aerospace Sciences Meeting and Exhibit*, Reno NV, January, 2000.
- T. L. Chen and A.R. Pritchett, Impact of Cockpit Decision-Aids for the Task of Emergency Flight Planning, Presented at the *AIAA Guidance, Navigation and Control Conference*, Montreal PQ, August 2001.

SUMMARY: STUDY #1

Piloting an aircraft is a demanding task, even during normal operations. It becomes much more difficult following an unexpected system failure. The pilot is required to make urgent decisions that may affect the mission, the condition of the aircraft, or the safety of all on board. He has to make these decisions in a short amount of time based on partial information he receives from the aircraft's warning/advisory system. This information is often limited in scope and may not give the pilot a clear indication of the actual problem. Critical time necessary for the recovery of the aircraft may be lost while the pilot diagnoses the problem.

In an attempt to improve this situation, the use of alerting systems in aircraft cockpits has increased steadily for many years. A debate has developed over the benefits of this evolution. Proponents of alerting systems contend that additional advisory systems can improve the capabilities of the pilot. The predicted benefits include reducing monitoring requirements, directing attention during emergencies, and reducing pilot workload as other responsibilities are offloaded to the alerting systems. Critics counter that increasing advisory system use actually increases pilot workload by adding additional cognitive and perceptual requirements. Another argument against increasing use of alerting systems contends that pilots are not using these systems as intended. They note that alerting systems designed to elicit immediate responses are sometimes used simply as attention directors, thus slowing expected response time. These issues relate to, and may generalize to, cockpit automation in general.

There are several major issues in this debate. One involves the level of assertiveness that an alerting system should have in the cockpit. Should the system simply present information to the pilot, alert him when the system determines a problem exists, or provide advice or directives on how to recover from a failure? Another is the knowledge level of the alerting system. How much knowledge is the system required to have to be useful? The sensors and large failure databases required by a smart system can have a substantial cost. The potential benefits of these smarter systems are not yet known. Finally, the question of pilot over-reliance and system dependability must be addressed.

This experiment examined some of the potential benefits of presenting system failure information to pilots using several levels of system knowledge and assertiveness. It was hypothesized that as the amount of information provided to the pilot and the level of system assertiveness increase, pilot use of the alerting system will increase. It was further hypothesized, in a test of pilot over-reliance, that pilots would ignore conflicting instrument indications and follow the alerting system.

Experiment Objectives

The objective of this experiment is to determine if alerting system knowledge and assertiveness affect pilot usage in diagnosing system failures. The experiment examined the following issues:

1. Ascertain how the level of knowledge of the alerting system affects pilot usage to diagnose system failures.
2. Ascertain how the level of alerting system assertiveness affects pilot usage to diagnose system failures.
3. Examine how pilots will respond to alerting system commands that are not supported by – or conflict with – other cockpit indications.

Experiment Design

Overview and Setup

A simulator evaluation was conducted using the US Army's UH-60 Simulated Flight Training System (SFTS) at Fort Rucker, Alabama. An additional system failure alerting display was added to the cockpit to provide varying levels of information to pilots.

The UH-60 SFTS is a full motion simulator with most of the system functionalities of the actual aircraft. The simulator operator can input system failures at specified intervals through a touch screen interface in the rear of the cockpit. The inherent cockpit warning system was used in its normal mode to alert subjects to applicable system failures. An additional warning system was added to the cockpit to allow the experimenter to provide the subjects with varying levels of additional information on the status of the aircraft. This warning system consisted of a laptop

computer connected to a flat panel display screen that was positioned in the cockpit in the center of the windscreen above the glare-shield. A speaker system was also connected to the computer to allow for auditory alerts in some test conditions.

Independent Variables

The experiment was designed as a two-factor experiment. These factors were the knowledge level of the system and the assertiveness of the system.

System Knowledge. The ability of the new warning system to diagnose system malfunctions was divided into six levels of knowledge. These levels of knowledge will determine how much information the system provides to the pilot on the status of the aircraft.

The levels of system knowledge are:

1. System diagnostics: General
2. System diagnostics: Some detail
3. System diagnostics: Detailed
4. System diagnostics and system implications
5. System diagnostics and aircraft implications
6. Recovery instructions: Recommendation/directive

System Assertiveness. The two levels of system assertiveness are informing/recommending and alerting/commanding. An informing warning was an indication to the pilot that a system failure had occurred through a textual readout on the cockpit LCD display as well as through normal cockpit indications (i.e., fuel gauge or oil pressure gauge). At the highest level of system knowledge, the informing system made a recommendation to the pilot on the best action to take in response to the existing malfunction.

An alerting warning provided an aural or visual signal to the pilot that a system failure had occurred, in addition to the information provided by the informing system. At the highest level of system knowledge, the alerting system directed/commanded the pilot to perform an action to correct for a system failure.

In the initial briefing, the additional display was presented to the each subject as either an informing or an alerting system. Each subject experienced only one mode of operation for the warning system and additional display. The briefing specified to the subjects the role of the additional warning system as either a secondary information source (for the informing system) or as a primary indicator of system malfunction (for the alerting system).

The twelve combinations of system knowledge and authority are listed in Table 1 as levels A-L.

		Levels of System Knowledge					
		1	2	3	4	5	6
Autho	Informing/ recommending	A	B	C	D	E	F
	Alerting/ commanding	G	H	I	J	K	L

Table 1.

Subjects

The subjects were twelve active duty Army helicopter pilots. The pilots were all qualified in the UH-60 helicopter and had between two and twenty years of operational helicopter flying experience. Total aircraft time ranged from 440 hours to 6800 hours, and UH-60 time ranged from 23 hours to 2500 hours.

Scenarios

For the experiment, fourteen system failure scenarios were presented to each subject. Each scenario presented the pilot with a different system malfunction or failure that required him to take some action and then make a decision regarding the completion of the mission. The system failures were introduced to the subjects concurrently with inherent cockpit systems and with the additional display added for the experiment. The additional display provided textual information to the pilots. The display provided information using one of the twelve levels of system knowledge and authority as described above. The list of scenarios is shown in Table 2.

- Scenario #1: Hydraulic failure
- Scenario #2: % Torque Split
- Scenario #3: Generator Failure
- Scenario #4: Engine Failure
- Scenario #5: Engine Oil Temperature
- Scenario #6: Fuel Pressure Loss
- Scenario #7: Rotor Vibration
- Scenario #8: Engine High-Speed Shaft Failure
- Scenario #9: Engine Fire Light Illuminated With No Fire
- Scenario #10: Airspeed Indications Incorrect
- Scenario #11: Crack in Tail Rotor Spar
- Scenario #12: Main Transmission Oil Pressure Slowly Decaying
- Scenario #13: Impending Main Rotor Blade Failure
- Scenario #14: % Torque Split (Indicates High Speed Shaft Failure)

Table 2.

Experimental Procedure

Initial Briefings The general scenario was described as an emergency mission to transport medical personnel to an aircraft crash site. The intent of this scenario was to add urgency to the mission to prevent the subjects from choosing to land and cancel the flight for minor malfunctions. The additional display was explained as an experimental, but functional, cockpit warning system. The subjects using the alerting system were briefed that they could use the additional system as a primary indicator of system malfunctions. The subjects using the non-alerting system were briefed that the system should be used as a backup/secondary system only. The tester reiterated to each subject the urgency of the mission and the functions and capabilities of the new warning system.

The subjects had three landing options during each scenario: the destination airfield one hour away, an alternate improved airfield 15 minutes away, or an unimproved emergency landing area (empty fields) in the immediate vicinity of the aircraft. The pilots were told that their mission was important, but passenger and aircraft safety was paramount.

The subjects were briefed that they were required to make all decisions in the cockpit. An experimenter acted as second pilot. The second pilot acted as the pilot on the controls and took appropriate actions, but only at the direction of the subject pilot. The experimenter never discussed the failure situations or alerted the subject pilot to any problems.

Main Experiment Pilots were initially responsible for maintaining a heading and altitude to reach an airfield one hour away. They were presented with enroute system failure scenarios that required them to respond to the malfunctions and then make decisions on the status of the aircraft and its ability to continue the assigned mission.

Failures were initiated at various periods (2-4 minutes) after the scenario began. When the operator input the malfunction to the system, the aircraft systems and instruments reacted with the indicated failure. Simultaneous with the initiation of the malfunction in the simulator, the test data was presented to the subjects on the LCD screen in the cockpit. The malfunction information appeared almost simultaneously on the LCD and on the aircraft instrumentation. The pilots were then to react to the malfunction by performing the appropriate emergency procedure. At the conclusion of any immediate action steps performed, the pilots informed the copilot of their landing decision (continue mission, divert, or land immediately).

At the conclusion of each scenario, either after a landing was made, the decision was made to continue, or the aircraft crashed, the scenario ended. The failure for that scenario was removed and the pilot either regained or was reset to his original flight parameters for the next scenario. This insured that the starting conditions were the same for each scenario.

Over-Reliance Test Conditions Two additional scenarios were added to the end of the experiment (scenarios 13 and 14) to test the tendency of the pilots to trust the new alerting system after using it for only a short time. The first scenario (#13) provided the pilot with information that was not available from the aircraft's inherent warning system. This information was correct, and if followed, prevented an incident.

Scenario #14 attempted to examine how pilots would respond to alerting system commands that were not supported by – or conflicted with – other cockpit indications. The actual malfunction presented in scenario #14 was an engine torque split, with the #1 engine failing to low side. In this malfunction, the #1 engine was failing, and the #2 engine was providing power to keep the aircraft flying. The LCD display indicated that the malfunction was a #2 engine high speed shaft failure. The procedure for resolving this malfunction includes an emergency shut down of the #2 engine. The LCD display for scenario #14 is shown in Figure 1. All pilots received identical information for scenario #14.

- #2 Engine High Speed Shaft Failure
- COLLECTIVE ADJUST.
- EMERGENCY ENGINE SHUTDOWN ON #2 ENGINE.

Figure 1.

The correct emergency procedure for the actual malfunction, an engine torque split is shown in Figure 2. The emergency procedure requires the pilot to execute either step 2 or 3, depending on the reaction of the engines. Step 2 was the proper action for this malfunction.

1. If TGT limit on either engine is not exceeded, slowly retard Engine Power Control lever on high % TRQ engine and observe % TRQ of low power engine.
2. If % TRQ of low power engine increases, Engine Power Control lever on high power engine – Retard to maintain % TRQ approximately 10 % below other engine.
- (OR)
3. If % TRQ of low power engine does not increase, or % RPM R decreases, Engine Power Control lever – Return high power engine to I.L.Y.
4. Land as soon as practicable

Figure 2.

The two malfunctions have some similar indications and some contradicting indications. The #1 engine has a lower RPM indication in both malfunctions, so a first glance may verify the malfunction shown on the LCD display. However, the torque indications show the opposite indications of what would be indicated in a high speed shaft failure. For the shaft failure (shown on the LCD display), the #2 torque indication would be low and #1 indication would be high. For the torque split (the actual malfunction), the #1 torque was low and the #2 was high.

The pilots, therefore, were required to examine the instruments closely to ascertain that the malfunction presented on the LCD display was not actually occurring, but that another failure was present. If the subjects followed the LCD display, and did not verify the malfunction with the instruments, they shut down the only good engine on the aircraft and crashed. If they attempted to verify the information on the display, they found that the instruments indicated an entirely different malfunction requiring the opposite procedure.

Debrief. After all 14 scenarios were complete, the pilots were asked to move to an adjoining room for a debrief. During the debrief, each scenario was discussed and a list of questions was answered. The videotape of the simulator period was available and was used to help the subjects recall specific scenarios when required.

Measures

Performance Measures. Three objective performance measures were selected for the experiment:

- Did the subject take the correct action in response to the malfunction?
- Did the subject make the correct landing decision?
- How quickly did the subject respond?

Subjective Measures (Debriefing). Each subject answered a series of questions in an extensive debrief following the simulation period. The debrief contained a list of questions for each scenario, followed by general questions on their opinions of the display and how they would improve it. These questions included:

- What was your first indication of a malfunction?
- What did you look at next to verify or gain more information?
- What was the primary indication you used to diagnose the system failure?
- Was the new warning system helpful?
- Did the new warning system make your decision process faster (through additional information) or slower (due to additional time spent verifying)?

Results

Performance Measures

Due to the consistent training level of the pilots, the performance measures did not result in any measurable data. The subjects were over 92% correct in their responses to the malfunctions (122 of 132 correct), so these measures did not provide any measurable differences. The response time also provided no usable data. Determining when the response actually occurred (when the pilot vocalized the procedure he was taking, when he grabbed a switch, or when he completed the action) was found to be too subjective. This made the response time performance measure unusable.

Subjective measures

Interesting results were found in several areas. The first was the in results due to the knowledge level and the amount of information on the LCD display. The second was the results due to the assertiveness of the system. The final area was the results of scenario 14, which looked at conflicting indications. This paper will focus on these three areas.

Results Based on Knowledge Level. As the knowledge level of the system increased from the lowest to the highest level, the amount of information on the LCD display increased. At the higher levels of knowledge, the new warning system had more knowledge than was realistically possible with current technology. This may have affected the responses of the pilots.

The responses to the question, “Was the new warning system helpful?” are shown in Figure 3. The percentage of responses indicating that the LCD display was helpful increased steadily from level one through level three. The responses then leveled off with no further improvement through level six. In a paired-comparison statistical test at the 0.05 level of significance, the number of responses indicating that the system was helpful at level three is significantly different than the number of responses indicating it was helpful at level one. Similarly, there is no difference between levels three through six. This seems to indicate that the increasing amount of information provided on the display was more helpful only through the first three levels. Increasing amounts of information beyond level three were perceived as adding no additional benefit to the pilots.

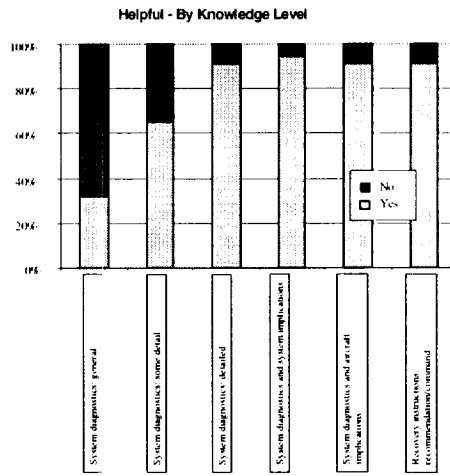


Figure 3.

The responses to the question, “Did the new warning system make your decision process faster (through additional information) or slower (due to additional time spent verifying)?” also revealed interesting results. (See Figure 4). The percentage of responses indicating that the LCD display made the decision process faster did not increase as the knowledge level increased, as anticipated. Instead, the responses show no particular pattern through level five. Then, comparing levels five and six, the responses for level six show a decrease in responses indicating that the display made the decision process faster, and an increase in responses that the display made the process slower. In a paired-comparison statistical test at the 0.05 level of significance, the number of responses indicating that the system made the decision process faster at level six is significantly different than the number of responses indicating it made the process faster at level five. The responses indicating that the system made the process faster are actually lower for level six than for levels three and four, and the percentage indicating that it made the process slower is the highest of all levels. These results indicate that the knowledge level of the system did not have a direct effect on making the decision process faster or slower for the first five levels. It seems to have had a detrimental effect at level six.

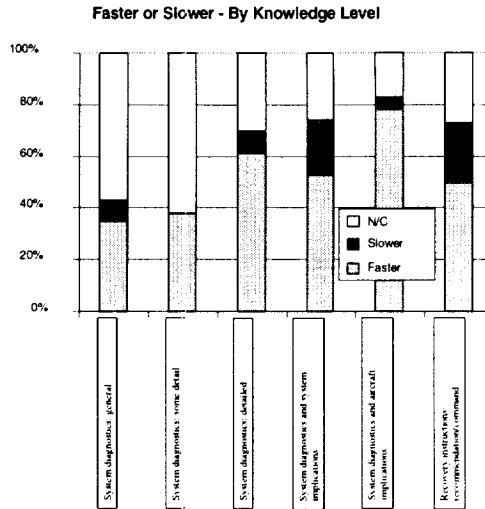


Figure 4.

Results Based on Assertiveness. Based on the assertiveness of the system, there was a significant difference in responses to the question, “What was your first indication of a malfunction?” In a paired-comparison statistical test at the 0.05 level of significance, the number of responses indicating that their first indication of a malfunction was the LCD display is significantly different than the number indicating that the cockpit instruments was their first indication. These results are shown in Figure 5. This result may be indicative of the effect of the directing-attention function of an alerting system compared to a non-alerting system.

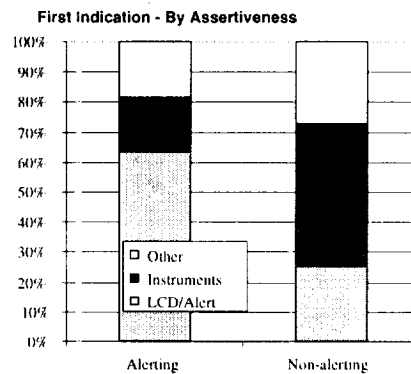


Figure 5.

Over-Reliance Test Results

12 subjects were presented with scenario #14. The results were:

Action	Number of Subjects
Disregarded LCD display	5
Followed LCD display	4
Initially followed LCD display	2
Landed with no action	1

Five subjects disregarded the LCD display and reacted to the actual malfunction using the information on the cockpit instruments. These subjects performed the correct procedure for an engine torque split and continued the mission safely. Four subjects followed the LCD display, shut down the good engine without confirmation from the other instruments, and crashed. Two subjects initially followed the LCD display, decreased power or idled the good engine, then recognized that the instruments were not confirming the LCD display information, and executed the correct procedure for a torque split. One subject initially followed the LCD display and decreased power on the good engine, then stopped, and landed with no action. He stated that he “could not resolve the conflict.”

Overall, seven of the twelve subjects performed the wrong procedure or were initially confused as to what action to take. The subjects had all experienced a similar malfunction earlier (torque split with the #2 engine failing), and all had performed the correct procedure. The subjects in each category were evenly divided by experience in total flight time.

DISCUSSION AND CONCLUSIONS

Several preliminary conclusions can be made from these results. First, the notion that “more information is always better” is disputed. Increasing usefulness of information to a pilot seems to stop at a limited amount of information. This was illustrated as the responses for the perceived helpfulness of the system leveled off at the third level. Also, further information provided beyond that point may slow the decision making process, as indicated by the decrease in the speed of the decision process at level six. This raises an interesting question: Is there a finite amount of information that is useful to pilots in an emergency situation, and can we ascertain what that level is?

The assertiveness level results indicated that the subjects used the alerting system more as an alerting signal than as a diagnostic tool. This indicates that the higher assertiveness levels are useful as an alerting tool or an attention-directing mechanism; however, it can not be assumed that the pilot will then follow the alerting system as the sole source of information.

No alerting system is correct 100% of the time. The over-reliance test in scenario 14 suggests that when an alerting system gives erroneous information that conflicts with other cockpit indications, serious mistakes can be made and pilots may have trouble correctly resolving the conflicts. Creating an alerting system whose commands can be easily assessed by the pilot remains a significant design challenge.

SUMMARY: STUDY #2

Responsibility for the safe completion of a flight rests primarily with the pilot-in-command. During emergencies onboard air transport aircraft, this responsibility can be demanding, due to the large number of tasks to which the pilot must attend, including: detecting and resolving failures in aircraft systems; continuing to monitor aircraft system health; coordinating with cabin crew, airline dispatchers and air traffic control; controlling the aircraft; and deciding upon (and then following) a course of action that will result in a safe landing. This inherent difficulty is compounded by a significant number of stressors, including physical danger, an uncomfortable physical environment (heat, smoke, noise, etc.), an overwhelming amount of information to consider, and the need to make in a short period of time. In addition, the aircraft may have degraded performance and handling qualities, limiting the extent to which the pilot’s past experience is relevant to the present problem.

The objectives of this research were to investigate how pilots generate and then follow a four-dimensional (4D) trajectory to the runway threshold during emergencies, and to examine the functions needed in pilot aids for these tasks. This paper first presents relevant research from a number of domains, highlighting the important aspects of these tasks, pilots’ needs in cockpit aids, and available technologies. Then, the design of a prototype aid is described. The results of a flight simulator evaluation with airline pilots are detailed. The paper concludes with a discussion of pilot performance at these tasks and design recommendations for future cockpit systems.

Background And Motivation

Once an emergency condition exists, effective generation of a safe trajectory (and then following this trajectory) becomes crucial to a safe landing. If done well, this can prevent a serious failure from evolving into an accident; if done poorly, a comparatively minor problem can lead to aircraft damage and fatalities. This trajectory must address multiple conflicting objectives including: minimizing to time-to-land; bounding stress on the aircraft

imposed by maneuvering; meeting airspace and regulatory limits, and flight envelope limits; and ensuring the plan is robust against uncertain and unpredictable elements of the environment.

In this paper, emergency trajectory generation is defined as the determination of a course of action with specific detail to describe all aircraft dynamic states for the remainder of the flight. This combination of a high level-of-detail and a long time-scale differentiates it from other types of trajectory generation and standard methods of flight planning. For example, strategic planning activities such as flight planning share extended time-scales with emergency trajectory generation, but utilize a low fidelity representation of the aircraft.¹ More specifically, plans generated through strategic planning are often described by waypoints and altitude crossings, not through a detailed trajectory. A common example of a strategic planning aid is the Flight Management Systems (FMS) currently found in modern air transport aircraft; air traffic control instructions and flight plans are also typically at this level of detail. Likewise, while time-critical planning requires the same detailed aircraft model as emergency trajectory generation, it does so over a time-scale on the order of seconds to minutes.¹ Due to the limited time-scale, such time-critical plans usually encompass only a single action or maneuver that meets a singular goal. Cockpit systems that provide this level of planning include the Traffic alert and Collision Avoidance System (TCAS), Ground Proximity Warning System (GPWS), and Rotorcraft Pilots Associate's Actions on Contact functions (RPA).²

Emergency trajectory generation instead falls under the definition of tactical planning proposed in¹. This type of planning requires both a high level of detail and a long time-scale in order to avoid generating a trajectory that is later found to be lacking. For instance, not including the detailed effects of aircraft dynamics may result in a delayed landing due to missed localizer or glideslope intercepts (when assumptions about turn rate, descent rate, etc. can not be met), or the execution of an overly extended flight path (when maximum performance maneuvering is not used by the flight plan). In an emergency, either of these situations can be a serious detriment to the safety of the flight.

The representation of a plan used in this study was that of a procedure. Specifically, a flight plan and its associated trajectory were defined and communicated as a series of actions (e.g. 'turn to heading 300' or 'descend to 8000 feet') initiated by discrete triggers and linked by the aircraft's continuously evolving dynamic states. This representation was chosen for several reasons. First, procedures are a common representation of tasks in high-workload, complex environments, including aviation.^{3,4} Second, trajectories are typically represented in civil aviation as procedures, with published charts dictating, for example, the turns, descents and speed changes demanded by specific arrival routes and approaches; therefore, a cockpit aid using this representation in emergencies would provide a familiar view to pilots and establish a flying task for which pilots are already highly trained. Finally, because this representation is so prevalent in nominal operations, autopilots and FMS have been designed to fly the aircraft by initiating distinct new control behaviors and target states at discrete points.

The time or place each action should be initiated, and its severity (e.g. the rate of a turn, descent rate, etc.) are dependent on the aircraft trajectory and states. For example, the time to start a turn onto the final approach course, and the required rate of turn, are dictated by aircraft speed through its impact on turn radius. As multiple actions are placed in series, a cascading effect ensues, with each action altering the aircraft trajectory and dynamic states at the time of subsequent actions. Continuing the example, a high-rate descent preceding the turn onto the localizer can increase the airspeed, which subsequently increases the turn-radius, and therefore may require changing the inbound course, which will subsequently affect the distance traveled and the descent rate need to reach glideslope intercept altitude, etc. This complex coupling prevents the decomposition of the trajectory into separate independent flight segments. Additionally, the coupling between such properties as descent-rate, speed, and turn-radius prevents the separation of the plan into lateral and vertical components. This makes it difficult to plan of a complete set of actions for the entire arrival and approach.

Generation of a detailed emergency trajectory can therefore be viewed as a task that may prevent problems such as taking too long to land (important in smoke and fire situations) or requiring extreme maneuvers to intercept the localizer and glideslope (important in situations with degraded aircraft stability and maneuverability). While several studies have examined re-planning in general,^{1,5,6,7} and military tactical planning aids in particular,² little experimental data exists on how air transport pilots plan a trajectory in emergencies. Likewise, cockpit voice recorder transcripts and accident reports provide only sparse and anecdotal evidence of how pilots perform this task.

The current literature on human decision making suggests that detailed trajectory generation is a very difficult task for pilots, as illustrated by two models of decision making and planning. A rational, analytic model of planning assumes the sequential process of (1) generation of alternatives, (2) imagining the consequences, perhaps through the process of 'mental simulation', (3) valuing (or evaluating) the consequences of the alternatives, and (4) choosing one alternative as a plan.⁸ Models describing observed human behavior in a variety of domains suggest

that experienced operators, such as pilots, rely substantially upon non-analytic strategies such as those defined by the Recognition-Primed Decision model.^{9,10} Through the use of pattern matching and recognition techniques, these non-analytic strategies have the advantage of rapidly providing a starting plan which may then be iteratively improved as circumstances allow. For pilots, this method works well in situations covered by their training and experience. However, the effective implementation of this method is reliant on three assumptions: (1) the pilots have sufficient experience, training, and intuition with very similar situations to select a reasonable initial plan of action; (2) the pilot is able to quickly and correctly evaluate the consequences of the plan; and (3) the detection of any bad decisions occurs early enough for the pilot to select and evaluate an alternate feasible course of action.

Both types of decision-making models note the need for pilots both to identify a reasonable initial plan of action, and to evaluate or predict the consequence of that plan of action. However, each emergency situation is highly unique: each occurs in a different place with a different underlying cause, different goals, and different obstacles to a safe landing. For example, one situation may demand a safe path to a nearby airport with a damaged aircraft; another situation may require the quickest trajectory to a far-away airport.

Some aspects of pilot training may be relevant to these tasks: specifically, in initial training on single engine aircraft in visual conditions, pilots are required to demonstrate the ability to execute a forced landing in a field in simulated engine-out conditions. However, more advanced training programs typically emphasize nominal operations, in which aircraft trajectory is dictated by published air routes and FMS calculations, rather than determined by the pilot. These programs also emphasize the procedural aspects of emergency responses, such as executing the correct procedures for specific emergencies; however, the common last step of emergency procedures is 'Land as soon as possible,' which does not provide detail as to what the landing trajectory should be. Extensively training pilots on all aspects of trajectory generation would be difficult, given the large number possible situations that would need to be covered.

Therefore, the task of identifying an initial feasible guess for a trajectory cannot be completely trained for, and instead presents pilots with an active and intensive task with only general guidelines as an aid. Likewise, the task of evaluating the performance expected of a planned trajectory is very difficult, given the magnitude of predicting all facets of a highly-detailed trajectory all the way to the runway and the aforementioned limits on decomposing the trajectory into manageable parts.

Unlike the time-critical and strategic planning aids mentioned earlier, no cockpit decision-aid exists that directly addresses the needs of emergency trajectory generation. Several cockpit aids intended for other purposes have some applicability. The first are charts and approach plates, which depict published air routes and approach procedures. The trajectories they present are not represented with a high level of detail and are formulated to meet criteria such as traffic flow which may not be relevant during an emergency; however, they still provide a baseline plan and act as a source of trajectory limits imposed by factors such as terrain. For pilots of transport aircraft equipped with glass cockpits, additional planning aids are available in the form of the trend-vector and the altitude range arc, providing accurate turn radius and bottom-of-descent information. However, these are of limited planning use as they are based solely on current aircraft states, and hence can neither depict the impact of future actions nor indicate whether current actions will ultimately contribute to a safe landing.

At this time, the 'level of automation' most appropriate for this task (i.e. which of the functions the aid should take over, and the ability of the pilot to override the system and/or modify its suggestions) is not known.^{11,12} The earlier discussions of decision making highlighted two functions that an aid may perform: identifying a reasonable initial plan of action, and evaluating the consequences of those actions. However, other issues must also be considered in assigning the role and function of the aid because of the impact they can have on the pilots' interaction with it. Studies of operator interaction with automated systems have repeatedly identified cases where automated or intelligent systems are not used because they do not bring sufficient benefits to the situation to warrant the time and effort required to use them, a condition commonly called under-reliance. Conversely, if the aid is capable of completely taking over a task, operators are prone to either completely rely on the system without verifying its accuracy and appropriateness to the immediate context (a condition commonly called over-reliance or mis-use), or to be biased by the output of the aid to the point that they can not reason independently (a condition commonly called automation bias).^{13,14,15,16} For example, in a study of a cooperative flight planning system (examining strategic planning), roughly 40% of pilots were induced to select poor flight plans by the introduction of faulty system information.⁷

This suggests that greater understanding is required of how pilots plan their flights in emergencies, and what interventions can be made to aid them and to encourage more-detailed trajectory generation. This study focused on the use of an intelligent cockpit system to examine both these research needs: interaction with such a

system in a flight simulator test provides a preliminary assessment of the qualities and functions pilots require from such a tool, and also forces pilot to actively demonstrate and verbalize their approach to planning.

It is envisioned that pilots will use trajectory-generation aids such as the one described in this paper after the decision to land is made. While the aircraft is heading to the destination airport, the Pilot-Not-Flying (PNF) will utilize the aid to plan a feasible set of actions for the arrival, approach and landing. At this point, before committing to any plan, the flight crew can review its consequences on the trajectory. After final acceptance of a plan, the pilots will then fly the plan, either manually using the aid as a reference, or through an automatic control system commanded by the aid. Pilots may also opportunistically improve the trajectory, or if the trajectory is found lacking, purposefully revert back to planning.

Beyond the benefits noted earlier in ensuring that near-term actions will lead to a safe landing, this emphasis on first planning and then flying has distinct advantages to the pilots given the cognitive demands they face.^{10,17} Planning is a highly cognitive activity demanding their full attention; as such, it is often limited to pre-flight and isolated (preferably low-tempo) periods of the flight. By generating the plan, the pilot then makes the subsequent flying task easier by producing a reference trajectory to follow without continuous involvement and re-planning.

This cockpit-decision aid complements other recent research efforts. For example, several studies have examined the fault-detection and fault-management processes also associated with emergencies.^{18,19,20} Likewise, several studies are examining the control technologies that can help a pilot fly a reference trajectory (or automatically control the airplane) when the aircraft's handling qualities have degraded.^{21,22}

Design and Development of A Prototype Planning Aid

This section outlines the development of a prototype called the Emergency Flight Planner (EFP). This prototype was intended to test the feasibility of providing pilots with a tool that could effectively predict the complex interactions between the actions of a plan. Since no such tool has been documented for this application, this prototype also serves as a means by which to assess the automatic functions and capabilities needed by pilots. A schematic of a complete planner system and the subsystems it requires is shown in Figure 1. The core functionality of the planner is the ability to predict the aircraft trajectory resulting from a given plan (i.e. list of actions). This implies the need for models of the aircraft's dynamics and the pilot's control behavior. A pilot interface is also required.

Because this study sought to assess the utility of the planner to pilots through a controlled flight simulator study, this prototype implemented the subsystems shown by bold blocks in Figure 6. In the simulator, exact knowledge of aircraft dynamics were used in lieu of aircraft model identification; in an operational flight planner, information regarding the performance degradations of the aircraft would need to be obtained through real-time system identification or directly from the aircraft controller that is compensating for the failure. Likewise, pre-scripted plans were used, as automatic plan generation would require further developments in current methods for hybrid-system analysis and optimization.²³ Specifically, standard methods of optimal trajectory calculation, such as numerical solutions of the Hamilton-Jacobi-Bellman equations, are not well-suited to the four-dimensional, hybrid dynamics created by the combination of discrete actions and continuously-evolving maneuvers. Likewise, existing solutions to discrete systems cannot accommodate the continuous trajectory segments, and the complex interactions between the discrete and continuous elements prevent their separation into two individual problems.

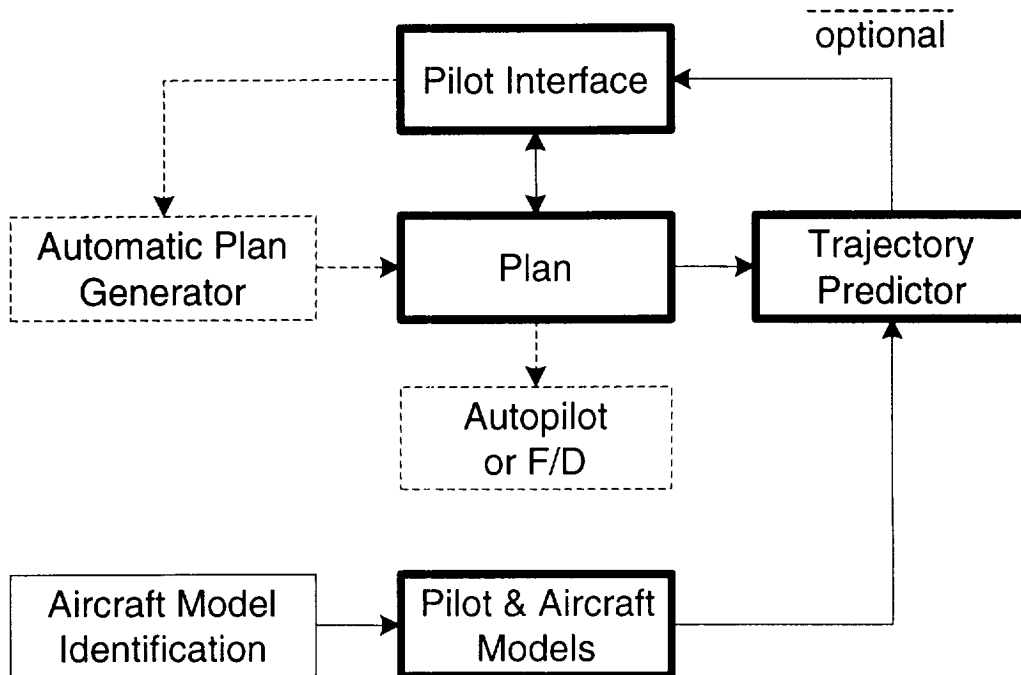


Figure 6 - Schematic of the Emergency Flight Planner

Actions and Trajectory Prediction

The trajectory-defining actions included in the EFP are those relevant to an arrival and approach to an airport, as shown in Table 3. Three types of discrete action triggers were available: elapsed time, aircraft location over a position fix, or elapsed time past a fix.

Heading	Vertical
Turn to Heading Fly to a Fix Intercept Localizer	Descend to Altitude Maintain Vertical Speed Intercept Glideslope
Speed	Miscellaneous
Set Speed Set Throttle	Set Flaps Set Gear

Table 3 - Arrival and Approach Actions Incorporated in the EFP

In predicting the future trajectory with the detail required of tactical plans, the discrete actions must be joined by accurate predictions of the continuously-evolving aircraft dynamic state. To meet these needs, the EFP used fast-time simulation to propagate the trajectory forward in time. The differential equations for the pilot-aircraft system are propagated forward, with the triggering of actions changing aircraft dynamics, commanded controls or target states at discrete points in time. For computational efficiency, the EFP utilizes a modified adaptive-timestep Runge-Kutta 4th order (RK4) algorithm. Standard adaptive-timestep RK4 algorithms maximize the time step of a continuous system while bounding numerical integration error; however, its timesteps may skip over the triggering

of new discrete actions. The modified algorithm therefore queries all active actions for an upper bound on the time step and compares it with that suggested by adaptive-timestep RK4. The EFP extrapolates most 30-minute trajectories in less than two seconds on a 450 MHz desktop PC computer.

Representing Pilot and Aircraft Behavior

The trajectory predicted by the fast-time simulation is a product of both the aircraft dynamic behavior and the control behavior expected of the pilot and/or aircraft control system. Research has shown that pilots adapt their control behavior in response to changes in the underlying aircraft dynamics to maintain a consistent closed-loop behavior; many adaptive controllers intended for flight following failures are intended to do the same.^{21,22} To replicate these control and dynamic behaviors, elaborate models of the aircraft dynamics and of control behavior may be sought for all failures over all flight conditions. However these models have obvious cost and complexity penalties; in addition, the behavior of an elaborate control models, if correct, would typically only serve to cancel out changes in the aircraft dynamic model. Therefore, the EFP used a static representation of control behavior and of aircraft dynamics that fits the stable closed-loop behavior achieved with adaptive control under a range of failures.

For the aircraft dynamics, the EFP prototype uses a stable four degree-of-freedom dynamics model: the longitudinal forces are thrust and drag; pitch and roll moments were governed by the ailerons and elevator; and coordinated flight was always assumed, thereby dictating side-force and yaw moment. Failures can be created by reconfiguring aerodynamic coefficients within the model; these effects were selected to represent predicted changes in aircraft performance, as opposed to changes in aircraft stability. Stability and control constraints were modeled as limits imposed on the pitch angle, bank angle and speed of the aircraft.

The aircraft control is handled by a collection of individual controllers for pitch, roll and throttle. These are swapped in-and-out in the same manner as autopilot modes. They control the aircraft towards the target states specified by the active actions, and keep the aircraft within the pitch, bank and speed limits demanded by the aircraft dynamic model.

Pilot Interface

Obviously, many pilot interface designs are possible; at a minimum, they must accept action and trigger information from the pilot and display the predicted trajectory to the pilot in such a way that the pilot can both assess the performance of the plan and then execute it. The pilot interface used with the EFP is shown in Figure 7. All action specific information is located on sidebar on the upper right, providing a chronologically sorted list of the actions and their triggers. The primary input device is a Control Display Unit (CDU), a common interface for air transport aircraft equipped with FMS. In the EFP, it provides a detailed textual display of a selected action, and is the entry device by which pilots can modify actions and select functions.

The predicted trajectory was displayed to the pilot on two spatial displays (the plan and vertical profile views) using a format analogous to that on pilot charts and approach plates. The trajectory is normally shown in white, except for any segments that violate flight envelope or stability constraints, which are shown in red. The current location of the aircraft is also displayed, allowing the pilot to monitor conformance to the plan. The plan view is a scalable and scrollable “North-Up” representation, with symbology based on the Boeing 747-400 Electronic Horizontal Situation Indicator (EHSI). While this view could be conceivably integrated with smaller existing EHSI displays, issues regarding clutter and resolution would need to be addressed. There is no widely used vertical profile display in air transport cockpits at this time, and no one ‘best’ display format has been experimentally demonstrated. Therefore, the EFP provides three pilot-selectable formats for the vertical profile display: the ‘time’ view displays trajectory altitude with respect to the elapsed flight time; the ‘distance’ view displays altitude along an unwrapped ground track; and the approach view provides a projection along the localizer beam, similar to that found on an approach plate.

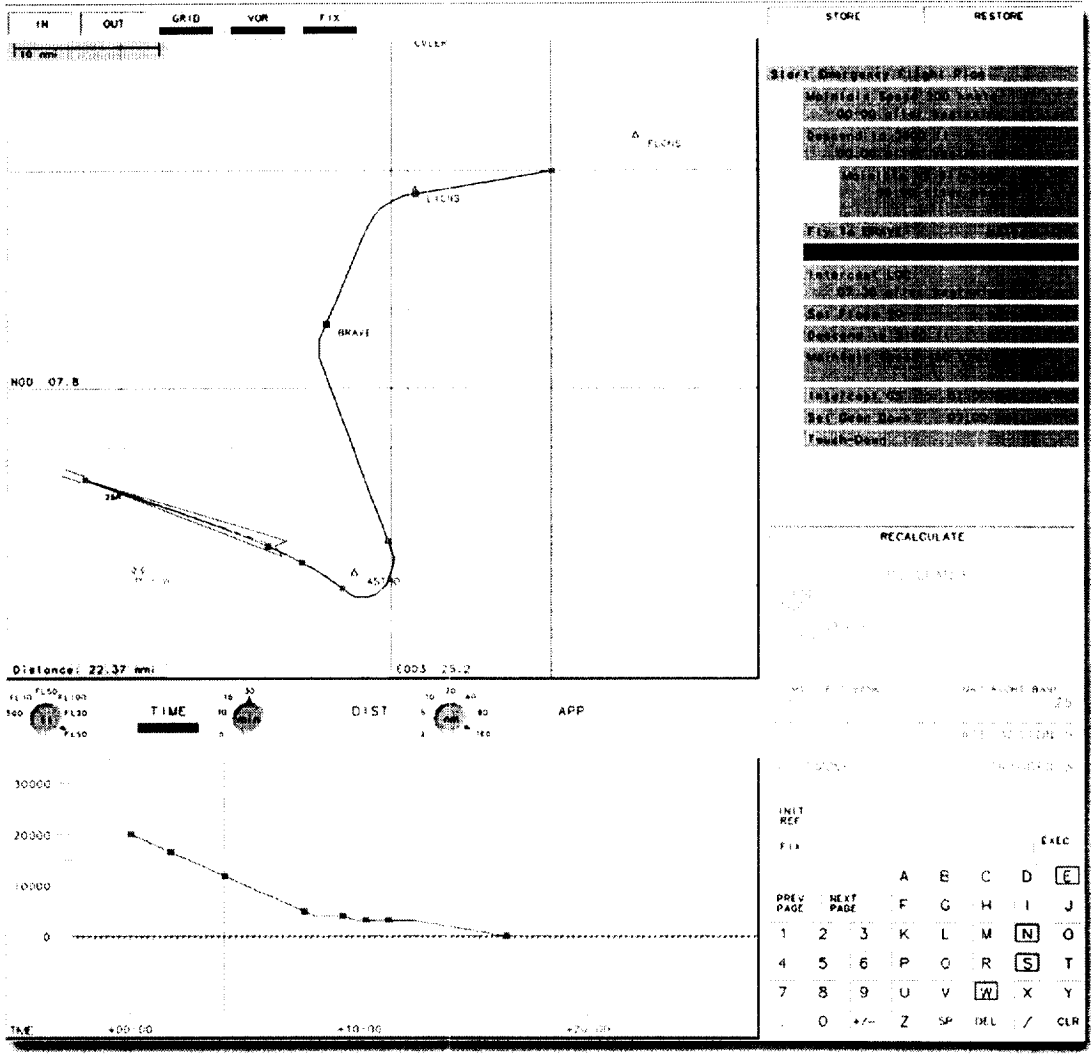


Figure 7. EFP Pilot Interface (Inverted Black & White View)

Because the trajectory has been simulated using reasonably-detailed dynamic models, the EFP can also display to the pilot a complete picture of aircraft state at any point in the future trajectory, including attitude, throttle settings, flight envelope limits, fuel status, airspeed, and aircraft configuration. The 'query view,' shown in Figure 8, displays this information at any point in the trajectory as selected by the pilot using a presentation similar to a glass cockpit Primary Flight Display (PFD). While planning, the pilot can select any point on the trajectory to see the aircraft state predicted there; while flying the aircraft, the query view can be set to automatically display the aircraft state at the point on the EFP trajectory closest to the current aircraft location.

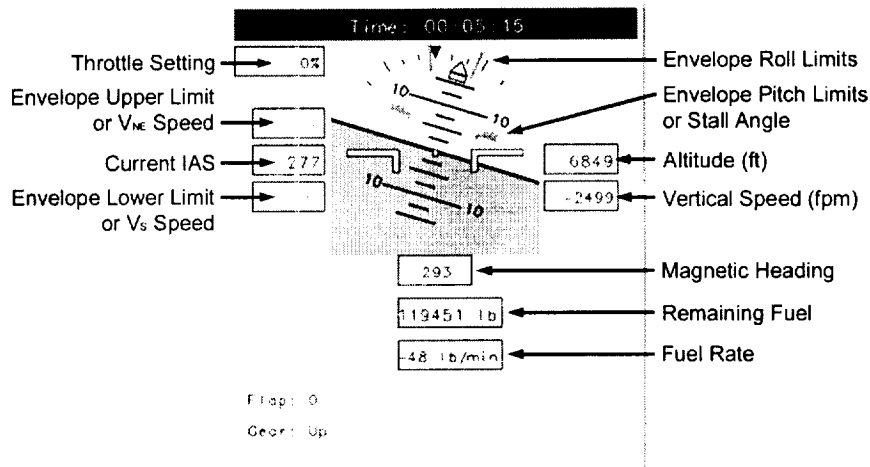


Figure 8. EFP Query View (Inverted Black and White View for Clarity)

Automated Functions

For the preliminary study described in the remainder of this paper, two variants of the EFP were created mirroring the two automatic functions discussed in the previous section. The 'Basic EFP' variant provides a mechanism by which pilots can enter a plan, from which the system then predicts the ensuing trajectory. The 'Pre-Loaded EFP' variant additionally provides automatic planning functions by presenting the pilots, at the start of planning, with a pre-loaded plan that they can accept, modify or delete. Both variants were otherwise identical, with the same interface, method of predicting the trajectory, etc.

Experiment Design

The EFP was tested in a part-task, desktop flight simulator with airline pilots as subjects. Each pilot participated in two consecutive experiments in one session. The goal of the primary experiment was to investigate how pilots approached the planning task with and without the EFP, to determine quantitatively whether either variant of the EFP aided the pilots in landing safely following a major system failure or emergency, and to gather the data needed to improve the design of in-flight planners. The secondary experiment comprised a single 'deviant' scenario in which the EFP had an erroneous model of the aircraft dynamics and hence made erroneous predictions of what a plan's associated trajectory would be. This tested the effect that such an error in the planner would have on the ability of pilots to execute a safe flight; given the sizeable evidence suggesting problems with automation bias, the hypothesis for the second experiment was that pilots would follow the erroneous trajectory prediction, with corresponding drops in performance.

Primary Experiment Independent Factors

In the primary experiment, the following two different factors were examined:

- **Planning Tool**

Charts-Only: In this baseline condition, pilots were provided with traditional paper en-route charts, STAR charts, and approach plates of the region of interest. An E6B type flight computer (a circular slide ruler) was also made available.

Basic EFP: This condition supplied the Basic EFP in addition to standard paper charts and E6B. Specifically, upon startup the Basic EFP presented the pilot with an empty action list, to which the pilot could enter actions to create a trajectory.

Pre-Loaded EFP: This condition supplied the Pre-Loaded EFP in addition to standard paper charts and E6B. Specifically, upon startup the Pre-Loaded EFP presented the pilot with a feasible trajectory, which the pilot was able to accept, ignore, clear, or modify as desired.

- Scenario Type

Performance Altering (PA) Scenarios: This type of scenario created conditions in which the pilot needed to plan (and then fly) a trajectory in which the aircraft had substantially different performance from nominal. The failures were: engine failure; stuck rudder; and inadvertent spoiler deployment.

Non-Performance Altering (NPA) Scenarios: This type of scenario created conditions in which aircraft performance was currently nominal, but a compelling need existed for an immediate emergency landing. The failures were: smoke in the cabin; cargo fire; and medical emergency.

Secondary Experiment Independent Factor

The secondary experiment had only one independent factor: the same three tool types as used in the primary experiment. The secondary experiment was restricted to a single performance-altering 'deviant' scenario (Asymmetric Loss of Outboard Aileron) in which the ability to turn to the left was diminished, but the EFP showed the opposite information, used this erroneous information in predicting the future trajectory, and, in the case of the Pre-Loaded EFP, suggested an erroneous trajectory.

Test Matrix

Each pilot completed a total of seven scenarios. The first six runs spanned all six combinations of independent factors (3 tool types X 2 scenario types) in the primary experiment; the final, seventh run used the secondary experiment's deviant scenario, with pilots equally divided among the three tool types. The orders of the runs were blocked by tool type to mitigate any learning effects due to increased familiarity with any tool.

Experiment Apparatus

The experiment was conducted at Georgia Tech utilizing the Reconfigurable Flight Simulator (RFS) software running on two networked desktop workstations, each with a 19-inch monitor.²⁴ One workstation and monitor set was dedicated to the EFP. The other workstation and its monitor provided the pilot with cockpit instruments, including a PFD, EHSI and Engine Indicating and Crew Alerting System (EICAS), all based on B747-400 displays. Additional envelope limits for roll, pitch, and speed were depicted on the PFD using the same format as the query tool, shown in Figure 3. Control of the aircraft was enabled through a side-stick and throttle while the EFP used a cursor controlled by a trackball.

Experiment Procedure and Scenarios

Following a briefing and two training runs, each pilot was asked to fly the seven data-collection runs specified by the test matrix. In each run, the pilot was told that he or she was the Captain of a Boeing 747-400, that an emergency had occurred, and that all relevant emergency checklists had already been performed. In all scenarios, the aircraft was in Instrument Meteorological Conditions (IMC) with no terrain or traffic considerations. Each run was split into two parts. During the first part, the pilot was asked to plan their approach to the airport for 15 minutes using the available tools; this period was described as an interval where the First Officer (not actually present at the experiment) was holding the aircraft in a descent towards a 'hand-off' point nearer the airport. The pilot was asked to verbalize the criteria and methods he or she applied in building each plan.

The second part then required the pilot to take control of the aircraft at the hand-off point, steer it onto the localizer and glideslope of the landing runway, and maintain the approach until 500 feet above the runway threshold. The aircraft dynamic model of the simulator was the same as that in the EFP with one exception: in the deviant scenario, the aircraft model underlying the simulator flown by the pilot utilized a different dynamic model from the EFP.

To avoid pilot familiarity with an airport, all scenarios involved fictitious airports. While all scenarios shared a common airspace structure and were intended to be of similar difficulty, slight differences in orientation and starting conditions were created to prevent learning effects. The starting conditions of all scenarios were

calibrated such that the Pre-Loaded EFP plan utilized similar amounts of aircraft maneuvering and programming effort. Additionally, the pre-loaded plans were constrained to be within a flight time of 13 to 14 minutes and a track distance of 55 to 65 nautical miles, while staying within all published attitude and speed limits; these plans had 13 or 14 actions each, including several configuration actions for extending the gear and each stage of flaps.

Subjects

Twelve airline pilots participated in this study. All had prior experience with FMS and “moving map” displays. Of the twelve pilots, eight were captains, and four were first officers. Average flight hours were 14000hrs and 8600hrs for the two groups respectively. Total flight hours ranged from 3800hrs to 25000hrs. All but one had received military flight training.

Primary Experiment Results

A total of 72 runs were performed in the primary experiment. Unless otherwise specified, the data sets were analyzed for tool and scenario type effects by fitting to a general linear model. The tool and scenario type were analyzed as fixed effects; pilots were analyzed as a random factor to allow the results to be generalized to the entire population of pilots. In addition, the general linear model also tested for interactions between the factors. Where significant variation was found, more specific tests identified significant differences, including one-way Analysis of Variance (ANOVA) and the Tukey multiple comparison procedure with 95% confidence intervals.²⁵ To test the residuals of the fit for the normality assumptions of these tests, the Kolmogorov-Smirnov normality test was applied.²⁶ In cases where the assumption of normality for the data did not hold, a non-parametric Kruskal-Wallis test was performed.²⁶

Pilot Performance in Planning and Flying Trajectories

The number of missed approaches (here defined as a situation where the pilot could not establish a stable flight path on both the localizer and glideslope by 500 feet above ground level) is an important measure of safety and pilot performance. A missed approach entails the aircraft having to circle for another approach, adding significantly more time and requiring additional low-altitude maneuvering. During the 72 runs, 6 instances of missed approaches were recorded. In 5 of these 6 instances, the pilot did not use the EFP as the primary reference. The only other instance of a missed approach occurred with a Pre-Loaded EFP variant in a PA scenario. In this case, the pilot did attempt to follow the plan given by the EFP.

While the small number of samples precludes any rigorous statistical analysis, further insight may be gained by observing the underlying cause of the missed approaches. Of the 6 missed approaches, 4 occurred during rapid-descent maneuvers in the time-critical NPA scenarios. In none of these runs did the pilot follow the plan in the EFP. One possible explanation for the high number of overly rapid descents is the lack of comprehension of the consequences of high descent rates and close-in ILS intercepts. The other two missed approaches were both PA scenarios with no apparent common denominator.

Another important metric of pilot performance is time to land; even in situations where time is not the highest priority, extending the duration of a flight is risky due to the unknown lifetime remaining in damaged aircraft systems. The average of the time to land measure (defined as the length of the pilots' flying time from the hand-off point to when the aircraft reached a height of 500 feet above ground level) is shown in Figure 9, with one outlier data point removed.

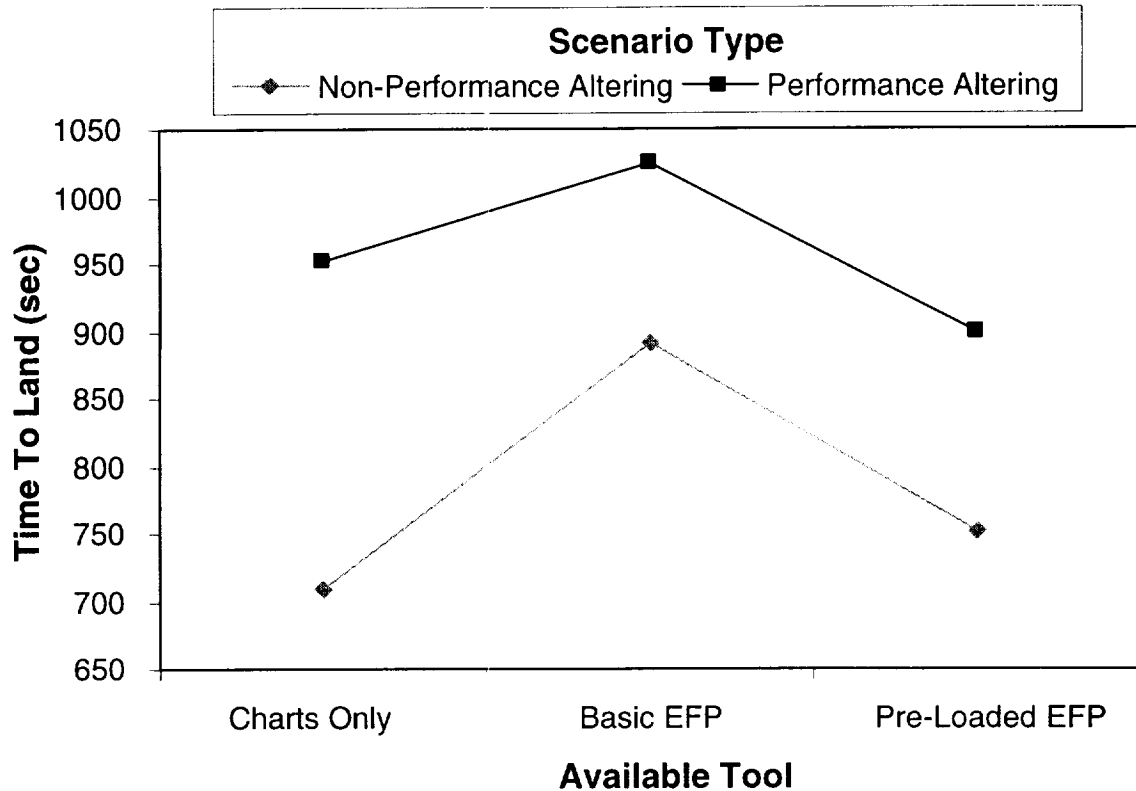


Figure 9 – Average Time-to-Land Categorized by Tool Type

An ANOVA and Tukey test found that the time-to-land for NPA scenarios is, on average, significantly lower than for the PA scenarios ($F=18.80$, $p<0.001$). The difference between NPA and PA scenarios' times can be attributed to the time-critical nature of NPA scenarios such as medical emergencies or fires. Conversely, pilots appear to be more conservative in PA scenarios for the sake of aircraft stability. In addition, analysis of the data found that the availability of the Basic EFP variant resulted in a greater time than the other two tool options, as shown by a Kruskal-Wallis test ($H=6.68$, $p=0.035$).

From experimenter observations and pilot comments, it was noted that pilots did not always follow the EFP's plans, most likely due to several factors such as difficulty in entering a plan (in the case of the Basic EFP) and concern regarding the adequacy of the pre-loaded plans (in the case of the Pre-Loaded EFP). This suggested that a more detailed factor could be used to provide more insight; results for both EFP variants were each broken down into two sub-categories, one for whether the pilots at least partially used the EFP, and the other for when the pilot did not follow the plan in the EFP at all. EFP usage was defined as situations where the pilot followed its plan for at least a portion of the flight, as judged by comparing track and vertical profile data from both the EFP plans and actual flight data. This created 5 distinct categories as shown in Figure 10. The time-to-land values are referenced to the length of the unmodified Pre-Loaded EFP plan around which the scenario was designed.

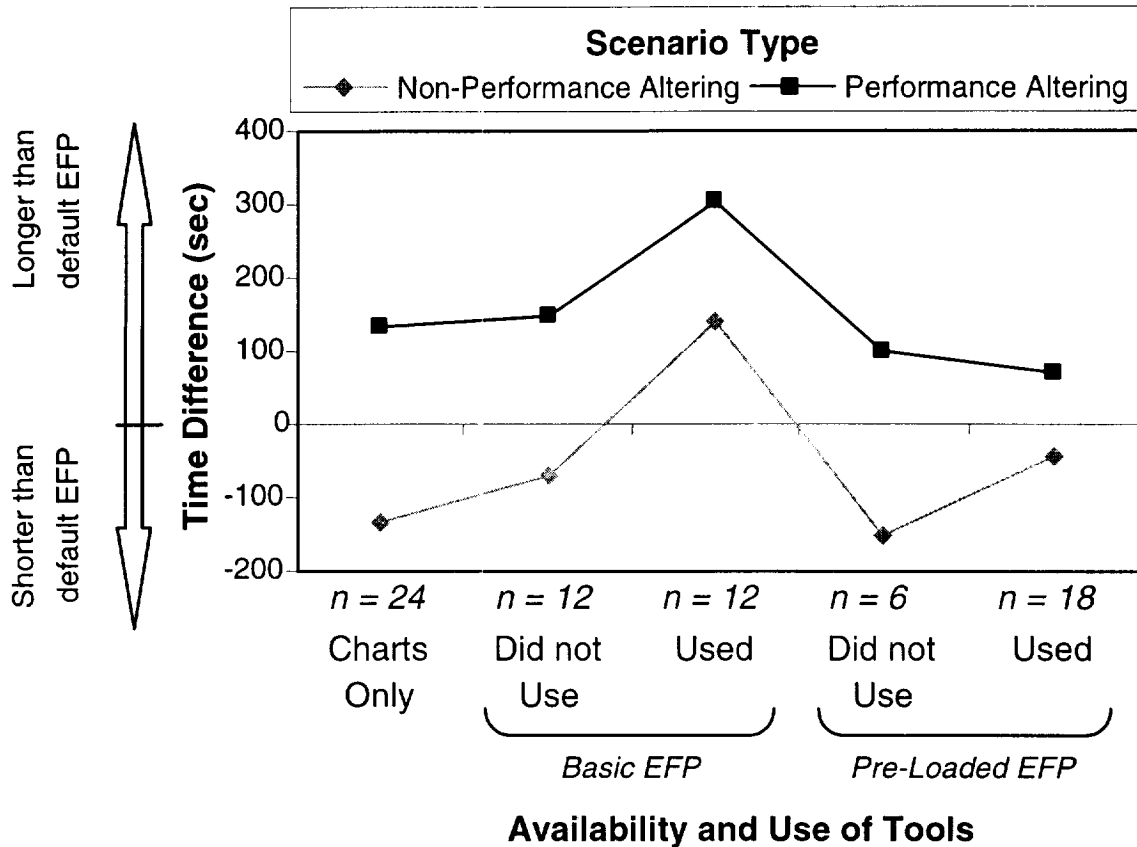


Figure 10 - Average Time To Land Categorized by Tool Type and Usage, Referenced to the Pre-Loaded Plan Duration

ANOVA found significant variation between these five conditions ($F=2.80, p=0.033$). A Tukey test with 95% limits identified significantly higher times in cases where the Basic EFP was used compared to the charts only condition. The same test with weaker 90% confidence limits shows an increase over all the other conditions (Basic EFP not used; Pre-Loaded EFP used and not used). Analysis of the duration of the plan created within the EFP provides a possible explanation. A Kruskal-Wallis test showed significant differences in predicted duration between cases using the two different EFP variants ($H=6.82, p=0.009$); specifically, the plans created in the Basic EFP were an average 1.5 minutes longer than the pre-loaded plans. Therefore, because the plans that pilots created in the Basic EFP were longer, adherence to them may have also caused a longer flight than required. No statistically significant differences were found between Pre-Loaded EFP and the baseline Charts-Only tool types.

Planning Constraints and Assumptions

Measures were also made into how pilots planned and flew in the different scenarios. Specifically, significant scenario effects were found in the number of violations of the placard flap and gear speed limits. In the NPA scenarios, where the emergency tended to be time-critical, several pilots opined that exceeding the flap speed limits was acceptable given the assumption that approximately a 10-knot safety buffer was incorporated into the listed value. The data mirrors their opinions, with a significantly higher number of flap violations in the NPA scenarios ($F=4.47, p=0.038$). However, the data also showed significant results for violations that were more than 10 knots over the listed value. With this revised limit, the NPA scenarios again had higher instances of violations with respect to the PA scenarios ($F=6.09, p=0.016$). In these cases, several pilots violated their own self-reported limits, apparently to land the aircraft as soon as possible.

This data provides two design insights. First, and most importantly, pilots' planning objectives change with the context of different emergency situations; correspondingly, flight envelope limits may also need to be relaxed in specific circumstances. Second, even with an undamaged aircraft, pilots may not fully realize the

dynamic interactions between trajectory-defining actions and therefore may not plan a trajectory that does not exceed aircraft limits.

Actions and Triggers Used by Pilots in Creating Plans

The types of actions and triggers in plans created by the pilots using the Basic EFP were recorded. Figure 11 shows the different types of actions with their cumulative total in all pilot-created plans, including plans that were ultimately not followed by pilots or were infeasible. A substantial number of “Fly to Fix”, “Maintain Speed”, and “Descend” actions were used. Relative to the default Pre-Loaded EFP plans, which contained flap actions for every flap interval on the placard, fewer flap and gear actions were in pilot-created EFP plans. Throttle and vertical speed actions were also lacking from the user created plans. While these results may indicate pilot-preferred actions, the ability to infer the necessity of the other actions is confounded by both the training provided to the pilots and the EFP interface.

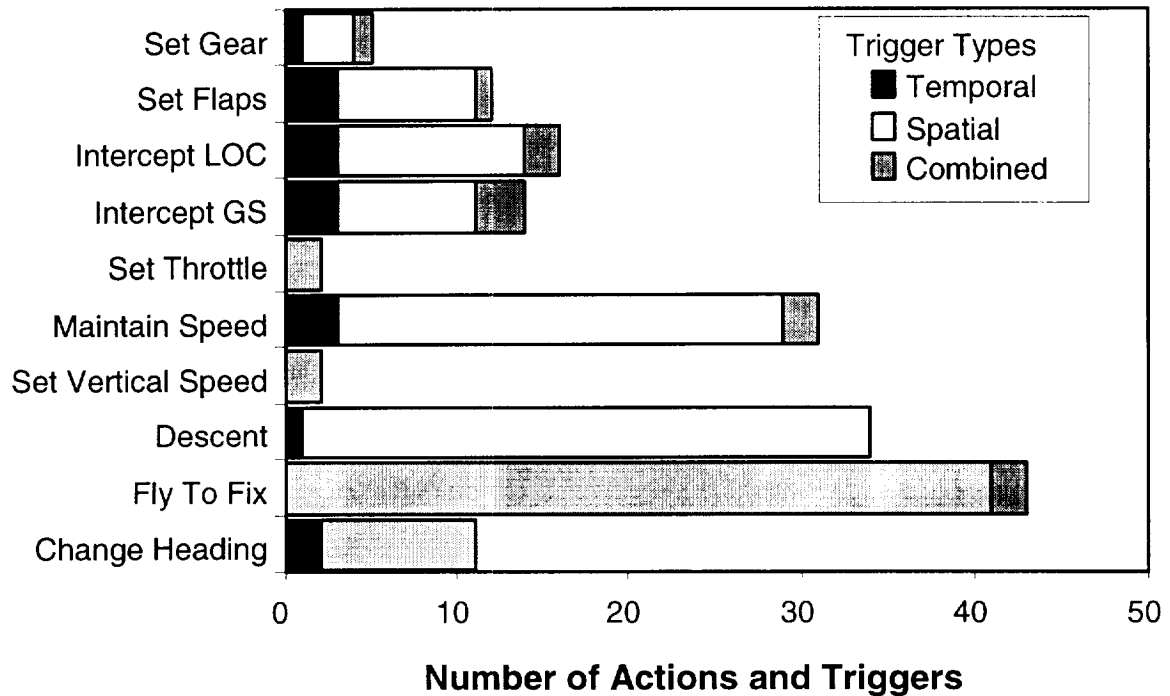


Figure 11 -Actions Used in Pilot-Created Plans, Subdivided by Trigger Type

The actions are subdivided by their associated triggering criteria. Most of the actions used a spatial trigger, such as when the aircraft passes over a certain location; pilots often created their own fixes to serve as triggers, rather than querying the tool to identify the corresponding time. The lack of use of the temporal triggers suggests that pilots may prefer spatial representations in conceiving and visualization plans. However, the spatial display of the trajectory itself may have encouraged the use of spatial triggers, as the only explicit portrayals of the time of any point in the trajectory were in the query view and in one mode of the vertical profile display.

Pilot Workload

In safety-critical tasks, performance measures are much more compelling than measures of pilot workload. However, workload can be taken as a measure of assistance that the cockpit aid provides to the pilot and as a contribution or detriment to pilot performance. Therefore, at the conclusion of each scenario, the pilots were asked to complete a NASA TLX evaluation of workload experienced in both the planning and flying tasks.²⁷ As indicated by the average ratings shown in Figure 12, the Basic EFP had higher workload ratings in each of the workload categories than either of the other planning tool types during the planning task; this result was found to be statistically significant by an ANOVA, Tukey test, and Kruskal-Wallis test to at least the 95% confidence level. A

similar analysis was performed on the data from the flying stage. However, no differences due to the tool provided were found. The temporal workload measure did have significantly higher values ($H=4.54$, $p=0.033$) in the NPA scenarios as opposed to the PA scenarios, as expected.

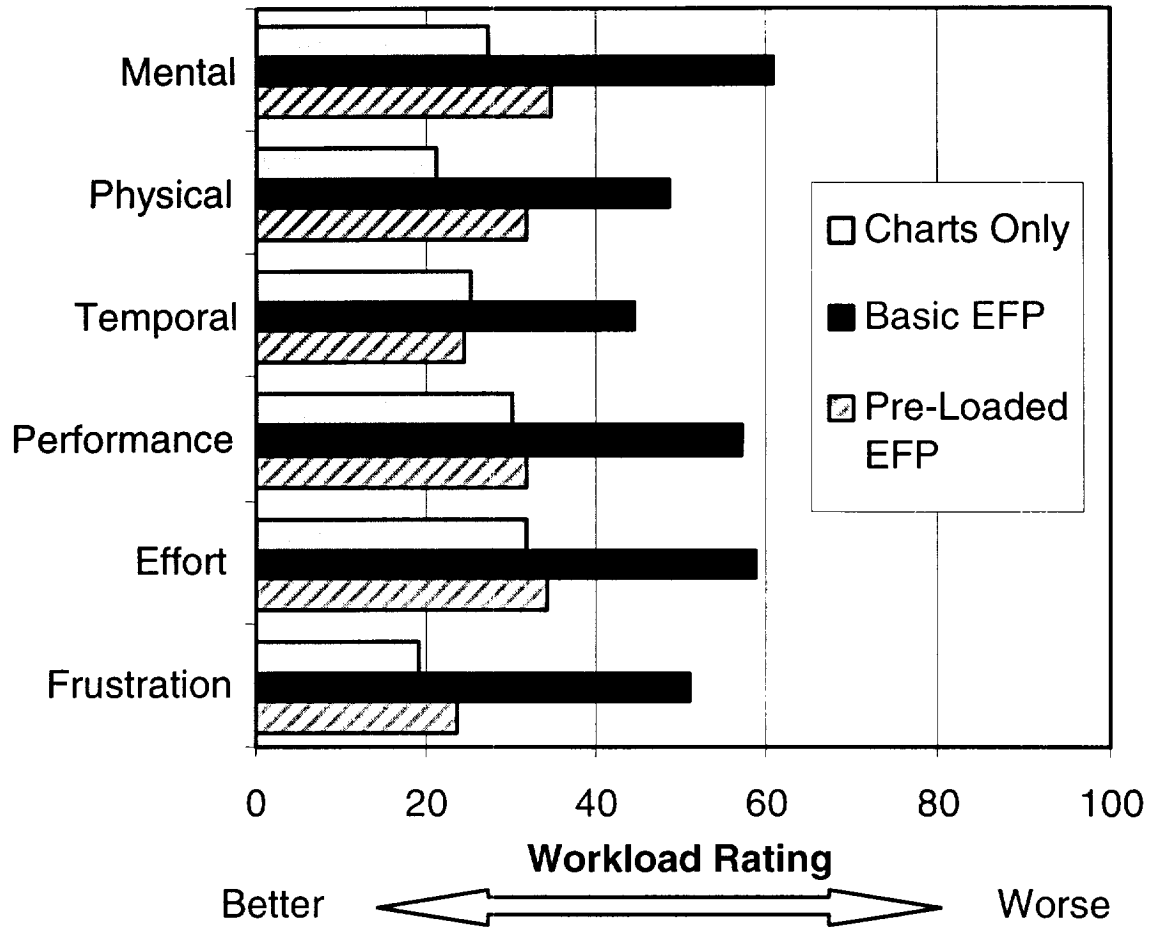


Figure 12 – Average TLX Workload Ratings for Planning Task

Secondary Experiment Results

A total of 12 runs were performed in the secondary experiment (one per pilot); each of the three planning tools therefore was provided to four pilots for one run. Due to the small sample size, statistical analysis was not appropriate. However, qualitative analysis of the aircraft track data noted interesting trends when comparing EFP usage (which would cause an infeasible trajectory) against EFP non-usage. Results were grouped by whether the pilots had an EFP variant available and followed its trajectory. Four pilots appeared to follow the EFP's plan; of these, three pilots initially overshoot the localizer similar to the sample track shown in Figure 13. Conversely, only 2 of the 8 pilots not using the EFP overshoot the localizer. Overshoots of the localizer often lead to additional maneuvering and unnecessary time and distance, with a corresponding frequent need for missed-approaches. In the cases where the EFP was used, the corresponding localizer overshoot added an average 178 seconds to the flight time and an average of 12.2 nautical miles to the track distance when compared with situations where the EFP was not used.

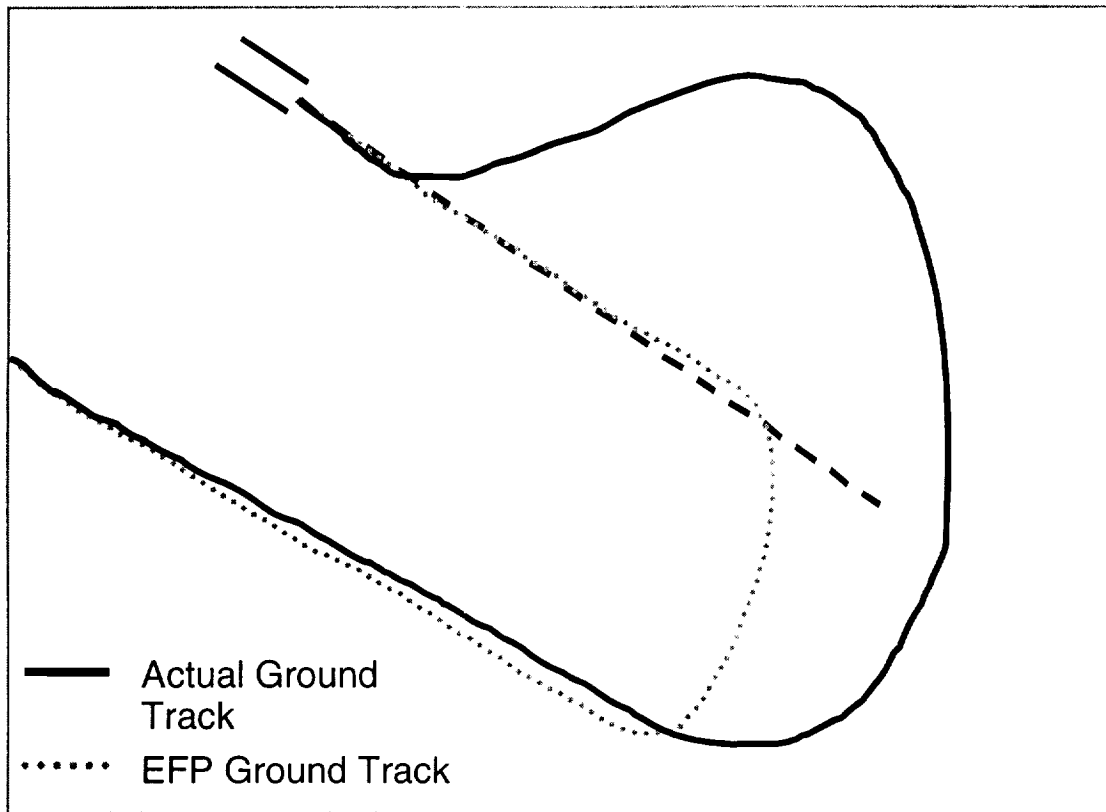


Figure 13 - Representative Deviant Scenario Track

Pilot Ratings of the EFP

At the conclusion of the two experiments, the pilots were asked to provide pairwise comparisons between the three different planning tools. The overall pilot preference shown in Figure 14 was determined through the Analytic Hierarchy Process (AHP).²⁸ The relative preference of any two tools can be obtained by taking the ratio of their respective areas. The Pre-Loaded EFP has a weak-preference over the Charts-Only condition (68% to 21%), and a strong preference over the Basic EFP (68% to 11%).

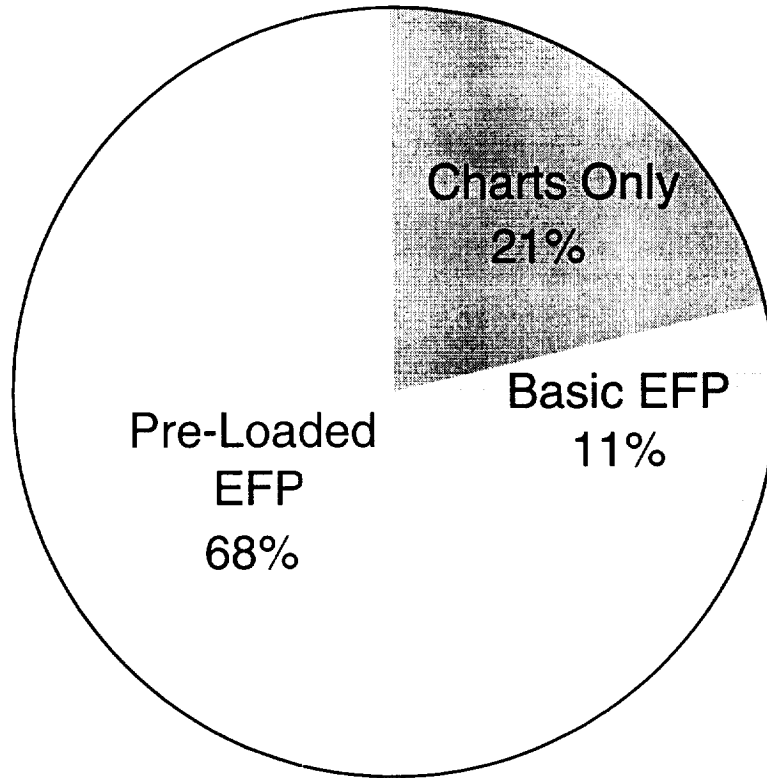


Figure 14 – Pairwise Comparisons of Tool Types, Analyzed with the Analytic Hierarchy Process (AHP)

CONCLUSIONS

In summary, this research has investigated the tasks of generating and then following a detailed trajectory to the runway threshold in emergencies. Little data currently exists into how air transport pilots perform these tasks, the difficulties they face, and the desired features of a decision aid. This study provided a preliminary investigation of these questions by using a prototype decision aid to examine tool design considerations directly, to gather quantitative evidence about the utility of a prototype aid, and gather data about pilots' planning activities and needs in an intelligent cockpit system for this task.

The results suggest that pilots face problems in creating and comprehensively evaluating a trajectory. In 6 of 72 runs pilots were unable to establish an approach course. Four of these occurred in aggressive rapid-descent maneuvers without guidance from the EFP. It is reasonable to hypothesize that, had the pilots been able to fully evaluate the adverse consequences of their current actions on their future trajectory, they would have decided to intercept further away from the airport with a slower descent rate. In addition, the fact that only one of the six incidents occurred when the pilot was using the EFP provides very preliminary evidence that such a tool may be useful in reducing such errors.

While such tools may be beneficial to pilots, problems found in the proof-of-concept prototypes tested in this study warrant further research and consideration during design. The first of these problems is related to the EFP's pilot interface, which primarily used a keyboard entry mechanism (through a CDU) that pilots described as being cumbersome and occasionally confusing. This suggests that merely attempting to leverage the existing cockpit systems such as the FMS by the addition of predictive routines for emergencies is not enough. A more streamlined interface is required that minimizes the amount of pilot workload required for this concept to be acceptable in an emergency environment.

The second problem associated with the prototype highlights potential issues with the functions the aid needs to perform. Significantly higher times to land were found in cases where the pilot was given the Basic EFP. Therefore, simply providing a planning tool that evaluates a pilot created plan may not be sufficient to guarantee

generation of the safest trajectory, although this issue may have been compounded by problems with the interface in this study. The Pre-Loaded EFP variant simulated a planner capable of suggesting plans to pilots. While its plans were not demonstrated to be optimal, it was found that the Pre-Loaded EFP still outperformed the Basic EFP by every measure, including performance, workload and pilot ratings.

Giving a cockpit system the ability to automatically generate and suggest plans to pilots raises several interesting research questions. In the deviant scenario, where the EFP provided the pilot with erroneous information, over-reliance on the displayed trajectory was common. Conversely, the fact that not all pilots followed the Pre-Loaded EFP's plans suggests that the potential also exists for under-reliance. Commensurate with studies of other automated systems, pilots in this study reported not relying upon plans suggested by the aid due to concern about their validity and the mechanism by which they were created. This suggests that not only does the suggested plan have to be in a clearly understandable form, but its underlying structure and objectives must also match those of pilots if over- and under-reliance are to be avoided.

Therefore, the underlying goals and criteria used in automatic trajectory generation must conform to those used by the pilots. However, this study found that these factors change with the context of the emergency. For example, in NPA scenarios the pilots tended to violate overspeed limits in an effort to minimize flight time; in PA scenarios, on the other hand, pilots were generally not as willing to overspeed or overstress the aircraft. Capturing these context sensitivities faces several challenges: accurately eliciting these criteria from pilots; capturing them into a machine-readable representation; giving the system an awareness of the current context; and establishing mechanisms for pilots and the cockpit system to communicate about their criteria and perceived context.

Likewise, methods of representing and displaying the plan need to be examined further. In this study, plans were represented as procedures listing a series of trajectory-defining actions. Pilot comments appear to support this representation; for example, pilot-suggested changes to the display included building in cues to the pilot of newly triggered actions while flying the trajectory. However, in using this representation, many unanswered questions remain: What 'actions' should be used to define the trajectory? What 'triggers' should initiate them? This study considered only a small list of actions and triggers – some of which pilots used heavily, and others which were used infrequently. Many other actions and triggers are possible, but to prevent overwhelming pilots with too many options, it will be important to identify those most relevant to the task at hand.

Other research questions address difficulties in automatically generating a plan. Common methods of optimizing trajectories typically require a clearly established objective function from which an absolute 'best' trajectory can be identified. However, in emergency flight planning, a clearly specified objective function may not always be obtainable – instead, the plan best meeting each several independent objectives and constraints must be found. Likewise, the objective function for these plans may include probabilistic concerns, such as finding a plan that is the most likely to meet all hard constraints in the face of future eventualities. Finally, the representation of a trajectory as being governed by discrete actions requires methods for rapidly optimizing complex hybrid systems.

A final research question examines this study's separation of the overall task into separate planning and flying stages. This delineation may be necessary for a pilot who is creating and flying a trajectory without automatic assistance. However, with the availability of intelligent aids, this distinction may no longer be necessary, as the system may be capable of continuously improving the trajectory. In implementing such a system, not only would the appropriate generation routines need to be determined and incorporated, but also its impact on the pilot would need to be studied for the possibility of decreased situation awareness (if the plan is constantly changing without their awareness) and of increased cognitive load (if the pilot is frequently asked to consider new potential plans, diverting attention away from other tasks).

REFERENCES

- ¹ Fan, T., Hyams, D. S., and Kuchar, J.K. "Study of In-Flight Replanning Decision Aids," *Proceedings of the Guidance, Navigation and Control Conference*, AIAA, Washington, DC, Aug. 1998, pp. 980-988.
- ² Baker, B. "Rotorcraft Pilot's Associate (RPA) Attack Mission Planner," *American Helicopter Society Forum*, American Helicopter Society, Alexandria, VA, May 1999, pp. 844-853.
- ³ Ockerman, J.J. and Pritchett, A.R. "A Review and Reappraisal of Task Guidance: Aiding Workers in Procedure Following," *International Journal of Cognitive Ergonomics*, Vol. 4, No. 3, 2000, pp. 191-212.
- ⁴ Degani, A., and Wiener, E. L. "Procedures In Complex Systems: The Airline Cockpit," *IEEE Transactions on Systems, Man, and Cybernetics--Part A: Systems and Humans*, Vol. 27, No. 3, 1997, pp. 302-312.

- ⁵ Patrick, N.J.M., "Decision-Aiding and Optimization for Vertical Navigation of Long-Haul Aircraft," Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge MA, Aug. 1996.
- ⁶ Dershowitz, A. "The Effect of Options on Pilot Decision Making in the Presence of Risk," Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, Oct. 1997.
- ⁷ Layton, C., Smith, P.J., McCoy, C.E. "Design of a Cooperative Problem-Solving System for En-Route Flight Planning: An Empirical Evaluation," *Human Factors*, Vol. 36, No. 1, 1994, pp. 94-119.
- ⁸ Johannsen, G. and Rouse, W. "Mathematical Concepts For Modeling Human Behavior In Complex Man-Machine Systems," *Human Factors*, Vol. 21, No. 6, 1979, pp. 733-747.
- ⁹ Klein, G.A. "A Recognition-Primed Decision (RPD) Model of Rapid Decision Making," *Decision Making In Action: Models And Methods*, Ed. Klein G.A. and Orasanu J., Ablex Publishing, Stamford, CT, 1993, pp. 138-147.
- ¹⁰ Bainbridge, L. "The Change in Concepts Needed to Account for Human Behavior in Complex Dynamic Tasks," *IEEE Transactions on Systems, Man and Cybernetics – Part A: Systems and Humans*, Vol. 27, No. 3, 1997, pp. 351-359.
- ¹¹ Guerlain, S.A. and Smith, P.J. "Interactive Critiquing as a Form of Decision Support: An Empirical Evaluation," *Human Factors*, Vol. 41, No. 1, 1999, pp. 72-89.
- ¹² Miller, C.A., Guerlain, S., and Hannen, M.D. "The Rotorcraft Pilots Associate Cockpit Information Manager: Acceptable Behavior from a New Crew Member," *American Helicopter Society Annual Forum*, American Helicopter Society, Alexandria, VA, May 1999, pp. 1321-1332.
- ¹³ Wiener, E.L. and Curry, R.E. "Flight-Deck Automation: Promises And Problems," *Ergonomics*, Vol. 23, No. 10, 1980, pp. 995-1011.
- ¹⁴ Parasurman, R. and Riley, V. "Humans and Automation: Use, Misuse, Disuse, Abuse," *Human Factors*, Vol. 39, No. 2, 1997, pp. 230-253.
- ¹⁵ Pritchett, A.R. "Reviewing the Role of Cockpit Alerting Systems," *Human Factors and Aerospace Safety*, Vol. 1, No. 1, 2001, pp. 5-38.
- ¹⁶ Schroeder, B. and Sarter, N. "Supporting Decision-Making and Action Selection Under Time Pressure and Uncertainty: The Case of Inflight Icing," *Proceedings of the International Symposium on Aviation Psychology*, Ohio State University, Columbus, OH, 2001.
- ¹⁷ Xiao, Y, Milgram, P. and Doyle, D.J. "Planning Behavior and its Functional Role in Interactions With Complex Systems," *IEEE Transactions on Systems, Man and Cybernetics – Part A: Systems and Humans*, Vol. 27, No. 3, 1997, pp. 313-323.
- ¹⁸ Menke, T.E. and Maybeck, P.S. "Multiple Model Adaptive Estimation Applied to the VISTA F-16 with Actuator and Sensor Failures," *NAECON '92: Proceedings of the IEEE 1992 National Aerospace and Electronics Conference*, Vol. 2, May 1992, pp. 441-448.
- ¹⁹ Rogers, W.H., Schutte, P.C. and Latorella, K.A. "Fault Management in Aviation Systems," in *Automation and Human Performance: Theory and Applications*, Ed. Parasuraman, R. and Mouloua, M., Lawrence Erlbaum, Mahwah, NJ, 1996.
- ²⁰ Deaton, J.E. and Glenn, F.A. "The Development of Specifications for an Automated Monitoring System Interface Associated With Aircraft Condition," *International Journal of Aviation Psychology*, Vol. 9, No. 2, 1999, pp. 175-187.
- ²¹ Rysdyk, R. T., Leonhardt, B. K., et al. "Development of an Intelligent Flight Propulsion & Control System: Nonlinear Adaptive Control," *AIAA Guidance, Navigation, and Control Conference*, Aug. 2000.
- ²² Bull, J., et al. "Piloted Simulation Tests of Propulsion Control as Backup to Loss of Primary Flight Controls for a B747-400 Jet Transport," NASA TM-112191, Apr. 1997.
- ²³ De Schutter, B. and De Moor, B. "Generalized Linear Complementarity Problems and the Analysis of Continuously Variable Systems and Discrete Event Systems," *Hybrid and Real-Time Systems*, Ed. Maler, O., Springer-Verlag, Berlin, 1997, pp. 409-414.
- ²⁴ Ippolito, C.A., and Pritchett, A. R., "Software Architecture for a Reconfigurable Flight Simulator," *AIAA Modeling and Simulation Technologies Conference*, Aug. 2000.
- ²⁵ Hogg, R.V. and Ledolter, J. *Applied Statistics for Engineers and Physical Scientists*, 2nd Ed., Macmillan, New York, NY, 1992.
- ²⁶ Siegel, S. and Castellan, N.J. *Nonparametric Statistics for the Behavioral Sciences*, 2nd Ed., McGraw Hill, Boston, MA, 1988.