

Development and Evaluation of an Automated Path Planning Aid

Robert Watts,¹ H. Claus Christmann,² Eric N. Johnson,³ Karen Feigh,⁴ and Panagiotis Tsiotras⁵
Georgia Institute of Technology, Atlanta, GA, 30332-0150, USA

Handling *en route* emergencies in modern transport aircraft through adequate teamwork between the pilot, the crew and the aircraft's automation systems is an ongoing and active field of research. An automated path planning aid tool can assist pilots with the tasks of selecting a convenient landing site and developing a safe path to land at this site in the event of an onboard emergency. This paper highlights the pilot evaluation results of a human factors study as part of such a proposed automated planning aid. Focusing on the interactions between the pilot and the automated planning aid, the presented results suggest that a particular implementation of the pilot aid interface, which uses a simple dial to sort the most promising landing sites, was effective. This selectable sorting capability, motivated by the anticipated cognitive mode of the pilot crew, improved the quality of the selected site for the majority of the cases tested. Although the presented approach increased the average time required for the selection of an alternate landing site, it decreased the time to complete the task in the case of emergencies unfamiliar to the pilot crew.

¹ Graduate Research Assistant; AIAA Student Member

Daniel Guggenheim School of Aerospace Engineering, 270 Ferst Dr

² Graduate Research Assistant; AIAA Student Member

Daniel Guggenheim School of Aerospace Engineering, 270 Ferst Dr

³ Lockheed Martin Associate Professor of Avionics Integration; AIAA Senior Member

Daniel Guggenheim School of Aerospace Engineering, 270 Ferst Dr

⁴ Assistant Professor; AIAA Member

Daniel Guggenheim School of Aerospace Engineering, 270 Ferst Dr

⁵ Professor; AIAA Fellow

Daniel Guggenheim School of Aerospace Engineering, 270 Ferst Dr

Nomenclature

α	= Significance level
p	= Probability of a mistaken rejection of the null hypothesis
β	= Probability of a mistaken acceptance of the null hypothesis (false negative)
$1 - \beta$	= Power; Probability of a correct rejection of the null hypothesis (true positive)
F	= F -distribution and statistic
Z	= Z -Test
η^2	= Effect Size
r_s	= Spearman's rank correlation coefficient
μ	= Mean value

I. Introduction

Modern air transportation has an excellent flight safety record. When failures do occur in flight, owing to the training and experience of the pilots almost always results in a safe landing. This is evidenced by a rate of only 1.35 accidents per one million hours flown in 2007 by US air carriers [1]. Despite this excellent record, the pilots' responsibility to land safely in case of an emergency can be very demanding. When an emergency situation occurs during a flight, the pilots' workload is very high and a number of tasks demand the pilots' attention. One of the important tasks is the planning and execution of a trajectory resulting in a safe landing. However, this task is complicated by multiple, often conflicting goals, including reducing time to land, staying within the flight envelope limits of the airplane, weather issues, as well as meeting any relevant regulatory requirements. Moreover, all these tasks must be accomplished in a stressful environment, often under severe time pressure [2].

Although fault tolerant adaptive automation is currently being developed, for the foreseeable future of civil transport aviation, pilots will be the ultimate decision makers, especially in cases of emergencies involving any type of aircraft performance degradation or flight envelope reduction. As a result, current research is being directed at pilot aids that aim at enhancing the pilot's Situation

Awareness (SA), as well as at supporting the pilot’s decision making process through the provision of relevant, situation-related information.

The purpose of this paper is to report on a human factors study related to efforts to develop an Automated Planning Aid (APA) (in terms of both an acceptable interface and control algorithms) that could assist pilots in generating a plan to safely land at alternative landing sites. In order to do so, the pilots must first determine the “best” landing site and then formulate an expedient and safe trajectory to the ground. This paper presents the results of an evaluation of an APA interface prototype by means of a human-in-the-loop test with commercial airline pilots, focusing on the selection of alternate landing sites during an emergency. Although the implemented APA in the simulator was also able to compute emergency paths to those sites, a detailed description and discussion of this part of the process is omitted in this paper, as it had no immediate effect on the APA interface evaluation.

The results of the study are evaluated in comparison to the opinion and judgment of a single subject matter expert. This expert had more than 20,000 hours of flight experience in over 20 years of service as a commercial pilot. The authors do acknowledge that this comparison might be improved by incorporating more experts, more test cases and a larger sample. However, given the fact that these scenarios were designed in cooperation with that expert to have an unambiguous “best” solution, it is likely that, given enough time to review each scenario, the vast majority of trained pilots would come to the same conclusion as to which landing site was the best alternative. As such, the authors do believe that the utilized metrics, and the comparison with the experts ranking of the landing sites, is valid for the evaluation of the APA for selecting an alternate landing site under tight time constraints.

II. Background

In an emergency situation, the crew must monitor the aircraft systems, detect and resolve any failures, control an aircraft with possibly degraded performance, and coordinate with the cabin crew, airline dispatchers, and air traffic control. In addition to these tasks, the pilots must also plan and execute a trajectory that will result in the safest landing possible. These tasks are made

even more difficult by the circumstances during an emergency. For example, the pilots may feel a sense of physical danger, or the cabin environment may be a distraction due to smoke, heat, or noise. Additionally, aircraft performance may be affected, resulting in degraded handling qualities. In order to understand some of the difficulties these circumstances present, a number of cognitive engineering models have been developed in the literature and are reviewed here.

A. Cognitive Considerations

During an emergency situation a number of contextual features change and alter the pilot's cognitive state. Cognitive Control Theory describes how the context of a situation influences cognition and behavior which change depending on the amount of control the person has [3]. The degree of control a person has is determined, in large part, by the amount of subjectively available time and the familiarity of the situation [4]. Subjectively available time refers to the amount of time that a person perceives that he or she has available to take action. The amount of time perceived may depend on the objective amount of available time, the predicted changes in the system, the person's level of arousal, as well as other factors. In Cognitive Control Theory the degree of control is discretized into four control modes: scrambled, opportunistic, tactical, and strategic. The relationship between the amount of subjectively available time, familiarity of the situation and the control modes is described in [4]. When both familiarity and available time are low, individuals are likely to exhibit behavior associated with a scrambled mode. With more time, but still without familiarity, individuals transition into opportunistic and then into strategic modes; a tactical mode is only expected at moderate to high levels of familiarity. As familiarity becomes greater and time remains low, individuals will transition into opportunistic mode and then to tactical mode. Individuals only transition into the strategic mode when familiarity is low to moderate; otherwise, they remain in the tactical mode regardless of the time available.

The most dangerous mode for a pilot to exhibit is the scrambled mode, which generally corresponds to a person in a state of panic. When a pilot is in this mode, he or she is not able to focus even on a single goal, namely, flying the aircraft. When a pilot has adequate subjectively available time in an emergency, the cognitive state may be better described by the opportunistic mode. In this

mode the pilot has a greater sense of control. The pilot is more likely to develop a plan or modify an existing plan in order to fit the current situation. The resulting plan may take into account the potential effects of candidate actions. This mode corresponds to “normal” performance. During an emergency situation, a pilot’s cognitive state will likely be somewhere between the scrambled and tactical modes, described by the opportunistic mode. In this mode, pilots are likely to use any plans and procedures available that are deemed to be sufficient; however, these may not be used correctly or most effectively.

The amount of subjectively available time perceived by a pilot may be influenced by a number of factors. The phase of flight during which an emergency occurs, the state and configuration of the aircraft, the type of emergency, the number of actions the pilot is required to complete, the availability of resources, as well as the initial stress level, all contribute to the subjectively available time. Additional stress may be caused by physical factors, such as smoke in the cabin or loud noises, or it may be purely psychological, such as the fear of impending danger. These stress factors affect the manner by which the pilot makes decisions. Although the pilot may be able to quickly develop a plan of action based on experience and intuition, stress can lead him or her to fixate on a single solution, and fail to compare alternatives [5, 6]. Additionally, the pilot may simply increase the speed with which he or she processes information, potentially leading to errors. The pilot may also reduce the amount of information that is sought and processed, known as filtration [7, 8]. These stress-related factors can cause pilots to make poor decisions, despite the fact that they would be able to make acceptable decisions under normal circumstances. These inferior decisions may cause incidents to become accidents.

In an emergency situation it is tempting to think that an automated system should be included. However, stress may also lead the pilot to either ignore or rely too heavily on an automated tool. He or she may assume that the plan generated by automation is best, without verifying its feasibility or exploring other options [9]. In addition, the pilot may seek only information which confirms the automation-generated solution as the best, while discounting other information (confirmation bias). Alternatively, rather than simply discount conflicting information, the pilot may attempt to rationalize and mentally force all available information to fit the automation-generated solution

(assimilation bias) [3]. Therefore, care must be taken when devising support systems intended for use in stressful situations.

In addition to the stress, the complex nature of the decision making task is also important. The design presented here is based on a hybrid of the Naturalistic Decision Making (NDM) and Rational, Analytic (RA) decision models. Taking the best of both worlds may allow the decision maker to reach the best result. The NDM framework is often used to describe how experts make complex decisions. Zsombok [10] describes NDM as,

“the way people use their experience to make decisions in field settings.”

Although experts are often able to make excellent decisions based on experience and intuition, many of the aforementioned effects of stress can negatively impact the quality of the decision. The RA model of decision making describes how a decision maker proceeds through a set of steps (generating alternatives, envisioning the consequences, evaluating the alternatives against a set of criteria and choosing the best plan) to reach a decision [11]. Although under nominal circumstances, a rational decision process may be helpful in determining the safest path to land, it may not be the most appropriate model of decision making during an in-flight emergency.

In the NDM/RA hybrid model chosen for the APA, the rational decision process can compensate for some of the weaknesses of NDM. For instance, the rational decision process generates a number of alternatives, which alleviates the tendency to fixate on a single solution. By automating the generation and the evaluation of alternatives, the process can be streamlined. It should be noted that, as Peter Simpson [12, p. 18] warns,

“a decision aiding system should not become a decision making system, and it should never simply dictate decision courses to the operator.”

However, by capitalizing on the automation’s fast lookup and simulation abilities, and human pattern recognition and intuition capabilities, the two decision models may be combined to make sound decisions more reliably.

B. Related Work

The APA must do two things: first, it must be able to accurately predict the most appropriate alternative landing sites, as well as the most suitable trajectory to land at these sites; second, it must provide an intuitive interface for the pilot crew that is appropriate to the task and anticipated context as well as the operator’s cognitive state. The completion of the first task requires that the aid determines the overall feasibility of a trajectory. A feasible trajectory must avoid obstacles, which may be static, such as a mountain, or dynamic, such as a severe weather system. The determination of such a trajectory is ideally made by taking into account the aircraft’s possibly abnormal characteristics, due to the emergency. Also, the trajectory usually must minimize time to land, which is important in many emergencies. For an appropriate solution to the second task, the aid must also provide an interface with the pilot through which information is effectively communicated in both directions. Most research to date has primarily focused on one or the other of these two tasks.

The landing site selection task has been suggested as a candidate for automation. Atkins, Portillo and Strube [13] have developed a method to complete this task. First, the footprint containing all feasible landing sites is calculated. Then the landing site list within this footprint is prioritized according to a number of weighted criteria, such as runway length, airport facilities available, etc. In their research, Atkins et al. chose example values for the criteria weights, but acknowledged that the criteria weights would ultimately be based on expert knowledge and would vary by emergency type.

The need for the pilot and the automated planning aid to interact with each other has also been investigated. The Emergency Flight Planner (EFP) by Chen and Pritchett [2] has been proposed as a prototype interface between the pilot and the pilot aid. The EFP allows the pilot to enter a (flight) plan, the ensuing trajectory is then predicted and evaluated. The EFP also provides an additional mode in which the pilot is presented with a preloaded trajectory, which can then be accepted, modified, or deleted. The results of testing with the EFP emphasized that generated plans must incorporate the structure and objectives used by pilots in order to be effective.

Layton, Smith and McCoy studied human-automation cooperative problem-solving for *en route*

flight planning in [14]. In that study, pilots and air traffic controllers were both used as subjects in the evaluation of three possible modes. The first mode was a sketching-only system, in which a plan devised by the subject was evaluated by the system and feedback was provided. The second was a sketching system with the additional capability for the user to specify constraints on the plan and allow the system to propose a solution, which matched those constraints. In the third mode, the system proposed a plan based on system-specified constraints. The results showed that in the second and third modes, users explored more possible options; however, they were also biased toward the system-generated alternative. The same study also highlighted the fact that the use of a fully automated aid could be detrimental if it performs suboptimally.

The previous results show that in order to increase the usefulness of an APA, the process by which pilots select an alternative landing location and plan a path to, it needs to be better understood. In addition, it needs to be better understood how the pilots' decision making processes can best be assisted by such a tool. It is expected that an aid that accepts and provides information in a manner that is most consonant with the pilot's mental process will be most effective.

III. Pilot Aid Tool Design

It should first be noted that pilots are currently not without some form of automated path planning assistance. The modern Flight Management Systems (FMS), which are used on board most major transport aircraft, include pages in the Control Display Unit (CDU) to help the pilot with the task of deciding on a divert landing site. For example, the alternates page **ALTN** of the Boeing 777 FMS displays four possible alternates at a time [15]. These may be input from a list that the pilot creates before the flight, from a database, or can be entered manually. The estimated time of arrival (ETA) and the predicted amount of remaining fuel are displayed. These four alternates are ordered by the ETA.

Although these pages are helpful, it is entirely possible the best choice will not be on the list. For instance, the nearest airport list only provides the landing sites at airports at which the aircraft is able to land normally, without taking into account the severity of the emergency. In the case of a severe emergency, the pilot may be willing to land at a runway that is not sufficiently long for

a normal landing with normal safety factors. Additionally, in the event of a performance altering emergency, such as a stuck elevator, the FMS cannot presently account for the post-failure flight dynamics of the aircraft. Thus, the plan generated may not be feasible, given the aircraft’s degraded performance. The pilot may alter the recommended plan by altering the waypoints used; however, this requires a non-trivial amount of time and work on the part of the pilot. In the case of an in-flight emergency, both time and cognitive resources may be limited due to the number of other tasks the pilots must address, which suggests the current FMS solutions could be improved for highly time-critical emergencies.

In order to address the shortcomings of the present system, a new APA concept has been developed. The proposed APA is linked to the aircraft’s health performance monitoring and alerting system and receives information about failures as they occur in the system. This data is then used to determine the post-failure performance of the aircraft. The updated flight dynamic characteristics of the airplane, combined with terrain and weather information, are then used to compute suited (e.g., time- or fuel-optimal) plans to reach a number of potential divert locations.

A. APA Design

Each of the paths to the alternate landing sites is permanently displayed graphically on the ND as well as textually on the CDU. Information about each landing site is collected from precompiled database information, such as data about airports and terrain, as well as live weather information. This additional information is made available to the pilot through the CDU. Based on the information collected, each site is associated with scores from 0 to 1 for different parameters, with a higher score representing a better fit. From these subscores, a cumulative score is calculated based on the system’s weighting of the different parameters. Although the simulated APA was capable of computing polygonal scores and sub-scores, for the purpose of this study, the utilized score values were hardcoded, bases on a subject matter expert’s option. (See III A.) The alternates are presented in descending order of the cumulative score in the CDU.

The APA design evaluated here was based on the results of a survey in conjunction with implications suggested by the Cognitive Control Theory survey in [16]. This study investigated the

pilots' tasks in the event of an in flight emergency, namely the tasks of choosing a safe landing site, and developing a safe trajectory to reach that site. This survey provides a useful perspective into the methods and priorities pilots use to accomplish these tasks. During an airborne emergency, the need to land quickly is always of high priority. Therefore, the most important factor considered by the pilots when selecting an alternative landing site is proximity in terms of time. Additionally, the weather at the airport, the length of the runway and the distance from the current location are also important criteria. The most important *en route* factors are the avoidance of severe weather and hazardous terrain.

One of the most important aspects to be considered in an emergency situation is the high workload, time-critical, stressful nature of the situation. Accordingly, one significant feature of any proposed aid is that it should reduce workload, rather than increase it. The aid must provide useful information in a coherent manner, without burdening the pilot with requests. Similarly, pilots view the aid only as a tool, not as a directive. Pilots will use an automatically generated plan in conjunction with their own experience and intuition. Ultimately, the pilot has the final decision-making authority.

A successful design must be closely integrated with interfaces that pilots are currently familiar with. The design of a new tool to be used in the cockpit is a very complex task, as the amount of information and controls available in a modern cockpit is quite large. Also, the physical area in which they must be contained is rather limited. All of the systems' displays and controls must be contained in a small and coherent cockpit layout. With this in mind, no single part of the entire APA system should be designed on its own. It does not exist as a standalone entity, but must work cooperatively with existing systems to allow the pilots to complete all of their responsibilities.

For these reasons, it was decided that the APA be integrated into the existing FMS, utilizing a CDU page which is based on the current *ALTN* page. The alternate routes are continuously displayed on the existing Navigation Display (ND) per survey respondents' preferences. Efforts were also made to ensure that the APA would not adversely contribute to the pilots' workload, but rather present relevant information in a coherent manner. With this in mind, it was determined that the most important function was to help pilots filter information. This was accomplished via the use of a

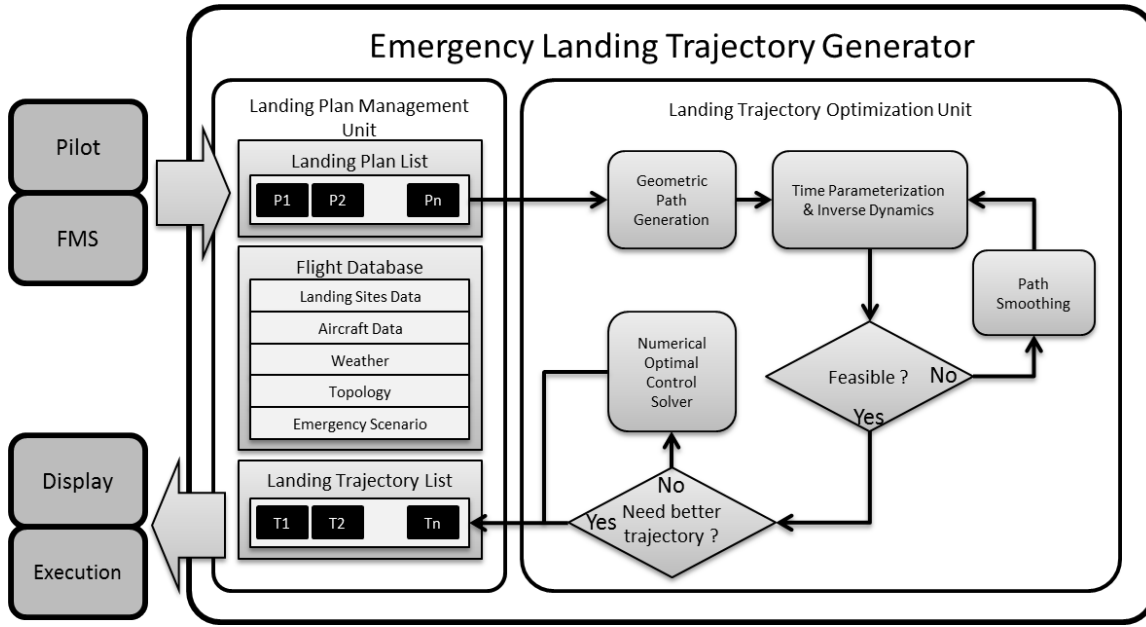


Fig. 1 Schematic of the APA's overall landing site selection and trajectory optimization process.

single dial in a slightly modified Navigation Display Control Panel (NDCP), by allowing the pilot to quickly indicate the severity of the emergency by limiting the types of alternates to be considered. Utilizing landing site types as differentiators for the severity of the damage allowed for an intuitive mapping from the pilots' situational analysis to a measure of urgency of the emergency.

In order for the ND to display the routes to the alternate destinations, it must have some method for determining these routes. There are various possible approaches to this research question, among them, for instance, [17, 18]. In the currently implemented design, the approach taken was to calculate these alternate routes in real time, starting from a Dubins path [19]. These Dubins paths serve as initial guesses for a high-fidelity trajectory optimization module whose output can be used by the pilot to get further information about the selected path, along with the corresponding control actions, and can be used to drive an autopilot and/or flight director [20]. The overall trajectory generation step of the APA is shown in Fig. 1. The research presented in this paper primarily focuses on the human-machine-interface between the APA and the pilot-not-flying, pictorially represented by the large arrows in the left part of the graphic.

An algorithm for determining appropriate criteria weights based on the type of encountered

emergency warrants a study of its own, and is not the focus of the current work. In order to avoid testing the specific criteria weight design, such as those derived from the prior survey results in [16], scores were hard-coded for every site in each scenario, following the advice from a single subject matter expert. This expert had more than 20,000 hours of flight experience in over 20 years of service as a commercial pilot. The expert was provided all information available about each landing site and, unlike the experiment subjects discussed below, was given an unlimited amount of time to consider each scenario thoroughly. This expert determined scores served to rank the landing sites; these were also the cumulative scores presented to pilots as a weighted combination of the criteria scores and as such presented the expert-determined ranking of the alternate landing sites.

Information from the APA is displayed to the crew through modifications of four displays: Navigation Display, Primary Flight Display (PFD), Navigation Display Control Panel (NDCP), and Control Display Unit. The APA prototype used during the evaluation was built using the Reconfigurable Flight Simulator [21]. Each of the display modules used in this simulation is roughly based on the Boeing 777 type displays.

B. Interface and Setup Description

The ND is on the right of the screen in front of the participant, replicating the setup familiar to pilots. (See IV C for an overall description of the apparatus.) The display is track up, i.e. the pilot’s own aircraft is centered at the bottom of the display and the immediate trajectory is displayed vertically extending from the pictorial representation of the pilot’s own aircraft (the triangular icon), with the current plan shown as a thin solid line (see ND detail in Fig. 2[33]). The graphical display of the routes to alternate destinations (the dashed lines in Fig. 2) allows the pilot to quickly assess the spatial arrangement of the available alternative landing sites.

The PFD was located to the left of the ND on the screen in front of the participant. (See IV C for an overall description of the apparatus.) The PFD provides information about the current state of the aircraft such as heading, flight speed, altitude, climb/descent rate, and pitch/roll attitudes. Because the participants were put in the role of a First Officer as the pilot-not-flying, the PFD was provided only as a reference to allow each participant to be aware of the corresponding aspects of the

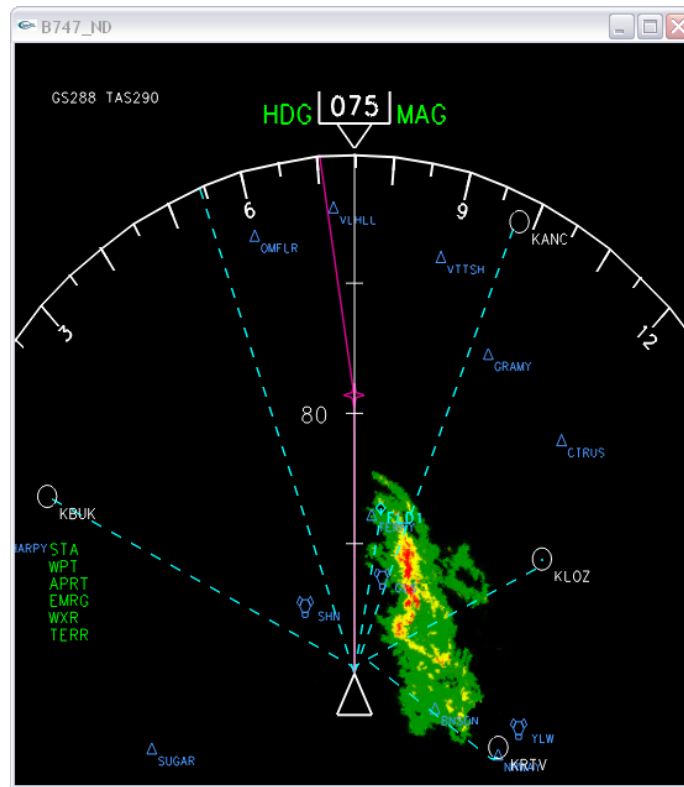


Fig. 2 Overlay of the proposed alternate landing sites on the Navigation Display (ND)

situation. This display provided information which corroborated the emergency that was described, such as loss of altitude as a result of engine failure.

The modified NDCP is the APA's primary input interface and includes seven buttons and two dials, as shown in Fig. 3. The buttons toggle data overlay options on the ND: WXR (shows weather systems in the area), STA (shows navigation stations), WPT (shows all waypoints in the area), ARPT (shows airports), EMRG (shows the candidate routes to alternative destinations), TFC (shows traffic), and TERR (shows the terrain). The dial to the right allows the user to set the range (in nautical miles) displayed on the ND (Fig. 3 and Fig. 2 both reflect a setting of 160 NM), the dial on the left was only present in one of the studied NDCP versions and allowed the pilots to filter possible landing locations. The dial allowed the pilot to quickly indicate the requirements of the landing site, as noted above. A focus of the study was the effect of the presence of this dial.

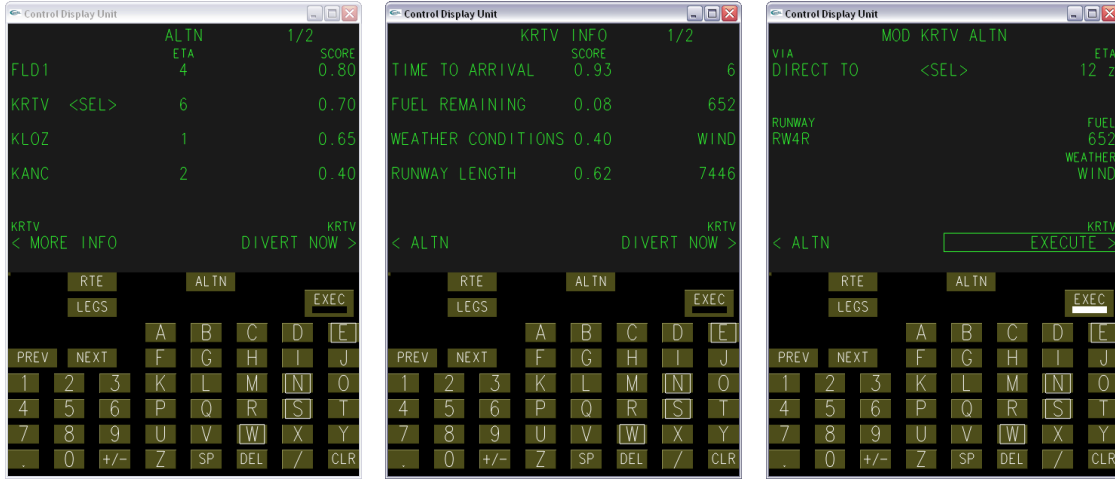
The Control Display Unit (CDU), see Fig. 4, provides a limited subset of the normal CDU functionality required for this evaluation, displaying landing site identification, estimated time of arrival, and the overall site assigned score. Pilots can get an overview via a (ranked) list of potential



Fig. 3 The modified Navigation Display Control Panel (NDCP)

alternate landing sites, access details for each of the trajectories, and command the execution of any of the proposed plans. The route (RTE) and legs (LEGS) pages provide information about the currently planned FMS route. The alternates (ALTN) page was redesigned to provide more information and support more effective use. The destination options, after being filtered by the left dial of the NDCP, were ranked according to the overall scores for each potential landing site. These are the same destination options which are displayed graphically on the ND and may include more than four destinations, in which case the NEXT button is used to move further down the list. The ALTN page allows the pilot to see additional information about each of the options, including time to land, distance, fuel remaining upon arrival, runway length, weather at site, medical services available, and maintenance services available.

Integrating these landing sites into the existing ALTN page allows the pilot to select among options in the same manner that is currently available on board modern commercial airliners. After selecting one of the destinations on the list (Fig. 4(a) shows the fictional landing site KRTV on the CDU's ALTN page as selected), the pilot was able to view more information about it by pressing MORE INFO , which brings up the corresponding MORE INFO page on the CDU (Fig. 4(b) shows the KRTV INFO page for a particular site, displaying the internal sub-score as well as the underlying data



(a) The modified ALTN page.

(b) The MORE INFO page.

(c) The DIVERT NOW page.

Fig. 4 Modified CDU pages provide an additional interface to the APA.

for relevant parameters). This page provides information about the landing site and the scores that are used by the ranking system for each of the criteria. After the execution routine has been armed, pressing the lit EXEC button (Fig. 4(c)) transfers the computed plan to the FMS, i.e. the autopilot and/or flight director in the PFD. Due to the implementation on a touchscreen, the subjects in this study had to click the actual field on the display, whereas a real CDU would provide buttons to the side of the text fields.

IV. Experiment Description

The experiment tested for differences in performance for pilots using two variations of the APA, focusing on the actual human-machine interaction, which is represented by the large arrows in the left of the schematic of the overall process, Fig. 1. One variation of the APA included the left dial shown in Fig. 3, which facilitates the filtering of landing sites; the other APA version did not include such a dial. The two variations were otherwise identical. In each run, the participant was presented with a scenario in which an emergency occurred. Emergencies that the pilot was expected to have been trained to handle, as well as unfamiliar emergencies, were presented. The pilot had the opportunity to use the aid, either with or without the dial, to consider the possible alternate landing sites, and finally select a plan to land. The participating pilots did not actually fly the simulated aircraft, but they participated as a First Officer, that is, the non-flying pilot. The simulation run

ended when the subject had selected a route and executed it by selecting the **EXEC** button, shown in Fig. 4(c).

A. Participants

A total of eight pilots participated in the experiment procedure. These pilots all hold an Airline Transport Pilot certificate and were experienced in a variety of aircraft, primarily Boeing, Airbus and Bombardier. One participant had recently retired, whereas all others currently fly with a commercial airline. The average number of flight experience for the participant pilots was 8194 h.

B. Independent Variables

The experiment included the variation of two independent variables. Each independent variable had two levels, creating four configurations with two replications. These variables were the aid type, and familiarity of the emergency.

The scenarios were run with two variations of the APA. In one mode, the pilot was able to use the previously described dial to filter possible landing sites (Fig. 3, left dial). In the second variant, this dial was not available. In the sequel, this variable is indicated by $()_{\text{dial}}$ and $()_{\text{no dial}}$, respectively.

The evaluation scenarios simulated two general types of emergencies: familiar emergencies and unfamiliar emergencies. In the familiar scenarios, the aircraft's performance was either unaltered or was altered in a manner that pilots had been trained to handle, such as a single engine failure. The second type of scenario was a performance altering emergency in which the failure was one which the pilots had not been specifically trained to manage, such as a stuck elevator. The scenarios were designed such that each scenario was comparable in terms of difficulty and number of options that the participants were expected to consider. The comparison of these two emergency categories is important because pilots may make decisions differently during a familiar scenario than they do during an unfamiliar one, indicated by $()_{\text{familiar}}$ and $()_{\text{unfamiliar}}$, respectively.

C. Dependent Variables

Two primary metrics were considered. As a first metric, the pilot's ability to choose the best landing site was assessed. As described in Section III A, a subject matter expert was consulted in order to provide aggregate scores for each landing site based on all information available, having unlimited time. These scores served to rank each of the landing sites in each scenario. The second metric was the amount of time pilots spent during the selection process. A reduction in time promotes safe flight by allowing the pilots to focus on other important tasks associated with handling the specific emergency, such as crew coordination, or alerting personnel on the ground. The time required for the pilot to select a path was used to determine the efficiency with which the pilot was able to develop a plan. This time was measured from the moment the emergency occurred, to the time that the pilot selected a plan for execution in the CDU.

In support of these measures, other secondary measures were used to assess the APA. The number of candidate landing sites the pilot reviewed and the number of times the pilot turned the filter dial (when available) were also measured. In addition to simply comparing the total time to complete the task, these measures allowed for a more granular analysis of the task. The following is an overview of all the dependent variables measured:

Time Time was measured from the time the emergency occurred to the time the pilot executed a path in the CDU.

Quality Each scenario involved a number of potential landing site options. These options were ranked according to the appropriateness for the given scenario. This ranking was enabled through an *a priori* evaluation through a subject matter expert (see Section III A). The quality aspect of the participant's performance was based on the rank assigned to the participant-selected landing site in the experts *a priori* determined ranking.

Number of alternates viewed The number of alternates viewed was the number of landing sites for which the pilot viewed the MORE INFO page. This was automatically recorded by the system and was the total number of alternates viewed before and after the emergency occurred. The number of alternates viewed was recorded in order to evaluate factors that influenced the

amount of time required to reach a selection.

Situation Awareness The participant’s situation awareness was assessed immediately after the completion of each run following the Situation Awareness Global Assessment Technique (SAGAT) method [22]. The displays were blanked and the participant was asked ten questions about the current scenario. The questions assessed all three levels of SA [23, 24]. Level 1 assessed the pilot’s perception of cues, Level 2 assessed the pilot’s comprehension of the situation, and Level 3 assessed the pilot’s ability to forecast future events. The ten questions contained five Level 1 questions, three Level 2 questions and two Level 3 questions. These were drawn from a pool of twelve Level 1, ten Level 2 and five Level 3 questions.

Workload The participants were asked to evaluate the perceived workload experienced in each scenario, in order to assess the feasibility of its use in a real emergency. The NASA Task Load Index (TLX) [34] was used in this study to assess the workload for six different sources; mental demand, physical demand, temporal demand, performance, effort, and frustration.

APA Assessment Upon completion of the experiment, the participants were asked to complete a questionnaire, which included questions about the pilot’s experience in each simulated flight, as well as an evaluation of both variations of the APA. These included subjective assessments as well as a rating from 1 to 10 based on the modified Cooper-Harper rating for displays [25].

Design of Experiments A 2x2 repeated measure, full-factorial design with one replicate was used to evaluate the effect of APA dial and experiment familiarity on the dependent variables. The design was blocked on the APA type, i.e. the presence or absence of the landing site type filter knob is one of the independent variables in the statistical analysis. As a result, each participant saw four emergencies with one variation, then four with the other variation, denoted by $()_{\text{dial}}$ and $()_{\text{no dial}}$, respectively. An additional “no failure” scenario was included in order to reduce the pilots’ expectancy of an emergency. Accordingly, each participant completed a total of nine runs, i.e. each of the four configurations paired with the eight scenarios plus the additional no-failure run.

Scenarios A total of eight emergency scenarios were created (not counting the no-failure straight-

and-level flight one). Each scenario was characterized by the emergency situation, the phase of flight during which the emergency took place, the alternate landing sites which were available and the time at which the failure occurred. The eight emergencies each occurred in one of three phases of flight: climb, descent or cruise. The emergencies which were repeated occurred in different phases of flight each time. Each scenario had a fixed number of potential landing sites, dependent on the phase of flight in which the emergency occurred. Emergencies occurring during climb had three sites, during cruise had six sites, and during descent had four sites. All identifiers of airports, waypoints, and navigation stations were fictional to prevent any effect due to location familiarity or lack thereof.

There were a total of six types of emergency situations, three familiar situations and three unfamiliar. In each case the Captain, who (as the more experienced aviator) was the flying pilot, described the emergency to the First Officer, who, as the non-flying pilot, was tasked with selecting an alternate landing site. The Captain's description was simulated through an audio playback. The PFD also showed any appropriate changes (such as descent) and the newly introduced alert display (Fig. 5(b)) annunciated the appropriate message. The three familiar emergencies used were engine failure, low fuel, and fire onboard. In order to allow for appropriate descriptions and understanding of the failures, flight control failures (for which pilots had not been trained for) were used as the unfamiliar emergency cases. These were: stuck rudder, stuck aileron, and stuck elevator. The amount of time passed after the simulation began until the emergency occurred was different for each scenario, and it was between 45 seconds and 105 seconds, in an attempt to create an element of surprise and (in conjunction with the no-failure run) minimize pilot readiness. Although real emergencies could happen hours into an otherwise normal flight, the metrics of interest for this study were assumed to not be affected by fatigue resulting from a prolonged period of normal flight time before the occurrence of an emergency.

Procedure Before entering the simulator, each participant had to read a briefing document. This document introduced the pilot to the features of the simulator and the procedures which would be used to conduct the experiment. The introductory material informed (erroneously)

participants that the compensation they would receive depended on how well they would perform the emergency planning task (but the compensation did not depend on correctly answering the SAGAT and TLX questions).

After reading this introductory material, the participant entered the simulator to complete at least two practice scenarios. In the first training scenario, no failure occurred. This run simply allowed the pilot to explore the interface and familiarize himself with the tools available. The second run presented a simple engine failure scenario. This allowed the pilot to gain an expectation of how emergencies would be presented and how the tools would allow diversion planning. After each run the participant was given sample situation awareness (SA) questions to familiarize them with the format and types of questions which would be asked. This was done primarily in order to avoid any differences between the first couple of runs and the later runs in terms of answering SA questions. Each participant was also asked to complete the TLX workload questions. The participant was then given the option to run either of the training scenarios again or to begin with the test runs.

After completing the training runs, each pilot completed nine test runs. Eight of these runs contained one of the aforementioned emergencies and one was completed without an emergency. The no-emergency run was included in order to slightly reduce the pilots' expectancy of an emergency. The pilot joined each flight in progress and was able to use the tools available to gain an understanding of the situation. Between 45 and 105 seconds after the simulator was started, an emergency situation was presented to the pilot through a recorded message. The participant then used the tools to determine the most appropriate landing site for the given situation.

After selecting the route to the alternate landing site for execution, the simulator was closed and the participant was handed a clipboard and asked to complete the ten SA questions. After completing these questions, the participant was asked to complete the TLX questionnaire on the touchscreen.

Upon completion of all nine runs, the pilot was given an additional set of questions pertaining to the experiment as a whole. These included subjective questions about the features and

usage of the APA, along with a Modified Cooper-Harper ranking sheet to rate the usability of the APA.

After the completion of all the runs, and after answering all the questions, the participants were made aware of the initial deception regarding the performance-tied compensation and were informed that all of them would receive the entire, same, amount as a token of appreciation for participating in this study.

Apparatus The overall experimental setup is shown in Fig. 5(a). An external frame is covered with black cloth to block ambient light sources and isolate the experiment setup. Inside, a mock up of a flight deck consisted of a large shelf in front of the pilot supporting the primary flight monitor and computer, and a center console separating the captain and first officer seats. Assuming the Captain to be the pilot-in-command during the emergency, the simulator aims at replicating a scenario in which the First Officer is tasked with the landing site selection. The dynamic interfaces the First Officer can access are a the PFD and the ND (conventional LCD display, top right in Fig. 5(b)) as well as the CDU and the ND Control Panel (touchscreen LCD, bottom left in Fig. 5(b)). The touchscreen was placed on the center console in order to allow the pilots to interact primarily with the CDU near its normal position.

The subjects were seated to the right of the center console in the First Officer's seat. Posters were included to provide the pilot with the look and feel of an actual cockpit. These images included the captain's seat, other displays which were not simulated, and a view through the windscreen. The computer screen in front of the pilot showed the PFD on the left and the ND on the right. The screen on the center console was an LCD touchscreen, where the pilots were able to interact with the system. This screen contained the CDU, the NDCP, an alert display and ancillary simulator controls.

V. Analysis of Results

The statistical analysis was performed using both parametric and non-parametric statistical analysis, using a repeated-measures Analysis of Variance (ANOVA) or a Spearman's signed rank test, respectively [26, 27]. The significance for all tests was set at $\alpha = 0.05$. In cases where the results



(a) First-Officer in a pilot-not-flying situation.

(b) The dynamic interfaces of the simulator.

Fig. 5 The overall experimental setup.

from the repeated-measures ANVOVA were found to be significant, the effect size η^2 is reported as a percentage of the overall variance attributed to each predictor. In cases where the results were found to be insignificant, the power for the test, $1 - \beta$, is reported.

The F -distribution and statistic utilized for the ANOVA is characterized by two parameters, the degrees of freedom in the numerator and denominator, respectively. The first represents the number of groups, for this work 1, the second the number of cases minus the number of groups, for this work $8 - 1 = 7$. In the following, $F(1, 7)$ represents the computed value for the observed data. This value is compared to $F_{critical} = F_{critical}(0.05; 1, 7) = 5.591$. If $F(1, 7) > F_{critical}$, the result is assumed to be significant at the 0.05 level. For details, see [28, 29].

A. Performance

Overall, the average time required to select a landing site was 110 seconds, with a standard deviation of 48 seconds. The unique combination of each phase of flight, failure type and airport availability in each scenario is an important factor which largely accounts for the difficulty in selecting a landing site (see Fig. 6). Each test run was completed in a suitable amount of time with the exception of one test run by one subject in which the engine failed during climb. The subject in this run selected an alternate landing site 218 seconds after the failure occurred, which was considered too long by the expert, given the aircraft's altitude at the time. All other test runs by this subject—as well as by all other subjects—were, according to the subject matter expert's opinion,

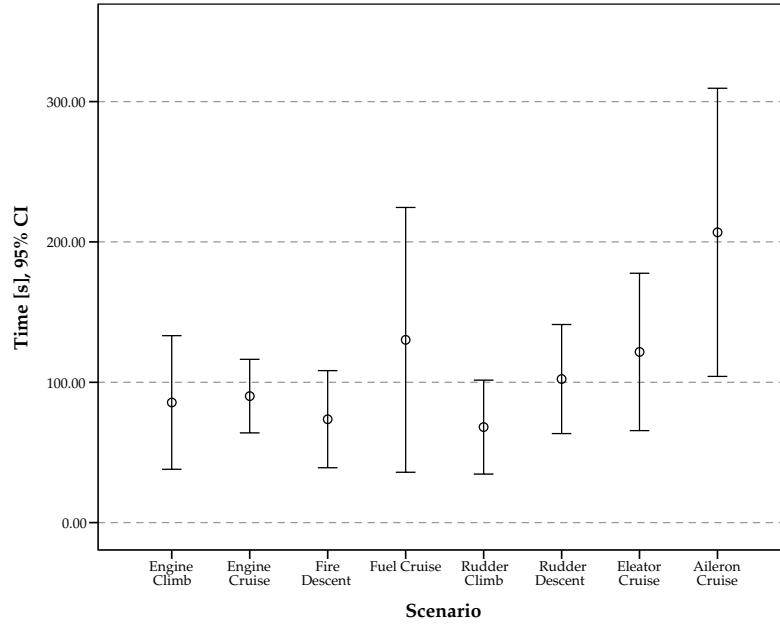


Fig. 6 The average Time-to-Complete by scenario and the corresponding 95 % confidence intervals.

completed within an acceptable amount of time. As shown in Fig. 6, the average times varied between the different scenarios. Although there was some variation between all scenarios, the times for the aileron failure scenario were significantly higher than for other scenarios. The average time of the aileron failure scenario was 206.8 seconds compared to an average of 96.0 seconds for all other scenarios.

The ability of pilots to select the best landing site (i.e., to pick the same landing site that the subject matter expert *a priori* determined to be the best pick of all alternates) was generally quite good. The median ranking of landing sites selected was 1, and the selected landing site was ranked highest in 57.8 % of runs. Figure 7 shows how the participants' selections were classified across scenarios. These results show that the scenarios may have differed in their difficulty. For the engine failure during climb and the rudder control failure in climb, every pilot was able to select the most appropriate landing site. For the fire emergency scenario, only one participant selected the most appropriate landing site. This may not be surprising, considering that time is most limited in the event of an onboard fire. For the elevator failure scenario, a number of different sites were selected, which may indicate that the differentiation between the best landing site and the other landing sites

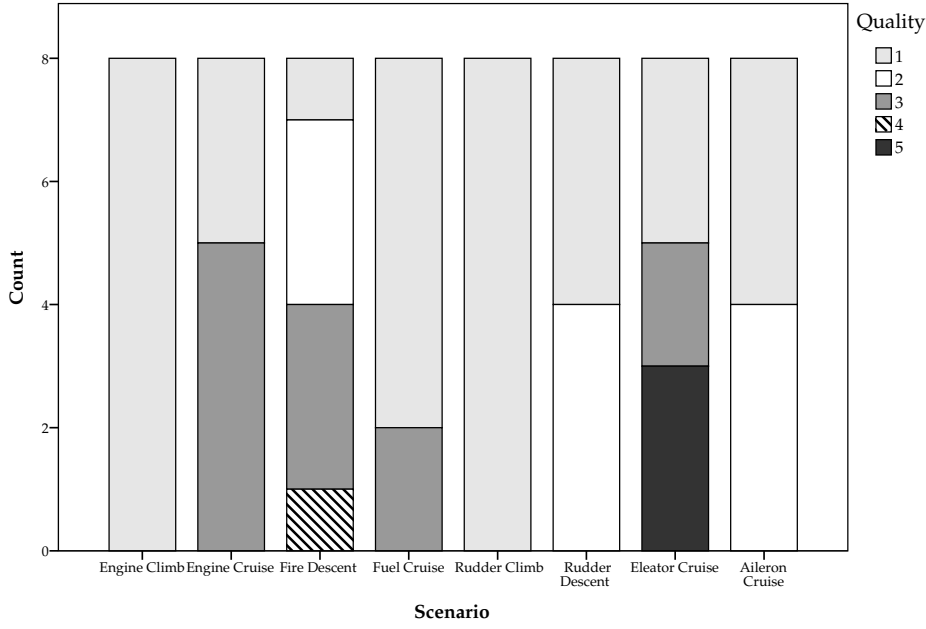


Fig. 7 Landing site quality by scenario. The quality metric is the rank of the selected option in the expert’s solution.

may not have been substantial enough.

In each scenario, there are some trade-offs between the two measures of performance, time and quality. It was expected that some pilots may spend more time deliberating and coming to a better decision, whereas others may make decisions more quickly at the expense of the quality of the decision. Spearman’s signed rank test [27] was used to determine the correlation between these two measures. The results showed that there was no significant correlation between time and quality ($r_s = 0.203$, $p = 0.107$). With the exception of four out of the total of 64 cases, all runs ended with a selection of a top three ranked site. Focusing on those results (i.e. quality levels 1-3), Figure 8 shows that runs which were ranked lower in quality, took slightly longer on average to make and had much greater variance than those classified as being the best, corresponding to a quality of 1.

A repeated-measures ANOVA was conducted to determine the effect of APA type and the familiarity of emergencies on the time to complete. The mean time for cases with the filter dial was slightly higher than those without the dial ($\mu_{\text{dial}} = 115.3$ s, $\mu_{\text{no dial}} = 104.4$ s). The time to select an alternate was lower in the case of familiar emergencies than in cases of unfamiliar emergencies ($\mu_{\text{familiar}} = 94.9$ s, $\mu_{\text{unfamiliar}} = 124.7$ s). The effect of APA type was statistically insignificant

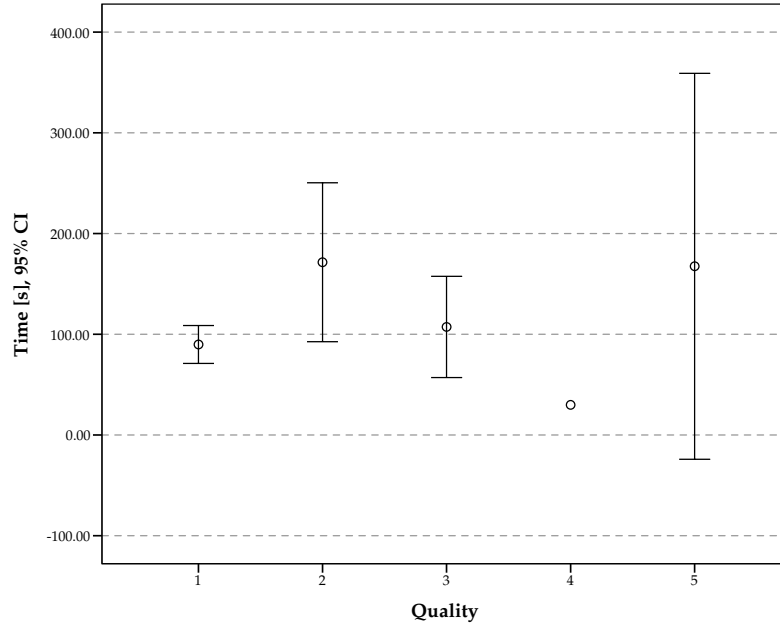


Fig. 8 Correlation between the average time to get to a decision and the decision quality (1 corresponds to the best quality).

($F(1, 7) = 0.217$, $p = 0.655$, $1 - \beta = 0.069$), whereas the effect of the familiarity of the emergency on time to complete was statistically significant ($F(1, 7) = 6.979$, $p = 0.033$, $\eta^2 = 0.499$). The interaction between APA type and emergency familiarity indicated that the effect of the dial was not statistically different in the familiar cases than in the unfamiliar cases, Fig. 9.

The effect of the APA variation was also tested in regards to the quality of the landing site selected, Fig. 10. Cases in which the filter dial was present resulted in slightly better quality of alternates selected. The selected landing site was ranked first or second in 84.4 % of cases which included the dial, compared to only 65.6 % of cases which did not include the dial. Because this dependent variable is ordinal, two Wilcoxon Signed Ranks tests [30] were used. The results from the Wilcoxon tests showed that the difference in quality of landing site selection was not significantly affected by either the APA type ($Z = -1.304$, $p = 0.192$), nor the familiarity of the emergency ($Z = -0.382$, $p = 0.703$). Thus the quality is robust regardless of the familiarity of the emergency or the ability to filter the suggestions.

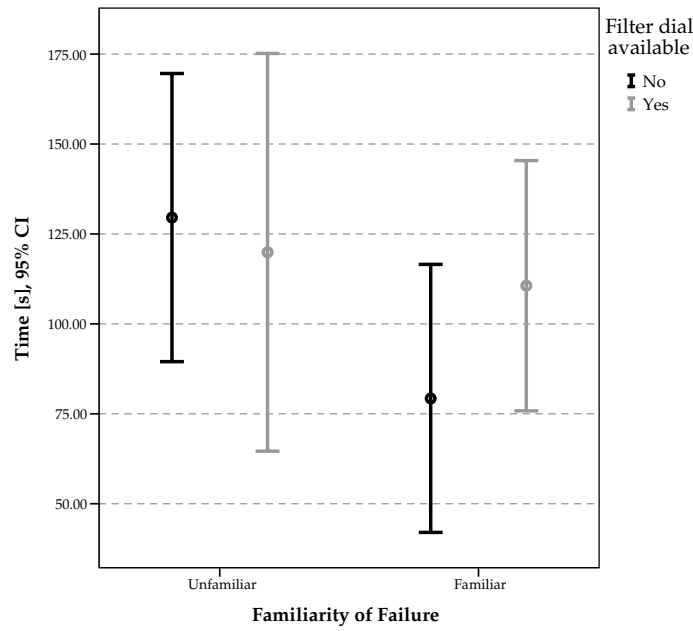
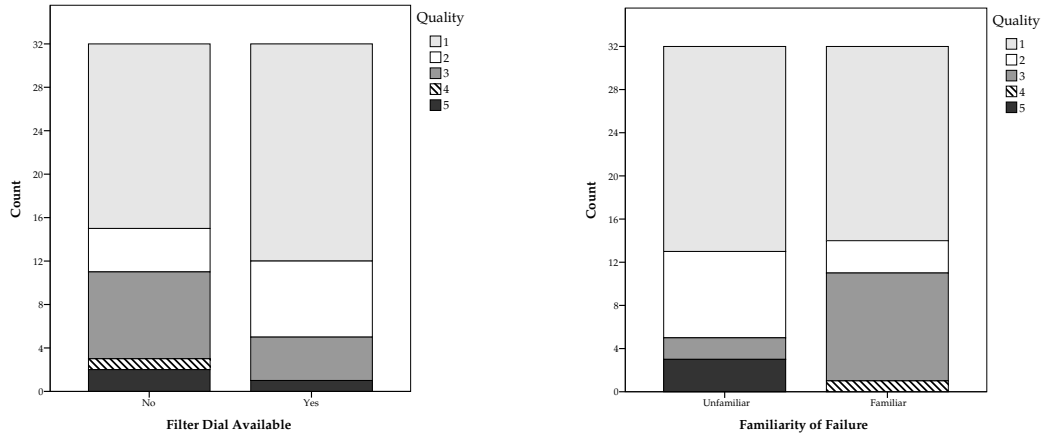


Fig. 9 Effect of APA variation and familiarity on the average time to make a selection as well as on the variance (indicated by the 95 % CI).



(a) With and without dial.

(b) Familiar and unfamiliar scenarios.

Fig. 10 Quality of the landing site selection by dial availability and failure familiarity.

B. Workload and Situation Awareness

As previously discussed, the cockpit environment can be very stressful in the event of an emergency. The pilots have a number of tasks which require their attention, and which must be completed in a timely fashion. Therefore, any automation added to the flight deck should not add to the already high workload. A repeated-measures ANOVA was used to determine the effect the

independent variables had on the pilots' workload. The effect of the addition of the dial slightly reduced the workload ($\mu_{\text{dial}} = 51.1$ s, $\mu_{\text{no dial}} = 54.9$ s), but this reduction was not statistically significant ($F(1, 7) = 1.878$, $p = 0.213$, $1 - \beta = 0.221$). The familiarity of the emergency did not have a significant effect on workload either ($\mu_{\text{familiar}} = 52.6$ s, $\mu_{\text{unfamiliar}} = 53.4$ s, $F(1, 7) = 1.277$, $p = 0.296$, $1 - \beta = 0.165$).

A correlation analysis was used to determine if there was indeed a correlation between the pilots' workload and their performance. Spearman's rank correlation coefficient was used to determine that the correlation between the workload and time ($r_s = 0.432$, $p < 0.001$) as well as between workload and quality ($r_s = 0.286$, $p = 0.022$) were both significant. As workload increased, the amount of time taken to determine a solution increased and the quality of the selection decreased.

In order to make a good decision, the pilot must be aware of the situation at hand. The results of the SA questionnaire were used to determine the effect of the aid variation on the pilots' understanding of each situation. Level 1 SA questions were correctly answered at 56 %, Level 2 at 52 % and Level 3 at 47 %. These low percentages of correct responses may be partly attributed to the type of questions that were asked. For instance, a number of the questions required that pilots recall airport identifiers in order to correctly answer the questions, which were unfamiliar to the participants. Also, pilots were able to keep only the important identifiers in mind, and these were kept only long enough to complete the scenario. These factors likely contributed to low numbers of correct responses.

Despite the low absolute scores, three separate repeated-measures ANOVAs were used to test for effects of the independent variables, one for each level of SA. For Level 2 and Level 3 situation awareness the effect of the APA dial and the familiarity of the experiment individually were marginally statistically significant. Level 1 situation awareness was overall marginally significantly affected by the APA dial ($F(1, 7) = 3.781$, $p = 0.093$, $\eta^2 = 0.351$) and the familiarity of the experiment ($F(1, 7) = 4.515$, $p = 0.024$, $\eta^2 = 0.539$). Level 1 SA was greater without the dial ($\mu_{\text{dial}} = 0.48$, $\mu_{\text{no dial}} = 0.58$) and greater in situations which were familiar ($\mu_{\text{familiar}} = 0.58$, $\mu_{\text{unfamiliar}} = 0.48$). Additionally, Level 2 situation awareness was significantly affected by the interaction of the APA dial and the familiarity of the experiment ($F(1, 7) = 9.00$, $p = 0.020$, $\eta^2 = 0.562$). In unfamiliar

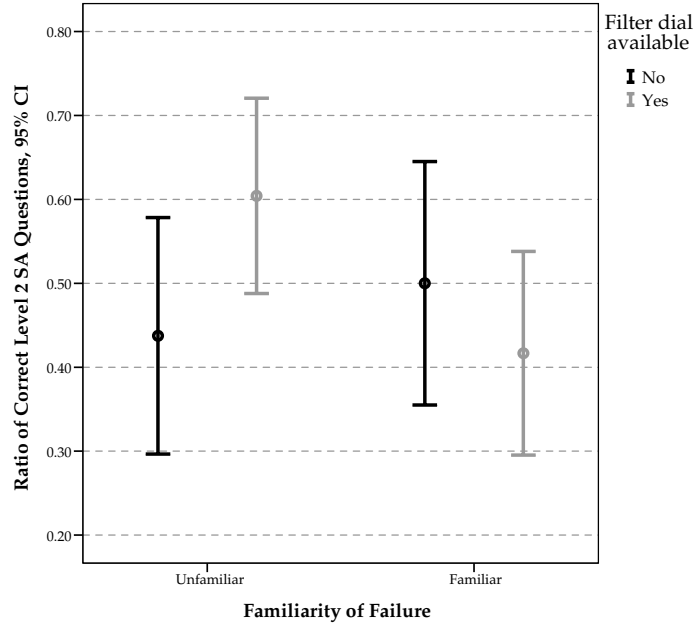


Fig. 11 Effect of the APA variation and scenario familiarity.

emergency situations, the Level 2 situation awareness (understanding perceptual cues) is improved with the inclusion of a filter dial ($\mu_{\text{dial}} = 0.604$, $\mu_{\text{no dial}} = 0.438$), whereas in familiar situations the Level 2 situation awareness is decreased ($\mu_{\text{dial}} = 0.417$, $\mu_{\text{no dial}} = 0.500$), Fig. 11.

As with workload, changes in situation awareness—regardless of the source—may have an effect on the pilots’ performance. In order to test for this correlation, the Spearman’s rank correlation coefficient test was again used. The results show that there was no significant correlation between performance and Level 1 or Level 2 SA. There was, however, a marginally significant negative correlation between Level 3 SA and the time to select a landing site ($r_s = -0.212$, $p = 0.092$). Thus the higher the Level 3 SA the more quickly the pilot was able to choose a landing site.

C. Secondary Measures

In order to understand the usage of the APA with and without the dial, both the number of landing sites for which the pilot viewed more information, and the number of times the dial setting was changed (for cases in which the dial was available) were analyzed. Viewing more options takes time but may lead to a better understanding of the alternates available and thus a better selection quality.

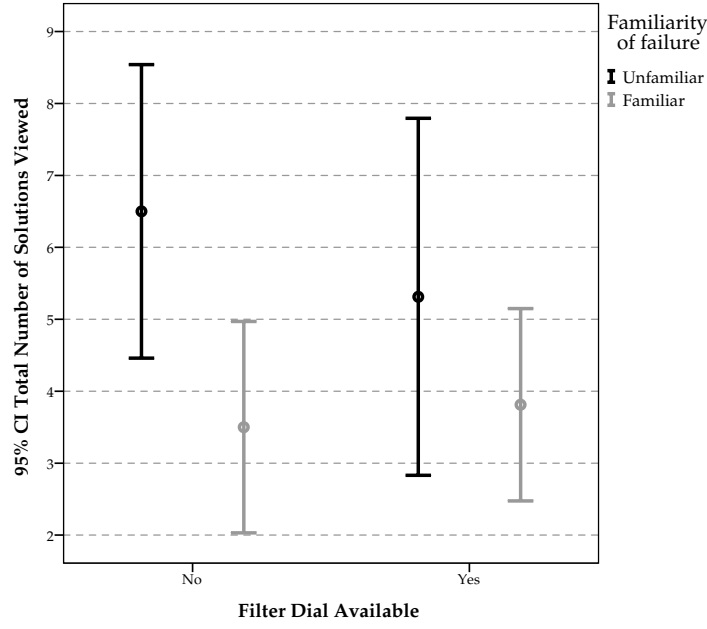


Fig. 12 Interaction of APA type and familiarity of failure on total number of solutions viewed.

A repeated-measures ANOVA was conducted on the number of landing sites viewed to determine the effect that the independent variables had on this measure. The average number of solutions viewed was slightly lower for the cases with the dial than for those without the dial ($\mu_{\text{dial}} = 4.5$, $\mu_{\text{no dial}} = 5.0$), however this difference is not statistically significant ($F(1, 7) = 0.329$, $p = 0.575$, $1 - \beta = 0.084$). The familiarity of the emergency had a significant effect on the number of solutions viewed ($F(1, 7) = 5.639$, $p = 0.031$, $\eta^2 = 0.273$). The average number of alternates viewed in familiar emergencies was lower than cases with unfamiliar emergencies ($\mu_{\text{familiar}} = 3.6$, $\mu_{\text{unfamiliar}} = 5.9$). A moderately significant interaction was also found between APA type and emergency familiarity ($F(1, 7) = 0.922$, $p = 0.058$, $1 - \beta = 0.147$). As seen in Fig. 12, in familiar emergencies, the addition of the dial slightly increased the number of solutions viewed from $\mu_{\text{no dial}} = 3.5$ to $\mu_{\text{dial}} = 3.8$. However, in unfamiliar emergencies, the dial reduced the average number of solutions viewed from $\mu_{\text{no dial}} = 6.5$ to $\mu_{\text{dial}} = 5.3$. This may indicate that in familiar cases the filter encouraged the pilot to examine more alternatives than he otherwise would have, and in the unfamiliar cases allowed the pilot to focus on a smaller but more promising set of alternatives.

In order to determine the effect the number of alternates viewed had on the pilots' performance, Spearman's rank correlation coefficient test [27] was used to determine the correlation between the

number of alternates viewed, the time to select an alternate, the workload reported and the quality of the landing site selected. There was a significant positive correlation between the number of alternates viewed and the time to complete the task ($r_s = 0.737$, $p < 0.001$). The slight positive correlation between the number of alternates viewed and the quality of the selection made was not significant ($r_s = 0.171$, $p = 0.178$). This implies that participants who viewed more solutions took significantly more time to make a selection, but did select significantly better alternatives. A significant positive correlation was found between the workload reported and the number of alternatives viewed ($r_s = 0.352$, $p = 0.002$). This positive correlation implies that for runs which pilots viewed more alternates, they also encountered a higher workload.

Following all of the data collection runs, the pilots were asked whether they subjectively felt that the filter dial was a useful addition to the APA system. Half of the participants responded that the addition of the dial made the system “much better,” whereas the other half responded that the dial made the system “better.” This feedback may be reflective of the fact that the dial may be used or ignored in any given situation. A general sentiment was expressed that a tool which can be used or ignored is desirable.

The Modified Cooper-Harper for Displays [25] was used to assess both variations of the aid. Table 1 provides the description associated with each level of the rating scale. Fig. 13 shows the results from both variations of the aid. Every participant rated the version with the dial the same or higher than the version without the dial. One notable case was the participant who assigned both versions of the aid a score of 10, which is described as “display is missing critical information; operator is unable to locate essential information...” This subject commented that “runway length is of critical importance and is too hard to find in the pages.” All other participants rated the variation with the dial as either “excellent and highly desirable,” (i.e., 1), or “good with negligible deficiencies,” (i.e., 2). Only half of the participants rated the variation without the dial in either of these categories.

Table 1 Modified Cooper-Harper Rating Scale for Displays.

Rating	Description
1	Excellent & Highly Desirable
2	Good with Negligible Deficiencies
3	Minor but Tolerable Deficiencies
4	Moderately Objectionable Deficiencies
5	Very Objectionable Deficiencies
6,7,8	Deficiencies Require Improvement: Major Deficiencies
9,10	Mandatory Redesign: Major Deficiencies

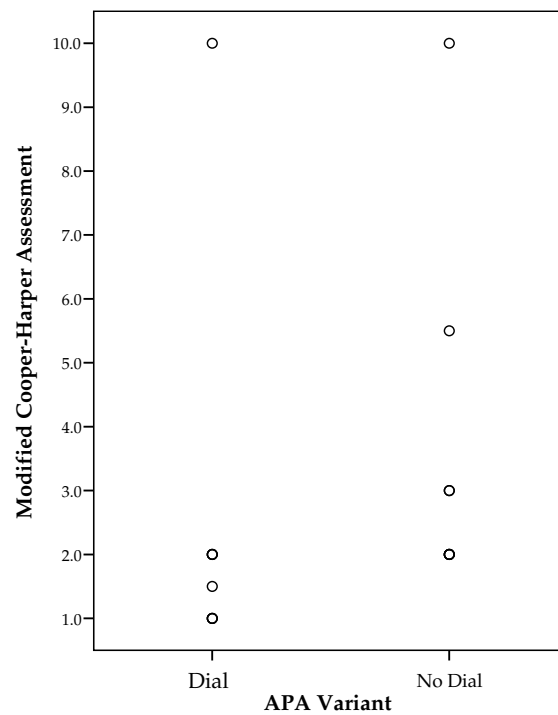


Fig. 13 Modified Cooper-Harper Ratings of the APA.

VI. Discussion

This experiment sought to measure pilots' performance in handling emergency situations when using an APA. The two primary measures of performance were the time required to select an alternate landing site and the quality of the landing site which was selected. It was expected that pilots would have to make a trade-off between these two factors. That is, to make a better decision, some pilots may more thoroughly consider their options, resulting in a long time to complete the task.

On the other hand, some pilots may choose to act quickly, without investigating all available options, resulting in a lower quality decision. The dial was expected to reduce this negative correlation by reducing the number of options pilots considered. However, the results showed that the correlation between these two measures was, in fact, positive. This suggests that runs in which the pilot took longer to complete the task actually resulted in poorer landing site quality. Though contrary to expectation, a number of plausible explanations exist. First the scenario design may have affected this relationship; specifically the number of sites which are similar to the highest ranked site. This similarity may have made some scenarios more difficult than others, requiring more time to consider the options and to differentiate between the highest ranked sites. Second, the benefit of automation may decrease as time pressure is relaxed, suggesting that automation may be less beneficial in lower time pressure scenarios [31]. Third, providing individuals with tools to aid with analytic decision making may result in increased decision time without the associated improvement in decision quality.

A difference in performance was expected for familiar and unfamiliar scenarios. The results supported this hypothesis and showed that pilots made their selections more quickly for familiar scenarios. This supports the results of the pilots' survey in [16], which showed that pilots were more likely to take immediate action in a familiar emergency. This result is also in line with the theory of recognition primed decision making [32] wherein experts upon recognizing a situation immediately understand the implications and are able to make decisions quickly. The effect of the addition of the dial was affected by the familiarity of the scenario. The time to complete the task was increased by the addition of the dial in the case of familiar emergencies, whereas the time was decreased by the dial in the case of unfamiliar emergencies. This may suggest that the dial proved more useful for filtering out inappropriate options in unfamiliar emergencies. Each participant used his own method to make the APA most useful. However, there were some comments that shed light on how the pilots used the APA. One point of interest is how pilots begin to narrow down the list of possible options. When the dial was available, the first step may be to adjust the filter to an appropriate setting.

In debriefing after the fact, most participants indicated that the ability to filter out unnecessary information and view more detailed information about each landing site were the two most useful

features. Pilots emphasized that a key attribute of these features was the speed with which information could be processed. A number of pilots commented that the “ALL FIELDS” dial setting (compare Fig. 3) was not useful. The ability to obtain critical information quickly was emphasized by pilots who suggested that more information should be encoded into the graphical display. The addition of a filter setting to show only sites with runways of sufficient length for the aircraft (as currently configured) was suggested by multiple participants. Finally, pilots commented that an APA-type aid should be linked with airline dispatchers. This ability could be used in a number of ways, such as live updates of airline preferred and weighting of evaluation criteria.

Pilots were asked to describe the types of scenarios in which an APA would be most useful, and those in which it would be the least useful. Most pilots expressed that the aid would be most useful in situations where there was high workload and high temporal pressure. These situations are characterized by the need to make a decision quickly. The ability to quickly access large amounts of pertinent information makes the aid a large improvement over current options. Pilots also commented that an APA would be useful in a less intense emergency, in which the aircraft is unable to reach its destination, but the situation is otherwise normal. The aid can be used more deliberately to assess all options and determine the most suitable landing site. The situations in which the aid would not be useful were varied. One participant responded that the aid would not be useful in dire emergencies such as a fire because, “the only piece of information necessary is the nearest runway, all other data is irrelevant.” However, another pilot did not identify a situation in which the aid would be least useful, but commented that, “information is always useful in formulating a plan, the more info, the better.”

Several caveats regarding realism and training apply to this work. First, although all of the pilots took the experiment seriously, the experimental conditions could not replicate all of the experiences of an actual in-flight emergency. The pilots were offered additional compensation for improved performance, but this does not entirely replicate the stress and pressures of a real emergency. Second, the pilots were given two training runs, which allowed them to familiarize themselves with the features available. Although this was sufficient for the pilots to gain a general understanding of the APA’s functionality, if fielded pilots would be trained to a much higher standard on the device.

VII. Conclusion

This work has sought to evaluate the efficacy of an automated path planning aid (APA), intended to help pilots plan a safe trajectory to land in the event of an in-flight emergency. A prototype was designed and implemented in a cockpit simulation. This simulator was used to test the aid and gather results and further feedback from pilots. The aid which was developed had to be compatible with existing cockpit designs. The aid was designed to be easy to use, without requiring unnecessary time and effort on the part of the pilot. A filter dial was added to allow the pilot to quickly focus only on alternates that were appropriate for a given emergency.

Comparisons between the two variations of the APA showed that the addition of the dial resulted in a small difference in the quality of landing site selected and longer times to select a site. The dial did not significantly reduce the number of alternates viewed, which was strongly correlated with the time metric. This may indicate that the dial did not simplify the task as much as anticipated. However, in the case of unfamiliar emergencies, the dial reduced both time to select a landing site as well as the number of solutions viewed. Every participant scored the variation with the filter dial more highly than the variation without the dial, indicating that they preferred to have the dial, despite the lack of improvement in performance measured in the experiment. The APA was tested using both familiar and unfamiliar emergencies in order to understand if the APA was more useful in one type of scenario than another type. Both survey and simulator results indicated that pilots are likely to act more quickly in a familiar emergency. Pilots found the filter dial and the consolidation of information about landing sites to be very useful features. The ability to quickly and easily access critical information is one of the most important characteristic of an emergency planning aid. This design facilitated the pilots' methods of assessing each landing site throughout a flight, before an emergency has occurred. The ranking system (though not always optimal) gave the pilots aggregate scores for each site and provided a more meaningful starting place when investigating the available options.

Acknowledgements

This work was supported under NASA Grant NNX08AB94A. The authors would particularly like to thank the pilots who participated in the experiment, especially David Pietruszewski who provided knowledge and expertise from an airline pilot’s perspective.

References

- [1] National Transportation Safety Board, “Accidents and Accident Rates by NTSB Classification, 1991 through 2010, for U.S. Air Carriers Operating Under 14 CFR 121,” 2010. <http://www.nts.gov/data/table2.html>.
- [2] Chen, T. and Pritchett, A., “Development and evaluation of a cockpit decision-aid for emergency trajectory generation,” *Journal of Aircraft*, Vol. 38, No. 5, 2001, pp. 935 – 943.
- [3] Cummings, M. L., “Automation bias in intelligent time critical decision support systems,” *AIAA 1st Intelligent Systems Technical Conference*, Vol. 2, Chicago, IL, United states, 20–22 Sept. 2004, pp. 557–562, AIAA-2004-6313.
- [4] Hollnagel, E., “Context, cognition and control,” *Co-operative Process Management*, edited by Y. Waern, chap. 4, Taylor & Francis Ltd., 1998, pp. 27–52.
- [5] Bearman, C. R., “Naturalistic Decision Making,” *Center for Decision Research Seminar Series*, Leeds, England, Jan. 2004.
- [6] Keinan, G., “Decision making under stress: Scanning of alternatives under controllable and uncontrollable threats,” *Journal of Personality and Social Psychology*, Vol. 52, No. 3, March 1987, pp. 639–664.
- [7] Maule, A., Hockey, G. J., and Bdzola, L., “Effects of time-pressure on decision-making under uncertainty: changes in affective state and information processing strategy,” *Acta Psychologica*, Vol. 104, No. 3, 2000, pp. 283 – 301.
- [8] Maule, A. J. and Edland, A. C., “The effects of time pressure on human judgement and decision making,” *Decision Making*, edited by R. Ranyard, W. R. Crozier, and O. Svenson, chap. 11, Routledge, London, 1997.
- [9] Mosier, K. L. and Skitka, L. J., “Human Decision Makers and Automated Decision Aids: Made for Each Other?” *Automation and human performance: theory and applications*, edited by R. Parasuraman and M. Mouloua, chap. 10, Lawrence Erlbaum Associates, Inc., Mahwah, New Jersey, United States, 1996, p. 201.
- [10] Zsombok, C. E., *Naturalistic Decision Making: Where Are We Now?*, chap. 1, Lawrence Erlbaum Associates, Inc., 1997, pp. 3–17.

- [11] Johannsen, G. and Rouse, W., “Mathematical concepts for modeling human behavior in complex man-machine systems,” *Human Factors*, Vol. 21, No. 6, Dec. 1979, pp. 733–47.
- [12] Simpson, P. A., “Naturalistic Decision Making in Aviation Environments,” Tech. rep., Defence Science And Technology Organisation Victoria (Australia), Aeronautical And Maritime Research Lab, Jan. 2001.
- [13] Atkins, E., Portillo, I. A., and Strube, M., “Emergency Flight Planning Applied to Total Loss of Thrust,” *Journal of Aircraft*, Vol. 43, No. 4, July–Aug. 2006, pp. 1205–1216.
- [14] Layton, C., Smith, P., and Coy, C. M., “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. www.hf.faa.gov/docs/coopprob/6enroute.pdf.
- [15] Honeywell, *Boeing 777 Flight Management System Pilot’s Guide*, 2001.
- [16] Watts, R., Tsiotras, P., and Johnson, E. N., “Pilot feedback for an automated planning aid system in the cockpit,” *28th IEEE/AIAA Digital Avionics Systems Conference*, IEEE, Orlando, FL, United states, 23–29 Oct. 2009, pp. 5.B.21–5.B.213.
- [17] Jain, S. and Tsiotras, P., “Trajectory optimization using multiresolution techniques,” *Journal of Guidance, Control, and Dynamics*, Vol. 31, No. 5, 2008, pp. 1424–1436.
- [18] Karelaiti, J., Virtanen, K., and Öström, J., “Automated generation of realistic near-optimal aircraft trajectories,” *Journal of Guidance, Control, and Dynamics*, Vol. 31, No. 3, 2008, pp. 674–688.
- [19] Dubins, L. E., “On Curves of Minimal Length with a Constraint on Average Curvature, and with Prescribed Initial and Terminal Positions and Tangents,” *American Journal of Mathematics*, Vol. 79, No. 3, 1957, pp. 497–516.
- [20] Bakolas, E., Zhao, Y., and Tsiotras, P., “Initial Guess Generation for Aircraft Landing Trajectory Optimization,” *AIAA Guidance, Navigation, and Control Conference*, Portland, OR, 8–11 Aug. 2011, AIAA 2011-6689.
- [21] Ippolito, C. A. and Pritchett, A. R., “Software architecture for a Reconfigurable Flight Simulator,” *AIAA Modeling and Simulation Technologies Conference*, Denver, CO, 14–17 Aug. 2000, AIAA-2000-4501.
- [22] Endsley, M. R., “Measurement of situation awareness in dynamic systems,” *Human Factors*, Vol. 37, No. 1, 1995, pp. 65–65.
- [23] Endsley, M. R., “Direct Measurement of Situation Awareness: Validity and Use of SAGAT,” *Situation Awareness: Analysis and Measurement*, chap. 7, Lawrence Erlbaum Associates, Inc., 2000, pp. 147–173.
- [24] Endsley, M. R., “Theoretical underpinnings of situation awareness: A critical review,” *Situation Aware-*

- ness: *Analysis and Measurement*, chap. 1, Lawrence Erlbaum Associates, Inc., 2000, pp. 3–32.
- [25] Cummings, M., Meyers, K., and Scott, S., “Modified Cooper Harper Evaluation Tool for Unmanned Vehicle Displays,” *Unmanned Vehicle Systems Canada Conference, Montobello, Quebec*, Unmanned Systems Canada, Ottawa, Ontario, Canada, Nov. 2006, pp. Civil-2.
- [26] Warner, R. M., *ANOVA*, [28], Sept. 2007, pp. 876–928.
- [27] Siegel, S. and Castellan Jr., N. J., *Spearman’s Signed Rank Test*, [29], 2nd ed., 1988, pp. 235–244.
- [28] Warner, R. M., *Applied Statistics: From Bivariate Through Multivariate Techniques*, SAGE Publications, Inc, Thousand Oaks, CA, USA, Sept. 2007.
- [29] Siegel, S. and Castellan Jr., N. J., *Nonparametric Statistics for The Behavioral Sciences*, McGraw-Hill, 2nd ed., 1988.
- [30] Siegel, S. and Castellan Jr., N. J., *Wilcoxon Signed-Rank Test*, [29], 2nd ed., 1988, pp. 77–78.
- [31] Johnson, K. E., Kuchar, J. K., and Oman, C. M., “Experimental Study of Automation to Support Time-Critical Replanning Decisions,” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 46, SAGE Publications, Thousand Oaks, California, USA, 2002, pp. 337–341.
- [32] Klein, G. A., “A Recognition-primed Decision (RPD) Model Of Rapid Decision Making,” *Decision Making in Action: Models and Methods*, edited by G. A. Klein, J. Orasanu, and R. Calderwood, chap. 6, Ablex Publishing, 1993, pp. 138–147.
- [33] The immediate trajectory line extends up from the aircraft representation, roughly until the “80 NM” distance marker, then slightly going left until intersecting the heading indicator at the marker representing a 70° heading. On the real system, this line is magenta to better identify it.
- [34] Human Performance Research Group, data available online at [http:// human-factors.arc.nasa.gov/groups/TLX/](http://human-factors.arc.nasa.gov/groups/TLX/). (Last retrieved Sep. 2012)