Integrated Modeling of Cognition and the Information Environment:

A Closed-Loop, ACT-R Approach to Modeling Approach and Landing

With and Without Synthetic Vision System (SVS) Technology

Michael D. Byrne
Rice University
Houston, Texas

Alex Kirlik
University of Illinois
Urbana-Champaign, Illinois

This page intentionally left blank.

1. Introduction

This report provides an overview of our ongoing research evaluating the impact of Synthetic Vision System (SVS) technology on pilot performance in commercial aviation in support of the NASA System-Wide Accident Prevention (SWAP) Human Performance Modeling (HPM) element (see Foyle, Goodman, and Hooey, 2003). We first provide a brief discussion in Section 2 of our theoretical and methodological perspective on this problem, and a brief discussion of lessons learned from our research last year modeling pilot performance in the T-NASA Taxi Navigation simulations. We present some of those lessons learned that have provided concrete implications for our current SVS modeling effort.

In the next section of this report, Section 3, we provide an extremely brief description of the NASA/Monterey Technologies, Inc. HPM-SVS Part-Task simulation and experimentation that provided the data set to support our current modeling effort. As indicated there, detailed information on both the simulation and experimentation can be found in Goodman, Hooey, Foyle, and Wilson (2003).

Section 4 of the report provides a description of an extensive set of statistical analyses we performed using the eye movement data collected in the part-task experiments. This section begins with a discussion of what aspects of the overall eye-movement data we selected to focus on for analysis purposes. Next, a set of analyses are presented consisting of, for example, breaking down the data set by phase of flight, approach event (e.g., nominal, missed, terrain mismatch), and of course, the SVS versus non-SVS experimental treatment. As the conclusion of this section demonstrates, these analyses were quite informative in terms of identifying the particular phenomenal that would serve as the focus for our closed-loop, ACT-R modeling.

Section 5 of the report discusses the additional (top-down or theoretical) sources of information guiding our modeling approach, including task analyses, the ACT-R approach, subject matter expert (SME) input, and extant theory of visual attention allocation from both engineering (e.g., Senders, 1964) and psychological (e.g., Wickens, 2002), perspectives.

Section 6 of the report goes on to describe how all the above information led us to focus on particular aspects of modeling pilot performance, and our detailed implementation approach. In particular, that section describes the three central phenomena around which our current efforts have been organized to date: 1) The desire to create a dynamic, closed-loop model of pilot cognition in interaction with the cockpit, aircraft, and environment; 2) The presumption that we are dealing with a relatively knowledgeable and adapted pilot, who is nevertheless presented with novel display technology; and 3) a focus on the allocation of visual attention as crucial to yielding important design- and training-related insights into the impact of SVS technology on cognition and performance. That section concludes with an overview of our detailed, computational implementation as it currently stands, and as we expect it to stand in the near future.

Section 7 describes our findings to date, and the implications of those findings for moving ahead in the near term. The findings are somewhat abstract at this point due to the fact that, although much work has been completed to date, we are only now grappling with the technical issues concerning coupling the ACT-R pilot model with the aircraft model we are using in order to provide a truly dynamic, closed loop account of attention allocation and pilot performance.

The report closes, in Section 8, with a distillation of progress made to date, lessons learned, and future directions. References follow Section 8.

2. Theoretical Perspective and Lessons Learned from Phase 1

2.1 Theoretical Orientation

Aviation incident and accident investigations often find both cognitive and environmental sources of human error. Environmental sources include factors such as flawed interface design, confusing automation, and unexpected weather conditions. Improved environmental design, such as the use of the Synthetic Vision Systems (SVS) that are the subject of our current research, often provide important leverage for reducing error and improving human performance. On the other hand, cognitive sources underlying the effectiveness and efficiency of performance include factors such as situation awareness, procedural compliance or non-compliance, and crew coordination. Many if not most significant incidents and accidents result from some combination of both cognitive and environmental factors. In fact, in a highly proceduralized domain such as aviation, with pilots who are highly trained and well-motivated, accidents rarely result from either environmental or cognitive causes alone. Training and experience are often sufficient to overcome even the most confusing interface designs, and the environment is often sufficiently redundant, reversible, and forgiving so that the vast majority of cognitive slips and mistakes have no serious consequences. Most highly consequential incidents and accidents result only when both environmental and cognitive factors collectively conspire to produce disaster.

Introducing new technology is a common approach to trying to reduce either the frequency, severity, or consequences of less-than-perfect pilot performance. Human performance modeling associated with evaluating the impact of technological interventions therefore requires giving consideration to both cognitive and environmental issues. This report describes the progress made to date on a research project in which dynamic, closed loop cognitive-environmental modeling, or more specifically pilot-vehicle-airport modeling, is currently being performed in order to shed light on both the positive and potential negative effects on the introductions of SVS technology in the commercial airline cockpit. Our current modeling consists of integrating a pilot model developed within the ACT-R cognitive architecture (Anderson, Bothell, Byrne, & Lebiere, 2002) with a commercial, off-the-shelf (COTS; Bowers and Jentsh, 2001) model of aircraft dynamics and the Santa Barbara airspace and airport which served as the basis for part-task experimentation. The overall objective of the NASA program in which we are participating is to develop computational human performance models with the predictive ability to aid designers and analysts in identifying likely vulnerabilities in human-machine performance in aviation.

Our current SVS modeling is actually the second stage in a longer term effort to meet this goal. Our research in the previous year focused on modeling to understand the causes of taxiing errors in a NASA simulation involving taxi-to-gate scenarios at a simulation of Chicago O'Hare (ORD) airport. To set the stage for the rest of this report, we briefly discuss the lessons learned from that effort which motivated our approach to the current problem. These lessons helped us get a bit of a head start regarding the selection of an initial modeling architecture for SVS modeling: a closed-loop, dynamically interacting dyad comprised of an ACT-R model of cognition and a commercially available aircraft-airport simulation package.

2.2 Lessons Learned From Phase 1

As we learned in our taxi modeling research, it is a nontrivial matter to apply scientific models of cognition, developed and validated primarily with psychological laboratory data, to applied contexts such as human-machine performance in aviation. Specific challenges include the following.

2.2.1 Communication Between Cognition and the Outside World

Experimental tasks are typically carefully designed in such a way that the inputs to, and outputs of, the cognitive system are readily identifiable. This is largely done by making the perceptual and motor demands associated with cognitive experimentation relatively trivial. Unfortunately, the perceptual and motor demands associated with aviation cognition can be extensive. This problem surfaced in Phase 1 research in the difficulties associated with coupling the ACT-R/PM model with the visual scene database, and also to some extent, with the aircraft model. The latter was not a severe problem because motor outputs could be considered to be relatively discrete in this instance, consisting of distinct settings of the throttle and brake. The SVS scenarios are similar in this regard although, as will be seen below, we have dedicated a good bit of effort to couple the perceptual mechanisms of ACT-R with both the cockpit and external scene provided by the flight simulator with which ACT-R is intended to interact.

2.2.2 Modeling Environmental Objects and Dynamics

As Phase 1 research clearly demonstrated, achieving a reasonable model of pilot cognition in dynamic, interactive contexts depends heavily on the availability of reasonable models of the visual, physical, and controlled environment, as well as its dynamics. The dynamics of human cognition and behavior is interleaved with, and occurs in concert with, the dynamics of environmental entities that also participate in the functioning of the integrated human-environment system. In our Phase 1 final report, we noted that cognitive modeling software packages can make better and more explicit provisions for representing objects and dynamics in the external environment so facilitate the task of modeling interactive behavior in contexts more complex than the desktop computer. As will be seen below, this issue has again resurfaced as a non-trivial issue in our current SVS modeling, if not theoretically, at least from a technical and data-communication perspective.

2.2.3 Timing Issues

On the basis of our Phase 1 research we concluded that modeling systems for running human-environment system models should provide separate clocks and processing resources for simulating cognitive and environmental dynamics. We noted that neither is subservient to, nor should be subsumed, by the other, nor should either model have to wait while the other is updating. We additionally noted that the current solution to this problem, in which processing resources are passed back and forth between cognitive and environmental components of the model, will be found to be increasingly unwieldy as more dynamic contexts are modeled. On the basis of this finding from our Phase 1 research, we made an early and explicit decision to implement our ACT-R pilot model and our aircraft/airport model on separate computers, connected by networked resources for 2-way data communication. With this approach, the processing demands associated with mimicking the dynamics of the two components of the total interactive systems do not compete. However, we have learned that neither of the two pieces of software we are using for modeling environments make explicit provisions for such inter-computer communication. As a result, we are currently at the stage of engineering our own solution to this problem.

3. Part-Task Simulation and Experimental Design

The focus of this research is on the evaluation of a new technology for the commercial airline cockpit (Foyle, Goodman, and Hooey, 2003). One of the factors that has long limited aviation is visibility; poor visibility conditions can substantially change the task of piloting an aircraft. However, with extensive and accurate computer-based geographic information systems, it is possible to generate the view of known terrain as long as the location of the observer is known. Modern GPS systems make it

possible to know the location of an airplane with high accuracy. Thus, the combination of the two systems makes it possible to render on a computer display the terrain that may not be visible due to adverse environmental conditions (e.g. fog, rain). This is the basis for NASA's Synthetic Vision System or SVS (for detailed information on the SVS design used in the current study, see Goodman, Hooey, Foyle, and Wilson, 2003).

Thus, a Synthetic Vision System (SVS) is essentially a computer generated display designed to provide the pilot with information that augments the out-the-window view, to better enable the pilot to fly safely, at low levels, through traffic, around terrain, and in low visibility conditions. Experiments conducted at NASA Ames Research Center by NASA and Monterey Technologies Inc. were performed to investigate the potential positive and negative effects of augmenting a cockpit with a prototype SVS display (Goodman, Hooey, Foyle, and Wilson (2003).

The purpose of these experiments was to collect data characterizing pilot performance and eye movement behavior during the approach and landing phase of flight using with both conventional and augmented displays under both Instrument Meteorological Conditions (IMC) and Visual Meteorological Conditions (VMC) conditions. The experimental plan, due to the cost and time required for studies of this complexity, focused on a limited number of pilots operating across a variety of conditions and treatments, as described in Goodman, Hooey, Foyle, and Wilson (2003).

4. Eye Tracker Data Collection, Interpretation, and Analysis

The first issue to be considered in this modeling effort is what to model. That is, what variance is there to be explained? In these scenarios and with the limited number of subjects available, there is not a rich corpus of speech or control manipulations, let alone errors, to model. Indeed, even if there were more errors, it is not clear how representative the small sample would be. The sample is also limited in terms of airports, weather conditions, scenarios, etc.

However, there is one extremely rich source of data: the eye movement record. Not only is there a great deal of data with which to work, there is substantial variance to be explained. Furthermore, we believe this eye movement record is the most generalizable component of the data. What we want to know is how the SVS, a visual display, affects pilots' visual behavior. For instance, if some higher-level metric of task performance (such as number of errors made) is not sensitive enough to show an effect of the SVS, this may be because the SVS is ignored by the pilots or because the difference it makes is compensated for by other factors. We can distinguish those cases via the eye-tracking data. Thus, we focused our empirical analysis on the eye data.

4.1 Collection and Coding Issues

Raw eye tracker data was provided by NASA at 20 Hz without any filtering or smoothing of the data. The data is noisy, which may be due to several sources such as blinking or a temporary loss of eye tracker calibration. The raw data are very detailed, containing raw X and Y coordinates, pupil dimensions, and other information. We looked primarily at one variable, region of interest (ROI), which was divided into eight sceneplanes, as follows (see also schematic in Figure 1):

Sceneplane 0 = Undefined or invalid data. Occurs when the eye cursor is centered on an area that is not defined as sceneplane 1 to 7 – i.e. the first officer, joy stick etc - or if the data is invalid (i.e. subject blinks).
Sceneplane 1 = Out-the-Window (OTW) View
Sceneplane 2 = SVS Display

Sceneplane 3 = Primary Flight Display
Sceneplane 4 = Nav Display
Sceneplane 5 = Mode Control Panel
Sceneplane 6 = Controls (Flaps, gears, speedbrakes, map scale)
Sceneplane 7 = Overlapping Area. The cockpit displays sit directly in front of the lower portion of the OTW view. Depending on the viewing angle of the subject (which varied slightly by subject, and over the day of trials), the eye tracker could not always determine whether the subject was looking at the black masking area around the displays, or the OTW view behind the masking. In these cases, the sceneplane was recorded as "7".
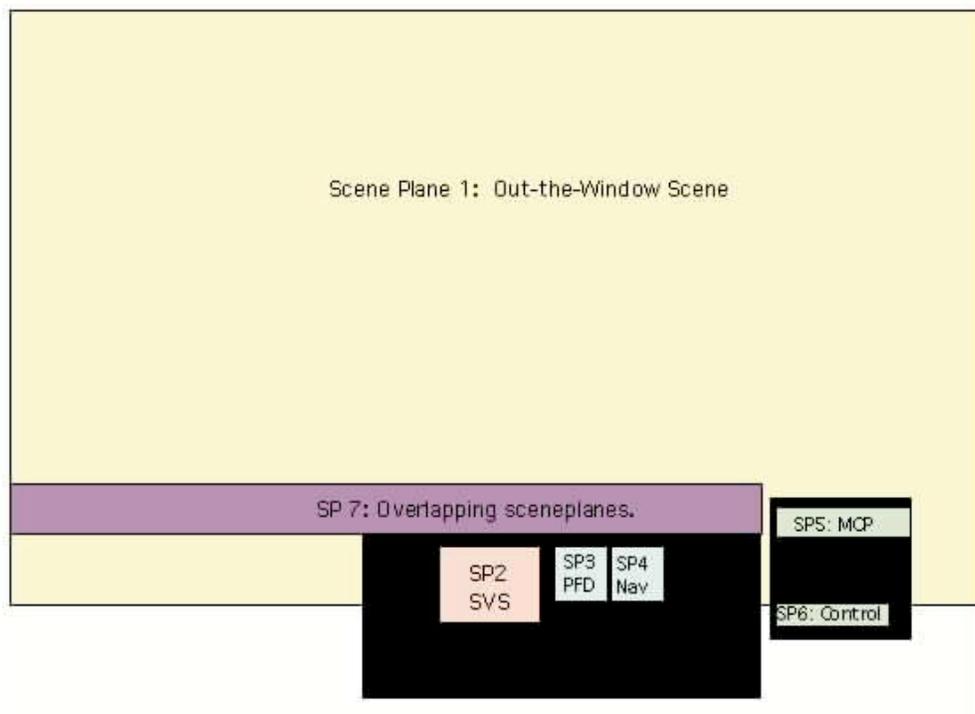
Figure 1. Overview of Sceneplane Layout.

In addition to the raw tracker data, NASA provided videotapes from two cameras:

Eye Tracker Camera. For each trial a videotape of the pilot's forward view was recorded from the head-mounted eye tracker. The pilot's point of gaze is shown by crosshairs superimposed over the visual scene. These tapes provide a fair representation of what the pilot was actually seeing at any given point in the trial.

Room View Camera. Additionally, for each trial an ambient audio and video recording was produced that depicts displays and control inputs and verbal communications. Three audio channels were recorded as follows: left channel was the Captain (subject), right channel was the FO (experimenter), and center channel was ATC (experimenter). It should be noted that the camera was mounted high and behind the pilot and that the visual perspective in the tapes is not that of the pilot.

These additional sources of data were useful in understanding what the pilots did and when they did it for the purposes of validating the task analysis, but were not quantitatively analyzed.

4.2 Eye Movement Analysis and Results

Fixations are distinguished from saccades (rapid voluntary eye movements used to move from one fixation to another) and very small involuntary eye movements of several types that occur during fixation. A "dwell" is defined as the time period during which a fixation or series of continuous fixations remain within a sceneplane or ROI.

Selected Data Analysis

To focus on SVS versus non-SVS cases in similar conditions, our initial data analysis focused solely on the conditions listed in Table 1.

Table 1. Selected conditions for data analysis

Display Configuration		Baseline	SVS
Visibility		IMC	IMC
Approach Event	Nominal Approach (nominal landing)	*Scenario #4*	*Scenario #7*
	Missed Approach (go-around)	*Scenario #5*	*Scenario #9*
	Terrain Mismatch (go-around)	*Scenario #6*	*Scenario #10*

Because we are primarily concerned with the allocation of visual attention, the primary variables of interest are those that index the amount of attention given to each sceneplane. Allocations can be counted in two ways, by count (number) of fixations or by total duration of dwell. Of course, if all fixations have the same duration, the relative proportions spent on each sceneplane will be identical, but this is an empirical question.

Additionally, there are a potentially large number of ways to segment the data. In this presentation we will limit our analysis to the following: we consider only four categories, namely, pilots' baseline eye movements, pilots' SVS versus baseline eye movements, pilots' eye movement by different flight phases, and pilots' eye movements by different approach scenarios.

4.2.1 Baseline Attention Allocation

The first thing to be established is the baseline allocation of attention. If the SVS is to serve as a proxy for looking out the window, it is useful to know how often pilots look out the window in the first place. In addition, since the SVS contains information that can be found on other displays (e.g., altitude, which is also on the PFD), it is important to know how often those other displays are accessed.

Figure 2 shows that overall pilots spent 40.2% of their total fixation dwell time on the PFD, and about 40% fixation duration on NAV. They only spent about 3% looking out the window, about 4.2% at the MCP, and about 6.5% at the Controls. This result suggests that the PFD and NAV displays are the major targets of attention, as they account for about 80% of the fixation dwell time.

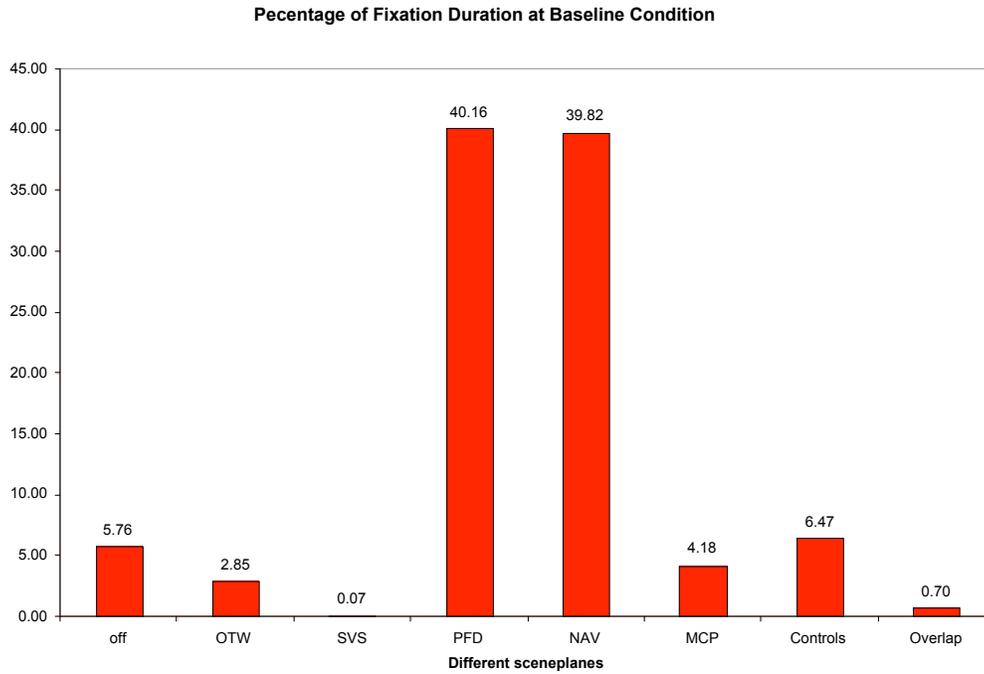**Pecentage of Fixation Duration at Baseline Condition**

Figure 2. Percentage of fixation dwell time in the baseline condition.

The fixation count measure shows a similar distribution, as depicted in Figure 3. Pilots spent 38.3% of their fixations on the PFD, and another 39.6% of their fixations on NAV, and relatively small amounts of time looking at other displays, most notably less than 3% out the window.
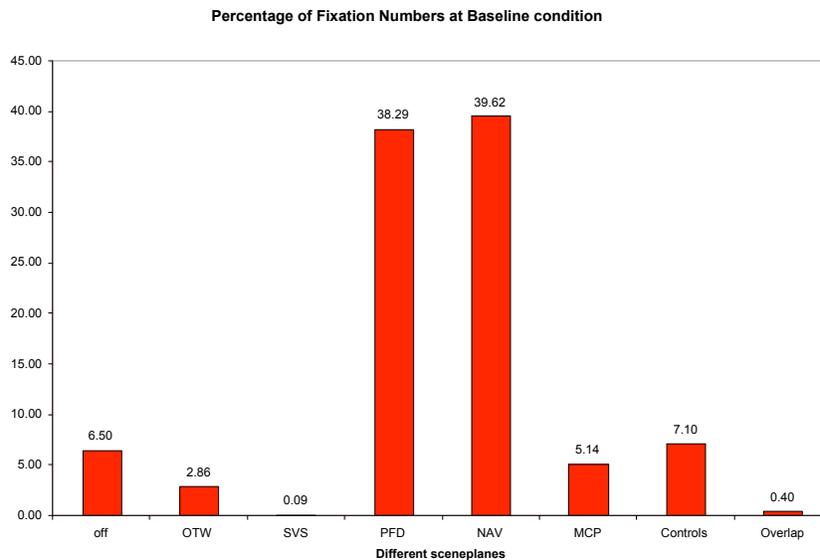
Percentage of Fixation Numbers at Baseline condition

Figure 3. Percentage of fixation count in the baseline condition.

4.2.2 Pilots Eye Movements at SVS versus Baseline Conditions

The obvious next question is how the SVS affects this distribution. Figure 4 shows that the distributions of pilots' fixation dwell time on different sceneplanes are similar for both the SVS and the baseline (No SVS) condition. In the SVS condition, pilots also spent most of their fixation time on the PFD (35.7%) and NAV (29.6%). Pilots also spent 2.6% fixation time looking out the window in the SVS condition. However, one notable feature in the SVS condition is that pilots displayed a large number of fixations (20.2%) on the SVS. This clearly shows that pilots did look at the SVS display quite frequently.

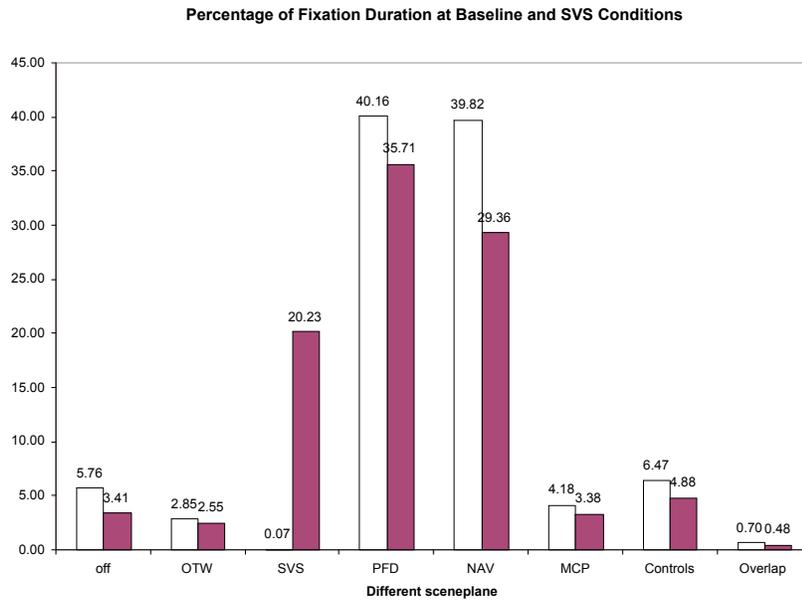**Percentage of Fixation Duration at Baseline and SVS Conditions**

Figure 4. Percentage of total fixation dwell time on each sceneplane in the baseline (light bars) and the SVS (dark bars) conditions.

Again, the distributions for fixation count follow those for time. Figure 5 shows the graph for fixation count for SVS and baseline conditions, and yields about the same results.

Percentage of Fixation Numbers at Baseline and SVS Conditions
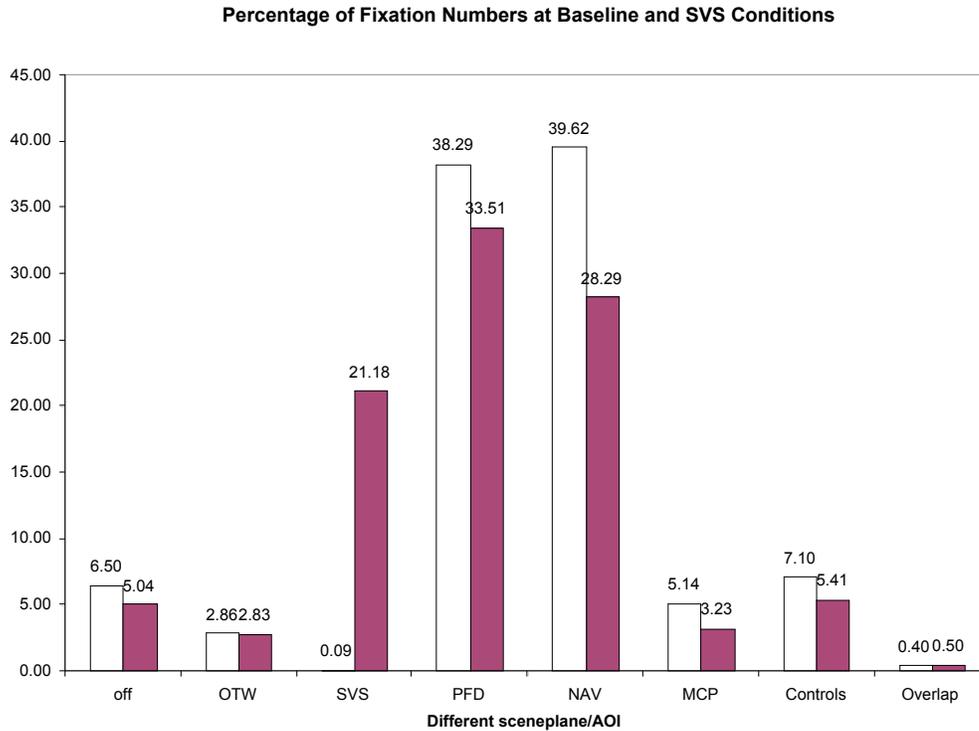

Figure 5. The percentage of fixation count on each sceneplane in the baseline (light bars) and the SVS (dark bars) conditions.

Clearly, pilots spend a fair proportion of their gaze on the SVS. An interesting question we wanted to address is from where they "stole" these gazes. Namely, they must reduce some portion of fixations associated with other sceneplanes. A natural suspect for the location from which fixations would be stolen was the OTW display, due to the inherent redundancy between the perceptual information gained from the SVS and OTW displays.

Surprisingly, however, the fixation dwell time and frequencies on different sceneplanes under SVS versus baseline (No SVS) conditions, as shown in the above two figures, shows no obvious difference in the amount of gaze directed out the window. Instead, it appears the SVS is associated with a reduction in the amount of gaze directed at the PFD and NAV displays. Thus, the SVS display was drawing attention away from other sources of information from within the cockpit, it was not acting as a "substitute" for the information provided by the OTW display.

We find this result both counterintuitive and quite interesting. The underlying rationale behind the SVS display is that it will act as a substitute for the OTW display when the information obtainable from the latter is degraded. Instead, these data seem to indicate that it did not act as a substitute source of environmental information (at least for these pilots). And as a possibly unintended result of the presence of the SVS, less attention was paid to other displays.

4.2.3 Analysis by Phase of Flight
All the above analyses combined all the flight phases together. During different flight phases, however, pilots may have different needs for different information. We therefore decided to break down the analysis of data by flight phase. The fight phases were defined as follows:

101

Phase 1. Start to Initial Approach Fix (IAF)
Phase 2. Initial Approach Fix (IAF) to Final Approach Fix (FAF)
Phase 3. Final Approach Fix (FAF) to Decision Height (DH)
Phase 4. Decision Height (DH) to End

Because the percentage of fixation dwell time and fixation counts on different sceneplanes provides similar information, to simplify the analysis we only dealt with the percentage of fixation dwell time on different sceneplanes in all subsequent analysis. Again, the natural starting point is the baseline condition, which is presented in Figure 6.. Note how the use of the PFD increases as the flight moves on, and the sharp drop in the use of NAV and sharp increase in OTW gazes in phase 4.
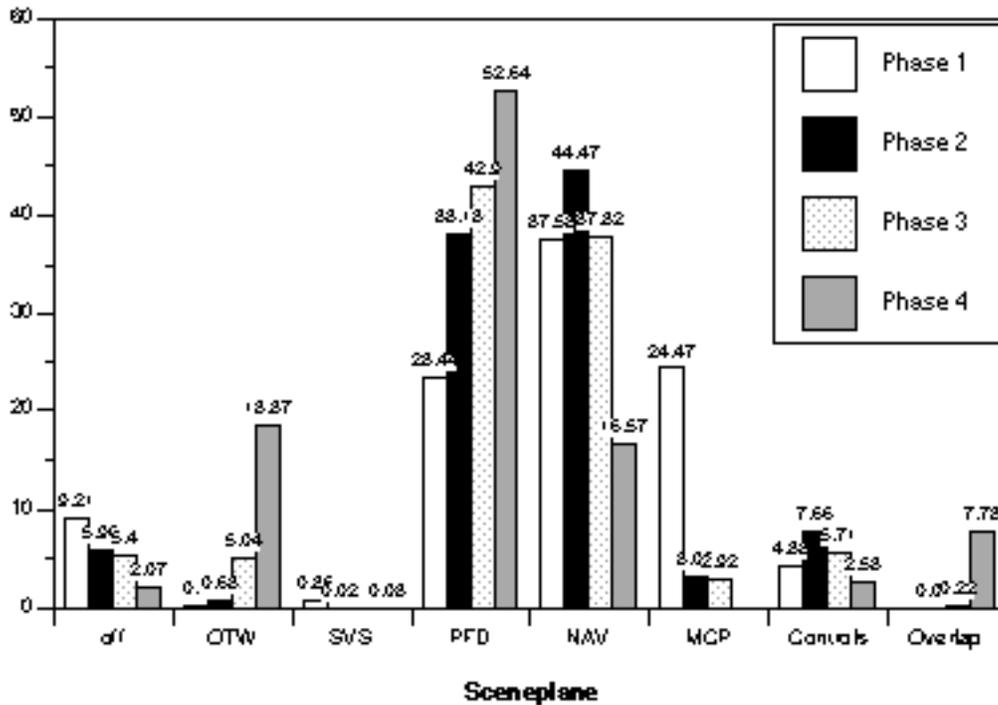

Figure 6. Percentage of fixation dwell time in the baseline condition by different flight phases.

Because pilots so rarely look out the window in phase 1, one might expect little use of the SVS in this phase. In fact, pilots do look a little at the SVS in this phase but overall the allocation of gaze is not substantially different in baseline vs. SVS in this condition, as shown in Figure 7.
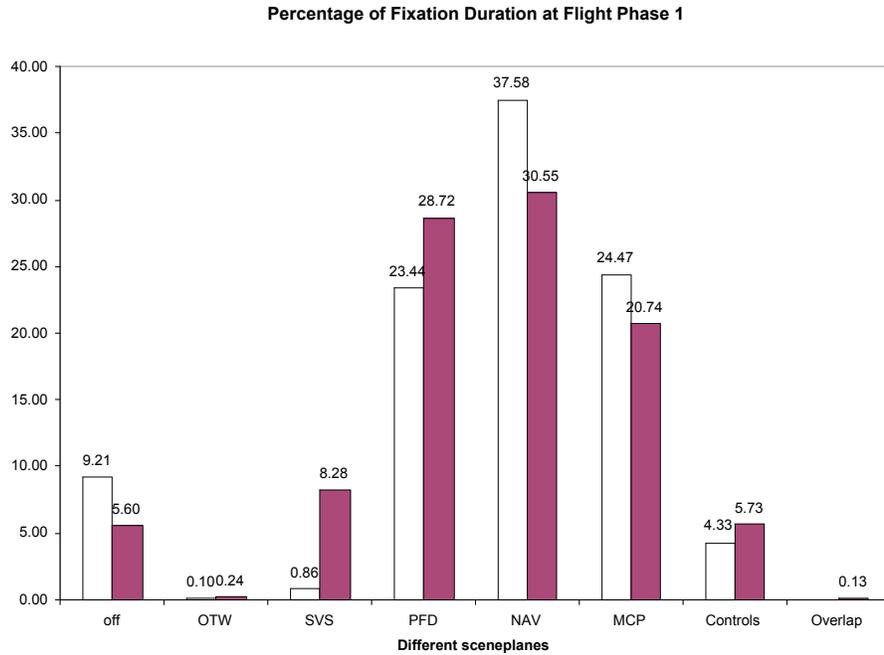
Percentage of Fixation Duration at Flight Phase 1

Figure 7. Percentage of fixation dwell time in the baseline (light bars) and the SVS (dark bars) conditions for flight phase 1.

.

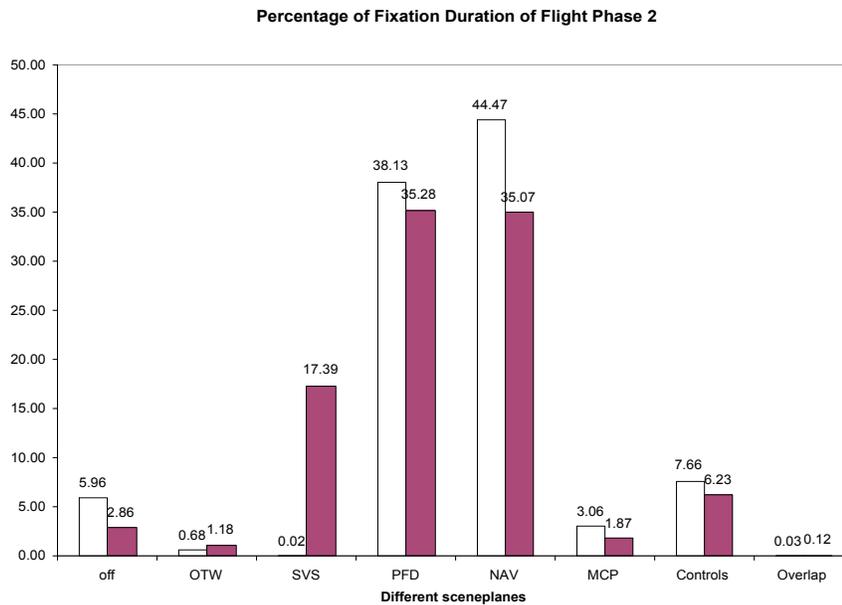**Percentage of Fixation Duration of Flight Phase 2**

Figure 8. Percentage of fixation dwell time in the baseline (light bars) and the SVS (dark bars) conditions for flight phase 2.

Phase 2 shows an increase in SVS use over phase 1, with most of the gaze time being "stolen" from PFD and NAV (see Figure 8). Note that in this phase of flight, there is virtually no use of OTW in either baseline or SVS conditions, but the SVS is still used when present. We believe this is due primarily to the symbology overlaid on the SVS. This trend continues through phase 3, depicted in Figure 9. Note the substantial use of SVS and reductions in use of PFD and NAV

Percentage of Fixation Duration of Flight Phase 3
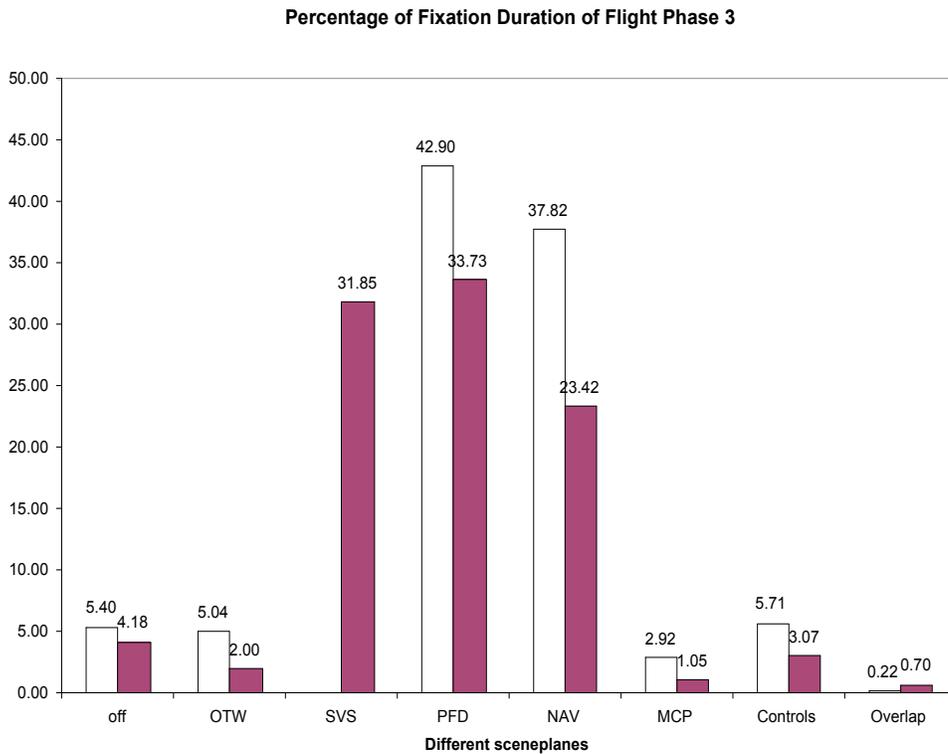

Figure 9. Percentage of fixation dwell time in the baseline (light bars) and the SVS (dark bars) conditions for flight phase 2.

Flight phase 4 shows the largest reduction of OTW looking as a result of the SVS (Figure 10), but this reduction in OTW gaze only accounts for about a third of the total SVS time, which appears to come primarily from the PFD. Overall, SVS gaze accounts for a fairly substantial proportion of total gaze time in phases 3 and 4, but this is not by simple reduction of OTW gaze. Instead, pilots seem to borrow gaze from the PFD.
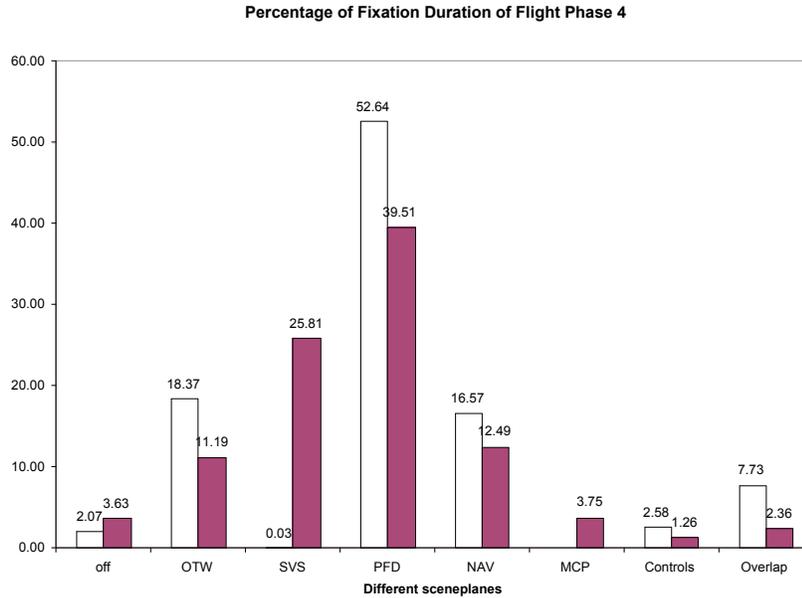
Percentage of Fixation Duration of Flight Phase 4

Figure 10. Percentage of fixation duration at the baseline and the SVS condition of flight phase4.

4.2.4 Late Phases of Flight Broken Down by Approach Events

Since the SVS versus non-SVS differences showed up most prominently in the final phases of flight, the next step in the data analysis focused on phase 3 and phase 4, further breaking down the data by different approach scenarios, namely, Nominal Landing, Missed Approach and Terrain Mismatch. This is depicted in Figures 11-13.
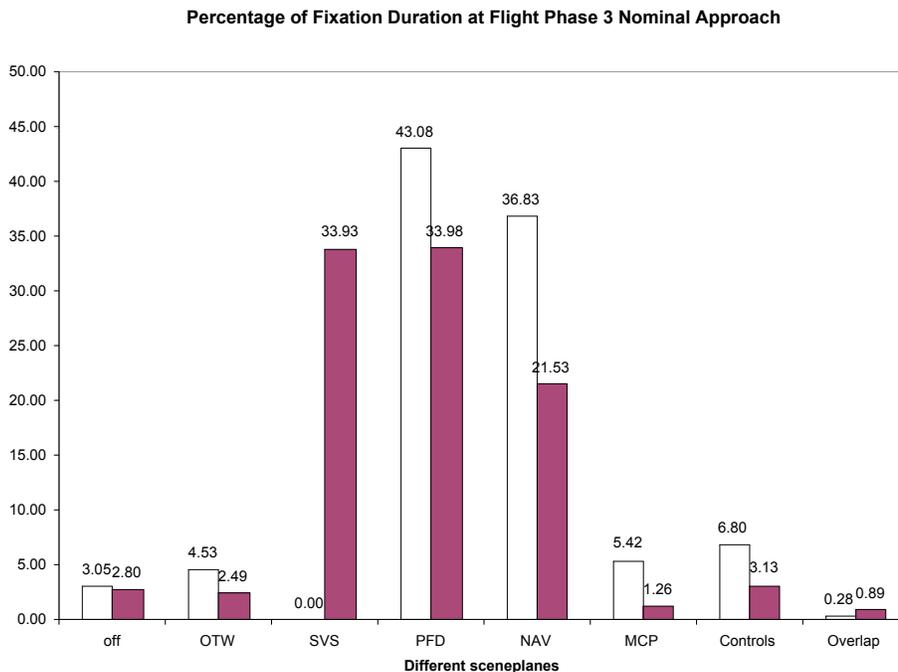
Percentage of Fixation Duration at Flight Phase 3 Nominal Approach

Figure 11. Percentage of fixation dwell time in baseline (light bars) and SVS (dark bars) conditions during flight phase 3 for the Nominal Approach scenarios.

105

Percentage of Fixation Duration at Flight Phase3 Missed Approach

Figure 12. Percentage of fixation dwell time in baseline (light bars) and SVS (dark bars) conditions during flight phase 3 for the Missed Approach scenarios.

Percentage of Fixation Duration of Flight Phase 3 Terrain Mismatch

Figure 13. Percentage of fixation dwell time in baseline (light bars) and SVS (dark bars) conditions during flight phase 3 for the Terrain Mismatch scenarios.

Note how similar the OTW and SVS usage is for the three scenarios during phase 3. This is in stark contrast to phase 4, presented in Figures 14-16.

Percentage of Fixation Duration of Nominal Landing at Phase 4

Figure 14. Percentage of fixation dwell time in baseline (light bars) and SVS (dark bars) conditions during flight phase 4 for the Nominal Approach scenarios.
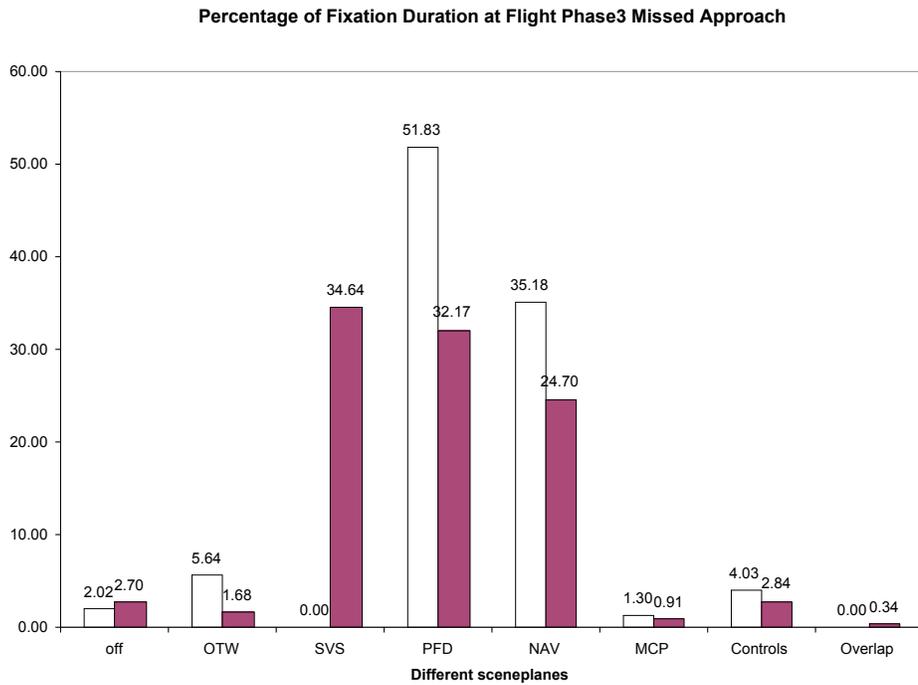
Percentage of Fixation Duration of Missed Approach at Phase 4

Figure 15. Percentage of fixation dwell time in baseline (light bars) and SVS (dark bars) conditions during flight phase 4 for the Missed Approach scenarios.

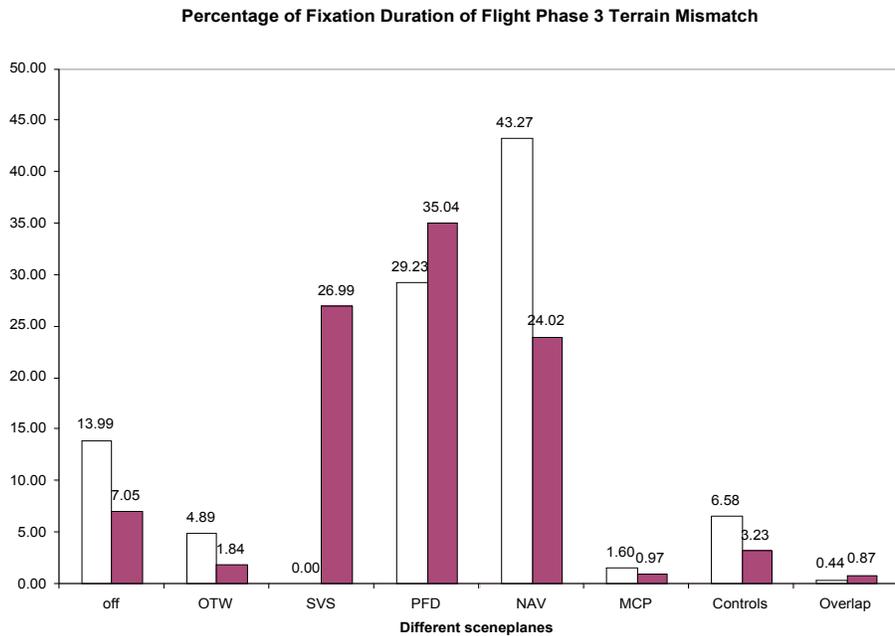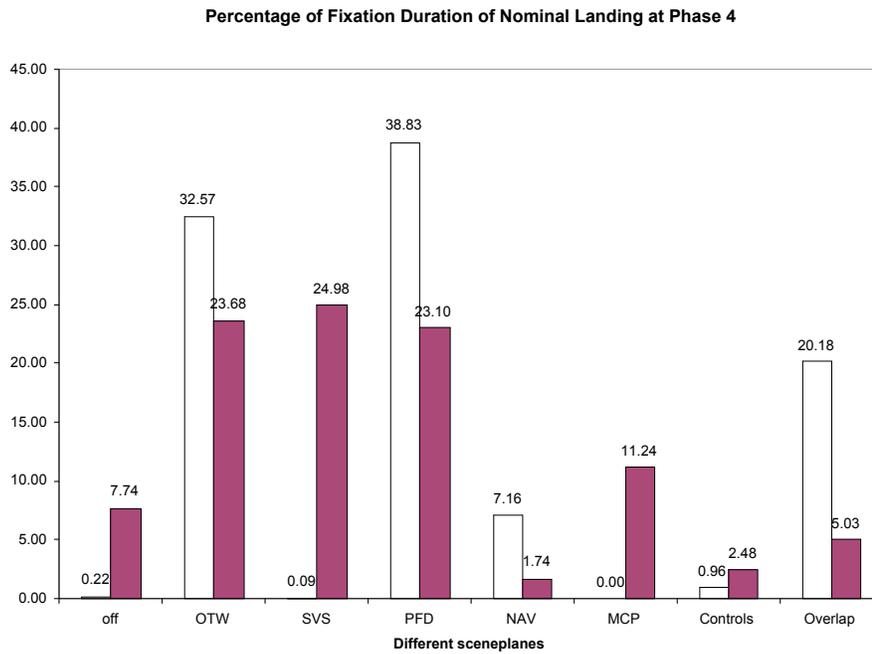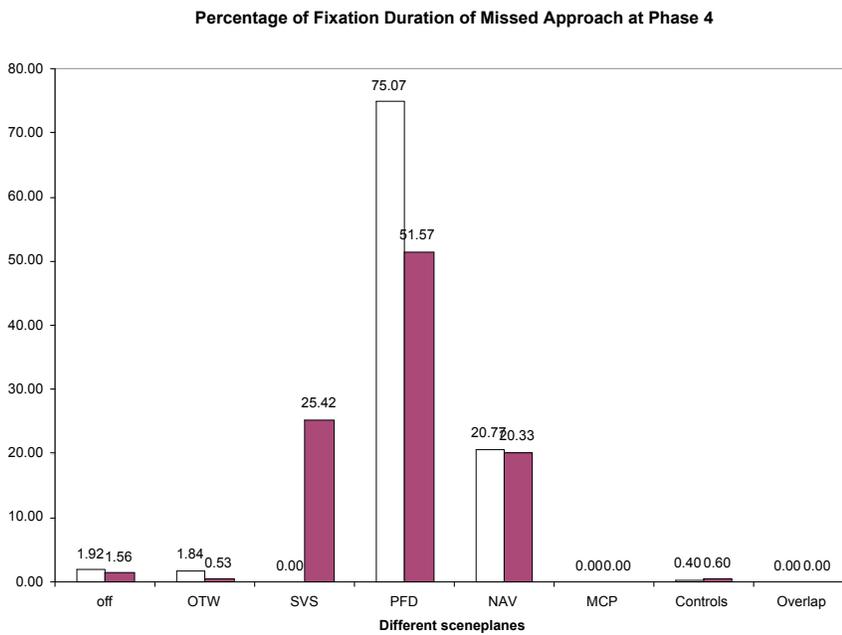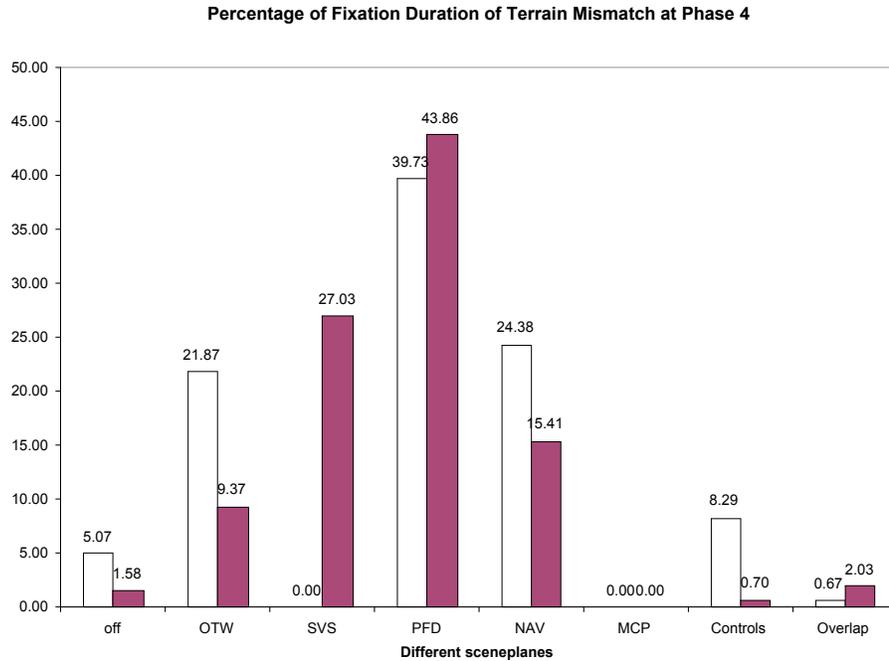**Percentage of Fixation Duration of Terrain Mismatch at Phase 4**

Figure 16. Percentage of fixation dwell time in baseline (light bars) and SVS (dark bars) conditions during flight phase 4 for the Terrain Mismatch scenarios.

The differences here are striking. In the nominal landing, pilots looked out the window quite a lot, even when the SVS was available. SVS use was certainly evident for the nominal approach in phase 4, but as seen in other graphs, this seems to come at least in part through less allocation of gaze to the PFD and possibly NAV. However, in the missed approach, OTW looking fell to virtually zero, both with and without SVS. Note that SVS use was roughly similar in nominal and missed approach, but in missed approach, this appears to have come almost entirely at the expense of gaze at the PFD.

For terrain mismatch conditions, the SVS use was again about the same (about a quarter of the total time), but this time it was *not* due to a reduction in looking at the PFD. Thus, use of the SVS is clearly conditioned on phase of flight and scenario. In addition, the SVS is evidently not simply a proxy for looking out the window. In particular, the SVS appears to serve many of the informational functions of the PFD, and perhaps also NAV, though to a lesser extent. This has other implications that will be discussed later.

The presence of the SVS may change not only how much the pilots look at various displays, but the strategies they use for acquiring information from their visual world. This should be reflected in changes to the order in which various pieces of information are acquired. Thus, we are in the process of expanding our data analysis to include sequential dependency information as well. That is, if fixation *n* is on the PFD, where is the most likely location for fixation *n*+1? Is this affected by the presence of the SVS, and if so, how? Since ACT-R produces a behavior stream, it should be possible to predict the various transition probabilities.

5. Description of Modeling Effort

A major focus of our modeling effort is to reproduce the trends observed on the basis of the detailed analyses presented in Section 4. We believe that to do so in a manner consistent with one of the major lessons learned from our Phase 1 research on taxi modeling, we must focus on the specification of cognition, the environment, and their interaction at a very fine grain of detail. We believe that this is especially crucial for modeling performance in a dynamic, interactive task. Highly detailed data and task analyses are critical, particularly for a cognitive architecture that operates at a very fine grain of temporal resolution, such as ACT-R. Thus, our focus has been on laying an appropriate foundation for the modeling effort. Our preference has been to eschew shortcuts for the promise of higher fidelity. While this has slowed certain aspects of our progress, we believe this will pay off later. Our approach has been to try to understand the major sources of both insight and constraint in generating our models. In addition to the detailed eye-movement analyses presented above, we identified three additional sources of information:

5.1 Task Analysis

Besides data analysis, our first order of business was to try to understand the task at a detailed level. This a challenge for this task because there is little overt action taken by the pilots; it appears to be primarily a supervisory control task, at least until the pilot takes manual control. However, the task is more complex than just that. To understand it, we have relied on three primary sources of information: the task analysis information collected and supplied by NASA Ames (Keller, Leiden, & Small, 2003); other related work in the human factors of aviation; and conversations with our subject matter expert (SME). We have synthesized these into the ACT-R formalism. An example of some of the resulting control structure appears in Figure 17.
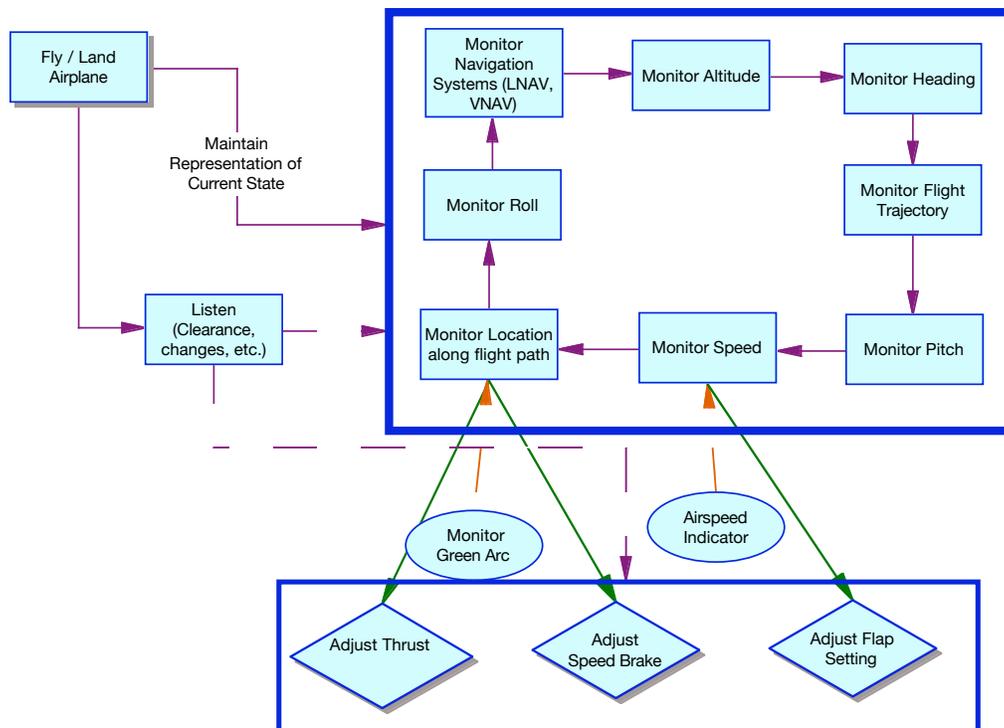

Figure 17. Flow of control resulting from task analyses.

The first insight from the knowledge engineering process is that the bulk of the task, particularly for the first two phases of flight, is primarily a monitoring task in which the pilot is engaged in maintaining his or her representation of the state of the aircraft. Additionally, we learned that pilots very actively check for a number of events and conditions which do not occur in the scenarios, such as late changes of wind direction that might lead to wind shear. Thus, for a lot of the time during the experimental trials, there is the appearance of little workload while in fact the pilots do still have a lot to do. This is fairly realistic for most landings which are, in fact, routine. However, pilots do have to monitor for non-routine conditions. In order to simulate the true workload accurately, we have included checks for many of these things in the model even though they do not occur in the scenarios.

5.2 ACT-R

The ACT-R architecture provides a great deal of constraint as well. Working within the parameters of the architecture sets certain boundaries and delimits scope. In particular, it means that we are modeling the task at a highly detailed level of analysis. ACT-R provides end-to-end modeling of the human operator side of the human-in-the-loop, from basic visual and auditory attentional operators to complex cognition and back down to basic motor movements. This impacts the strategies that are even possible and the way in which knowledge about dynamic state has to be updated to be maintained. A thorough review of ACT-R is far beyond the scope of this presentation. However, it should be noted that we are now using the most recent version of ACT-R, termed ACT-R 5.0 (see Anderson, Bothell, Byrne, & Lebiere, 2002 for a detailed description). ACT-R 5.0 incorporates the perceptual-motor extensions found in ACT-R/PM and provides for even more aggressive parallel execution of cognitive, perceptual, and motor operations than did the ACT-R/PM version of the system used for the taxi model work.

5.3 Extant Accounts of Relevant Phenomena

Because the eye-movement data are the primary focus of the modeling effort, we have examined other data and models in the "allocation of attention" domain in the human factors literature (e.g., Senders, 1964; Wickens, 2002). These are high-level (relative to ACT-R) accounts of how operators choose which objects to visually sample and at what frequency. The basic findings are that the rate at which particular displays are sampled depends jointly on the task importance of the displayed information as well as the rate of change of the information. As one might expect, more important information is sampled more often, and more dynamic information is sampled more often. We believe that these accounts provide a useful high-level starting point; we hope to provide the explanation for how these high-level phenomena emerge from a combination of task and environmental constraints and relatively low-level cognitive-perceptual capabilities. In other words, we recognize and have learned from theories that predict the percent of fixation time on particular information sources through mathematical modeling. Our goal, in contrast, is to create a process model from which such higher-level descriptions emerge as a function of the lower-level mechanisms in the model-environment system.

6. Focus and Intent of Modeling Effort

We had three major foci in the present effort:

6.1 A Dynamic, Closed-Loop Approach

One of the things which distinguishes an analysis at the level of a cognitive architecture such as ACT-R is that it is possible to "close the loop" of the human-machine system. That is, both the human and the evaluated system are modeled dynamically and in detail, and the two sub-models are coupled,

yielding a model of the complete dynamic system. Work on the taxiing model revealed that fidelity of the machine/environment model was critical in understanding the performance of the human model; in particular, many of the "higher-level" decisions ultimately depended on "low-level" properties of the human-environment system. For example, the decision of "which strategy should I use to choose which direction to go?" often depended on things like the distance between the sign and the intersection as well as when the cognitive system was free to sample that part of the visual environment. Because ACT-R is fundamentally a non-linear system, small perturbations in the dynamic state of the human-environment system at one time can often lead to large differences in state or behavior further down the road.

Thus, we feel it is critical to continue with this rather complete, closed-loop approach. As previously mentioned, this means we have to contend with a great deal of detail in modeling the pilots' behavior, but ultimately we believe that path will lead to the best model.

6.2 An Adapted Pilot

Present efforts are based on modeling a pilot who is both knowledgeable about the task and well-adapted to it. We are neither modeling novice pilots nor the acquisition/development of piloting expertise. This limits the scope of the model but has other implications as well.

In particular, this means the task analysis information is, in some sense, "contaminated" by the fact that the pilots come into the task with a pre-existing strategy for how to sample the relevant displays. Because they know which information is most important and have a clear model of which information will be most dynamic, their strategies reflect this knowledge. That is, the relevance and rate of change for properties like altitude are known in advance by the pilots, so the pilot does not have to figure out how often to sample that information, he or she already knows how often it needs to be sampled. However, we believe that this has certain implications which we may want to relax later, see the section on later efforts for more details.

6.3 An Attention Allocation Focus

As mentioned previously, we believe the primary phenomenon to be explained here is how the pilots deploy their visual attention across the visual array and how this is (or is not) affected by the SVS. While this appears straightforward, there are some subtle issues here which we are exploring. For example, the ACT-R model produces time stamped individual shifts of visual attention (saccades) to small targets; we believe it is a mistake to attempt to map these directly to the individual saccades made by the pilots. Rather, such data can be analyzed at different levels of abstraction. For example, one could reasonably be interested only in more gross performance measures, such as the proportion of fixations on each scene plane, for which we have human data. We can run the model, which produces data at a much finer level of detail, but then extract these higher-level measures from the model run. In fact, this extraction can be performed with more or less the same set of analysis tools that were used on the human data.

An important research question is: What level of analysis is appropriate to guide design decisions? Did we want only the more gross measures such as proportion of fixations on each scene plane, or was it worthwhile to attempt to match the exact sequence of fixations generated by a model run with the exact sequence generated by one human trial? While the answer is somewhere in between, this is still an empirical question. Because ACT-R produces behavior at a fine grain size, we had the option of potentially examining behavior at multiple levels.

6.4 Implementation Approach

Many of the details of the implementation have already been discussed. The primary inputs to the cognitive model come from the task analysis; this is the source of the procedural knowledge and the bulk of the initial declarative knowledge given to ACT-R. The output of the model is a time stamped series of behaviors including individual attention shifts, speech output, button presses, and the like. The primary point of comparison for the model output is the human eye-tracking data, which can be examined at various levels of abstraction. One piece that has not been described in detail thus far is the other half of the simulation: the simulation of the aircraft. We have mocked up the primary displays (NAV, PFD, MCP, etc.) in the language of ACT-R so that it can directly "view" those pieces of the display. However, this is not enough; ACT-R requires a dynamic environment with which to interact. For instance, if the flap setting is changed by the model, there are certain expectations about downstream effects on flight performance. To make those happen properly, a simulation of the airplane is required. We have purchased the commercial software package X-Plane for this purpose and are in the process of linking X-Plane to ACT-R (note that X-Plane has been certified by the FAA for training pilots, see http://www.x-plane.com/FTD.html). Figure 18 presents a picture of X-Plane in action.

Figure 18. The X-Plane flight simulation package.

This linkage process is not trivial; we are writing a network interface (based on the UDP protocol) between the two programs from the ground up. X-Plane natively supports sending certain kinds of information such as altitude and heading via the network interface, but other things cannot be sent, including the view out the window. This represents something of a problem since the ACT-R model needs something to "see" out the window (and on the SVS). However, we believe this problem can be solved relatively straightforwardly by abstracting out only what the model would need to look for when it looks. For example, because we know the plane's absolute position and orientation with respect to the airport, we can determine whether whatever piece of information the model was seeking would be available. This task-oriented solution may have uses in other domains as well.

In addition, we have to supply X-Plane with the aircraft specifications (a 757) and the appropriate approach/navigation and FMC programming (e.g., fix points) for Santa Barbara. Fortunately, the 757 specifications and the airport and geography for Santa Barbara were freely available and could simply be plugged in. Figure 19 presents a diagram describing the system. System runs will involve initializing both ACT-R and X-Plane appropriately, running them, and collecting a trace of the output. X-Plane is designed to run in real time, so generating multiple simulation runs will be time-consuming. (However, there may be some workarounds for this and we are hoping to get X-Plane to run 2x or 4x real time.)
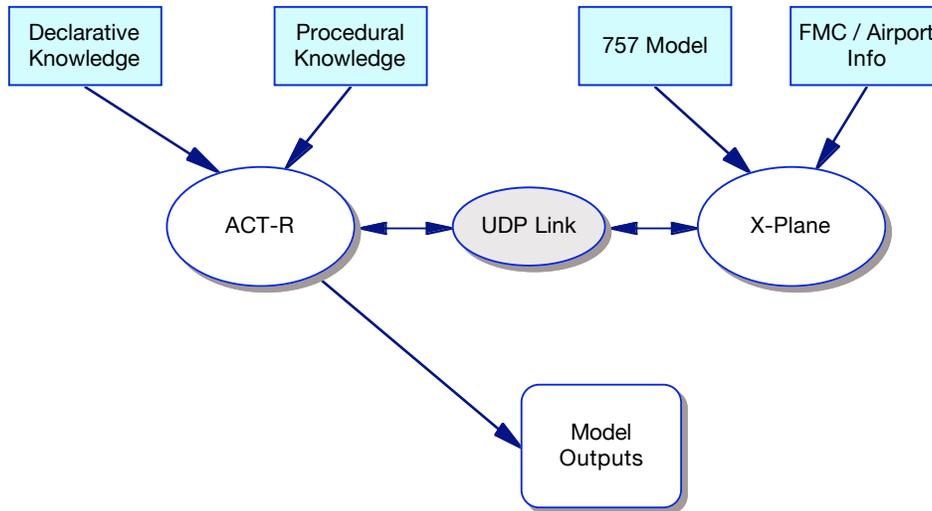
Figure 19. System overview.

7. Findings

7.1 Preliminary Results

Because the fully-coupled simulation is not yet completely operational, our findings are currently somewhat preliminary. However, we believe that we have still made substantial progress and, more importantly, gained significant insight. First, our initial data analysis shows that the SVS does indeed affect attention allocation, and that this is conditioned on phase of flight. Consider Figure 18, which shows the percentage of the dwell time by region of interest (ROI) for flight phase 1 (start to initial fix). Note the similarity between the non-SVS and SVS conditions. Contrast this with Figure 20, which presents the same data for phase 3 (final fix to decision altitude). Note how the pilots make little use of the SVS in phase 1, but in phase 3 their eyes are aimed at the SVS nearly a third of the time. As mentioned earlier, that the SVS is not simply a proxy for looking out the window in phase 3; pilots rarely look out the window at this phase. Instead, pilots look at the SVS and look less at the PFD and NAV displays.

At a high level, the model has a clear story for these data. The model predictions are based on the number of times a piece of information must be found and where the model will look for that piece of information. The model proportion presented here is simply the number of times attention will be directed to any particular display divided by the number of times attention will be directed to all relevant displays. When a piece of information could be found on the SVS as well as somewhere else

(the PFD or OTW), the weak assumption was made that the model would look for that information from the SVS half the time and from the other source (PFD, OTW) the other half of the time.

Table 2 presents the overall (that is, not conditioned by phase of flight) data for both the human subjects and the model. This is essentially a static approximation of the dynamic ACT-R system, which may vary from this somewhat in final form. However, the initial analysis is encouraging; the fully-dynamic model should certainly be able to capture the patterns found in the data.

Table 2. Human data and analytic model predictions for fixation frequency

Region of Interest	Data, no SVS	Model, no SVS	Data, with SVS	Model, with SVS
NAV	0.39	0.30	0.28	0.27
PFD	0.38	0.44	0.33	0.30
MCP	0.07	0.19	0.03	0.20
OTW	0.03	0.07	0.03	0.03
SVS	-	-	0.21	0.20

What is important to note here is that the predictions for the SVS condition, in particular, are sensitive to local properties of the display. We currently use a rough estimate that if an item is available on the SVS then the model will look at the SVS half the time; this is currently a baseline assumption. In fact, if the model needs to look for a particular piece of information that is available in multiple locations (e.g., altitude, which is on both PFD and SVS), where it will look will be conditioned on where it is currently looking. ACT-R models are sensitive to local costs, and looking further away takes longer, so the model will prefer to look for the altitude on the SVS if it is already looking at the SVS.

Essentially, we see high-level properties of the model, such as its overall attention allocation behavior, as emergent from the combination of lower-level mechanisms and the structure of the task and environment. This should allow us to make predictions about even very small changes of the display; for instance, the model predicts that the overlay of airspeed and altitude on the SVS is a major factor in determining the degree to which the pilots will look at the SVS, as detailed below.

7.2 Implications

Given the previous section, some of the predicted implications for SVS design are fairly straightforward. For example, at the HFES conference this past October, a field test of an SVS system was described (Prinzel, et al., 2002). This SVS, however, was different from the SVS used in the Ames study for which we have data. Specifically, the SVS which was field tested had altitude and airspeed displayed in moving bars (the way they are displayed on the PFD) overlaid, which the modeled SVS does not. Our model would suggest that this is will lead to increased SVS usage because it makes rate of change of altitude and airspeed easier to obtain from the SVS. Our task analysis shows that the rate of change of altitude and airspeed are quite critical quantities at many points in the task, so making them available on the SVS should definitely lead to increased SVS usage.

However, this is not true for all the overlaid symbology on the SVS. For example, while it is true that pilots do need heading information, this is not as critical to place on the SVS because heading information generally does not change as fast, or is needed as often, as rate of change of altitude. On the flip side, there is other symbology that could be added that would likely lead to increased SVS use. The flight path predictor (FPP) is, we believe, a good bit of symbology, but we believe, based on

the frequency with which the model needs to look at the NAV display, that this would be improved if the SVS also rendered the way points. Thus, flying to a way point would, (only in a visual sense), "simply" means keeping the FPP lined up with the way point indicator.

In general, the model predicts that the symbology overlaid on the SVS is a key component in determining how the SVS is integrated into the pilots' attention allocation strategies. This is almost certainly conditioned on phase of flight as well, suggesting perhaps that the symbology be configurable depending on phase of flight. Furthermore, we expect that the ACT-R model should be able to make predictions about the effects of adding or removing specific pieces of information.

We have gained other insights as well. First, even from the three subjects for whom data was provided by Ames, there were substantial individual differences, particularly at the more local levels. Because of this, and because such differences are likely to exist in the wider population of potential SVS users, we believe it would be a huge mistake to try to fit every aspect of these individuals' behaviors. Attempting to fit the complete scan path for any one subject would not only be laborious, it would almost certainly be an instance of fitting a great deal of noise. Just because the model is capable of generating fine-grained behavior does not mean that should be the basis of evaluation; rather, we believe more abstracted measures will do a better job of "smoothing out" individual difference noise and thus should constitute the model's criteria. We are not yet certain exactly what the best measures should be, but we believe this is an important question that we likely would not have considered without the combination of our model and the data we have in hand.

8. Progress and Lesson Learned

8.1 Progress and Advances

While there is still much more work to be done and many things to learn, we believe we have generated several advances. First, the model is not tied to any of the specifics of the scenario. If the FMC is pre-programmed correctly and the model is given a relatively modest amount of knowledge about the airport, the model should be able to run through the approach fixes for any approach and landing scenario, as long as no serious maneuvers are required. This could potentially be a win for future aviation safety research. Second, the network interface we are developing could have wide applicability, as many simulation environments (e.g. video games) use similar communication protocols; this may make it possible to connect ACT-R to a wider range of environments. This should be particularly powerful when combined with the task-oriented solution we have generated to the out-the-window vision problem.

In addition, we believe that we may have some leverage on some other high-level and abstract human factors constructs, such as "situation awareness." There is no box or section of the ACT-R architecture that one could point to as being situation awareness. Rather, we have observed that the model has to keep a number of pieces of information available at various times (some things, like altitude, all the time); the accessibility of the set of needed information about the aircraft's state might be termed the model's situation awareness, but it is not a unitary thing. It is both distributed, in that it lives in multiple declarative memory elements, and dynamic, in that different pieces are needed and "refreshed" by checking the environment at different rates. We hope ultimately that this work will lead toward more formal definitions, at least in an ACT-R context, of a number of terms from the human factors literature (e.g., situation awareness, workload) that are currently somewhat vague.

Furthermore, we are excited by the idea that previous results in the attention allocation area might be explained by lower-level mechanisms in our ACT-R framework. For instance, consider the effects of rate of change of a display item on the human sampling rate. We believe this effect falls very naturally out of ACT-R's memory system. When the model looks at the display for a particular piece of information, a representation of that information is created in declarative memory. However, ACT-R's memory decays over time, which creates a need to re-sample the environment. If the environment is re-sampled and yields the same result, rather than creating a new representation of say, airspeed, the activation of the extant representation will be incremented. This means it will take longer for that piece of information to decay, which means it will not have to be sampled as often. Thus, information that changes slowly will more often yield the same value when sampled, and thus will decay more slowly, requiring less frequent sampling.

8.2 Challenges Remaining

Doing a detailed simulation of human-in-the-loop performance in a domain this complex is fraught with challenges; many of them have already been described. Probably the biggest thing that could have gone more smoothly and should be considered for future efforts is to give the modeling teams direct access to the simulator code; the X-Plane solution we believe will ultimately work, but it has been slow going. On the positive side, there should be a downstream payoff for future efforts to link ACT-R to other systems.

Nonetheless, we realize sometimes there are limitations of time and energy for what can be provided to modeling teams, and it is clear that the rich eye-tracking data is more important than providing such linkages.

8.3 Future Directions

Obviously, there is still a great deal of work to be done to completely "close the loop;" this is currently our top priority. Once that is done, we hope to explore the design space for the SVS a little, and will try variants of the current SVS symbology to assess their impact on the model's performance. We are hoping this will lead to greater insight into the evaluation of SVS technology.

In addition, we would like to explore "de-adapting" the task analysis. One of the issues with many task analyses as they currently stand is that they include the operator's attunement to the constraints of the environment and may not be terribly useful at predicting how performance would be if the environment were different. We hope to produce a more abstract model, possibly more complex than is needed to mimic, from an input-output perspective, the over-learned routines that underlie skilled adaptation to an existing cockpit design. To us at least, a more abstract model seems to be necessary to allow us to predict the cognitive implications of novel changes to the cockpit design from a first-principles perspective.

9. References

Anderson, J. R., Bothell, D., Byrne, M. D., & Lebiere, C. (2002). An integrated theory of the mind. Manuscript sumitted to *Psychological Reivew* and available at http://actr.psy.cmu.edu/papers/403/IntegratedTheory.pdf.

Bowers, C. A. and Jentsch, F. (2001). Use of commercial, off-the-shelf, simulations for team research. In E. Salas (Ed.), *Advances in Human Performance and Cognitive Engineering Research, Vol. 1*. Mahwah, NJ: Erlbuam. (pp. 293-318).

Goodman, A., Hooey, B.L. Foyle, D.C., & Wilson, J.R. (2003). Characterizing visual performance during approach and landing with and without a synthetic vision display: A part task study. *Proceedings of the 2003 NASA Aviation Safety Program Conference on Human Performance Modeling of Approach and Landing with Augmented Displays* (David C. Foyle, Allen Goodman & Becky L. Hooey, Eds.) NASA Conference Proceedings NASA/CP-2003-212267.

Foyle, D.C., Goodman, A., & Hooey, B.L. (2003). An overview of the NASA Aviation Safety Program (AvSP) System-Wide Accident Prevention (SWAP) Human Performance Modeling Element. *Proceedings of the 2003 NASA Aviation Safety Program Conference on Human Performance Modeling of Approach and Landing with Augmented Displays* (David C. Foyle, Allen Goodman & Becky L. Hooey, Eds.) NASA Conference Proceedings NASA/CP-2003-212267

Keller, J., Leiden, K., Small. R. (2003). Cognitive task analysis of commercial jet aircraft pilots during instrument approaches for baseline and synthetic vision displays. In, Proceedings of the 2003 NASA Aviation Safety Program Conference on Human Performance Modeling of Approach and Landing with Augmented Displays, (David C. Foyle, Allen Goodman & Becky L. Hooey, Eds.). NASA Conference Proceedings NASA/CP-2003-212267.

Prinzel, L. J., Kramer, L. J., Comstock, J. R., Bailey, R. E., Hughes, M. F., Parrish, R. V. (2002). NASA Synthetic Vision EGE Flight Test. In *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting* (pp. 135-139). Santa Monica, CA: Human Factors and Ergonomics Society.

Senders, J. (1964). The human operator as a monitor and controller of multidegree of freedom systems. *IEEE Transactions on Human Factors in Electronics, HFE-5*, 2-6.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Science*. (pp 1-19).