Flight Simulation Study of Airplane State Awareness and Prediction Technologies

Steven D. Young, Taumi Daniels, Emory Evans, and Evan Dill
NASA Langley Research Center, Hampton, VA

Maarten Uijt de Haag
Ohio University, Athens, OH

Tim Etherington
Rockwell-Collins, Cedar Rapids, IA

Abstract—Airplane state awareness (ASA) is a pilot performance attribute derived from the more general attribute known as situation awareness. Airplane state alludes primarily to attitude and energy state, but also infers other state variables, such as the state of automated or autonomous systems, that can affect attitude or energy state. Recognizing that loss of ASA has been a contributing factor to recent accidents, an industry-wide team has recommended several Safety Enhancements (SEs) to resolve or mitigate the problem. Two of these SEs call for research and development of new technology that can predict energy and/or auto-flight system states, and intuitively notify or alert flight crews to future unsafe or otherwise undesired states. In addition, it is desired that future air vehicles will be able to operate with a high degree of awareness of their own well-being. This form of ASA requires onboard predictive capabilities that can inform decision-making functions of critical markers trending to unsafe states. This paper describes a high-fidelity flight simulation study designed to address the two industry-recommended SEs for current aircraft, as well as this desired self-awareness capability for future aircraft. Eleven commercial airline crews participated in the testing, completing more than 220 flights. Flight scenarios were utilized that span a broad set of conditions including several that emulated recent accidents. An extensive data set was collected that includes both qualitative data from the pilots, and quantitative data from a unique set of instrumentation devices. The latter includes a head-/eye-tracking system and a physiological measurement system. State-of-the-art flight deck systems and indicators were evaluated, as were a set of new technologies. These included an enhancement to the bank angle indicator; predictive algorithms and indications of where the auto-flight system will take the aircraft and when automation mode changes will occur or where energy-related problems may occur; and synoptic (i.e., graphical) depictions of the effects of loss of flight critical data, combined with streamlined electronic checklists. Topics covered by this paper include the research program context, test objectives, descriptions of the technologies under test, platform and operational environment setup, a summary of findings, and future work.

Keywords—Energy state, autonomy monitoring, loss of air data, predictive capability, electronic checklists, synoptics

I. INTRODUCTION

In the period 2010 to 2014, the Commercial Aviation Safety Team (CAST) sponsored a study of 18 commercial aviation events that occurred within ~10 years prior to the study kickoff. Results identified 12 recurring problem themes involving loss of ASA, and suggested a number of intervention strategies [1]. CAST assessed these strategies for effectiveness and feasibility, and then recommended Safety Enhancements (SEs) for the industry to implement [2]. Of these, five were deemed to require research to enable the enhancement. CAST also published plans that articulated the necessary outputs from the research. As part of its role in CAST, NASA initiated subprojects within the Airspace Technology Demonstrations project to work with industry to produce many of these outputs. The Automation and Information Management Experiment (AIME) is the first in a series of experiments intended to address or achieve outputs defined within the plans for SE-207 [3] and SE-208 [4]. These outputs are:

- **SE-207, Output 3,** “Systems that predict the future aircraft energy state and/or autoflight configuration if the current course of action is continued and provide appropriate alerting” [3]
- **SE-208, Output 1a,** “Displays that present the current and future expected state of automated systems in an intuitive manner; displays should provide information for both the mode currently selected and impending mode transitions expected per design of these systems” [4]
- **SE-208, Output 1b,** “Displays that show, in a simple, integrated manner (e.g., a synoptic), the aircraft flight-critical data systems in use by automated systems and primary flight instruments, for both the mode currently selected and any impending mode transitions expected per design of these systems” [4]

Together, the technologies under development and evaluation as part of AIME are intended to enable improved energy, automation, and/or system state awareness, as well as to provide new predictive capabilities with respect to these aspects of ASA. The remaining outputs required for SE-207
and SE-208 research are being developed by other activities within the project and are not reported here.

II. TEST OVERVIEW AND OBJECTIVES

AIME was conducted over an 11-week period, completing January 28, 2016. As discussed in [5], the objectives were four-fold. First, to raise the Technology Readiness Level (TRL) for new technology via testing in a high-fidelity flight simulator environment; this includes confirming that technologies perform as intended across a span of targeted conditions (e.g., representative of accident circumstances). Second, to evaluate the usability of new technology concepts; thereby helping to decide whether the project is on the correct path, or needs a change of direction to produce the CAST-desired research outputs. Third, to expose design characteristics that require refinement for future work; and discover previously unknown issues related to SE-207/208 goals. Fourth, to evaluate the use of the eye-tracking and physiological measurement systems for their potential to help determine design effectiveness, as well as for detecting attention-related issues.

A. Technologies Under Evaluation

One important factor to consider when evaluating new technology concepts is the reference platform into which the new technology will be inserted. Based on guidance from the CAST, it was decided that the Boeing 787 (B-787) flight deck [6] would be used for this purpose. However, as most of the relevant B-787 indicators, displays, and functions are also available on other recently developed aircraft, findings should be applicable. Retrofit of these technologies onto older aircraft may also be possible.

The B-787 is a highly advanced aircraft, representing in many ways state of the art. The many indicators, displays, and functions relevant to SE-207/208 objectives are summarized in [5]. This list includes: the Primary Flight Display (PFD), the Head-Up Display (HUD), the Flight Mode Annunciator (FMA) in the upper center of the PFD, airspeed and altitude “tapes” on the left and right side of the PFD, the Flight Path Vector (FPV) on the PFD, the Navigation Display (ND), the Vertical Situation Display (VSD), the Engine Indicating and Crew Alerting System (EICAS), the synoptic displays (7 types) available on Multi-Function Displays (MFDs), the Lower MFD which serves as the pilot interface to the Flight Management System (FMS), Electronic Checklists (ECL), the Electronic Flight Bag (EFB), the Air Traffic Control (ATC) display that lists ATC data link messages that have been received. These surfaces associated with these displays were organized to mimic the B-787 layout as shown in Figure 1.

As extensions to this reference set of displays and functions, four new technology concepts are evaluated. These are referred to as: (1) ME – Maneuver Envelope display and effects estimation function, information provided on the PFD; (2) TP – Trajectory Prediction function, information provided on the ND and VSD; (3) SIS – System Interaction Synoptic, information provided as an additional tab on the Synoptics Display of the upper MFD with associated simplified checklists within the Electronic Checklist (ECL) function; and (4) MHP/PAE – Multiple Hypothesis Prediction function with Predictive Alerting for unsafe Energy states, information provided on EICAS, ND, and VSD.

ME and TP are the result of previous research [7][8] and have matured to a higher TRL. For these technologies, AIME sought to evaluate their span of applicability across a broader set of conditions, such as those encountered in the CAST events. SIS and PAE are less mature, with AIME being a first opportunity for them to be demonstrated and assessed for usability. Below, these four technologies are briefly explained. Additional descriptions of their designs for this study, and findings resulting from their use can be found in [9] and [10].

As shown in Figure 2, the ME function’s predominant effect on the PFD is the bank angle indicator enhancement, which provides an additional (and visual) method to maintain awareness of the changing relationship between airspeed, bank angle, and lift. The ME function uses aerodynamic modeling and current conditions to adjust maneuverability bands (amber and red/black segments) dynamically during flight [8].

As shown in Figure 3, the TP and MHP/PAE functions provide predictive information on the VSD related to where automation mode changes will occur, or where energy-related problems will occur, respectively, if the current course of action is continued. The TP function’s prediction is indicated as a green line of constant length (5 min); this is where the auto-flight system will take the aircraft if the pilot makes no additional inputs. The green circle symbol indicates where a mode change will occur. The label is the name of the mode
change that will occur. These same cues are also provided on the ND where lateral-related predictions are more relevant [7].

The MHP/PAE prediction is indicated as a cyan circle and label on the TP-generated green line. In conjunction with this graphic indication, an EICAS message is provided to further explain the type of energy-related problem being predicted. For this test, six predictors were evaluated: over-speed, stall, vertical speed limit exceeded, unstable approach, high-and-fast on approach, and low-and-slow on approach [5][9].

Fig. 3. Sample TP and PAE indications on Vertical Situation Display
ALT hold transition (top); VNAV transitions (middle); PAE high-and-fast (ENERGY) example (bottom)

As shown in Figure 4, the SIS display provides a graphical illustration indicating the validity of flight-critical data and how this data affects other systems. The display has five segments of information. The top segment represents the Mode Control Panel (MCP); changes here can indicate elements of the auto-flight system that may become inoperative due to invalid data. Below this, the head-down displays are illustrated, along with how they can be affected. In the center segment, the flight-critical data types are listed along with their status (green = good; white = switched to backup source; amber = invalid). For this implementation, the flight-critical data types are: altitude, airspeed, attitude, heading, and position. The fourth segment (near the bottom) represents the underlying system architecture that provides the data. This example uses the facility’s architecture consisting of two Air Data Computers (ADCs) and three Inertial Reference Units (IRUs). This segment would look different for other aircraft. Finally, the bottom segment indicates the flight control mode, which typically changes from “primary” to “secondary” when flight-critical data becomes degraded or otherwise invalid. This overall layout and style of the SIS is designed to be consistent with training materials provided by the manufacturer, as well as the other standard synoptics.

The SIS synoptic does not replace the checklist; rather it supplements it, providing much of the same information in a different form. For this SIS evaluation, all associated off-nominal checklists were shortened to remove information that was conveyed on the SIS. For the situation shown in Figure 4, a blockage/icing of the pitot-static ports has triggered four EICAS messages in short succession and each has an associated checklist to complete.

Fig. 4. EICAS (left) and SIS (right) for loss of air data situation

B. Facility and Environment

AIME utilized the Research Flight Deck (RFD) within the Cockpit Motion Facility at NASA’s Langley Research Center. Although initially designed as a B-757 flight deck, RFD was modified to create the CAST-recommended reference test condition (i.e., B-787-like displays, interfaces, and functions). Several operational complexities were also enabled for AIME. These include both current complexities (e.g., RNAV and closely-spaced parallel approaches) and emerging ones (e.g., NextGen’s DataComm and net-enabled information services). This provided a realistic operational environment, albeit a complex one, in which flight crews may have little prior experience or training. However, previous similar experiments have shown crews can learn and perform well with limited exposure to this environment [11]. Specific environment capabilities were provided to support these operations and to use as independent variables that span a set of test conditions for TRL assessment; these included: two traffic cases, two airport configurations, five weather conditions, three turbulence levels (none, light, moderate), and two storm conditions (with 3D winds and regions of convection based on NOAA-provided data from past events).

C. Scenarios

Given the test objectives, one of the most challenging aspects of the AIME setup was defining a relevant set of flight
scenarios for the crews to experience. Advancing TRL requires testing across a good range of conditions, particularly any conditions that drive the need for the new technologies being developed. Similarly, effective usability assessment by the crew can only be done if the evaluation conditions span situations where usage would occur. With this in mind, several sources were used to develop an initial set of situations to be considered. The primary ones are listed below.

- CAST ASA study events [1][2]
- FAA Flight Deck Automation study [12]
- Loss of Control Accident Analysis study [13]
- Accident reports from recent events (e.g., Asiana-214 [14], Air France-447 [15], UPS-1354 [16])
- Landing automation workshop [17]
- ME, TP, and MHP previous research [7][8][18]
- Complexity and automation studies [11][19]
- Project interviews and workshops with industry Subject Matter Experts (e.g., pilots)

The highest priority was given to the 18 accidents and incidents investigated by the CAST ASA study team. While it would be impossible to replicate all conditions leading up to any of these accidents, there are many precursors that could be emulated, and these were used in the AIME scenarios.

Based on these sources and the resulting analyses, six types of scenarios were defined for AIME: nominal (type 1); loss of flight-critical data (type 2), increased risk of low energy (type 3), increased risk of high energy (type 4), unanticipated automation interaction (type 5), and distractions affecting workload management (type 6). These scenario types were not designed to be mutually-exclusive (e.g., a loss of flight-critical data (type 2) may, and did, induce unanticipated automation interaction (type 5)). Following multiple workshops and SME consultations, 2-4 scenarios of each type were designed resulting in ~20 unique scenarios to use for the run sequences for each crew. These are summarized in Table I. Weather cases (Wx) are coded as:

1. VMC, Ceiling 1000 ft, Visibility 3 nmi; Calm; Dusk
2. IMC, Ceiling 500 ft, Visibility 2400 ft; Storm, light turbulence and winds from Feb-25; Night
3. IMC, Ceiling 200 ft, Visibility 1800 ft; Storm, moderate turbulence and winds from Apr-27; Night
4. Clear-and-visibility ok (CAVOK); Dusk

### TABLE I. AIME SCENARIO SUMMARIES BY TYPE.

<table>
<thead>
<tr>
<th>Type</th>
<th>Wx</th>
<th>Off-nominal(s)</th>
<th>Reference Event(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>None, follow published STAR (KMEM) or SID (KDEN), and use airline standard operating procedures</td>
<td>None</td>
</tr>
<tr>
<td>2</td>
<td>1,3</td>
<td>(a,b) Failure within the pitot-static system due to icing/blockage of the pitot and/or static ports, and/or failure of the pitot heat system; airspeed and/or altitude become unreliable; (c,d) Failure of the Inertial Reference System (IRS), the Inertial Reference Unit (IRU), and/or the Attitude Heading Reference Unit (AHRU)</td>
<td>Air France (2009); Adam Air (2007); Midwest Express (2005); IcelandAir (Baltimore) (2002)</td>
</tr>
<tr>
<td>3</td>
<td>0,3</td>
<td>(a) Radar altimeter reports -8 ft at ~2000 ft and after glideslope captured; this causes A/T to go to ROLLOUT FLARE mode, returning throttles to idle position even if the pilot pushes them forward; (b) Aircraft in front slows to allow leading aircraft to clear runway; ATC asks owshp to slow to maintain in-trail spacing; (c) after turn-to-final, set commanded A/T setting to zero; (d) ATC issues vectors for an ILS approach, then “descend to 3000, slow to 180”, then after “Contact tower”, issues “Runway change to 18R” (farther parallel runway); after A/T retracts to idle, A/T disconnects</td>
<td>Turkish Air (2009); Thomsonfly (2007)</td>
</tr>
<tr>
<td>4</td>
<td>0,4</td>
<td>(a) NOTAM: “Glideslope Out of Service”; ATC issues clearances that sets crew up high (above glideslope), then clears them for visual approach; (b) ATC issues two “direct-tos” causing crew to be high and fast at 4000 ft, ATC tells them to expect runway 36L; shortly after turning onto final, ATC issues clearance to switch to the parallel (closer) runway; if crew doesn’t go-around by 300 ft, ATC calls for go-around, “fly runway heading, level at 2000 ft”; There is a tailwind of 4 knots on final approach</td>
<td>Asiana (2013); IcelandAir (Oslo) (2002)</td>
</tr>
<tr>
<td>5</td>
<td>0,1,2</td>
<td>(a) Tailwind encountered is significantly higher than forecast wind data in the FMS, causes unexpected VNAV transitions during downwind leg; (b) After passing 6000 ft, ATC issues “Level at 4000, hold at present position”, shortly after clearance, A/T disconnects without audible annunciation; (c) during an RNAV approach, ATC calls for unexpected go-around at low altitude (~300 ft AGL)</td>
<td>IcelandAir (Oslo) (2002)</td>
</tr>
<tr>
<td>6</td>
<td>1,2</td>
<td>(a) EICAS message – “HYD SYS PRESS (CENTER ONLY)” shortly after start; later at 3000 ft, ATC issues windshear advisory, “Low level wind shear, +/- 10 knots, 1000 ft on approach to Runway 18C”; (b) ANTI SKID warning on EICAS shortly after start; Use shorter runway, with runway conditions reported as ‘poor’</td>
<td>None</td>
</tr>
</tbody>
</table>
D. Test Variables and Procedures

The facility and environment were set up to fly the defined scenarios within the Memphis International Airport (KMEM) terminal airspace for arrivals, and at the Denver International Airport (KDEN) for departures. Independent variables were weather conditions (5); airport configuration (2); traffic conditions (3); flight path/procedure (9); and off-nominal conditions (17). As discussed previously, each scenario had unique settings for each of these variables to cover the desired span of conditions. The simulator was also set up to enable/disable the four new technology concepts so they could be evaluated individually or as an aggregate.

For arrivals, flights began after top-of-descent, but before 10,000 ft MSL, and on the published standard arrival (STAR) procedure. A mix of ILS and RNAV approaches were flown. Prior to the start of each flight, the crew conducted an approach briefing and assigned Pilot Flying (PF) and Pilot Monitoring (PM) roles. During flight, pilots were asked to follow normal procedures as if a revenue flight, with caveats discussed during training or induced by the facility’s differences from line operations. Pilots were instructed to go-around if unstable or if otherwise deemed necessary; and to continue flights until the researcher called “end of run.”

For departures, flights began while holding short of the active (departure) runway and awaiting takeoff clearance. While holding short, all relevant checklists were accomplished as well as briefing NOTAMs and weather conditions. Takeoff and climb-out followed the appropriate published standard instrument departure (SID) procedure.

After each flight, a short questionnaire was completed by each crewmember using a researcher-provided tablet computer. For flights involving new technology, this included the ten usability questions taken from [20]. Both quantitative and qualitative data was recorded for AIME to support post-hoc analysis. Data types include:

- Interview and questionnaire responses (both pilots)
- Head-/eye-tracking data (both pilots)
- Physiological measures (both pilots)
- Flight data (e.g., aircraft state, pilot button presses)
- Cockpit audio and video recordings
- Researcher and pilot observations

E. Summary of Completed Tests

Table II provides a summary of the number of flights completed by the 11 crews during the testing across the various display conditions and scenario types. For each scenario type, one goal was to balance tests using the baseline (BL) display condition versus the new technology conditions. Table III provides a summary of the number of technology usability assessments completed by the 22 pilots. Analysis of these assessments is provided in [10].

Figure 5 provides an overview of the profiles flown at KMEM during the testing. For brevity, extended flight segments when diverting are not shown. The departure segments from KDEN are also not shown.

![Image](image_url)

Fig. 5. Flight paths for approaches to KMEM
Lateral (top); Vertical (bottom)

Participants were all Air Transport Pilot (ATP)-qualified and currently flying for major carriers (i.e., United, Delta, and American Airlines). Crews were paired as Captain and FO from the same airline. Type ratings covered B-787, B-777, B-
767, B-757, B-747, B-737, A-330, A-321, A-320, and A-319. Flight experience ranged from 5,000 to 39,000 flight hours, with an average experience of 15,500 hours.

III. SELECTED FINDINGS

As a result of the testing, the research team confirmed many general hypotheses regarding the two relevant CAST-defined safety enhancements, the technologies under evaluation, and how best to proceed with the research. For example, each of the new technologies was deemed usable and useful; and efficiency when using the new technology was apparent (e.g., path and speed deviations, time to detect problems, auto-flight system interactions, determining whether approaches are stable). Workload and general situation awareness did not significantly change (except for the Type 2 scenarios), and usage of the new technologies corresponded well to the designer’s intent, albeit with only minimal time spent on training. Performance and comments regarding the baseline (BL) (787-like) displays and interfaces were highly correlated to each participant’s background and experience.

The following sub-sections provide selected evidence supporting these hypotheses and claims. Additional results can be found in [9][10]. In some areas, data analyses continue and will be published in future reports.

A. Predicting Auto-Flight System Trajectories, Mode Changes, and Energy-Related Problems

Two of the three desired research outputs given in Section I are addressed by the combined TP/MHP/PAE technology evaluated during AIME (SE-207 Output 3; and SE-208 Output 1a). Care was taken to make sure the pilot indicators associated with these technologies are simple, intuitive, and useful. This consisted of the five-minute look-ahead “green line” on the VSD/ND, the circle symbols on this line representing where mode changes or energy-related problems are predicted to occur, and the predictive text messages indicated on the EICAS. Usability results are detailed in [10], but as shown in Figure 6, an average System Usability Scale (SUS) rating [20] of 80.35 and 78.95 was given to the TP technology for the pilot-flying (PF) and pilot-monitoring (PM) roles, respectively. This is not unexpected as the TRL is relatively high due to extensive previous research, testing, and subsequent refinement.

In terms of performance, several metrics were considered for both pilot/system interactions and the underlying algorithms/automation. For interactions, these include: time to complete task, time to notice or detect context-relevant information, the number and type of interactions, adherence to flight path-related restrictions, and whether there were errors of omission or commission. For the algorithms and automation, these include path and speed profile tracking, predictive accuracy, false alarm and missed detection rates, and the number and type of system failures. Results of analyses of each of these measures will be published separately. For brevity, here we amalgamate many of these measures by discussing correspondence to established stable approach criteria.

Per [21], as well as airline-trained procedures and policies, the aircraft must be “stable” by 1000 ft in order to continue the approach. If not stable, a go-around must be initiated. Of the 102 approaches flown during AIME that proceeded to 1000 ft, 85 were found to be stable and resulted in a safe landing. Of the remaining 17 flights, nine were correctly identified by the crew as unstable and a go-around was executed. These flights were predominantly those involving a clearance to change runways late in the flight. After post-hoc analysis, eight flights were found to be unstable even though the landing was continued; each of these were in VMC and became stable no later than 800 ft. Although there is no direct correspondence between the number and type of BL flights to the number and type of BL+TP flights, the relative ratio of stable flights to total flights was found to be higher when the new technology was available (92% vs. 81%).

Several scenarios were designed in an attempt to induce energy-related problems. Of these, four included an unexpected manipulation of the auto-throttles (A/T) by the airplane: two were A/T disconnects, two forced the A/T to go to IDLE mode. Figures 7, 8, and 9 illustrate the main effect for three of these; the fourth is presented in Section III-C. Without A/T support (after the black circle symbol in the figures), airspeed generally decreases until at least one pilot notices, or a stall warning is issued. For the flights shown in Figure 7, the A/T disconnect occurred just after an ATC clearance to hold present position at ~4900 ft, and with airspeed at ~230 knots. In two flights, airspeed dropped more than 50 knots before either pilot reacted. In contrast, for the flights shown in Figure 8, the A/T disconnect occurs later in the descent when airspeed is ~180 knots, and ATC has requested a runway change. Here, the effect on airspeed varies significantly across the crews; but loss of airspeed due to not noticing is less significant. Finally, for the flights shown in Figure 9, the aircraft is very late in the approach and at a relatively slow speed already. Crews here are already concerned about being too slow, so airspeed dropping further does not occur. Most of these flights ended in go-around and pilots commented on loss of A/T support.

It should be noted that the new technology concepts tested here were not designed specifically to address such A/T mode changes. However, one such technology is the subject of research by one of our partners, and is one of a set planned for future evaluation in the AIME series [22].

Data such as above reinforces previous findings regarding the trend toward over-reliance on automation or autonomous

---

**Usability**

System Usability Scale, ≥ 80.3 = A

<table>
<thead>
<tr>
<th>Usability</th>
<th>TP</th>
<th>TP</th>
<th>SIS</th>
<th>SIS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>80.35</td>
<td>78.95</td>
<td>80.00</td>
<td>86.50</td>
</tr>
<tr>
<td>PF</td>
<td>A</td>
<td>A-</td>
<td>A</td>
<td>A+</td>
</tr>
<tr>
<td>PM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 6. Average usability scores for TP and SIS [10][20]
functions. When an unexpected system behavior occurs (e.g., A/T disconnect), pilots have come to expect to be informed in some way by the aircraft. If such notifications are not part of the design, or if they fail, the resulting change of system state will likely not be detected unless there is a significant secondary effect (e.g., stickshaker warning). These behaviors should not be considered pilot errors, but rather a predictable consequence of thousands of hours of flight experience where such off-nominal events have not occurred. Effectively combatting this phenomena must be done on three fronts: system design, procedure design, and training.

Fig. 7. A/T disconnect flights (scenario 5b)

Fig. 8. A/T disconnect flights (scenario 3d)

Fig. 9. A/T disconnect flights (scenario 3c)

B. Displaying the State of Flight-Critical Data Systems and the Effects of Loss of These Data

The SIS display, combined with a streamlined checklist, addresses SE-208, Output 1b, enabling more effective response to loss of flight-critical data during flight. Design of the graphical layout and display elements were derived from pilot training materials such that it would be both useful and familiar to pilots. The color scheme and style were aligned with those of the existing synoptics already on the aircraft (e.g., for hydraulics and fuel). As shown in Figure 6, average usability scores of 80.00% and 86.5% were given to the SIS display for the PF and PM roles, respectively. On the SUS scale [20], this rating corresponds to an ‘A’ and ‘A+’. This high grade was somewhat unexpected, as it is lower TRL with only minimal prior research and testing. The higher usability score given for the PM role is likely a reflection of the increased usage of the system by the PM.

Similar metrics to those given in Section III-A are relevant; however, the most important are probably time to complete task, time to notice or detect context-relevant information, and whether there were errors of omission or commission. For the underlying algorithms and automation, most do not apply as the graphic elements change based on relatively simple logic only (i.e., if a specific EICAS warning occurs, then change the SIS elements accordingly). False alarm and missed detection rates should be no worse than those of the EICAS itself.

Based on the testing, time-to-complete task was perhaps the most compelling measure. For example, for the loss of air data test condition shown in Figure 4, four EICAS warnings are triggered. Each has an associated checklist for the pilots to complete using the ECL system. This is representative of industry standard operating procedures. By referencing the new SIS graphic and associated information, the checklists for this message can be shortened – removing many of the textual “notes” given in the standard (i.e., BL) checklists. Figure 10 illustrates page #1 of the revised checklist for the “NAV AIR DATA SYS” condition, reminding pilots to consult the SIS for the information regarding affected systems. For this test condition, the SIS allowed for a reduction of the original four checklists by a total of 432 lines of text. Individual checklists were often reduced over 50% and were able to fit on one page. The combination of the pictorial information, indicating what systems were working and what had failed, combined with the brevity of text significantly reduced time to complete the procedure. In addition, there was much less confusion within the crews as to the effects of these types of failures, and much more effective discussion and decision-making with respect to appropriate actions to take (e.g., frequent pointing to the SIS graphic versus reading aloud lines of text on the ECL).

Fig. 10. Revised checklist with SIS note
Figure 11 illustrates one aspect of aircraft behavior after the loss of flight-critical data for the situations tested during AIME. In these situations, the crew was in a descending holding pattern. Just after the event occurred, flight path variations were significantly different with the SIS and ME technologies available (BL+SIS+ME) versus without these technologies (BL). First, there is less variability while trying to remain in the holding pattern (probably due to ME support). Second, less time is spent in the holding pattern while trying to work through the checklist, discuss divert options, contact ATC, then break from the pattern (probably due to SIS). Keep in mind that once the failure occurred, the PF was forced to hand-fly the aircraft for the remainder of the IMC flight. As a result, workload remained high for both crewmembers.

For the PM, ND attention increased by 5%, while EFB attention decreased by 4%. This latter is largely due to the reduced checklists enabled by the SIS technology, where the delta is much higher during the time spent responding to the air data system or IRS/IRU failures.

C. Applying Visual Attention

Previous research by our team has illustrated how demands for visual attention in the modern flight deck are significant, due largely to growing complexity [11][23]. Eye-tracking data collected during AIME reinforces this finding. Using the system employed for [23], an AIME-specific set of Areas of Interest (AOIs) and sub-AOIs were defined (Figure 12).

Figure 13 shows how attention was distributed across the major AOIs in the previous study [11][23] versus AIME. The display surfaces and AOIs were essentially the same for each, although some of the functions changed. For example, in the previous study, the EFBs were used primarily for ATC datalink communications (i.e., DataComm), weather information, and charts, while for AIME, they were used primarily for checklists and charts. The larger number of off-nominal situations in AIME that required checklist use accounts for the increased attention to the EFB. Likewise, PFD usage in AIME by the PF increased, primarily due to the off-nominals, which occurred in bad weather where there were no external visual references. Regardless, both studies show how many demands there are for visual attention, and why indicators can be missed, or misunderstood, even by a two-pilot crew. In the previous study Inattentional Blindness (IB) was observed regarding a traffic display; in AIME, IB seems to have manifest with the FMA (as is discussed below).

For the AIME testing, Figure 14 illustrates the overall effect of adding the new technology. For the PF, PFD and ND attention increased by 3%. EFB attention decreased by 2%.
The eye-tracker also provides insight into behavior following the radar altimeter failure condition utilized in one of the Type 3 scenarios. As with the reference CAST event (Turkish Air, 2009), the only indicators of this condition were the small numeric indicator of radar altitude near the center of the ADI, and the auto-flight system mode change indicated as “ROLLOUT FLARE” on the FMA (at the top of the PFD). If neither pilot looks at, or notices, one of these indicators, it’s very unlikely they will diagnose the problem and its effect (e.g., on the auto-throttles). Largely because none of the 22 pilots had ever experienced such a failure, reactions varied widely. Figure 15 illustrates some of the behaviors measured during AIME, with significant variations in airspeed and pitch as the pilots “fought” the auto-flight system, rather than immediately decoupling/disconnecting.

Interviews after these tests revealed that none of the pilots correctly diagnosed the source of the problem, yet all of them did recognize the odd behavior of the auto-flight system and manage to mitigate the effects (to various degrees) while executing a go-around.

With respect to indicator effectiveness, eye-tracking data suggests some pilots may have experienced IB when seeing the radar altitude indication and/or the FMA indication. Figure 16 shows the variability in time-to-notice and dwell time when crews first looked at the FMA after the failure. For five of the flights, the PMs did not look at the FMA for more than 18 seconds after the failure (and mode change) occurred. In general, PFs were more attentive to the FMA, with 8 of 11 looking at it in less than 5 seconds. The PF for flight #4 seemed to realize most effectively the odd auto-flight system behavior, looking at the FMA in less than 5 sec, and dwelling there for nearly 15 seconds.

Interestingly, for the flights shown here the crews did not mention the mode change in some cases, although the eye-tracker suggests they were looking at the FMA. Crew conversations following the failure were more along the lines of “what just happened?” and/or “why is it doing that?” Manual disconnect of the A/T and/or the A/P was the most effective way to mitigate the problem; this occurred in all of the flights, but with large variability in time to take this action.

D. Technology Readiness Levels

As stated in [3] and [4], one of the goals of the safety enhancement research is to advance TRL for technologies that show potential. In particular, a TRL of five or greater should be reached in order for effective transfer to manufacturers. Each of the four new technology concepts evaluated during AIME have elements that meet this threshold; while each also have elements that require further advancement.

Because of the prior research and development of TP and ME technology, maturity was already relatively high prior to the AIME evaluation (TRL 4 to 5). AIME advances these to the 5 to 6 range; with the remaining challenges to overcome:
(1) the performance of TP should be checked with the function either residing within a state-of-the-art FMS, or interfaced to such an FMS using the latest interface specification (i.e., includes the trajectory intent data bus); and (2) both TP and ME should be checked against additional realistic (and challenging) flight path management scenarios, where mode changes are known to be more frequent (for TP), or where maneuvering margins are reduced (for ME). Of course, each could also benefit from additional testing across different aircraft types; but this could occur at higher TRLs.

Although AIME was an initial evaluation of the SIS concept, it seems relatively simple in terms of technological complexity (i.e., high TRL can be achieved quickly); however, it’s more challenging implementation issues are associated with the effect on procedures and philosophy regarding the use of checklists, and the respective role synoptics and EICAS notifications can play. These types of issues are often referred to during advancement of Operational Readiness Levels.

While the MHP/PAE technology shows perhaps the most promise with respect to enabling crews to “stay ahead” of energy-related problems, the required predictive algorithms can be quite complex, particularly in the presence of uncertainty associated with required inputs. Research will continue during follow-on studies to raise its TRL (3 to 6).

Finally, although each of the four technologies was designed to be commensurate with relevant regulatory policies and standards (e.g., Advisory Circulars), TRL advancement must ensure this as the designs mature.

E. Pilot Observations and Suggestions

For each crew, there were several opportunities to collect qualitative comments and suggestions regarding both the technologies under evaluation and the test conditions. Following each flight, each pilot completed a short questionnaire and verbally recorded comments about the flight just completed. At the end of the two-day session, a semi-structured interview allowed for more general observations and suggestions to help guide the research going forward. Some of this feedback is paraphrased below, with more specific data given in [10]. A more comprehensive listing will be published in the future.

ME – Several pilots felt the bank angle indications and warning bands should ‘pop-up’ when within say 10 degrees of amber band; then turn off once back in nominal bank. Supporting this, some described cases where the band can overlay important information. Multiple pilots observed that this indicator could be very useful when the flight director is not functioning. Two pilots said the bank angle element was extraneous; speed bands are sufficient. There were a few concerns that it may confuse pilots since stall is more related to angle-of-attack and pitch angle, than it is to bank angle. Training would have to clearly define the true intent of this indicator vice other indicators. Most pilots commented it would be very useful at altitude, where maneuverability margins can be small.

TP – All pilots had generally positive comments about the function and indicators as aids that improve awareness of automation state and intentions. Knowing where the auto-flight system intends to take the aircraft and when/where mode changes will occur was deemed by all as a way to help avoid confusion that can sometimes occur. Several advised designers to be careful not to create clutter; and to look for ways to better distinguish, and train differences between the green line predictor and the white line predictor. Many felt it could be more useful for less experienced pilots who may not have seen some of the possible mode changes that can occur. Some commented specifically on the benefit of knowing ahead of time where the aircraft is predicted to intercept the localizer and glideslope. Most found it a useful cross-check to the FMA for awareness of automation states and state changes.

SIS – Pilots were unanimous in feeling the display and the associated reduced checklists make relevant tasks much less demanding. Most commented that providing the big picture of the situation at a glance, allowed for quickly realizing what systems are still available vs. those that are inoperative. Several appreciated its role in providing a back up to the EICAS messages for better understanding of malfunctions. Comments were along the lines of: EICAS only gives a title to the problem; SIS provides the rest of the story to understand why. All felt that SIS increased situation awareness and avoided the PM becoming “buried in” checklists. Specific comments from some described the difficulty that currently exists when working through lengthy checklists; SIS can help to mitigate this difficulty without compromising safety. Several commented they would like to see how it can be applied to additional, and more complex emergencies. Several also commented on the issue of using synoptics in an essential sense, something not done to date and hence may require changes to certification and/or procedure philosophies.

PAE – Because this technology was in an early developmental stage, its performance was not consistent across the flights, making it difficult for pilots to comment on utility for some of the test conditions. However, they did provide suggestions for future versions. Namely, the indicators and warnings should be more salient and tied to aural messages. The latter was in the design, but not tested in AIME. In general, pilots felt it was a good system concept, but judged usability low in its present state of development.

IV. Next Steps

Based on findings from AIME, and related efforts by research partners in industry, subsequent studies are planned over the next three years to achieve the full set of specified SE-207 and SE-208 research outputs. During this time, important metrics of success will be achieving the TRLs needed for transition to the fleet, and finding the most cost-effective solutions. Collaboration with the FAA is also planned, particularly in areas where new policies or standards may be beneficial or required. Beyond these safety enhancements, findings are also helping to define new NASA research plans regarding safety assurance and the application of new forms of autonomy to enable aviation transformation in the decades to come.
ACKNOWLEDGMENTS

The authors would like to thank Stefan Schuet and Kimberlee Shish of NASA Ames Research Center for their contributions to the design, implementation, and execution of this study. Thanks also to Tom Slotter and his team at Honeywell for their efforts to enable the trajectory intent interface bus on a legacy Pegasus FMS. Thanks to the Boeing Flight Services group for their help as we transitioned to the B-787-like environment. Thanks to Capt. Daniel Kiggins of Flight Services group for their help as we transitioned to the interface bus on a legacy Pegasus FMS. Thanks to Boeing Honeywell for their efforts to enable the trajectory intent this study. Thanks also to Tom Slotter and his team at American Airlines for his help with scenario development. Thanks to NASA LaRC’s simulation development team, led by Victoria Chung and James Barnes, for their creativity and dedication to completing a complex set of tests.

REFERENCES


