SCENARIOS FOR “RARE EVENT” SIMULATION AND FLIGHT TESTING

Richard L. Newman

TR-02-07A

September 2002

CREW SYSTEMS
Post Office Box 963
San Marcos, Texas

Summary

During testing of cockpit systems and displays, a frequent objective is to evaluate a system's ability to maintain the pilots "situation awareness." That maintenance of situation awareness (SA) is critical to flight safety is a truism. While SA is, no doubt, required for pilots and crew members to make correct operational decisions, it has been difficult to identify precise objective test criteria.

SA has been defined as correct perception of the current operational environment, comprehension of the effect of the environment on vehicle status, and the projection of vehicle status to the future. From an operational point-of-view, we need to determine what the effect is on operational decisions — What does the pilot do when confronted as circumstances develop. This means that many test scenarios involve "rare events." How do we develop experimental scenarios that will yield suitable data and will, at the same time, make efficient use of experimental facilities.

The approach taken is to compare experimental studies over the past several years with various human error models. This review shows four different experimental approaches to rare event test and evaluation. These match up well with the various situation awareness and human error taxonomies. Several proposed test plans are reviewed and changes in the scenarios made to match the appropriate human error model.

Contents

List of Tables

Scenarios for “Rare Event” Simulation and Flight Testing

Richard L. Newman

Background

During testing of cockpit systems and displays, we are often called upon to evaluate a systems ability to maintain the pilots “situation awareness.” That maintenance of situation awareness (SA) is critical to flight safety is a truism. While SA is, no doubt, required for pilots and crew members to make correct operational decisions, it has been difficult to find precise definitions for use in developing test criteria.

It is important to keep our discussion of SA in an operational setting with the aim of developing valid test criteria. To this end, *Situation Awareness* (SA) may be defined as correct perception of the current operational environment, comprehension of the effect of the environment on vehicle status, and the projection of vehicle status to the future.*

What we are really interested in, from an operational point-of-view is what the effect is on operational decisions — What does the pilot do when confronted as circumstances develop. Does the pilot make correct decisions? Does he have the correct information to make these decisions? Besco (2) described the attributes of good pilots by asking what should they do. A similar approach should be our criteria for developing SA test criteria — Are the decisions made by the pilot correct?

Loss of Situation Awareness (LOSA)

Situation awareness (SA) may be defined the correct perception of the environment, the current situation, and the projection of future status. Jones and Endsley (1) describe three levels of LOSA:

- ✈ Level 1: Fail to perceive information or mis-perception of information
- ✈ Level 2: Improper integration or comprehension of information
- ✈ Level 3: Incorrect projection of future actions of the system.

Perhaps the most fundamental form of LOSA happens when the pilot becomes unaware of his geographical position. In the extreme form, this can be described as “lost.” However, in a less extreme form, he may simply be unaware of the aircraft’s proximity to terrain. This can lead to controlled flight Into terrain (CFIT) in which the pilot files the airplane into the ground under complete control. The classic example of such an accident was the flight of a Boeing 727 into Mount Weather while on final approach to Dulles Airport in December 1974.(3)

* This definition was adapted from Jones and Endsley (1) who defined SA as “perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future.”

A loss in awareness of aircraft position relative to the track indicated by instrument flight procedures is also important. Coyle and Krolak (4) discuss the issues related to maintenance of geographical situation. In addition to citing the usual array of LOSA accidents, they indicate that 30% of initial instrument rating flight checks are failed, due to loss of situation awareness, mostly in while flying published procedures.

The pilot may have good terrain awareness, but not know his location relative to obstacles. This type of accident is referred to as "collision with obstacle."

Collisions with terrain or obstacles may also be a result of the pilot's inability to assess the aircraft's flight path relative to the obstacle. In these accidents, the pilot may be well aware of the presence of the obstacle (or terrain), but either believes the aircraft will clear the ridge or is unaware of aircraft performance.

An incorrect flight path assessment can also lead to a short of the runway. This is often true when adverse conditions make such assessment difficult. Typical adverse conditions include darkness, restrictions to visibility, or precipitation. Night circling approaches are particularly vulnerable. A recent MD-80 accident at LaGuardia (5) is an example of such an accident caused by incorrect flight path assessment. In incorrect performance assessment accidents, the pilot is aware of the terrain or obstacle, but simply misinterprets the aircraft trajectory relative to the obstruction.

Loss of geographic awareness can lead to increased pilot workload during diversions or other flight plan changes. If the pilot is confused to his geographical location or is not clear where his destination or proposed route is, he may make an incorrect decision for his initial heading or altitude. This was evident in the Boeing 757 accident near Cali in late 1995.(6) In this accident, the crew accepted a revised clearance to a different runway than originally planned. The increased workload during the reprogramming of the flight management system led to an incorrect turn away from the fix, rather than toward the fix.

Systems Monitoring

It can be difficult in many aircraft for the pilots to monitor their systems status during high workload portions of the flight. In particular failures of the flight instruments themselves may lead to misinterpretation and subsequent loss of control. In a 1974 accident, the crew inadvertently left the pitot heat off. When the pitot tubes iced over during the climb, the airspeed indications became erroneously high. Unfortunately the crew failed to detect the discrepancy and ultimately stalled the airplane resulting in a fatal crash.(7).

It can be very difficult to deal with failed or invalid flight instruments even if the pilot realizes that there is a problem.(4) In addition to detecting a failure, it is important for displays to provide sufficient cues to allow crewmembers to diagnose and correct system failures.(8)

Spatial Disorientation (SDO)

Spatial disorientation of pilots has been a persistent problem in aviation since the first flight into (and probable spiral dive out of) a cloud. It has long been known that pilots are unable to maintain straight and level flight without visual cues.(9) Benson (10) defines SDO as "an experience occurring inflight in which there is a defect in the aviator's perception of the attitude or position of his aircraft* or where conflicting perceptions give rise to confusion or uncertainty†."

* Equivalent to Jones and Endley's (1) Level 1 LOSA

† Could be either Level 1 or Level 2 LOSA

Test and Evaluation (T&E)

There are several elements of situation awareness that flight tests must be designed to evaluate. These elements can be divided into continuous events and rare events. Continuous events are those that happen throughout the flight, such as

- ✈ Monitoring position/trajectory relative to flight plan
- ✈ Assessing aircraft performance

Continuous event evaluation, particularly awareness of position/trajectory relative to complex flight paths is a common experimental scenario for SA evaluations. Several recent published results are available.(11, 12)

Rare events include occurrences that do not happen often, but have the potential for major accidents, such as

- ✈ Detecting proximity relative to terrain
- ✈ Detecting external threats (obstacles or traffic*)
- ✈ System failure detection

Experimentally, these three are more difficult to handle. The crux of the problem is "how to we test for pilot response to rare events?" Two issues are key: pilot recognition and pilot reaction. Many T&E studies have concentrated on pilot reaction (such as the typical unusual attitude recovery studies); fewer have examined the ability of pilots to recognize a situation. Even fewer have examined pilots' reaction to a "rare event."

In a way, studies of these rare events are similar to vigilance studies, where subjects are exposed to hours of boredom waiting for a singular event. As such, they are not very efficient in terms of time. Unlike vigilance studies, it can be very expensive to place pilots in a simulator for many trials until the "rare event" happens.

Let us examine how rare events in aviation operations have been studied. In the discussions that follow, we will concern ourselves with the scenarios. We are not intending to discuss the relative merits of the experiments or with the validity of the results.

System Failures

Pilot recognition to a system failure is usually conducted by asking the subject pilot to fly a particular pattern and subjecting them to some pre-determined failure. Normally, the subject pilot is not aware of the real purpose of the test, but is given a bogus reason.

* Detection of airborne traffic in visual conditions is a continuous event. Detection of airborne traffic during instrument conditions is a rare event. Runway incursions are rare events. This may partially explain differences in traffic detection between Fisher, Haines, and Price (13) and Anderson and Newman(14), *vide infra*.

Newman and Quam (15) examined general aviation pilot reaction to an unexpected vacuum gyro failure. The subject pilots were told the purpose of the flight experiment was to measure their tracking performance using conventional VOR navigation. The real purpose was to determine if they could recognize and cope with a vacuum system failure. At a predetermined (but unknown to the subject pilot) time, the safety pilot closed a valve shutting off the vacuum system and causing an attitude and directional gyro spin-down. In a post-experiment debriefing, the subject pilots reported that they were surprised by the failure and were not expecting it.

In a simulator experiment, Smith and Chappelow (16) initiated total engine failure to a group of helicopter pilots. The subject pilots were informed that the exercise was a training sortie. The simulator instructor initiated the failure at unannounced times and the recovery monitored. Smith and Chappelow measure reaction times and total recovery times, thus separating recognition from recovery. They did not mention a debriefing question relating to the surprise of the failure, but it is likely that pilots undergoing a simulator training exercise would expect engine failures.

In both cases, indeed in most failure detection experiments, only one valid data point per subject pilot is possible.*

Singer and Dekker (8) studied multiple failures presented to subject pilots during high workload instrument approaches. Their study was less concerned with failure detection than with the ability of the various displays tested to assist the pilots with diagnosis and corrective actions. The scenario was an instrument approach in a part-task simulator. Each subject pilot flew sixteen approaches with four different systems display suites. During each approach a systems failure was presented and the subject pilot tasked with taking appropriate corrective action.

Unusual Attitude Recognition and Recovery

Perhaps the most common “rare event” tested is the unusual attitude (UA) recognition and recovery. However, most, if not all, evaluations place the subject pilot in a series of UAs in a series of rapid-fire recognize and recover with no element of surprise. Simulator evaluations normally reset the initial conditions in one of a series of extreme attitudes and require the subject pilot to recover using standard techniques.

During flight evaluations, the evaluation display is blanked (or the subject pilot asked to close his eyes) and safety pilot flies the aircraft into a predetermined unusual attitude at

* Although in one study, pilots were tasked with a cross-country route in the western United States. During the flight, the subject pilots were given a radar vector toward a mountain range. The purpose was to determine if the subject pilots could detect the anomaly. Following this trial, the display was changed and the same task flown with the same radar vector. Finally, a third display was used and the identical route and radar vector flown! The results showed a high correlation of detecting the impending ground collision with the trial sequence.

which point the display is restored and the flight controls are given back to the evaluation pilot for recovery. Newman and Greeley (17) outline conventional test techniques for UA recovery.

Such evaluations are useful for comparing variations in flight displays for recovery, but do not measure the ability of the pilot to monitor aircraft attitude or recognize a rare upset. Gallimore, Patterson, Brannon, and Nalepka (18) modified the entry conditions by having the subject pilot fly as a wingman. During the run, the lead aircraft flew into an unusual attitude and was removed from the visual scene. At this point, the subject pilot was to recover from the unusual attitude.

Unusual Situations

Most reports of experiments dealing with unusual situations have embedded one or two rare events into another study. Often the evaluation of unusual events is the primary purpose for the study and, for example, tracking performance measurement a fall-out of the study. Sometimes the reverse is true.

Typical experiments, described below, embedded the rare event into another study. In other words, the subject pilot thought he was being examined for one purpose and was unaware of the real purpose of the experiment.

Runway incursions: In the late 1970s, Fisher *et al.* (13) examined the ability of pilots to detect runway incursions while using head-up displays. In this experiment, each subject pilot flew 19 ILS approaches for data in a simulator. Twelve of these approaches were flown using a HUD, seven without a HUD. During one of the HUD approaches and one of the no-HUD approaches the visual scene showed an airplane on the runway. Each pilot experienced the simulated incursion twice, once with a HUD and once without. There was a difference in subject pilot reaction with faster reaction times during the second exposure.

In Fischer *et al.*, the difference in reaction times between first and second encounter makes using more than one encounter highly questionable.

In a civil HUD certification, attention was paid to traffic detection. During this flight test (14), the pilot flying by reference to the HUD tended to spot traffic before the safety pilot (who had been assigned responsibility for looking for traffic). While looking for traffic is not a rare or unexpected event, there is apparently a difference in pilot behavior between simulated and actual flight.

HUD misalignment: Naish (19) conducted a HUD tracking experiment in a simulator. During one run, the runway and the HUD were purposely mis-aligned. All subjects ignored the HUD commands and flew visually to the runway.

During a flight evaluation of a HUD to aid in midair retrieval, a gyro error created an inadvertent misalignment between the HUD symbology and the real world. In this case, the subject pilot ignored the real world and flew by reference to the HUD causing a hazardous situation.(20) In a follow-on study, the gyro was purposely mis-aligned, creating similar situation. Again, the subject pilots flew using the display, not the real world.(21)

While there are HUD-related issues, the point for immediate discussion is the difference in pilot response when a misalignment is encountered in a simulator and in-flight. In the simulator, Naish's subject pilots ignored the HUD and followed the external cues. In flight, Newman's subject pilots ignored the external cues and followed the HUD cues.

In both cases, each subject pilot encountered the situation once during a series of several experimental runs.

Unexpected threats: There have been a couple of attempts to overcome the difficulty of only allowing one (or, at most, two) "events" per subject. Hofer, Braune, Boucek, and Pfaff (22) had twelve subject pilots fly a series of instrument takeoffs and instrument approaches. During each takeoff/approach sequence, each pilot experienced an "event" for a total of 16 "events" per pilot. The events were divided into display events (such as a frozen airspeed indication), a scene event (such as a runway incursion), or a display+scene event. Of the 192 events (for all pilots), 72 were serious enough to be considered a hazard. Overall, 36% of the HUD and 26% of the HDD events were missed.

More recently, McKay, Guirguis, Zhang, and Newman (23) evaluated enhanced vision systems to aid in pilot detection of "anomalies." Their anomalies were arranged in similar fashion to those of Hofer *et al.* Again, significant "anomalies" were missed, in spite of the pilots being warned that there would be one per approach.

Human Error Taxonomy

Situation Awareness Error

Endsley (24) describes situation awareness as the perception of the elements in the environment, their comprehension, and the projection of their status in the near future. She developed a model where the environment and objectives are fed through a situation awareness filter to lead to a decision. Table I summarizes the SA Error Taxonomy.

The recommended evaluation called for frequent interruptions with pilot perception of the elements, comprehension of the elements, and projection determined from specific questions.

Table I: Situation Awareness Error Taxonomy

Error Type	Description
Level 1 Failure to Correctly Perceive Element	
Data not available	Data is not available due to failure of the system design
Data hard to detect	Examples: inadequate lighting or resolution
Failure to observe data	Data not perceived due to omission, attentional limitations, distractions, etc.
Misinterpretation	Data misperceived because of prior expectations or misunderstood because of distraction
Cognitive failures	Caused by high workload
Level 2 Failure to Comprehend Elements	
Poor mental model	Poor mental model does not allow combining for information needed to meet goals
Incorrect mental model	Leads to incorrect assessment
Over reliance on default model	Routine expectation of the system is assumed
Other	
Level 3 Failure to Project Future State	
Poor mental model	Poor mental model does not allow for projection into future state
Over projection of current trends	Projection further into future than current data warrants
Other	

Cognitive Failures

Rasmussen (25) developed an approach for diagnosing cognitive failures. He considered information errors, diagnostic errors, goal errors, strategy errors. Rassmussen's taxonomy is shown in Table II

Table II: Internal Human Error Taxonomy

Error Type	Description
Information error	Did not detect cues
Diagnostic error	Did not accurately diagnose system state
Goal error	Did not choose reasonable goal
Strategy selection error	Did not choose appropriate strategy
Procedure error	Did not select procedures
Action error	Did not execute procedures as intended

Unsafe Acts

Reason (26) classifies human action into three levels: skill-based, rule-based, and knowledge-based actions. Skill-based actions are routinely practiced and automatic, such as keeping a car in its lane. Rule-based actions are a combination of conscious and unconscious processes, such as routine takeoffs and landings. Knowledge-based actions involve cognitive conscious thought to accomplish a process. Within each level, errors are further categorized.

Table III: Unsafe Acts

Human Action	Error or Violation Type	Description
Skill-based action	Attentional slip	failure to monitor progress
	Memory lapse	omission of steps from plan
	Perceptual error (Note a)	failure to recognize situation
	Routine violation	actions that cut corners
Rule-based action	Misapplication of good rule	overlooking some aspect of rule, such as misapplication of brakes on icy surface
	Application of bad rule	use of incorrect rule, such as failure to wear seat belt
	situational violation	action taken intentionally to save time.
Knowledge-based action	Mistake	error resulting from inaccurate mental model, etc.
	Exceptional violation	situation not covered by training or procedure
Note a	Maurino, Reason, Johnston, and Lee (27)	

Approach to "Rare Event" Evaluations

What do all of these evaluations have in common? The experiments were based on operational tasks. The tasks were generally aimed at recognition, not on procedures following recognition.* From this generalization, we can develop a generalized approach to situation awareness studies.

Operational Tasks and Scenarios

During the development of flight displays, mission task elements (MTEs) should be flown first. MTEs are stylized maneuvers that are similar to maneuvers used during actual missions. MTEs include instrument tasks, such as vertical S's, Yankee patterns, and instrument tracking, and visual tasks such as air-to-ground visual tracking. These maneuvers are designed to ensure that the various building blocks of maneuvers used during operations fit with the display and aircraft characteristics. MTEs are primarily concerned with aircraft/display dynamics.

Mission related tasks put these building blocks together to form sequences of tasks that will be flown in an operational setting. This increases the cognitive demands on the pilot. The assessment of these increased cognitive demands is a major objective of these evaluations.†

Generally, these tasks will be developed from the intended mission for the aircraft/display being tested. It is not possible to list all possible mission related tasks for all aircraft.

During the course of these evaluations, embedded tasks may be included at unexpected (to the evaluation pilot) times. These embedded tasks will include engine and system failures, unplanned diversions and threats. Sufficient numbers of such embedded tasks should be included to allow for some statistical validity. The need for statistical significance, however, must be weighed against creating an artificial number of "rare events" with the effect of reducing the realism of the environment.

Develop Test Objectives

The test objectives should reflect one of the Human Error Taxonomies (*vide supra*). For example, if the SA Taxonomy is used in developing the display model, the display might be designed to assist with Level 1 errors (failure to perceive). In this case, the experiment

* Most UA recoveries and some system failures (8) experiments are exceptions to this generalization

† To distinguish between MTEs and mission tasks, consider instrument approaches. Flying a simple instrument approach from intercept to tracking down to minimums would be considered an MTE. This task is primarily concerned with aircraft/display interactions. If the task includes self-navigation to the final approach and requires a cognitive decision to land or go-around, then it becomes a mission task.

would examine the ability of the pilot to detect, for example, external threats. The experiment would concentrate on detection, less on decisions following detection.

If, on the other hand, the display were designed to assist with Level 2 errors (failure to asses), the experiment would concentrate on the decision following detection. Similarly, displays addressing Level 3 errors would concentrate on the actions following the detection/decision.

Develop Objective Test Criteria

Objective test metrics should be developed, if possible. Such metrics could include reaction time, percent correct decisions, etc. Sometimes, only quasi-objective data can be obtained, such as Cooper-Harper ratings (28) or SAGAT responses (29).

Sample Experimental Scenarios

Two sample scenarios were obtained from the NASA Ames Website. They are attached as appendices. Additional scenarios were obtained from published sources or experiments known to the author. The following discussions do not deal with equipment suitability or experimental design, but concentrate on the "appropriateness" of the scenarios.

We will use Endsley's (24) situation awareness (SA) taxonomy for consistency. Either Rasmussen's (25) or Reason's (26) taxonomies could be used, although, they are probably more appropriate for studying the ability of pilots to use displays to recover from situations.

Commercial CFIT Scenario

Tasks & Scenarios: This scenario (See Appendix A) involves a non-scheduled flight from Montreal to a fictitious Caribbean island. The scenario is suggestive of the Korean Air Lines accident at Gaum.(30) The scenario description suggests that a number of system failures can be introduced to ensure a CFIT situation.

The proposed scenario is certainly appropriate and would be effective in evaluating enhanced/synthetic vision (E/SV) and other systems designed to prevent CFIT accidents. The crew workload is kept high by a combination of time pressure and poor weather. Fatigue is likely to be a factor. However, it is not clear how crew fatigue can be kept high without requiring an excessive non productive simulator flight time.

Such a scenario can only be used once since surprise is a major factor. Further, to be completely realistic, the approach should be flown once to replicate non-familiarity with the approach and terrain.

Test Objectives: The scenario does not state the desired test objectives. Detection of hazardous terrain is implied, but not clearly stated.

Objective Test Criteria: The scenario description does not state test criteria.

Overall Assessment: The proposed scenario appears to be suitable for CFIT related studies. However, we would recommend that flying the entire flight would be an inefficient use of resources (simulator, experimenters, and subjects). We would recommend using the terminal area operations only as part of a series of approaches.

Human Performance Modeling - I

Tasks & Scenarios: This proposed scenario (Appendix B) uses a series of approaches to the same airport/runway, but adds complications to the approaches. Some approaches have weather complications, some ATC complications, and some terrain mismatch. The repeated approaches, most with some non-standard event, eliminate the surprise factor.

Such an approach is quite suitable for studying the effect of detection of events or Level 1 LOSA* in Endsley's (24) taxonomy.

Test Objectives: However, this scenario is more suited to evaluating the display's ability to assist the pilot in determining "escape maneuvers" -- i. e. for Levels 2 and 3 LOSA.* The ATC complications could also be used to evaluate different guidance displays, such as highway-in-the-sky, for workload reduction or decision making assistance.

This scenario could allow for a few embedded "threats," such as non-charted towers or runway incursions to be added to the simulation. Such threats would evaluate perception of threats (Level 1 SA) and, if kept to a low frequency of trials, should not interfere with an experiment to evaluate the decision-making characteristics of the display.

Objective Test Criteria: The scenario description lists a large number of objective and subjective data, but does not state test criteria.

Overall Assessment: The proposed scenario appears to be suitable for E/SV related studies. We would recommend adding threats and near-threats.†

Human Performance Modeling - II

Tasks & Scenarios: Keller and Leiden (31) describe approach scenarios in very general terms. In fact, the report reads like a primer for terminal area operations. They describe the tasks required during an automatic RNAV/FMS approach to Oakland, California Airport. The task descriptions include transition from automatic to manual flight and monitoring tasks aircraft flight path, aircraft systems, and air traffic situation.

Keller and Leiden list a number of problems and errors. These errors include errors in setting RNAV/FMS parameters (waypoint programming, altitude constraints). Surprisingly, they do not discuss internal RNAV/FMS failures, such as failure to capture either the new course or altitude.

There also discuss "generic" errors, which are general situation awareness errors. Many of these generic errors are based on pilot techniques (speed control, flight path control). This is surprising since the previous discussion concerns automatic (i. e. coupled) approaches.

Test Objectives: The objectives are not clear. The stated objective is to model the human performance model to use in future SV experiments. It is not clear how to relate the approach description to SV evaluation.

Objective Test Criteria: The report does not state test criteria.

* See table I on page 8

† Near threats could include aircraft holding short of the runway or uncharted towers off to the side of final.

Overall Assessment: The scenario described is a generic coupled-RNAV instrument approach. It is difficult to assess the suitability since the goals are not adequately defined.

Highway-in-the-Sky Scenario - I

Tasks & Scenarios: In this scenario, reported by Williams (12), the subject pilots flew a typical closed traffic pattern using one of several advanced displays, including a highway-in-the-sky (HITS). During three of six patterns, airplanes flew within the field-of-view (FOV) and the subject pilot instructed to make a radio call identifying the aircraft. These aircraft appeared during course intercept (high workload) and during straight course tracking (low workload).

The runs were terminated in midair by blanking the displays. At the end of two of six patterns, the subject pilots were asked to estimate airspeed, heading, and altitude.

Test Objectives: Williams states three types of SA as applying to HITS. The first is knowledge of position of the aircraft's flight path relative to the intended flight path. This would be a continuous event.

The second type of SA is the ability to detect other aircraft. This is a "rare event," albeit a rare event of relatively high probability. The third type of SA, according to Williams, is knowledge of secondary information, such as airspeed, altitude, and heading.

All three types of SA are Level 1 in scope, although the flight path information could be considered Level 2 or 3.

Objective Test Criteria: The experiment used latency (reaction time) as the test criteria for traffic detection. A SAGAT questionnaire was used to determine perception of aircraft state.

Overall Assessment: The scenario chosen is marginally appropriate for the study. There are two difficulties. First, the flight path was fairly simple, consisting of a standard traffic pattern, not some of the more complicated flight paths proposed by other experimenters. Second, the workload was fairly low. Including marginal weather would have increased the workload and could have been used as an additional independent variable.

Having half of the runs with "rare events" is suitable for detection of other aircraft in flight. Such a proportion would not be appropriate for normal rare events, such as runway incursions or uncharted hazards.

Highway-in-the-Sky Scenario II

Tasks & Scenarios: This scenario was used by the NASA Synthetic Vision Program at Dallas-Fort Worth Airport (KDFW) in October 2000. The subject pilots flew approaches using various synthetic vision concepts.* Most of the "data" taken was in the form of preference scales. However, one approach per pilot involved a sidestep maneuver from one runway to another. The subject pilot was briefed on this maneuver during the pre-experiment briefing.

Test Objectives: This particular maneuver was intended to evaluate the ability of the display to provide information to aid a pilot during a sidestep -- Level 3 SA according to Endsley (24). The question of surprise or detection was not an issue.

Objective Test Criteria: The experiment used pilot preference as the criteria. No objective test criteria were employed.

Overall Assessment: In this regard, the sidestep scenario was an appropriate maneuver being operationally significant and well fitted to the experimental objective.

Vector Toward Mountains

Tasks & Scenarios: In one recent, as yet unpublished study, pilots were tasked with a cross-country route in the western United States. During the flight, the subject pilots were given a radar vector toward a mountain range. The purpose was to determine if the subject pilots could detect the anomaly. Three identical experimental runs were made with same route and the same radar vector. (See footnote on page 5.)

Test Objectives: Flight toward an unknown mountain range is a rare event worth evaluating The type of SA being evaluated is Level 1 SA -- dealing with perception of the hazard.. Since detection of rare events is strongly correlated with the number of times the subject pilot has seen the event in the current scenario, we would question the validity of repeating the identical event three times over.

Objective Test Criteria: The experiment used latency (reaction time) as the test criteria.

Overall Assessment: The scenario is a realistic one. However the rare event should be modified to present different threats with each run. we would recommend using a threat in one run (hitting a ridge) and a near-threat in another (i. e., clearing the ridge by a few hundred feet) To maximize data obtained, all runs should fly in relatively close proximity to terrain with some runs well clear of terrain/obstacles, some close to terrain/obstacles, and some guided into terrain/obstacles. Both vertical and horizontal proximity to threats should be used.

Multiple Systems Failures

Tasks & Scenarios: Singer and Dekker (8) studied multiple failures presented to subject pilots during high workload instrument approaches. Their objective was to determine abil-

* Newman, RL, Unpublished debriefing notes from NASA Synthetic Vision Tests at DFW Airport, October 2000

ity of the displays tested to assist the pilots with diagnosis and corrective actions. The scenario was an instrument approach in a part-task simulator. During each approach a systems failure was presented and the subject pilot tasked with taking appropriate corrective action while maintaining a stabilized instrument approach.

Test Objectives: The objective matches with Endsley's (24) Level 3 SA -- the ability to project to future times. As such, the experiment is less concerned with detection of the failure than with the corrective action. The scenario chosen was entirely appropriate for this study and maximized the data taken with respect to the equipment and personnel utilization.

It might be argued that providing one approach with no failure would provide a baseline. This, however, assumes that one of the results of the experiment would be ILS tracking performance with multiple failures. This was not a stated objective, nor is it useful data.

Objective Test Criteria: The criteria used was measuring the latency (reaction time) following the failure. In addition, Singer and Dekker verified the accuracy of the pilots' assessments using a questionnaire.

Overall Assessment: The overall scenario was realistic and appropriate to the test objective.

Recommendations

General Recommendations

Several recommendations for creating experimental scenarios for display and pilot-vehicle interface studies can be made. These include:

- ✈ Develop Operational Scenarios
- ✈ Model Human Error
- ✈ Develop Test Objectives
- ✈ Develop Objective Test Criteria

Develop Operational Scenarios: First, it is important to develop operationally relevant experimental scenarios. These scenarios should reflect the intended operational use of the equipment under consideration

For novel navigation and SA displays, such as highway/tunnel-in-the-sky displays, the scenarios will likely be some navigation route involving convoluted routes similar to those described above, although not necessarily at low altitudes.

Traffic displays, such as TCAS or CDTI, will likely use high traffic density terminal area navigation. Other applications will need to develop appropriate scenarios.

Model Human Error: The likely human-involved errors must be modeled and will, again, depend on the specific application. Endsley's (24) or Rasmussen's (25) taxonomies are examples of how to characterize the issues. For example, Endsley lists perception, comprehension, and prediction as the key elements. Rasmussen lists detection, diagnosis, goal setting, strategy, procedure, and action as the elements. The experimental design must reflect the issues taxonomy chosen by the experimenter.

Develop Test Objectives: The test objectives should follow directly from the human error model used coupled with the intended use.

Develop Objective Test Criteria: It is important to set objective metrics to the extent possible, such as reaction time, accuracy of decisions, accuracy of tracking, etc.

Situation Awareness Testing: The chapter on "Situation Awareness Testing" from Reference (17) is attached as Appendix C.

Enhanced/Synthetic Vision System (E/SVS) Testing

Scenarios: For enhanced/synthetic vision systems, the scenarios should be some low-level phase of flight. Almost certainly these would be terminal area and airport operations (terminal navigation, approach-and-landing, take-off-and-departure, etc.)

Human Error Model: For E/SV applications, the initial human error model will likely involve detection and recognition of the external object or threat. Once the detection and recognition issues are resolved, the decision-making issues follow, then the procedural issues.

Test Objectives: The test objectives follow directly. For target detection, appropriate threats and near-threats should be used. McKay *et al.* (23) used a variety of threats involving uncharted towers, and runway incursions. Other threats could include CFIT and near-CFIT encounters. These could be produced by uncharted terrain, experimenter-induced altimeter errors, or experimenter induced navigation blunders.

A decision must be made whether the singular "rare event" should be used with a resulting inefficient use of resources or whether the experiment should use a variety of "events" some of which are of interest and some of which are present merely to keep the subject looking. McKay *et al.* used the following anomalies:

- ✈ Below landing minima
- ✈ Hazardous obstruction
- ✈ Non-hazardous obstruction
- ✈ Inappropriate bank indication
- ✈ Inappropriate heading indication
- ✈ Runway incursion
- ✈ Near runway incursion
- ✈ Sensor boresighting error
- ✈ Synthetic vision misalignment
- ✈ Nothing

Some of these anomalies were scene-related, others were internal aircraft related. Some were hazardous, some were not. The pilot was never sure where to look. This approach should allow for discrimination between E/SV alternatives. Hofer *et al.* (22) used a similar set.

Test Criteria: Typical objective metrics include reaction time following event/object appearance, accuracy of assessment as an hazardous or a non-hazardous event.

Advanced Navigation Displays

Scenarios: For novel guidance displays, such as highway-in-the-sky, the scenarios should involve high traffic density terminal areas with complex patterns. The phases of operations should concentrate on terminal area and airport operations (terminal navigation, approach-and-landing, take-off-and-departure, etc.), although enroute operations may need to be considered in follow-on testing

Human Error Model: For E/SV applications, the initial human error model will likely involve procedural issues and the ability of the pilot to comprehend the current situation. Follow-on studies may examine the pilot's ability to detect unsafe situations, such as navigation blunders or loss of terrain separation.

Test Objectives: The test objectives follow directly. Procedural issues predominate. Some complicated pathways reconfigure the displayed flight path depending on, among other things, engine failure or aircraft maneuvering. One must measure the pilot's ability to keep track of these. It may be desirable for the experimenter to introduce navigation "blunders" to determine if the pilot can detect these. Diversions should also be introduced. If CDTI is included, stray aircraft can be introduced.

Test Criteria: Typical objective metrics include reaction time following event/object appearance or flight technical error, particularly following turns. SAGAT (24) probes may be beneficial to determine if the subject truly can comprehend the overall situation.

Use of Blunders or System Errors

One complaint about highway-in-the sky displays is that they are too compelling. Some pilots* feel that it will be very difficult for pilots to detect blunders. We agree that pilots' ability to detect blunders should be evaluated. Since it is unlikely that subjects will create blunders during an experimental session, the experimenter should introduce blunders, such as mis-programmed routes or incorrect altimeter settings. Since blunders are normally accomplished unintentionally, the subject should be neither more nor less likely to detect them.

* Newman, RL, Unpublished debriefing notes from NASA Synthetic Vision Tests at DFW Airport, October 2000

Abbreviations

ACRC	Aeronautica Civil of the Republic of Columbia
ATC	Air Traffic Control
CDTI	Cockpit Display of Traffic Information
CFIT	Controlled Flight into Terrain
E/SV(S)	Enhanced/Synthetic Vision (System)
FMS	Flight Management System
FOV	Field-of-View
HDD	Head-Down Display
HUD	Head-Up Display
ILS	Instrument Landing System
IMC	Instrument Meteorological Conditions
KDFW	Dallas-Fort Worth Airport
LOSA	Loss of Situation Awareness
MTE	Mission Task Element
nm	Nautical Miles
NTSB	National Transportation Safety Board
RNAV	Area Navigation
SA	Situation Awareness
SAGAT	Situation Awareness Global Assessment Technique (24)
SDO	Spatial Disorientation
T&E	Test and Evaluation
TCAS	Traffic Alert and Collision Avoidance System
UA	Unusual Attitude
USAF	United States Air Force
VHF	Very High Frequency
VOR	VHF Omnirange

Definition

Blunder An unintentional mis-programming of flight plan information in an onboard navigation system.

Situation: Awareness: Perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future.(1)

References

1 Jones, DG. and Endsley, MR, 1996, "Sources of Situation Awareness Errors in Aviation," Aviation, Space, and Environmental Medicine: 67, 507-512

2 Besco, RO, 1996, "What All Good Pilots Do: Does Anyone Know Or Care?", Business and Commercial Aviation, November 1996, pp. 108-114

3 NTSB, 1975, Trans World Airlines, Inc., Boeing 727-231, N54328, Berryville, Virginia, December 1, 1974, National Transportation Safety Board AAR-75-16

4 Coyle, S and Krolak, W, 1999, Common Sense Navigation for the 21st Century, Transport Canada White Paper

5 NTSB, 1997, Descent Below Visual Glidepath and Collision with Terrain: Delta Air Lines Flight 554, MD-88, N914DL, LaGuardia Airport, New York, October 19, 1996, National Transportation Safety Board AAR-97/03

6 ACRC, 1997, Controlled Flight Into Terrain: American Airlines Flight 965, Boeing 757-223, Near Cali, Columbia, December 20, 1995, Aeronautica Civil of the Republic of Columbia, Bogota

7 NTSB, 1975, NWA B-727-251, N274US, Near Thiells, New York, December 1, 1974, National Transportation Safety Board AAR-75-13; Ironically, this accident occurred on the same day as another situation awareness accident.(3)

8 Singer, G and Dekker, SWA, 2000, "Pilot Performance During Multiple Failures: An Empirical Study of Different Warning Systems," Transportation Human Factors: 2, 63-76

9 Ocker, WC and Crane, CJ, 1932, Blind Flight in Theory and Practice, San Antonio: Naylor Publishing

10 Benson, AJ, 1965, "Spatial Disorientation in Flight," in Aviation Physiology, Gillies, JA (ed.), London: Pergamon Press, pp. 1086-1129

11 Snow, MP and Reising, JM, 1999, "Effect of Pathway-in-the-Sky and Synthetic Terrain Imagery on Situation Awareness in a Simulated Low-Level Ingress Scenario," Proceedings 4th Annual Symposium on Situational Awareness in the Tactical Air Environment, Piney Point, pp. 198-207

12 Williams, KW, 2000, Impact of Aviation Highway-in-the-Sky Displays on Pilot Situation Awareness, Federal Aviation Administration DOT/FAA/AM-00/31

13 Fisher, E, Haines, RF, and Price, TA, 1980, Cognitive Issues in Head-Up Displays, National Aeronautics and Space Administration TP-1711

14 Anderson. M and Newman, RL, 1994, "HUD Flight Testing: Lessons Learned," Presented at Southeast Section Symposium, Society of Experimental Test Pilots, Stone Mountain, Georgia, May 1994

15 Newman, RL and Quam, DL, 1982, "Pilot Reaction to Attitude Gyro Failure," Canadian Aeronautics and Space Journal: 28, 303-310

16 PR Smith and Chappelow, JW, 1995, "Pilot Intervention Times," in Human Factors in Aviation Operations, Fuller, R, Johnston, N, and McDonald, N (eds.), Aldershot, UK: Avebury

17 Newman, RL and Greeley, KW, 2001, Cockpit Displays: Test and Evaluation, Aldershot, UK: Avebury

18 Gallimore, JJ, Patterson, FR, Brannon, NG, and Nalepka, JP, 2000, "The Opto-Kinetic Cervical Reflex During Formation Flight," Aviation, Space, and Environmental Medicine: 71, 812-821
19 Naish, JM, 1964, "Combination of Information in Superimposed Visual Fields," Nature: 202, 641-646
20 Newman, RL, 1977, CH-3E (MARS) Head-Up Display Evaluation, Crew Systems TR-77-2, 1977
21 Ross, RE 1976, Head-Up Display Evaluation, 6594th Test Group LTR-76-8
22 Hofer, EF, Braune, RJ, Boucek, GP, and Pfaff, TA, 2000, Attention Switching Between Near and Far Domains: An Exploratory Study of Pilots' Attention Switching with Head-Up and Head-Down Tactical Displays in Simulated Flight Displays, Boeing Report D6-36668
23 McKay, D, Guirguis, M, Zhang, R, and Newman, RL, 2002, Evaluation of EVS for Approach Hazard Detection, Presented at NATO RTO Systems Concepts and Integration and Sensors Electronics and Technology Symposium, Enhanced and Synthetic Vision Systems, Ottawa, Canada, September 2002
24 Endsley, MR, 1988, "Design and Evaluation for Situation Enhancement," Proceedings of the Human Factors Society 32nd Annual Meeting, pp. 97-101
25 Rasmussen, J, 1982, "Human Errors: A Taxonomy for Describing Human Malfunctions in Industrial Installations," Journal of Occupational Accidents: 4, 311-335
26 Reason, J, 1990, Human Error, New York: Cambridge Press
27 Maurino, DE, Reason, J, Johnston, N, and Lee, RB, 1995, Beyond Aviation Human Factors, Aldershot, UK: Ashgate
28 Cooper, GE, and Harper, RP, 1969, The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities, NASA TN-D-5153
29 Endsley, MR, 1988, "Situation Awareness Global Assessment Technique (SAGAT)," Proceedings of the National Aerospace and Electronics Conference, pp. 789-795
30 NTSB, 2000, Aircraft Accident Report: Controlled Flight Into Terrain, KAF 801, Boeing 747-300, HL7468, Nimitz Hill, Gaum, August 6, 1997, National Transportation Safety Board AAR-00-01
31 Keller, J, and Leiden, K, 2002, Information to Support the Human Performance Modeling of a B757 Flight Crew During Approach and Landing: RNAV, Micro Analysis & Design Report, June 2002

Appendix A

Human Error Modeling-Commercial CFIT Scenario

Appendix B

Human Performance Modeling-Simulation Specification

14 Situation Awareness Testing*

This Chapter describes several flight test techniques for assessing the situation awareness aspects of cockpit flight displays. Background for these discussions is presented in Chapter 6 (page 53).

Direct Evaluations

Attitude Awareness

This flight test complements the standard unusual attitude recovery where the pilot is placed in a UA and tasked to recover. This test is designed to evaluate the ability of the display symbology to convey spatial orientation awareness to the pilot.

a **Objectives**: The objectives are to evaluate the ability of the display and symbology to convey spatial orientation awareness to the pilot and to evaluate the display to provide cues to allow pilots to recognize and recover from inadvertent UAs.

b **Requirements**: These test should be flown in ground based simulators in appropriate aircraft configurations. These tests should be performed in all display modes, including declutter levels.

c **Action**: The evaluation pilot shall be tasked to follow another aircraft in trail formation. The lead aircraft or symbol will fly a series of maneuvers within the capability of the aircraft under evaluation similar to the method of Gallimore *et al.* (*114*)
The evaluation pilot will be instructed to follow the aircraft, but not to exceed normal limitations in bank and pitch angle, minimum and maximum airspeed, and minimum altitude. These values should be clearly defined. The lead aircraft will, during the course of the test, approach these limits without exceeding them as well as fly beyond them.
When any limit is exceeded, the pilot will be instructed to make a verbal call and recover to straight and level flight using standard procedures from instrument flight manuals (*139-140*) or other published standard. Following recovery, she shall complete displays ratings (DRR and DFR).
If the evaluation pilots fails to note penetration of the envelope limits, the lead aircraft will be removed from the scene after two seconds forcing her to initiate an unusual attitude recovery.
Several replications of these flight envelope limit data points will be required.

d **Data**: Aircraft performance data should be recorded as functions of time:

* from Newman and Greeley, Cockpit Displays: Test and Evaluation, Aldershot, UK: Ashgate, 2001, Reference (17), © Richard L Newman and Kevin W Greeley, 2001

Airspeed
Altitude
Pitch and bank angles.

Target acquisition and tracking accuracy shall be recorded, preferably using visual image recording. Pilot ratings (DRRs and DFRs) should be hand recorded. Voice recording may be used to supplement the hand recorded data.

e **Performance criteria**: Performance criteria are the same as unusual attitude recoveries.

i *Desired performance*:

- Abandoned pursuit at aircraft limitations (±5 knot, ±10 deg bank tolerance).
- Abandoned pursuit at altitude floor (±100 ft tolerance).
- Reaction time to initial correct control input < 1.5 sec.
- No initial control reversals.
- Initial control input in accordance with (IAW) published standards. Published standards include instrument flight manuals (*139-140*) aircraft-specific manuals, or equivalent documents.
- No overshoot on recovery.

ii *Adequate performance*:

- Abandoned pursuit at aircraft limitations (±10 knot, ±20 deg bank tolerance).
- Abandoned pursuit at altitude floor (±500 ft tolerance).
- Reaction time to initial correct control input < 2.0 sec.
- Single initial control reversals.
- Initial control input IAW published standards. Published standards include instrument flight manuals (*139-140*) aircraft-specific manuals, or equivalent documents.
- Single overshoot on recovery.

f **Safety considerations**: These tests are flown in the simulator because of the hazards involved with aggressive formation flight in IMC.

Embedded Evaluations

The following are proposed situation awareness (SA) test suitable for display evaluation. In some cases, these have been used in previous programs. There are a number of SA evaluation methods used by psychologists, most of which seek subjective assessment of SA.(*88*) The tests proposed here are based on the presumption that the best measure of SA is the effect on performance. SAP-like assessments are used as well.

Failure Detection:

a **Objectives**: These tests are designed to verify that the pilot can recognize critical failures and take appropriate corrective action while maintaining aircraft control.

b **Requirements**: These test shall be flown in conjunction with other evaluations, particularly the mission task testing.

Navigation, sensor, and display system failures shall be introduced during mission related tasks to determine the reaction time and responses of the pilot. Appropriate procedures for operating the aircraft following these failures shall be established prior to conducting these tests.

It may be required to use a special software version to allow the flight test engineer to introduce some system failures such as simulating attitude gyro failures. If such software versions are used, the response of the system to the simulated failure shall be verified prior to conducting these tests.

Typical failures (and representative simulation means) should include

- Radar altitude failure (pull circuit breaker, C/B)
- Navigation failure (pull C/B)
- Attitude gyro failure (software simulation of both hardover and slow failures)
- Flight director failure (software simulation of both hardover and slow failures)
- Compass failure (pull C/B or software simulation)
- Airspeed or altitude transducer (pull C/B or equivalent)
- Head tracker failure (pull C/B or software simulation).

c **Action**: For tests involving operations in IMC, the evaluation pilot shall be hooded or otherwise prevented from using outside visual cues. System failures shall be introduced during operations at unexpected times by the safety pilot or flight test engineer.

Upon recognition of the failure, the evaluation pilot shall take appropriate action as described in the aircraft/system operating procedures. The evaluation pilot will be briefed to make a verbal call upon recognition of all system failures as a means of determining detection time. Following recovery, the evaluation pilot shall complete displays ratings (DRR and DFR).

d **Data**: Aircraft performance data should be recorded as functions of time:

Airspeed
Altitude
Pitch and bank angles
Control positions
Parameters appropriate to failure
Event (for start of maneuver).

Pilot ratings (DRRs and DFRs) should be hand recorded. Voice recording may be used to indicate the start of the UA and to used to supplement the hand recorded data.

e **Performance criteria**:

i *Desired performance*:

for attitude failures

- Failure detection time within 1 sec.
- Maintain altitude within ±200 ft..
- Maintain airspeed within ±10 knots.
- Maintain heading within ±10 deg.

for heading failures

- Failure detection time within 2 sec.

- Maintain altitude within ±100 ft.
- Maintain airspeed within ±5 knots.
- Maintain heading within ±10 deg.

for other failures
- Failure detection time within TBD.
- Maintain altitude within ±100 ft.
- Maintain airspeed within ±5 knots.
- Maintain heading within ±5 deg.
- Course deviations < 1 dot.

for glideslope failures above localizer MDA
- Minimum altitude = MDA +20/-0 ft.

Note: Slowly developing failures, such as autopilot softovers may require time to detect. In such cases, it is better to describe performance in terms of flight path deviations rather than time.

- Pilot actions in accordance with (IAW) published standards. Published standards include instrument manuals,(*139-140*) aircraft-specific manuals, or equivalent documents.

ii *Adequate performance*:

for attitude failures
- Failure detection time within 2 sec.
- Maintain altitude within ±500 ft.
- Maintain airspeed within ±20 knots.
- Maintain heading within ±20 deg.

for heading failures
- Failure detection time within 5 sec.
- Maintain altitude within ±200 ft.
- Maintain airspeed within ±10 knots.
- Maintain heading within ±20 deg.

for other failures
- Failure detection time within TBD.
- Maintain altitude within ±200 ft.
- Maintain airspeed within ±10 knots.
- Maintain heading within ±10 deg.
- Course deviations < 2 dots.

for glideslope failures above localizer MDA
- Minimum altitude = MDA +40/-20 ft.

Note: Slowly developing failures, such as autopilot softovers may require time to detect. In such cases, it is better to describe performance in terms of flight path deviations rather than time.

- Pilot actions IAW published standards. Published standards include instrument manuals(*139-140*), aircraft-specific manuals, or equivalent documents.

f **Safety considerations**: These tests involve simulated failures. Suitable precautions to ensure the safety pilot retains adequate instrument references and that ATC clearances are not jeopardized.

Geographic Awareness

The recommended scenarios duplicate situations, encountered during normal flight operations where the pilot must maintain geographical orientation.

Typical scenarios should include

- Deviation from a planned flight plan caused by intentional deviations around weather, obstacles, or threats, followed by a return to the original flight plan.
- Deviation from a planned flight plan caused by intentional deviations around weather, obstacles, or threats, followed by a modification to the flight plan (such as bypassing the next waypoint).
- Changing the flight plan route.
- Changing to another, unplanned instrument approach.
- Diverting to an alternate.

These tests are intended to determine if the pilotage display permits the maintenance of geographical awareness. At the same time, other parallel evaluations should be conducted to ensure that the navigation display is suitable and the pilot can enter the necessary data to generate a new flight plan during those tests which involve flight plan modification.

a **Objectives**: These tests are designed to verify that the pilot can maintain geographical situation awareness, determine her orientation relative to desired course, and determine bearing and time to the next waypoint.

b **Requirements**: These test shall be flown in conjunction with other evaluations, particularly the mission simulations conducted under Chapter 13.

c **Action**: The pilot shall be given a typical mission profile to fly. During the course of flying this profile, he will be given diversion instructions by the safety pilot. The diversions should be of the following types:

- A series of heading instructions around a hypothetical weather build-up, an obstacle, or a threat. The evaluation pilot will then be told to return to his original flight plan route.
- A series of heading instructions around a hypothetical weather build-up, an obstacle, or a threat. The evaluation pilot will then be told to fly directly to a waypoint on her original flight plan. This need not be the next sequential waypoint; some evaluations should involve flying to another waypoint..

  These instructions should include terminal area navigation simulating ATC radar vectoring followed by a clearance to join the published instrument approach procedure, either a planned procedure or an alternate procedure.
- An instruction to fly directly to the a subsequent waypoint, bypassing the active waypoint.

- An instruction to divert to an alternate destination.

For tests involving operations in IMC, the evaluation pilot shall be hooded or otherwise prevented from using outside visual cues.

Following recovery, the evaluation pilot shall complete displays ratings (DRRs and DFRs).

d **Data**: Aircraft performance data should be recorded as functions of time:

Airspeed
Barometric altitude
Radar altitude
Heading
Course deviation
Glideslope deviation.

Pilot ratings (DRRs and DFRs) should be hand recorded. It is important to ensure, for the purposes of evaluating the pilotage display, that the evaluators rate the pilotage display not the ease of data entry (although such data entry ratings are important to the overall evaluation of the cockpit). Voice recordings may be used to supplement the hand recorded data.

Pilot estimates of bearing should be obtained from his initial course decisions. Once established on the new course, the safety pilot/simulator operator should ask for estimates of distances and course line aspect.

i *Desired performance*:

- Maintain geographical awareness: enroute
  Bearing to waypoint ± 5 deg
  Distance to waypoint ± 0.5 nm
  Distance to course line ± 0.2 nm
  Course line aspect ± 10 deg.
- Maintain geographical awareness: ingress to target
  Bearing to waypoint ± 2 deg
  Distance to waypoint ± 0.2 nm
  Distance to course line ± 0.1 nm
  Course line aspect ± 5 deg.
- Maintain geographical awareness: terminal area
  Bearing to waypoint ± 2 deg
  Distance to waypoint ± 0.2 nm
  Distance to course line ± 0.1 nm
  Course line aspect ± 5 deg.
- Maintain geographical awareness: instrument approach
  Bearing to waypoint ± 2 deg
  Distance to waypoint ± 0.1 nm
  Distance to course line ± 0.1 nm
  Course line aspect ± 2 deg.
- Maintain geographical awareness: during low-level NOE
  Bearing to waypoint ± 5 deg
  Distance to waypoint ± 600 ft

Distance to course line ± 300 ft
Course line aspect ± 10 deg.

ii *Adequate performance*:

- Maintain geographical awareness: enroute
  Bearing to waypoint ± 10 deg
  Distance to waypoint ± 1.0 nm
  Distance to course line ± 0.5 nm
  Course line aspect ± 30 deg.
- Maintain geographical awareness: ingress to target
  Bearing to waypoint ± 5 deg
  Distance to waypoint ± 0.5 nm
  Distance to course line ± 0.2 nm
  Course line aspect ± 10 deg.
- Maintain geographical awareness: terminal area
  Bearing to waypoint ± 5 deg
  Distance to waypoint ± 0.5 nm
  Distance to course line ± 0.2 nm
  Course line aspect ± 10 deg.
- Maintain geographical awareness: instrument approach
  Bearing to waypoint ± 5 deg
  Distance to waypoint ± 0.5 nm
  Distance to course line ± 0.2 nm
  Course line aspect ± 5 deg.
- Maintain geographical awareness: during low-level NOE
  Bearing to waypoint ± 10 deg
  Distance to waypoint ± 1200 ft
  Distance to course line ± 600 ft
  Course line aspect ± 30 deg.

f **Safety considerations**: These tests have no particular hazard other than the high workload for the safety pilot. Standard precautions for instrument training flights should be observed.

Collateral Evaluations

Traffic or Threat Detection

a **Objectives**: These tests are designed to verify that the pilot can maintain adequate awareness of the traffic/threat situation and maintain adequate lookout for threats and other traffic.

b **Requirements**: These tests shall be flown in conjunction with other evaluations.

c **Action**: The evaluation pilot will be briefed to make verbal calls for all observed traffic or threats. Both cooperative aircraft and targets of opportunity may be used.

d **Data**: The safety pilot or observer will make a note of the relative location of the reported traffic/threat at the time the evaluation pilot makes her verbal call. This position should be in relative terms (i. e. one o'clock, two miles). The safety pilot/observer will also note the delay, if any, between his own sighting of the traffic and that call by the evaluation pilot.

e **Performance criteria**: There is no consensus for performance criteria. One civil HUD program compared sighting times between evaluation pilot and safety pilot with acceptable performance being no significant difference in sighting times.

f **Safety considerations**: There are no additional hazards resulting from this evaluation.

Performance Assessment

One issue that was advanced in early HUD certification was the ability of a pilot to assess flight performance.(*9*) This task has been developed from this issue.

a **Objectives**: Check the ability of the pilot to assess flight performance.

b **Requirements**: These tests shall be flown in conjunction with all evaluations, including MTE testing.

c **Action**: During the course of the evaluations, the pilot will complete display readability and display flyability ratings. These ratings require the pilot to evaluate both his workload and his task performance.
The subjective task performance will be recorded and compared with objective performance data. For example, in a pirouette, the pilot may state his altitude performance as "meeting desired criteria" when, in fact, it fails to meet these criteria.

d **Data**: Aircraft performance and position data should be recorded electronically, preferably from the display databus. Subjective performance estimation shall be obtained from the pilot ratings (DRRs and DFRs).

e **Performance criteria**: Agreement between subjective and objective performance constitutes desirable results. Actual performance better than subjective performance is undesirable. Actual performance worse than subjective performance is unsatisfactory and will be unacceptable if hazardous. The degree of agreement will depend on the aircraft and flight task.
Prior to developing the test plan, agreement must be reached on the degree of agreement required to meet these criteria. In a previous program,(*9*) an comparison was made of the evaluation pilot's assessment of ILS performance and the actual value. In that program, agreement was deemed satisfactory if the two values were within one dot deviation or five knots in airspeed as that was thought to be the limit of the pilot's recollection of his performance following the approach.
i *Desired performance*:
- Agreement between subjective performance and actual performance within a value dependent on aircraft and flight task.
- Subjective performance assessment conservative.

ii *Adequate performance*:

- Agreement between subjective performance and actual performance within a value dependent on aircraft and flight task.
- Subjective performance assessment non-hazardous.

f **Safety considerations**: There are no additional hazards resulting from this evaluation.

References

9 Anderson, M. W., French, D. D., Newman, R. L., and Phillips, M. R., “Flight Testing a General Aviation Head-Up Display,” *Journal of Aircraft*: 33, 1996, 235-238

114 Gallimore, J. J., Brannon, N. G., Patterson, F. R., and Nalepka, J. P., “Effects of FOV and Aircraft Bank on Pilot Head Movement and Reversal Errors During Simulated Flight,” *Aviation, Space, and Environmental Medicine*: 70, 1999, 1152-1160

139 *Instrument Flying*, AFI 21-216; formerly AFM-51-37

140 *All-Weather Flight Manual*, NAVAER 80-80T-60, 1957