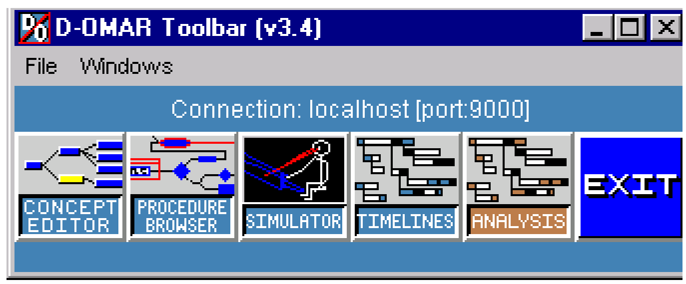
BBN Report No. 8399

Modeling the NASA SVS Part-task Scenarios in D-OMAR

Stephen Deutsch
Richard Pew



July 2004

Prepared by: BBN Technologies
10 Moulton St.
Cambridge, MA 02138

Prepared for: David Foyle
Aviation Safety Program
NASA Ames Research Center
Moffett Field, CA 94035

Table of Contents

Table of Figures

Table of Tables

1. Introduction

Human performance modeling (HPM) is a technology that has the potential to help address flight deck and air traffic controller (ATC) workplace design issues and support aircrew procedure development and evaluation. It can be used early in the workplace development process before prototypes or full-scale simulations are available or later in the design cycle to support on-going development. Unfortunately, the technology is not yet sufficiently mature that it can be applied easily and routinely. However, it is reasonable to undertake exploratory studies to evaluate aircrew procedures for employing new flight deck technologies and to evaluate their role in promoting aviation safety. At the same time, these investigations will improve the architectures and software tools that are available to support the human performance modeling process.

Under the aegis of the Aviation Safety Program, NASA Ames Research Center (ARC) initiated a program element concerned with modeling human performance. The element's goal is "to develop and demonstrate cognitive models of human performance that will aid aviation product designers in developing equipment and procedures that support pilots' tasks, are easier to use, and are less susceptible to error" (Foyle, Goodman, & Hooey, 2003). Within the framework of modeling human performance, we have sought to achieve a better understanding of the sources of human error, identify system design elements and procedures to mitigate error, and thereby contribute to improving aviation safety. The program element is a multi-year, multi-contractor endeavor that is now completing its final year tasks.

For the current phase of this research effort, NASA asked that the modeling teams examine approach and landing operations, and compare aircrew performance for baseline and Synthetic Vision System (SVS) equipped flight decks. BBN's research effort has focused on the further development of captain and first officer human performance models and the approach and landing procedures that employ baseline and SVS-equipped flight deck configurations. Human performance models for the approach, tower, and ground controllers have been developed to interact with the aircrew models to faithfully represent the airspace. The human performance modeling effort paralleled and profited from of the NASA part-task simulation studies (Goodman, Hooey, Foyle, & Wilson, 2003) that collected pilot performance data in baseline and SVS-equipped flight deck operations at the Santa Barbara Municipal Airport (SBA).

The modeling effort was accomplished using BBN's Distributed Operator Model Architecture[1] (D-OMAR) to represent the behaviors of the aircrews and air traffic controllers. D-OMAR was also used to model the aircraft and their flight decks, the ATC workplaces, and the essential features of the Santa Barbara Municipal Airport and the local airspace. Our goal has been to produce models that appropriately represent robust approach and landing performance similar to that exhibited by the pilot subjects in the recently conducted NASA part-task simulation experiments. The D-OMAR models have successfully executed nine of the part-task scenarios.

Our findings with respect to the scenarios provided the starting point for the just completed stage of our research effort. The findings were derived from our evaluation of the results of the part task experiments, the process of generating the modeled version of the experiments, and the comparison of the model results with the human subject results. A key finding from the part-task simulation study suggested that when provided with an SVS as a second attitude instrument, pilots shift the balance between attending to the attitude instruments and the navigation instrument toward the attitude instruments (Deutsch & Pew, 2003). In seeking to mitigate this effect, we "designed" an Enhanced-SVS that combines traditional PFD and the SVS functionalities in a single attitude instrument and used the modeling environment to explore the aircrew's use of the instrument in the SVS scenarios. The goal of the experiment was to provide

[1] The D-OMAR software with user manual is available as OpenSource at http://omar.bbn.com.

the basis for restoring the baseline attitude- verses navigation-instrument scan-allocation while preserving the advantages of the SVS.

A second finding related to the subject response to the part-task simulation rule to execute a go-around if the SVS is found to be misaligned. That two of the three human subjects did not immediately elect a go-around was not a complete surprise. In modeling the terrain mismatch scenario, we explored two scenario outcomes: following the rule to immediately initiate a go-around and continuing the approach and landing.

In this document, the BBN Final Report on the SVS modeling portion of the program, Section 2 traces the development of the human performance models, the aircrew and air traffic controller procedures that they execute, and the strategy for building the emulation of the NASA part-task scenarios. Section 3 traces the design and evaluation of an Enhanced-SVS, and the motivation for and the modeling of alternative pilot responses to the detection of the SVS misalignment. Section 4 provides background on the D-OMAR modeling environment, a brief discussion of the cognitive architecture for the models, and additional detail on the multi-tasking behaviors of the models.

An important NASA goal for the program element (Foyle et al., 2003) was to foster improvements in the underlying technologies supporting the development of human performance models. Section 5 provides a discussion of the areas in D-OMAR that were extended to meet the requirements of the modeling tasks. Section 6 provides a discussion of conclusions derived from the model research effort and it contributions toward improving aviation safety.

In the discussion that follows, we will at times revisit the same materials from different perspectives as we discuss: (1) What the NASA part-task simulation data said; (2) The development of the models for the scenarios; (3) What was learned from the models as they were developed; (4) What was learned from the follow-on exercises exploring the findings using the model, (5) What was it about the models' cognitive theory and architecture that served to support what was accomplished, and (6) What improvements in the models were motivated by the modeling activity.

2. Modeling the NASA Part-task Scenarios

The initial tasks of the research effort focused on the use of the D-OMAR simulation framework to build the necessary modeling elements to recreate the NASA part-task simulation environment. We then developed and executed scenarios corresponding to those explored in the NASA experiments (Goodman et al., 2003). Existing human performance models for the aircrews were refined with the goal of replicating the performance of the NASA experiment's human subjects. In this section, we describe the approach to developing the aircrew models, the principal elements for the simulation environment, the model runs that reproduced the part-task scenarios, and the assessment of the data from the model trials to support the validation of the models. We return to the issue of model validation again in Section 3.1.3.

2.1. NASA Support for the Modeling Effort

NASA Ames made available several important resources to facilitate the modeling effort for the approach and landing trials at Santa Barbara Municipal Airport. Unlike the ILS approach used in the earlier O'Hare scenarios (Deutsch & Pew, 2001: Deutsch & Pew, 2002), the part-task simulation dictated an RNAV approach. To assist the modeling teams in the transition to the approach for SBA, a cognitive task analysis (Keller & Leiden, 2002a) provided a detailed description of aircrew and air traffic controller procedures for the RNAV approach. The document included information on procedures and flight deck systems used to support an RNAV approach. An addendum to the document (Keller & Leiden, 2002b) extended the task analysis to cover the aircrew's use of an SVS during the approach and landing. These documents provided much of the information necessary to construct the D-OMAR goal, sub-goal, and procedure network that represent the standard aircrew and air traffic controller procedures for the approach and landing. Goodman et al. (2003) included the RNAV approach plate for SBA runway 33L (Figure 1) developed by NASA for the part-task experiment.

Our modeling effort relied heavily on the documentation of and data from the NASA baseline and SVS part-task scenario trials. Goodman et al. (2003) provided a detailed description of the simulated flight deck, the design for the ten scenario trials, and a description of the scenario trial data collected for the three subjects. Trial data included simulation output for each run, eye tracker data, and video (with audio) recordings based on an eye-tracker camera and a room-view camera. The eye tracker data included both fixation sequence and dwell sequence data files. The data were made available by trial with summary statistics available by flight path segment. A summary dwell sequence spreadsheet was assembled to provide duration percentage for viewing each instrument across all trials and all subjects. The spreadsheet made it possible to compare and contrast individual subject eye tracking behaviors across the trials. In summary, these data provided a comprehensive view of aircrew behaviors essential to modeling the baseline and SVS-assisted approach and landing trials.

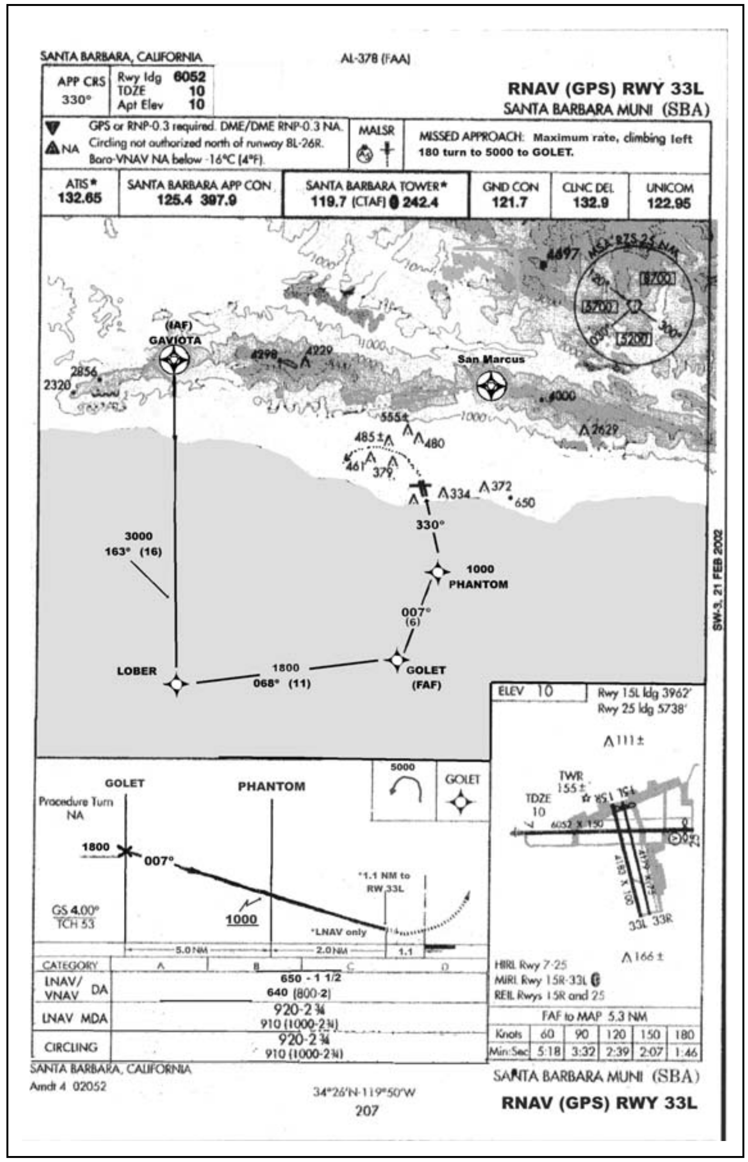


Figure 1 RNAV Approach Plate for SBA 33L

The BBN team also participated in and profited from an SVS/SWAP information-sharing workshop held at NASA Langley Research Center late in 2001. The workshop provided presentations on the development of the NASA SVS system as well as a review of the flight test experiments in using the system. The SVS section of the NASA Aviation Safety Program (AvSP) web pages provided access to a number of publications that included the Concept of Operations for Commercial and Business Aircraft Synthetic Vision Systems (Williams, Waller, Koelling, Burdette, Doyle, Capron, Barry, & Gifford, 2001) and the Synthetic Vision System (SVS) Concept Assessment Report (Norman, 2002).

2.2. Strategy for Developing the Modeled Scenarios

The NASA SVS part-task simulation study (Goodman et al., 2003) included ten scenarios that selectively covered three variables of interest: display configuration, visibility conditions, and approach events. There were two display configurations: a baseline flight deck consisting mainly of a primary flight display (PFD), a horizontal situation indicator (HSI), a side stick controller, and a mode control panel (MCP), and a second configuration in which the baseline configuration was augmented with an SVS display. The SVS was a 16.7-inch wide by 7.5-inch high head-down display with a 31 degree horizontal and 23 degree vertical field of view—a moderately wide-angle with respect to unity (Goodman et al., 2003). Landing gear, flaps, speed brakes, throttle settings, and map scale for the HSI were set by the first officer using a separate display panel.

Visibility conditions included visual meteorological condition (VMC) and instrument meteorological condition (IMC). The light haze in VMC allowed visual flight rules. Under IMC, the approach controller provided the aircrews with a *reported* 800-foot cloud ceiling. The approach event conditions included a nominal approach, a "late reassignment" scenario with a runway situation requiring a side-step to a closely parallel runway, a missed approach scenario condition dictating a go-around, and a "terrain mismatch" scenario in which the SVS was found to be laterally misaligned from the actual runway as the aircraft emerged from the cloud cover. The "go-around" scenarios took two forms: in the VMC case, traffic on the runway made the go-around necessary. In the IMC case, in spite of the reported 800-foot ceiling, the aircraft was still in the cloud cover at the 650-foot decision height; hence, the captain was unable to acquire the runway making the go-around necessary.

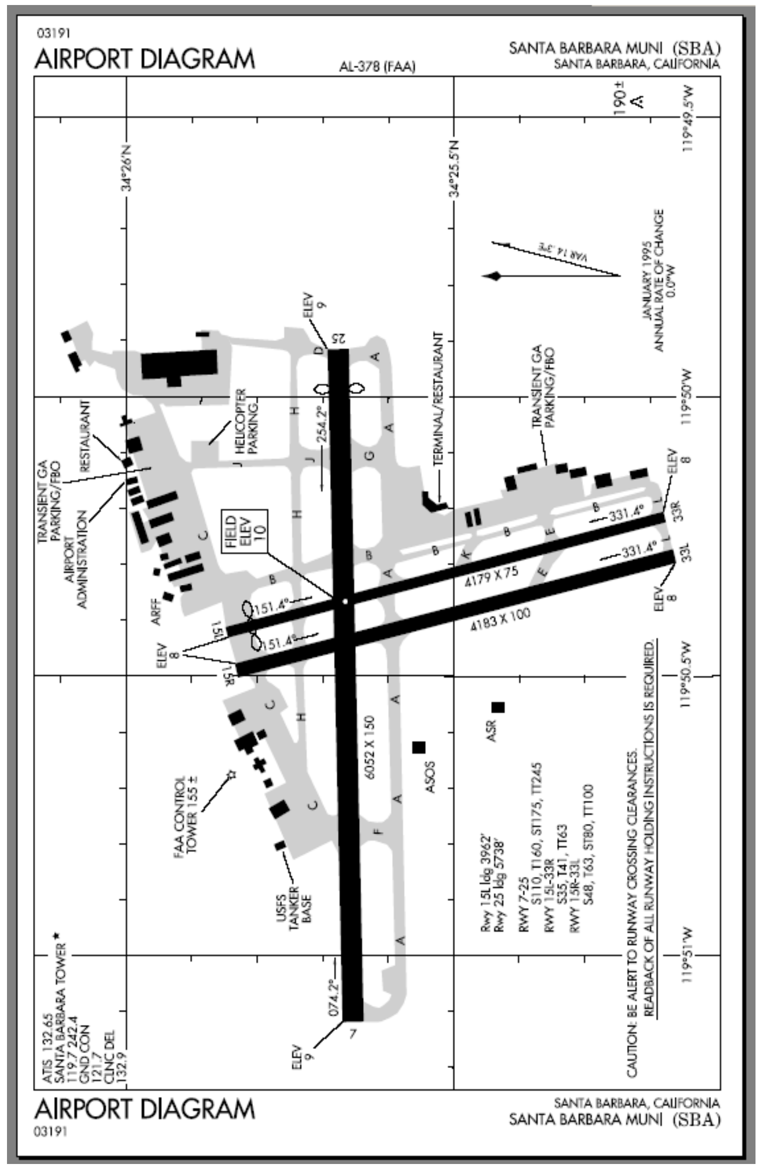


Figure 2 Airport Diagram for Santa Barbara Municipal Airport

Constructing the SBA airport model and local airspace model required data from several sources. The data structures used in the model were based on those developed for the previously modeled O'Hare scenarios. Information on Santa Barbara Municipal Airport was obtained primarily from the AIRNAV.COM web page for the airport. The web page included latitude-longitude information for the endpoints of the runways, information on radio frequencies for the controllers, and a link to the airport diagram (Figure 2). The approach plate for SBA runway 33L (Figure 1) provided information on key elements of the local airspace. The AIRNAV.COM pages also provided information on the radio navigation aids for the approach to SBA runway 33L. The navigation aid information obtained included location latitude and longitude, and radio frequencies. The availability of an airport physical description database would certainly help the process of constructing an airport model. The developers and maintainers of AIRNAV.COM have provided an important first step in this direction.

As modelers, we were asked to focus initially on the "late reassignment" scenarios using the baseline and SVS-equipped flight decks. Our approach was to first develop the basic aircrew and air traffic controller procedures to support the RNAV approach using the baseline flight deck under VMC. With those basic scenario elements in place, we then added cloud cover to the model to represent IMC. For the modeled IMC condition (excluding the go-around condition) scenarios, the *actual* cloud ceiling was at 800 feet; hence, the breakout from the cloud cover occurred 150 feet above the 650-foot decision height designated by the approach plate.

With the nominal VMC and IMC scenarios operational, we were ready to address the "late reassignment" condition. We developed an extended scenario with two closely spaced aircraft on the approach to SBA runway 33L. Our goal was to create a realistic situation that forced the late reassignment option. As the modeled *lead* aircraft landed, it blew a tire and temporarily held on the active runway. The aircrew notified the ATC of the problem and the ATC then addressed the situation by offering the *trailing* aircraft the opportunity to side step to the parallel runway 33R. The captain of the trailing aircraft as pilot-flying readily directed the first officer to communicate the acceptance of the side-step offer to the controller and completed the landing on 33R.

To model the IMC "go-around" condition, the reported 800-foot cloud cover was modeled as actually as extending all the way to ground level. Under these conditions, neither the airport nor the runway could be visually acquired out the window as the aircraft descended through the established decision height and the captain was forced to elect a go-around.

Lastly, we added the SVS to the flight deck. The modeled SVS included the basic information elements of the NASA part-task experiment SVS as shown in Figure 7. Detailed information on the part-task experiment SVS is available in Goodman et al. (2003). As in the NASA part-task scenarios, a single SVS was provided for use only by the captain. Based on the NASA-provided cognitive task analysis (Keller and Leiden, 2002b), the aircrew procedures were extended to include the use of the SVS. With the SVS-equipped flight deck and aircrew procedures in place, we were then able to run the four IMC scenarios using the SVS-equipped flight deck.

At this point, the modeled aircrews have successfully executed nine of the ten NASA part-task simulation scenarios (Goodman et al., 2003):

- Scenario 1 – nominal approach using the baseline configuration in VMC
- Scenario 2 – late reassignment approach using the baseline configuration in VMC
- Scenario 4 – nominal approach using the baseline configuration in IMC
- Scenario 5 – missed approach using the baseline configuration in IMC
- Scenario 6 – terrain mismatch using the baseline configuration in IMC
- Scenario 7 – nominal approach using the SVS configuration in IMC
- Scenario 8 – late reassignment approach using the SVS configuration in IMC
- Scenario 9 – missed approach using the SVS configuration in IMC
- Scenario 10 – terrain mismatch using the SVS configuration in IMC

The missed approach in VMC (Scenario 3), the one scenario that has not yet been replicated using the model is currently a work in progress. The tactic being employed for setting up the condition is to use a variation on the late reassignment scenario. For Scenario 3, there will be a NOTAM indicating that runway 33R is closed for repairs. With the lead aircraft temporarily blocking runway 33L due to a blown tire, the controller will have no option other than requesting that the trailing aircraft execute a go-around.

The modeled scenarios differed in two respects from the part-task scenarios. The first difference was in the role played by the first officer. In the part-task experiments, the first officer role was played by a surrogate who by design adopted a passive stance. In adapting the passive stance, the surrogate acted in response to specific requests from the experiment subject, the captain, and delayed necessary interventions as long as reasonable possible. In contrast with the passive role adopted by the first officer in the part-task simulation, the modeled first officer assumed an active stance much like that expected of an experienced commercial pilot.

The second difference between the part-task and the modeled scenarios related to the termination of the scenario trials. For those scenarios that would otherwise have lead to a successful landing, the part-task trials were terminated as the aircraft descended to an altitude in the vicinity of 50 to 100 feet above field elevation (AFE). In our modeling of the scenarios, we carried the scenarios through the landing and the taxi operations. The scenarios were terminated as the landing aircraft approached its designated concourse.

In the late reassignment scenario in particular, we used the landing and taxi capabilities for the lead aircraft to create the situation that forced the late reassignment option. Adding the landing and taxi capability was a comparatively straightforward process. We had built the aircrew and aircraft procedures for the landing and taxi operations as part of our earlier research into aircrew error during taxi operations at O'Hare (Deutsch & Pew, 2001; 2002). With the aircrew and ATC landing and taxi procedures already in place, it was largely a matter of providing the data required to support the execution of the procedures. We had already developed an SBA airport model with the runways for the scenarios. The taxiways were added based on information from the airport diagram. Information on runway and taxiway signage was not available and had to be constructed as needed to support taxi operations. The ground controller was provided with appropriate taxi directives to pass to the aircrews as they exited the active runway. The aircrews were then able to act on the taxi directives based on the runway and taxiway signage. With these extensions in place, we were able to set up and execute the late reassignment scenario. The changes made it possible to continue all of the scenarios (with the exception of the go-around scenarios) through landing and taxi operations as was done for the O'Hare scenarios.

2.3. Scenario Execution and Model Validation

The D-OMAR aircrews, much like real aircrews in similar situations, readily accomplished the nine modeled scenarios. For the baseline flight deck scenarios in VMC (Scenario 1) and IMC (Scenario 4) conditions at SBA, the modeled aircrews successfully executed the approach and landing using RNAV procedures much as the human subjects did in the part-task simulation. The story was much the same for the nominal approach in IMC conditions using the SVS-equipped flight deck (Scenario 7). When on the baseline VMC approach and in the SVS-equipped IMC approach, the tower controller requested that the aircrew side step from SBA runway 33L to the closely parallel runway 33R (Scenarios 2 and 8), the captain instructed the first officer to accept the request and then successfully executed the side step maneuver to runway 33R. For the missed approach scenarios using the baseline flight deck (Scenario 5) and the SVS-equipped flight deck (Scenario 9), the cloud cover persisted through the 650-foot decision height making it impossible to visually acquire the airport and runway, and in each case, the captain elected to go-around. As specified on the approach plate for SBA runway 33L, go-around procedures required a climb to 5000 feet and a 180-degree turn to the left to GOLET. The go-around scenarios were terminated half way through the turn.

In the terrain mismatch scenarios (Scenarios 6 and 10), the modeled aircrew sometimes followed the instructions that dictated a go-around on detecting the SVS misalignment and sometimes made the flight path adjustments necessary to successfully complete the landing. The terrain mismatch scenarios and the motivation for developing the alternate outcomes are discussed in Section 3.2.

The successful of execution of the scenarios on the part of the aircrews and the air traffic controllers was an important milestone in the *validation* of the human performance models. It established the basic capability of the models' goals and procedures to address a range of scenario events in a pilot-like manner. In the process of developing the models and then refining the models once the scenarios were being successfully executed, the performance of the models was reviewed at several levels of detail. The review of model performance took place either during scenario execution or by reviewing data collected during each simulation run. A time-tagged on-line trace (Figure 3) tracked the aircrew's conversation on the flight deck as well as the exchanges with the controllers managing the airspace. The trace also tracked flight deck actions taken by the aircrew that followed from this discourse. A more detailed view of aircrew performance was reviewed using the Gantt-style task timeline display (Figure 4) of goal and procedure execution by the captain and first officer. This display was used to review and evaluate aircrew and air traffic controller performance at the task level by examining the timeline for goal, sub-goal, and procedure execution leading to task completions. The task timeline display also made possible to evaluate the multi-tasking behaviors of the aircrews and controllers. In the following sections, we first review the details of the aircrew and air traffic controller procedures as they were developed for the scenarios and then discuss the validation of the human performance models.

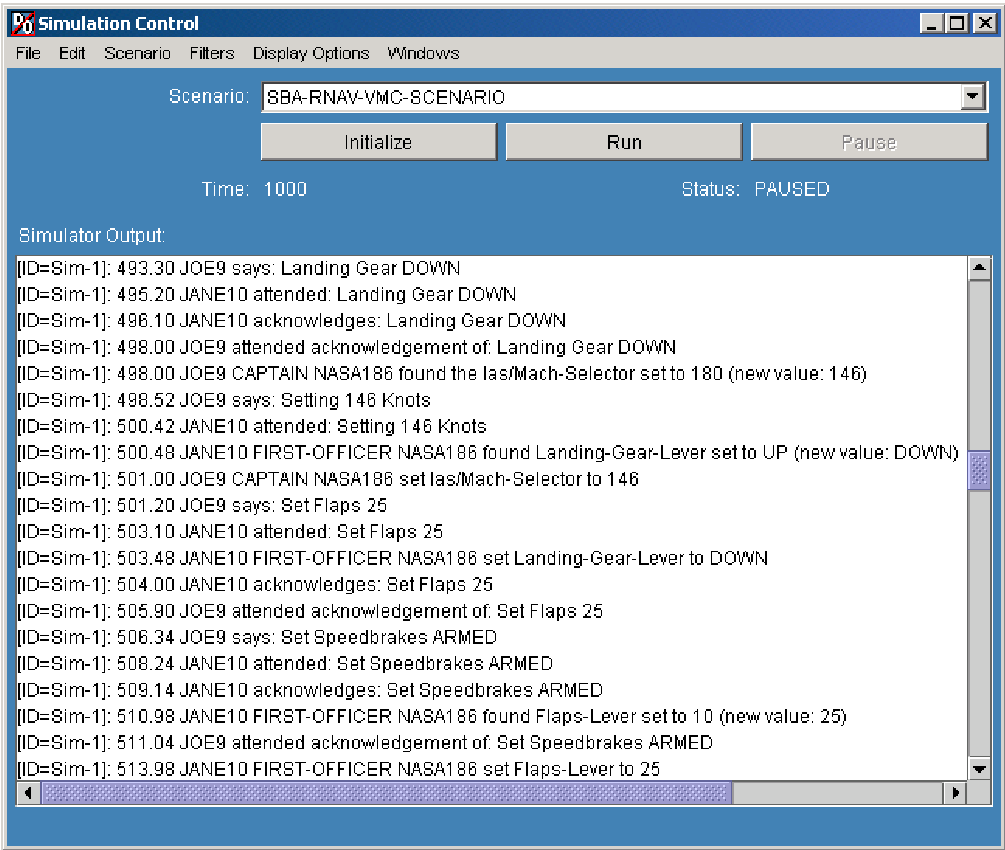


Figure 3 Simulation Control Panel and On-line Trace

2.3.1 Modeling the Aircrews, the ATCs, and their Operating Procedures

In the modeled scenarios, as in real-world operations, the captain and first officer must work together to execute the approach-and-landing procedures. In responding to a series of controller directives, actions must be prioritized, appropriate aircrew communications must be generated to coordinate the execution of these actions, and interrupts in the form of further directions from the controllers must be handled. The interrupts are not unexpected, but rather meet expectations consistent with the local air traffic and weather conditions. Reactive behaviors are determined within the framework of the aircrew's active goals and procedures. In meeting their responsibilities, the captain and first officer have a significant number of tasks in process, each of which requires a coordinated mix of perceptual, cognitive, and motor skills. The scenarios create situations in which the response to demands must be appropriately prioritized to achieve acceptable performance.

Each aircraft is populated by human performance models for the captain and first officer. The aircrew models are extensions of the models that executed the ILS approach, landing, and taxi procedures in the O'Hare scenarios (Deutsch & Pew, 2001; Deutsch & Pew, 2002). The new RNAV procedures that they employ at SBA are based on the Keller and Leiden (2002a; 2002b) cognitive task analysis. In the modeled scenarios, each controller is represented by a human performance model.

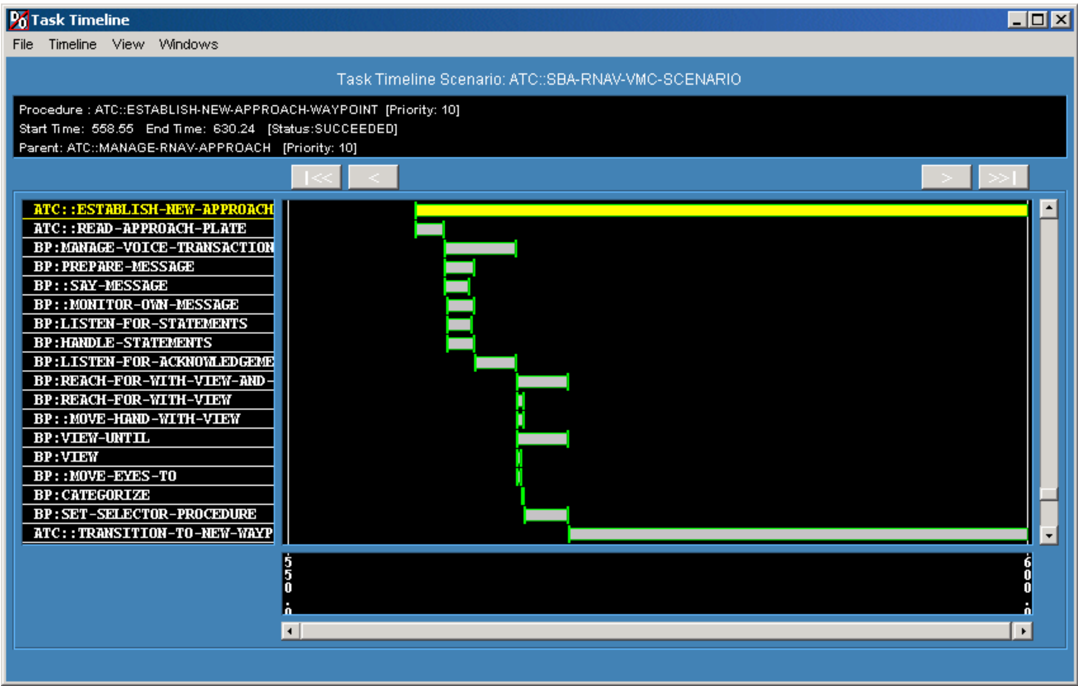


Figure 4 Task Timeline Display

The modeled approach and landing scenarios follow the standard progression in the airspace transiting from approach controller to tower controller to ground controller and terminating as the aircraft completes the landing and taxies to its assigned concourse. In the scenarios in which a go-around was required, the scenarios terminated shortly as the climb out with the required left turn was initiated. The scenarios as modeled more closely follow real world operations than was possible in the part-task simulation.

The aircraft model includes the instruments and controls necessary for the crew to execute the required approach and landing scenarios. The flight deck portion of the model, as originally developed, closely emulated the actual 757 flight deck. Figure 8 provides a view of a portion of the 767 flight deck. As the sister aircraft to the 757, the 767 flight deck closely resembles that of the 757, making the 767 web pages a reasonable source for information on the flight deck instruments. The NASA part-task simulation flight deck is presented in Figure 7. While different in physical appearance, the 757/767 and part-task simulation flight decks are similar at an information level, the level at which the model was implemented.

The principal flight deck instruments developed for the model include the PFD, air speed and altitude instruments, the HSI, and the SVS. The PFD includes annunciators for LNAV and VNAV. Modeled flight controls include the MCP, switches for the autopilot, and levers for the throttles, flaps, landing gear, and speed brakes. The central instrument panel includes lights providing landing gear status.

The aircrew models make use of a modeled approach plate for SBA runway 33L (see Figure 1 for the approach plate used by for the part-task experiment on which the approach plate model was based) for information related to the RNAV approach. The information used by the captain to brief the go-around contingency at the beginning of the approach is derived from the approach plate. The approach plate is subsequently used by the aircrew as a reference to track the sequence of fixes along the approach path. As the aircrew monitors the aircraft's transition through each of the approach path fixes, the approach plate is the aircrew's source for the heading and altitude for the new leg. Later in the scenarios, following the landing, as the aircraft departs the active runway, the first officer uses the SBA airport diagram (Figure 2) to support taxi operations leading to the concourse.

Voice communication between the captain and first officer is used to coordinate the execution of approach and landing procedures. Party-line radio communication is modeled with the aircrew resetting the radio frequency as they transition from one controller to the next. Even in a scenario with one or two aircraft, it is not uncommon for an intra-aircrew conversation to be interrupted by an air traffic controller directive. Careful attention has been paid to the fine details of interleaving of aircrew and air traffic controller conversations and to handling air traffic controller interruptions to aircrew conversations. Section 5.1 describes enhancements made to the conversation model that were required to met the demands of SBA scenario situations.

As each scenario begins, the captain is the pilot-flying and the first officer, as the pilot-not-flying, handles communications with the air traffic controllers. The approach controller clears NASA186 (and the lead aircraft, NASA277 in the two aircraft scenarios) for the approach to SBA runway 33L and provides information on VMC or IMC depending on the scenario. Using information from the approach controller and the approach plate for runway 33L, the captain continues the approach by reviewing the runway and weather information with the first officer. The captain briefs the go-around procedures based on the information read from the approach plate. For the RNAV approach, the captain asks that MCP modes LNAV and VNAV be set and checks the PFD annunciator mode lights to verify the first officers actions in setting the proper modes.

The aircrew then focuses on navigation as the aircraft proceeds along the flight path from one fix to the next under the control of the flight management computer (FMC). As they approach each fix, the captain calls for a new MCP altitude setting for the next leg based on information read from the approach plate. The first officer sets the MCP altitude selector and the captain verifies the setting. The aircrew then monitors the aircraft's heading and altitude changes (information derived from the HSI and the PFD) as the aircraft transitions onto the flight segment to the next fix. They continue to monitor the heading until the new desired heading is fully established. They monitor altitude to assure that they hold at the selected target altitude. As the approach progresses, the captain calls for a series of flap settings consistent with their speed and position along the approach path.

The handoff from the approach controller to the tower takes place along the leg to GOLET (see Figure 1). The approach controller provides the radio frequency for the tower, the captain and first officer set the required frequency, and the first officer contacts the tower. The tower controller immediately grants the aircrew the clearance to land. The captain asks for the final flap setting, that the landing gear be lowered, and that the speed brakes be armed. At this point, the captain asks for execution of the *landing checklist*, which is then acted on with the first officer. Reading from the checklist, the first officer cycles through checks with the captain for the landing gear status, the flaps setting, and the speed brakes setting.

As the aircraft continues its descent, the first officer monitors the aircraft's altitude and makes call-outs at 1000 feet AFE, as they approach decision height, and at 100 feet AFE. The captain is responsible for the out-the-window sighting of the airport and the runway, and making the decision to land. Under VMC, the captain can readily acquire the airport and the runway out the window well before decision height. For the SVS-equipped, IMC-condition scenario, the captain uses the SVS to acquire the airport and runway before they break out of the cloud cover, but must still visually acquire the runway out the window to support the decision to land. In IMC, based on the weather briefing from the controller, the captain anticipates the break out from the cloud cover at 800 feet and has adequate time to acquire the runway before the 650-foot decision height. Depending on the outcome of the decision to land (based on visually acquiring the runway before decision height), the captain will either take “manual” control of the aircraft to complete the landing or elect the preprogrammed go-around maneuver.

Figure 5 provides a D-OMAR plan-view display of two aircraft on their approach to SBA 33L in the “late reassignment” scenario. At the point presented in the plan view, the lead aircraft has blown a tire on landing causing it to hold temporarily on the active runway. The top panel on the right in Figure 5 records the conversations among the approach and tower controllers and the two aircraft. NASA277’s communication with the tower controller related to the blown tire shows up in the recorded dialog. The tower controller addresses the situation by asking the trailing aircraft, NASA186, to side step to runway 33R.

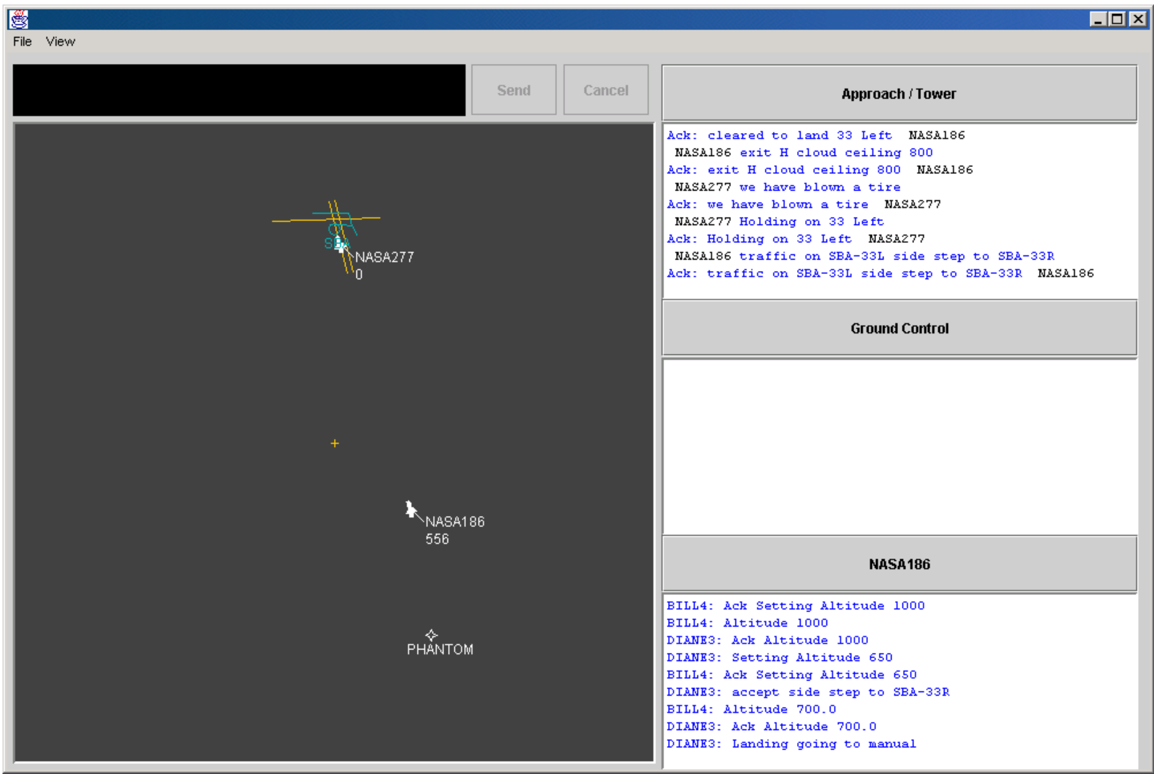


Figure 5 Screen View from the Late Reassignment Scenario

The lower panel on the right records details of the conversation on the flight deck of the trailing aircraft, NASA186. Diane, the NASA186 captain, tells her first officer that she will accept the request to side step to 33R. The first officer communicates the captain’s acceptance of the side step request to the tower controller. The first officer’s response completes the side step transaction with the tower controller as related in the upper display panel. The captain then announces her decision to proceed with the landing. As they land, the captain manages speed brake and reverse thrust settings, while the first officer calls out current ground speed following weight-on-wheels. Subsequent to the landing, NASA186 notifies the tower as the aircraft leaves the active runway, contacts ground controller, and receives directions to taxi to its assigned concourse. The scenario is terminated as the aircraft approaches the concourse.

2.3.2 Validation—Assessing Model Behaviors

D-OMAR simulation tools provide explicit measures of model behaviors. They are essential in the *assessment* and *validation* of model performance just as they are essential for managing the complexity required to create the models. Aircrew behaviors are frequently multi-task behaviors—the successful, or occasionally less successful integration of the demands of several ongoing tasks. Each of these tasks may be made up of several steps that require the coordinated interaction of several human functional capabilities (e.g., maintaining an intra-crew conversation and the coordination of hand-eye actions to set the altitude selector). The execution of a checklist interrupted by an ATC communication is at once a common occurrence and a challenging event sequence to faithfully model. As the detailed model behaviors are developed, the assessment process is used assure that the varied situations encountered across the scenarios are addressed by reasonable, human-like aircrew model performance. The models may be said to be validated to the extent that the assessment yields satisfactory results. Many elements of the assessment are quantitative; some are qualitative. Multiple levels of visibility into model behaviors are essential both to develop the aircrew procedures and to conduct the assessment to support the validation of model performance.

D-OMAR graphical display tools, as illustrated earlier in this section, each provide a unique view into model behaviors. A plan view, as illustrated in Figure 5, allows an observer to monitor the progress of the aircraft along its flight path. The plan view display has recently been supplemented by a similar HSI-like display (Figure 6). A Gantt chart display (see Figure 4) provides detailed information on goals and procedures as executed by the captain and first officer. An event timeline (not illustrated) provides detailed insight into the behaviors of the publish-subscribe protocol used to coordinate procedure execution. Lastly, a detailed event log is recorded for each simulation run with key events displayed in the trace pane of the simulation control panel (see Figure 3) as the simulation progresses. The event log serves as the data source for each of the aforementioned display types.

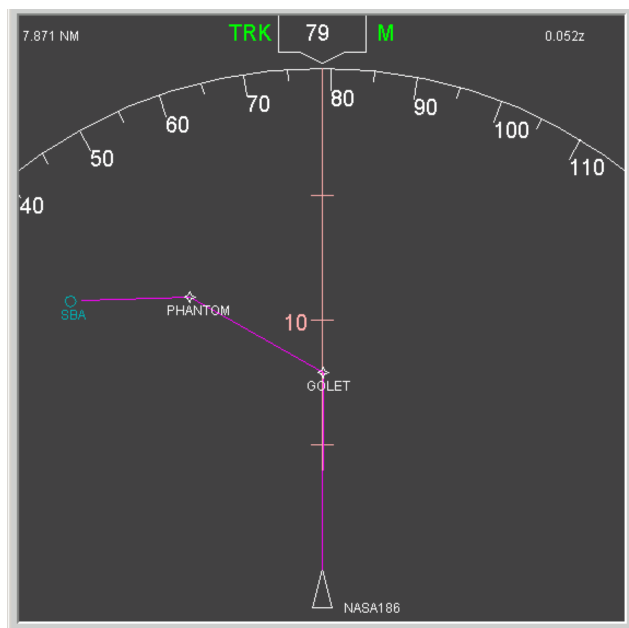


Figure 6 Horizontal Situation Indicator (HSI)

Some of the evaluation tools operate concurrently with the simulation; others are used while the simulation is paused or once a simulation run has completed. The plan views and the simulation trace operate concurrently with the simulation, the task and event timelines are available once the simulation has been paused. An event recording system is used to capture the data that supports the evaluation tool presentations. Several event types are basic elements of the D-OMAR simulator, others are more specialized and created to address the data capture requirements of a particular domain or scenario.

Particular event types have been created to track the behaviors of the human performance models for the aircrew and air traffic controllers. One of the event types records flight deck actions taken by each aircrew member (and workplace actions for the air traffic controllers); for example, the setting of an MCP selector for altitude or the movement of the lever to establish a particular flap setting. Since in-person conversations on the flight deck and party-line conversations with the air traffic controllers are so important, conversation events have been defined to record these conversations. The aircrew and air traffic controller conversations presented in the right hand panels of Figure 5 are generated by “after methods” on the event recording process.

The event types for flight deck actions and conversations are key elements providing data for the on-line trace. They record and print the actions taken by the aircrew and the air traffic controllers. As such, they represent the outcomes of the execution of the goals and procedures whose development was based on the cognitive task analysis for the RNAV approach and landing (Keller & Leiden, 2002a).

The simulation control panel (Figure 3) provides an interface for the investigator to select and manage the execution of a scenario. During scenario execution, the panel provides an on-line trace of selected scenario events. The “Scenario” line in the panel provides for the selection of the scenario to be executed. The “Initialize,” “Run,” and “Pause” buttons enable the investigator to control the flow of scenario execution. The D-OMAR simulator is capable of both real-time and fast-time operation. In fast-time mode, the simulator is very efficient taking about 40 seconds to complete 1150 seconds of real-time for the two-aircraft “late reassignment” scenarios. On a newly acquired laptop computer, the time is now down to less than 25 seconds.

The sample control panel presented in Figure 3 includes a short section of the trace from the nominal VMC approach scenario, the SBA-RNAV-VMC-SCENARIO. At this point in the scenario, the captain has just asked that the landing gear be lowered, the flaps be set to 25 degrees, and the speed brakes be armed. As recorded in the trace, the first officer attends to each request and executes each requested action, first checking the current setting and then adjusting the setting as necessary. Part way through the sequence of requests for actions on the part of the first officer, the captain adjusts the MCP IAS/mach speed selector and consistent with established procedures, announces the change as it is made.

The proportionally sized slider at the side of the Figure 3 provides a hint at the extent of the trace for the nominal VMC scenario. The trace is lengthy, but manageable in size, in large part, because the printing for most event types is turned off. For example, procedure event data that is essential to the task timeline described below, would if printed, make it far more difficult to isolate the important events currently presented. The ability to capture a complete set of scenario event data, yet tailor the trace content is important in making the trace a useful model assessment and validation tool.

The on-line trace has been specialized to provide insight into critical actions taken by aircrew and air traffic controller models in the scenarios. The actions are the products of the execution of clusters of a captain’s or first officer’s goals and procedures that each has duration in time. The task timeline display is used to provide insight into how an agent’s goals and procedures play out concurrently in time to generate these actions. The display provides detailed information on procedure execution that includes the start and stop time for the procedure, the procedure’s priority, the procedure’s parent and child procedures, the success or failure status on completion, and information on interruptions to procedure execution. Figure 4 is an example of the timeline display from the baseline VMC scenario. The slider at the right of the figure is quite small indicating that the captain has quite a large set of on-going goals and procedures beyond those currently visible in the display. Some of these goals and procedures represent preparedness to respond to anticipated events; some are currently active in the timeframe covered by the display. The selected timeframe for the display, identified in the display’s bottom panel, is fifty seconds, from 550 seconds into the scenario to 600 seconds in the scenario. The aircraft is approaching GOLET and about to transition to the leg to PHANTOM on the approach to SBA 33L (see Figure 1).

The procedures that appear in the timeline in Figure 4 are a subset of the captain's procedures as s/he prepares for and then monitors the transition to PHANTOM, the next leg of the approach. Each line of the display represents one of the captain's goals or procedures. The goal or procedure's name appears in the panel at the left, the bar represents the duration of the procedure's execution. Time periods for which a procedure is interrupted (there were none for these procedures) are indicated within the bar representing the procedure's duration. The mouse (not shown), over the first procedure, causes the top panel to be filled with that procedure's primary attributes: the procedure's name and priority, its start time, its end time if it has completed, its current status, and the name of its parent goal or procedure.

In the procedures shown in Figure 4, the captain is concerned with the basic actions for establishing the aircraft on the new leg. The captain first reads the approach plate to verify the altitude for the next waypoint (executed as a single procedure). The captain then tells the first officer that s/he is setting the MCP with the new desired altitude (the next seven lines of the display cover the verbal transaction: speaking the message and listening to the acknowledgement). The captain then sets the MCP IAS/mach speed selector (a coordinated hand-eye activity accomplished by the next eight lines of the display). In the last line shown in the timeline display, the captain is concerned with monitoring the aircraft's transition to the new heading and altitude. Information on current heading is derived from the on-going scan of the HSI display. Information on current altitude is derived from the PFD and the SVS for the scenarios in which it is available. The monitoring procedure obtains this information by "subscribing" to scan procedures' that "publish" the information. The publish-subscribe protocol provides a basic mechanism to coordinate procedure execution and move information among procedures. It is more fully described in the Section 4.3.

Each of the assessment tools described has opened a new window of visibility into model performance and with each new window, our experience has been that while we see new aspects of successful model performance, we also find new layers of model shortcomings. When our model was first connected to a simple anthropometric model, we found that it would reach for a selector to set a new value and happily leave its arm extended to maintain the reach to the selector until the next operation requiring the use of that hand. In viewing the textual trace, this very clear shortcoming in the model had never been detected. There is now a background procedure by which a model will return its arm to a suitable rest position and indeed, move its arm among rest positions when not otherwise called upon.

Validation will be as good as the capabilities of the tools used to assess behaviors; hence, considerable effort has been devoted to these tools. Validation will be useful to the extent that the assessment covers a range of situations appropriated for the intended use of the models. Up to this point, the focus of our validation discussion has been on assuring that the models for the captain and first officer collaborate effectively, respond appropriately to air traffic controllers, and *robustly* execute the broad range of specific approach and landing procedures much as the skilled human crews did in the part-task simulation and in real-world situations.

The behaviors of application-critical aspects of model behavior deserve a special role in model validation. Pilot scanning patterns are such a behavior. Data relevant to the comparison of human subject and model scanning patterns are presented in Table 1 and Table 2. As discussed in Section 5.2, much of the approach and landing sequence was temporally under-constrained, which partially explains the variations in averaged instrument dwell times across subjects. Average instrument dwell times by flight segment for the model generally fell within the range of the human subject times and thereby supported the validation of the model. In a recent related experiment, Wickens, Alexander, Thomas, Horrey, Nunes, and Hardy (2004) similarly noted a "broad range in looking behavior which can sustain roughly equivalent levels of performance." The visual behaviors of the model adequately support task execution in a human-like manner, however the broad range of suitable human scan behaviors implies that this is not a stringent validation criteria for the model.

3. Following up on the Initial Findings

To emulate the NASA part-task simulation study we assembled models of the airport and airspace entities as the basis for the necessary test bed. The aircrew and air traffic controller models, the aircraft and their flight decks, the ATC workplaces completed the simulation environment. It was then possible to set up the conditions that were necessary to examine model behaviors in each of the part-task simulation scenarios. The assessment conducted in the course of building the scenarios served as the initial validation of the human performance models for the aircrews. The findings developed in reviewing the part-task simulation data and in the work of reproducing the scenarios laid the foundation for the further research effort that is reported on here.

The first finding pursued was the concern that the addition of the SVS as a second attitude instrument appeared to alter time-tested pilot instrument scan patterns. While the underlying tasks to be accomplished remained the same, when the SVS was added to the flight deck, the balance in time that subjects devoted to the attitude displays (the PFD and the SVS) increase relative to time devoted to the navigation display (the HSI). To address this situation, we “designed” a single attitude instrument, an Enhance-SVS that combined the essential features of the PFD and the SVS in a single attitude instrument. We then ran the SVS part-task scenarios using the Enhanced-SVS with the goal of restoring the original pilot scan pattern.

The second finding pursued was the concern over potential pilot response to the rule that dictated a go-around when the SVS runway display was found to be misaligned with the actual runway. Only one of the three subjects in the NASA part-task scenarios immediately commanded a go-around when the SVS was found to be laterally misaligned with the runway as seen out the window. The model of pilot behaviors for the misalignment scenario was extended to reflect the possibility that rather than immediately electing a go-around, some pilots might pursue the landing in the misalignment condition.

A finding that was not pursued related to the balance in pilot time devoted to the out-the-window view and the use of the SVS immediately before landing. In the part-task experiment, the addition of the SVS display to the flight deck augmented the out-the-window view while at the same time providing much of the same attitude information as the PFD. During the flight phase from decision height to landing, there were individual differences in the behaviors of the three subjects in the part-task experiment. While Subjects 4 and 5 reverted to the use of the out-the-window view when the aircraft descended below the 800-foot cloud cover, Subject 3 continued to rely heavily on the SVS, using the out-the-window view for only five percent of the flight phase. Wickens et al. (2004) also noted in their related experiment, “that certain pilots are more susceptible to display-induced tunneling than others.” Subject 3’s very limited use of the out-the-window view during the final phases of the landing and the related finding from the Wickens et al. experiment strongly suggest that pilot susceptibility to tunneling deserves further attention.

3.1. The SVS as a Second Attitude Display

When the SVS was available on the flight deck, the captain, as modeled, alternated between the SVS and PFD in the scan for aircraft attitude, speed, altitude, and altitude rate information. The basic scan then included the out-the-window view, the PFD, the SVS, and the HSI. One impact of the modeled scan of the two flight deck instruments with an overlap in functionality was that less time was devoted to the HSI display and the navigation function that it supports. The finding with respect to the model triggered a further review of the human subject data. The same effect was then found in human subject data for the part-task simulation. Thus, a finding identified in the modeling process, was further supported by human subject data from the part-task simulation. Our goal in following up on this finding was to provide a means to restore what appeared to be the more efficient scan pattern as used by the pilots for the baseline flight deck configuration.

3.1.1 The SVS as a Second Attitude Display has Workload Implications

The modeled aircrews and the subjects in the part-task experiments tended to spend less time attending to the HSI display when the SVS was available even in the early phases of the approach where they were principally monitoring their progress along the flight path. The presence of the SVS appeared to reduce the time allocated to the HSI even when the HSI was the information source most relevant to the current flight phase. One explanation might be that the aircrews had sufficient time to accomplish their navigation task and were simply using the HSI to the extent that it was needed. Nevertheless, it is possible that there is an underlying problem.

On the SVS-equipped flight deck, the captain effectively has two attitude instruments. In scanning these two separate instruments, it appears that they may be drawn to spend more time attending to attitude-related information than would be necessary when using a single attitude-instrument configuration. With the dual instrument configuration, the *habituated* pattern may become a largely redundant two-instrument scan with attitude information derived from the two similar but not identical displays. In situations where time pressure is high, having two instruments from which to obtain required information can impose an additional workload burden on the pilot. Does the pilot stick with the time consuming habituated scan or try to save time by switching to a single instrument, but less practiced scan? Even electing to consider the options has a cost. Moreover, the extra cognitive effort required to switch to a single instrument scan may negate its potential inherent advantage. Changing a habituated two-instrument scan pattern when change is most difficult is an imposition on the aircrew that should be avoided if possible.

3.1.2 An Enhanced-SVS as the Single Primary Flight Display

The addition of the SVS to flight deck with its capability to provide a “clear day” view at all times has the potential to lead to greater operating safety particularly during IMC and night time operations. Therefore, the potential impact of the additional instrument on pilot scan patterns is one for which mitigation should be sought. One possible solution is to consider a single attitude display combining PFD and SVS functionality—an Enhanced-SVS. The goal is to restore the original single-PFD pilot scan pattern. We postulated that the habituated scan would then be a single instrument scan that readily avoids the choice of scan pattern that the two-instrument configuration can potentially impose, just when choice is most difficult.

Modeling provided an ideal means to explore pilot performance using the Enhanced-SVS. It was necessary to define the information requirements to be met by the Enhanced-SVS instrument and to refine pilot behaviors to use the new single PFD instrument. With these changes in place, it was then possible to rerun the SVS scenarios using the new flight deck configuration and examine the impact on pilot scan patterns. Based on the development of the refined pilot procedures it was not too hard to predict that we would find a scan pattern much like that when using the standard PFD, while at the same time putting in place the advantages inherent in having the SVS capabilities present—the best of both worlds.

During much of the approach, the pilots are concerned primarily with heading, speed, and altitude along each flight leg. In the model, each aircrew member performs an essential background instrument-monitoring task to assure that the critical values are being maintained. As the aircraft transitions from one flight leg to the next, the aircrew is concerned with monitoring the changes in these values as the aircraft transitions through waypoints established in the flight management computer. Each aircrew member executes procedures to actively monitor the aircraft’s progress until the new flight leg is fully established.

For the purposes of the modeling environment, we were concerned with the informational capabilities of the instruments to the extent that they were to be used by the pilot models. As we have seen, the visual form in which the information as presented on the actual 757 flight deck is quite different from that of the part-task simulation flight deck. However, at the informational level, the flight deck developed for the part-task simulation trials was very similar to that of the actual 757 flight deck. Our goal has been to look at pilot strategy with respect to accessing information to support the approach and landing tasks and not to look at the tradeoff between one particular information presentation and another. Further development of the vision component of our models would be necessary to evaluate tradeoffs in the visual presentation of individual information items.

Figure 7 provides a view of the principal NASA part-task simulation flight deck instruments: the SVS, the PFD, and the HSI. Figure 8 provides a view of the corresponding instruments on the 757/767 flight deck. The physical appearance of the part-task simulation flight deck and the 757/767 flight deck are quite different. Values for air speed and altitude are presented in air speed and altitude tape format in the simulation environment and as circular dials on the actual flight deck instrumentation. The tapes are incorporated as part of the PDF in the simulator, but are present as separate instruments adjacent to the PFD on the actual flight deck.

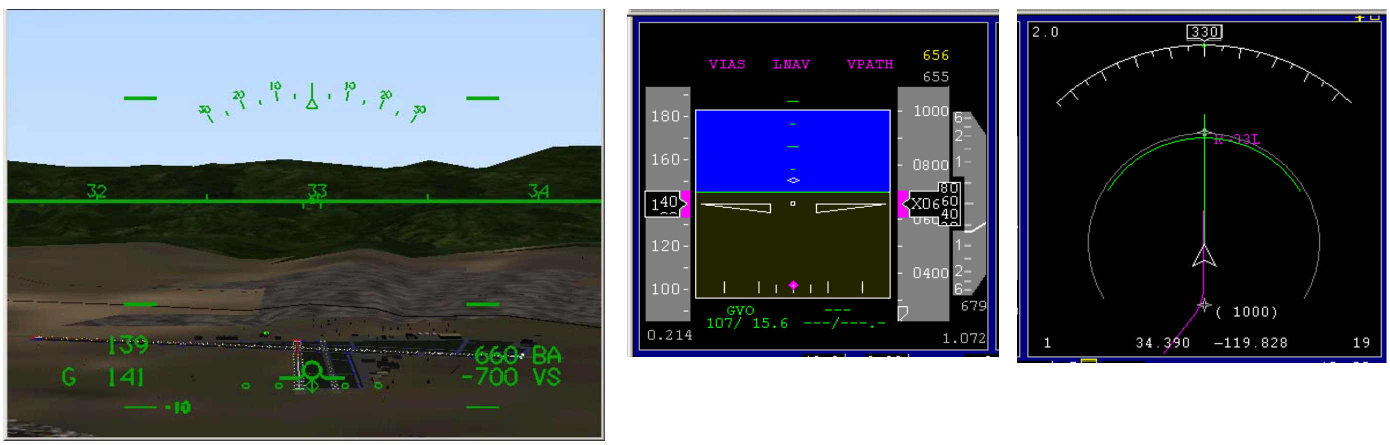


Figure 7 NASA Part-task Simulation Flight Deck Instruments[2]

To obtain heading, airspeed, and altitude information from the part-task instruments a pilot would get heading from the HSI, and airspeed and altitude from the respective airspeed and altitude tapes on the PFD. On the 757 flight deck, heading is available immediately below the PFD on the HSI, and airspeed and altitude are available on either side of the PFD—the basic scan is actually a scan of sites peripheral to the PFD. At the information seeking level, the scans are quite similar. While the timing of the scan to access, say, the altitude tape versus the altitude dial may differ slightly, for the purposes of this investigation, these differences were not expected to be a factor.

Completing the design for an operational Enhanced-SVS would be a significant undertaking. Fortunately, at the information seeking level, "designing" an Enhanced-SVS becomes a relatively simple task. If we are to remove the standard PFD from the flight deck, we need to assure that each of its information items continue to be available in an appropriate manner. To start with, the SVS, unlike the other part-task flight deck instrument and the actual 757 flight deck instruments, immediately provides the three values of interest for the basic scan: heading, airspeed, and altitude. As described in Goodman et al. (2003), the SVS also provides a vertical speed indicator,

[2] The illustration of the NASA part-task simulation flight deck instruments has been assembled from individual figures in Goodman et al. (2003). Additional information on the instruments is available in that document.

a roll indicator, a pitch ladder, localizer dots, and a flight path predictor. The mode annunciators are the only missing instrumentation. For the purposes of running the modeled SVS scenarios using the Enhanced- SVS as the single attitude instrument, it was sufficient to add the mode annunciators to the SVS as already modeled. With this simple change accomplished, the essential PFD functionality was established in the Enhanced-SVS and the PFD could be removed from the flight deck configuration.

Figure 8 757/767 Flight Deck Instruments[3]

For the part-task experiments, it was only the captain that was provided with an SVS. With the captain as the experiment subject and the first officer role played by one of the experimenters there would have been no purpose to making an SVS available to the first officer. In the modeling environment, we had a full two-person aircrew with symmetric workplaces. Since the Enhanced-SVS became a “replacement” for the PFD, we decided to also replace the PFD on the first officer’s side with an Enhanced-SVS to maintain the flight deck symmetry and thereby enable the first officer to perform SVS-assisted operations. If the current SVS concept and design works out as expected, the symmetric employment of an Enhanced-SVS is a likely longer-term operational configuration. Hence, the model provides an early look at how an aircrew might perform with that configuration. Furthermore, an SVS that has the nominal PFD display behavior as an

[3] The illustration of the 757/767 flight deck instruments was constructed from individual instrument images available at www.meriweather.com.

alternate operating mode would allow the aircrew to select their preferred mode of operation and provide a possible fail-safe mode for the SVS component of the display.

3.1.3 The Enhanced-SVS Scenario Trials

Our goal in pursuing the design of the Enhanced-SVS was to provide an SVS instrument that would restore the balance in the pilot PFD-HSI scan pattern as seen on the baseline flight deck. To obtain the scan data through which to assess the success of the design we set up scenarios to repeat each of the four NASA part-task SVS-equipped trials using the Enhanced-SVS as the single attitude instrument replacing the PFD and SVS. The SVS-equipped trials included the nominal approach, the late reassignment requiring a side-step to a closely parallel runway, the missed approach, and the mismatch in alignment between the out-the-window view and the SVS view of the runway as the aircraft emerged from the cloud cover at 800 feet. As in the original scenarios, the modeled aircrews readily executed the approach and landing or go-around as required in each of the scenarios.

Table 1 and Table 2 provide eye-tracking data for representative flight segments for the human subject and model trials. The human subject data for the baseline and SVS trials was provided by NASA (Goodman et al., 2003). The model data for the baseline, SVS, and Enhanced-SVS trials was derived from the D-OMAR simulation trials. The columns of the table present the percentage of dwell time that each subject, human or model, devoted to the out-the-window view (OTW), to each of the principal flight deck instruments—the HSI, the PFD, and the SVS when present, and to “other” identified areas in their field of view. For the human subjects, the row labeled “off” accounts for the percentage of dwell time for which the eye cursor was centered on an undefined area or for which the data was invalid (e.g., subject blinks) (Goodman et al., 2003).

The data in Table 1 covers the flight segment between the initial approach fix (IAF) and the final approach fix (FAF) on the approach to SBA runway 33L. The data in Table 2 covers the flight segment between the FAF and decision height (DH). Decision height for the approach, established by the approach plate at 650 feet, was transited shortly after passing through the PHANTOM waypoint at 1000 feet. The approach plate provided to the subjects for the NASA part-task experiment (Goodman et al., 2003) is shown in Figure 1. The IAF, FAF, and PHANTOM waypoints are identified on the approach plate. Columns 2 through 4 provide data from the baseline-IMC approach for the three human subjects. Column 5 provides data from the baseline-IMC approach for the D-OMAR model. Columns 6 through 8 provide data from the SVS-equipped IMC approach for the three human subjects. Column 9 provides data from SVS-equipped IMC approach for the D-OMAR model. Column 10 provides model data for the IMC scenario using the Enhanced-SVS attitude display that made it possible to eliminate the PFD.

As can be seen in the tables, when the SVS is added to the flight deck configuration for the nominal IMC trials, the percentage allocation of dwell time devoted to the navigation display, the HSI, generally decreases while the allocation to the two attitude displays, the PFD and the SVS, increases. For the trial using the Enhanced-SVS, the percentage dwell time allocation closely approximates the percentage dwell times for the baseline IMC trial using the single attitude display. The model trials demonstrate that when using the Enhanced-SVS as the single attitude display, the allocation of time devoted to the attitude display and the navigation display reverted to that seen in the baseline configuration. The balance varies through the phases of the scenarios, particularly around the maneuvers at flight path waypoints, but is once again consistent with the pattern in the baseline condition.

Evaluating alternate flight deck instrument configurations and their implications for aircrew performance can be very expensive to pursue in time and in cost. In this instance, modeling enabled us to quickly design and evaluate the proposed Enhanced-SVS. The modeling outcomes suggest that the Enhanced-SVS might well avoid the potential workload implications of the PFD-SVS configuration.

Table 1 Human Subjects and Model Eye-tracking Data: IMC Nominal Approach - IAF to FAF

	Subject 3	**Subject 4**	**Subject 5**	**Model**		**Subject 3**	**Subject 4**	**Subject 5**	**Model**		**Model**
Condition	IMC	IMC	IMC	IMC		SVS	SVS	SVS	SVS		Enh-SVS
off	2.24	3.90	5.15			2.40	4.76	1.53			
OTW	0	1.00	0.35	6.31		0	4.82	0.06	6.15		6.46
SVS						**24.02**	**15.25**	**11.84**	**26.03**		**36.57**
PFD	**43.63**	**55.95**	**34.42**	**37.00**		**27.93**	**31.36**	**42.57**	**26.11**		
NAV	**32.62**	**33.52**	**54.71**	**47.44**		**28.96**	**39.55**	**37.39**	**34.03**		**47.45**
other	21.51	5.49	5.36			16.69	4.26	6.62			

Table 2 Human Subjects and Model Eye-tracking Data: IMC Nominal Approach - FAF to DH

	Subject 3	**Subject 4**	**Subject 5**	**Model**		**Subject 3**	**Subject 4**	**Subject 5**	**Model**		**Model**
Condition	IMC	IMC	IMC	IMC		SVS	SVS	SVS	SVS		Enh-SVS
off	2.22	4.08	2.84			1.80	5.07	1.53			
OTW	0	11.57	2.03	10.39		0.52	4.37	2.57	8.55		10.79
SVS						**19.22**	**33.53**	**49.04**	**28.37**		**37.60**
PFD	**39.98**	**46.72**	**42.54**	**39.60**		**41.21**	**35.96**	**24.76**	**29.53**		
NAV	**33.45**	**32.40**	**44.64**	**39.62**		**26.90**	**19.53**	**18.16**	**28.66**		**39.57**
other	24.35	5.22	7.94			10.36	1.55	3.95			

3.2. Alternate Responses to SVS Misalignment

A second finding was derived directly from the human subject trial data and was reflected in the subsequent model development. Our related work on aircrew decision-making (Deutsch & Pew, 2004b) pointed to a very strong bias on the part of aircrews to complete an in-process approach and landing in spite of cues suggesting that a go-around may be the better course of action. Orasanu, Martin, and Davison (1998) have referred to this type of response as a plan continuation error—the execution of the current plan is continued in spite of significant evidence suggesting an alternate course of action. For the NASA part-task trials, the instructions to the subjects were quite explicit: If they detected a terrain mismatch between the SVS view of the runway and out-the-window view they to declare the to and execute a go-around.

With this as background, we elected to review the flight paths of the aircraft in the terrain mismatch scenarios following the point at which the subjects recognized the SVS-mismatch condition. We were interested in the subjects' response following the descent through the cloud cover and the detection of the SVS misalignment. NASA (Goodman et al., 2003) provided room-camera view and eye-camera view VCR tapes that included an audio recording for each of the scenario trials. They also provided part-task simulation trace data on aircraft position and attitude for the trials. In reviewing the eye-camera tapes, we found that shortly after the first officer notified Subject 5 that the aircraft was approaching decision height, the subject looked out the windscreen and recognized that "we are not aligned visually, …" and immediately decided that "we're going around." The subject acted quickly in following the specified response to the SVS misalignment situation by calling for and executing a go-around.

When Subject 3 was informed by the first officer that the aircraft was approaching decision height, he responded with "airport in sight, runway in sight," followed shortly by "SVS does not line up with the picture out the window," and then "I'm going to take it down a little bit lower, but it's putting me in a bad position." The subject subsequently proceeded with the approach and it was only when the first officer announced "200 feet" that the subject responded, "Okay, I'm off center, going around."

When the first officer informed Subject 4 that the aircraft was approaching decision height, he responded with “runway in sight” and subsequently announced contrary to the misalignment rule, “Landing.” It was only a short time later when the first officer notified the captain that it “looks offset to me” that the subject responded: “It is” and initiated the go around. The intent on the parts of Subjects 3 and 4 was to proceed with the landing. Had the simulated aircraft been more realistic in terms of the operation and effect of the controls, and the ability to monitor control settings (monitoring throttle settings, in particular, was difficult), Subjects 3 and 4 might well have succeeded in aligning their aircraft for the runway and proceeded with the landing.

As seen in the subject responses, on finding that the SVS is misaligned, the decision to pursue the landing or elect a go-around, is one that is not straightforward. Two of the three subjects elected to set aside the “go-around rule” and pursue the landing. Several factors may have played into their decisions. The situation was not unlike the side-step scenario in which they were explicitly asked if they would side step to the parallel runway. The offset and approach timing were similar to the misalignment scenario and the subjects (and the models) were readily able to execute the necessary maneuver when requested to do so. It is also possible that they may have been reluctant to climb back into the cloud cover when dependent on an instrument that was then known to be misaligned.

While we were not surprised by the finding that two of the three subjects pursued the approach and landing well past when the decision to go-around might have been made, we did initially develop pilot procedures that faithfully followed the decision rule that dictated the go-around. We subsequently changed the decision-making step to more accurately reflect the subject behaviors that were observed. Cues considered in making the decision included the detected offset from the runway, distance to the runway threshold, and the “priority” that the modeled pilot assigned to pursuing the landing. With the evaluation of these cues in place, the model sometimes elected to go-around and sometimes successfully completed the landing. Developing procedures for the mitigation of the potential SVS misalignment problem deserve more attention—it is a point at which there is very little time for thinking through solutions to problems and well-practiced responses need to be readily available.

3.3. Cross-checking SVS Misalignment

We conclude our discussion of the use of the Enhanced-SVS flight deck configuration with one last observation related to providing an SVS to the first officer as well as the captain. In the NASA part-task simulation trials and the model trials, when it was only the captain who was provided with an SVS, the response to the detection of an SVS misalignment was the captain’s decision alone to make and was one that could be made and acted upon very rapidly. In the Enhanced-SVS scenarios, since the Enhanced-SVS was a replacement for the PFD, we provided the first officer as well as the captain with an SVS. In refining the aircrew procedures to accommodate the second Enhanced-SVS for the first officer, we took the natural step of including a cross-check when the misalignment was detected. The time to accomplish the cross-check can be relatively short, but in a situation where the interval between emerging from the cloud cover and reaching decision height is short, the cross-check time itself can be a factor that needs to be attended to. The decision to land can be a time-critical decision with significant factors on each side of the binary choice. For the present, we call attention to the additional time associated with the cross-check of the SVS misalignment as a timing factor that can impact the decision to land.

4. Background on the D-OMAR Human Performance Models

The development of human performance models in D-OMAR (Deutsch & Pew, 2002; Deutsch & Pew, 2001; Deutsch, 1998; Deutsch & Adams, 1995) has been based on research in experimental psychology, cognitive science, cognitive neuroscience, and recent cross-disciplinary work in the theory of consciousness. As with most human performance models, the complexity of the models makes it difficult to provide a description that is both brief and complete. There is a theoretical framework that underlies the architecture; there is a broad range of individual human functional capabilities that must be represented; there are complex interactions among these capabilities, that taken together, generate the observed behaviors, and there are practical compromises that inevitably must be made in producing a working model. To provide additional insight into the D-OMAR human performance model behaviors used in the NASA SBA approach and landing scenarios, we introduce a few of the central features of the models and of the architecture that provides the framework for their development. We briefly discuss architecture for the models, how multiple task behaviors are developed within the architecture, how a distributed model of working memory fits into the architecture, and the role of vision in supporting the multiple task behaviors of the models.

4.1. Cognitive Architecture for the D-OMAR Models

It is common to speak of a *cognitive architecture* for a human performance model when in fact, what is required is an architecture to support perceptual, cognitive, and motor functions. When we use the "cognitive architecture" shorthand, we are speaking of an architecture that encompasses perceptual and motor functions as well as cognition. A cognitive architecture typically constrains and guides the construction of a model. It provides the operational framework for the computation aspects of the model. The architecture should reflect the main tenets of a *theory* for how operational perceptual, cognitive, and motor capabilities interoperate to produce human-like behaviors.

Most human performance modeling environments are framed by a particular architecture, hopefully an instantiation of a theory. D-OMAR[4] is very different in this respect. Our sense is that there is still much to learn about cognitive architectures and that there is much to be gained by leaving room to explore alternative approaches to their construction. Rather than a specific cognitive architecture, D-OMAR is a framework in which to explore alternate architectural constructs in which to develop models. It is a discrete-event simulator and a powerful suite of representation languages that may be used selectively to instantiate an architecture and construct human performance models.

4.2. Representation Languages in D-OMAR

D-OMAR representation languages provide the foundation for establishing a cognitive architecture and instantiating a human performance model within that architecture. They are also the basis for building the simulation environment entities with which the human performance models must interact. A frame language, the Simple Frame Language (SFL), is typically used to define the agents and objects in the simulation environment. SFL is a direct descendent of KL-ONE, developed during the 70s and early 80s at BBN (Schmolze & Brachman, 1981; Brachman & Schmolze 1985). It provides capabilities very similar to OWL, currently under development by the Web Ontology Working Group as part of the W3C Semantic Web Activity. In SFL, we speak of *concepts* and the *roles* that define concept slots. Concepts are closely analogous to object classes with slots in traditional object-oriented programming languages.

While the most critical element in model development is creating human-like model behaviors, the environment must also support the behaviors of the non-human agents as well. Model behaviors are defined in the Simulation Core (SCORE) language, a language for expressing the goals and procedures of an agent. In this framework, agents are just a special type of SFL object with the capability to execute SCORE language procedures—they are the active players within the simulation environment. The agents may be models of human players (e.g., the aircrews and the air traffic controllers), models of workplaces (e.g., aircraft flight decks or ATC

[4] Detailed information about the representation languages and the simulator is available in the documentation for the Lisp version of D-OMAR at the web site http://omar.bbn.com.

workplaces), or airspace entities (e.g., the aircraft or radar systems). The agent types for a simulation typical form a sub-tree within the multiple inheritance hierarchy of SFL concepts.

A rule language, the Flavors Expert (FLEX) (Shapiro, 1984), is included to complete the D-OMAR suite of languages, but has not been used in any of the D-OMAR models to date. While many human performance modeling environments (e.g., ACT-R (Anderson & Lebiere, 1998), SOAR (Laird, Newell, & Rosenbaum, 1987)) include production rule languages as central elements of their architectures, there is little motivation to use rules in an environment in which a procedural language plays a central role. Most of the complexity in rules is in the predicates necessary to define the situations in which they are applicable. This complexity in rule predicates is further aggravated by the need to manage sequencing in ruled-based behaviors. This complexity falls away when rule-like decisions are imbedded in procedures—the procedures define the context and control behavior sequencing.

The representation languages have been used to implement and refine the particular architecture now employed to address the NASA task. D-OMAR has also been used to explore variations on elements of the current architecture and to implement an entirely different architecture at a quite different level of granularity. This level of flexibility in the implementation of model architectures seems essential to the effort to improve the capabilities of human performance models.

4.3. An Architecture for Multiple Task Management

One of the principle areas of research in the development of D-OMAR has been in the modeling of human multi-task behaviors. In developing D-OMAR, we have sought to define an architecture and a computational framework in which to assemble functional capabilities that operate in parallel, subject to appropriate constraints, and that taken together exhibit the multiple task behaviors of human operators—aircrews and air traffic controllers. The desired behaviors have a combination of proactive and reactive components. That is, the operators have an agenda that they are pursuing, but must also respond to events as they occur. Consequently, within the proactive agenda, there may be newly motivated tasks for which on-going tasks must be deferred. The bounds on what can be accomplished concurrently take several forms. A typical behavior may be to set aside a flight deck conversation in order to respond to an ATC communication, while at another level, two competing tasks may each require the use of the pilot’s eyes to guide a manual operation. In the first instance, it is a matter of protocol that must be established, in the second, it is contention for a physical resource that must be arbitrated.

The core of a D-OMAR model is a network of procedures whose signal-driven activation varies in response to events that are proactively channeled to achieve aircrew goals. From a bottom up perspective, there is an assembly of individual perceptual, cognitive, and motor capabilities that are recruited as procedures to address current goals and sub-goals. Neumann’s (1987) functional view of attention, and the localization of mental operations in the brain, as put forward by Posner, Petersen, Fox, and Raichle (1988) are important contributions supporting this capabilities-based approach to modeling human behaviors. Taken together, they point to the functional components in task execution as taking place at particular local brain centers with the coordinated operation of several such centers required to accomplish any given task.

The form that the coordination in the execution of computational elements might take is of particular importance in the architecture for developing model behaviors. Specific functional capabilities are modeled as procedures that represent the work of the various perceptual, cognitive, and motor centers acting in support of the completion of a task. We need to provide for coordination among the procedures necessary to accomplish a task. We also need to provide for the movement of information among the procedures. A publish-subscribe protocol is used to address of these requirements. The unit of information transfer is a message identified by a name specifying the message *type* and containing related information elements.

As a procedure develops informational elements, they are assembled into an appropriately type-named message that is then *published*. Procedures requiring informational elements of a particular message, *subscribe* to the message by message type. By subscribing to a message type, the receiving procedure is alerted to publication of the message and the message content may then be processed by the receiving procedure. The publish-subscribe protocol at once provides for the movement of information among procedures while its triggering mechanism supports the coordination of procedure execution. The publish-subscribe protocol thereby creates dynamic links in the network of computational elements, the procedures that represent the model's collective functional capabilities.

Goals and their sub-goals and plans are used to establish the network of procedures. Together, they represent the things that a person knows how to do: basic person skills (e.g., coordinated hand-eye actions to set a flight deck selector) and domain specific skills (e.g., the captain making the decision to land). Active goals represent the operator's proactive agenda for managing his or her tasks. These top-level goals typically activate a series of sub-goals and procedures. The goals and sub-goals represent the objectives of the actions to be taken; the procedures are the implementation of the actions to achieve the goals and sub-goals. The procedures each may include decision points to address variations in the local situation. Hence, the operator's overall agenda is implemented by the dynamic network of procedures established by the goal-procedure hierarchy and linked by the publish-subscribe protocol. A subset of the procedures is active; most are in a wait-state. The procedures in a wait-state represent the capabilities to complete actions currently underway and to respond to impinging events.

With goals and procedures as its major elements, the architecture that we created in D-OMAR is very much an architecture of *process* (Edelman, 1987; 1989), with declarative memory relegated to a supporting role. The model's expertise, represented as goals and procedures, constitutes the capability of the model to perform in a human-like manner. Logan (1988) and Bargh and Chartrand (1999) have extended the traditional view in thinking about human skills as restricted to perceptual and motor skills by suggesting that there are *cognitive skills* that play a significant role in human expertise. The cognitive skills, the expertise of the aircrew, are thus represented as process—goals and procedures. Glenberg (1997) identifies process memory, memory for what we know how to do as memory's primary role. Domain expertise, what the model knows how to do, is represented in long-term memory, in large part, as the cognitive skills of the model. In the next section, we describe how the declarative aspect of working memory fits into the process-based architecture.

4.4. The Distributed Model of Working Memory

On the flight deck, each aircrew member will typically have several tasks are at various stages of completion. In pursuing the completion of each ongoing task, there will usually be several goals and procedures concurrently active. As the aircraft proceeds along the flight leg to PHANTOM (see Figure 1) in IMC with an 800-foot cloud ceiling, at least two of the captain's tasks are concerned with current altitude. For our example, let us say that the altimeter is currently reading 1021 feet AFE. The more immediate task will be concerned with monitoring the descent to the target altitude of 1000 feet for the flight leg to PHANTOM. The second task concerned with tracking the current altitude, has as it goal, the decision to land to be made as the aircraft descends through the decision height of 650 feet. There is other work in process, but these two tasks are sufficient to suggest how the altimeter reading—a working memory element—is processed in the model.

For all scenarios, an altitude reading is available from the PFD. For those scenarios that include an SVS, the SVS provides a second source for an altitude reading. The captain and the first officer each periodically scan the PFD (and separately, the captain scans the SVS when present). In our example, we will follow the captain's reading of the PFD. Each scan provides at least the heading, altitude, and speed in numeric form. In the vernacular of the modeling architecture, upon completion of the reading of the PFD the numeric altitude value of “1021” is *published*—that is, the labeled value is part of an attitude scan message that is the product of “scanning” the PFD (or the SVS), where “attitude-scan” is the message type. The published message is then available to any subscribers to the message type.

In our example, at least two of the captain’s procedures *subscribe* to attitude scan messages: the first is concerned with monitoring the aircraft’s descent to the target altitude of 1000 feet for the current flight leg to PHANTOM; the second is preparing to make the decision to land at or near the 650-foot decision height. The active scan procedures for both the PFD and the SVS publish attitude messages, which in effect, “wake” the procedures that are subscribed to the message type. The publishing of the products of the proactive visual process triggers cognitive processes requiring that information to facilitate their next actions.

From the captain’s perspective, the value “1021” is processed in a unique manner in each procedure. For the task of monitoring the descent to the target altitude, “1021” immediately becomes “I need to attend carefully to the aircraft’s altitude over the next few seconds to assure that we level off at 1000 feet.” For the decision to land task, “1021” becomes “I’ve got a little time, but I should be breaking out of the cloud cover shortly and I then need to acquire to the airport and runway to support the decision to land.” In each case, the published altitude value of “1021” is immediately, significantly, and uniquely *reinterpreted* by the subscribing procedures. The value “1021” is just an intermediate value in a rapid and concurrent succession of transformations.

The model’s publish-subscribe protocol is designed to mirror the movement of information through the brain’s visual centers and on to cognitive centers for further interpretation, action planning, and action execution. The model glosses over the many processing steps in the visual center that produce the symbol “1021,” but then more faithfully represents the further processing of “1021” that leads to the captain’s actions related to assuring that the aircraft levels off at 1000 feet and to preparations to look for the break in the cloud cover and the sighting of runway 33L needed to support the decision to land.

“1021” is certainly a working memory item, but it is simply one local value at one stage in a multi-branching process of transformations and interpretations starting in the visual system and moving through cognitive areas and then to motors areas that drive the resultant actions. Working memory items need a home in the architecture for a human performance model, but they cannot be properly captured in a single box in an architecture diagram. In the D-OMAR model, we posit that working memory, as we have just seen, is widely *distributed* across brain centers. Moreover, we assert that these working memory items do not have a separate existence as database items, but rather are each encapsulated by local processes, the procedures that operate on them at each stage of their migration, transformation, and interpretation.

4.5. Vision as a Component within Multi-task Behaviors

Demands on the pilot’s visual system are varied and sometimes complex. A broad range of these visual capabilities is included in the models. Some of a pilot’s actions are purely visual. There are basic processes that take in information from the major flight deck instruments and the view out the windscreen. The viewing of a flight deck instrument can be a generic guidance or navigation status check, or the monitoring of an instrument for a specific target value. The view out the windscreen can be to assure that there are no traffic conflicts (i.e., seeing *nothing* can be the desired outcome), acquiring a specific object (e.g., sighting runway 33L on the approach to support the decision to land), or tracking the aircraft’s alignment with runway 33R while executing the side-step maneuver. Acquiring a specific value occurs at a brief instant in time; in tracking the runway, the viewing may extend over a fair period of time with interruptions to address other visual tasks.

For some actions, the visual component plays a supporting role. The execution of flight deck operations require coordinated hand-eye actions to set switches, adjust selector settings, and reposition control levers. A slightly different form of hand-eye coordination is needed in using the tiller to guide aircraft taxi operations.

Reading and more broadly, the interpretation of graphical display information are further important visual tasks. The approach plate and airport diagram are used as sources of information to support approach, landing, and taxi operations. Some of the information from these sources is purely textual; some involves the interpretation of annotated graphical information. Elements of the PFD, HSI, and SVS require graphical interpretation. Each of these visual operations is an important component of model behavior.

Vision plays a central role in the modeling of a pilot's multi-task capabilities. The pilot's procedures for each of these visual activities has a priority associated with it. Within this framework, the vision system is a resource and the tasks that require its capabilities compete based on their assigned priority. As modeled, the scans of the windscreen and the individual flight deck instruments, the PFD, the HSI, and the SVS when it is present, are modeled as separate procedures, all with the same priority. The scan of the 757's separate air speed and altitude instruments (Figure 8) is included in the scan of the PFD. The *background* scan pattern thus produced, moves smoothly to the windscreen and from instrument to instrument. At particular points in the approach, instruments may be dropped from the scan pattern (e.g., the scan of the HSI after the decision to land). Clark (1999) describes similar variations in purpose directed saccade patterns as first reported by Yarbus (1967). To date, it has not been necessary to dynamically adjust the priorities of specific elements of the basic scan procedures to produce a reasonable scan pattern.

The action to scan out the windscreen for the runway or to support coordinated hand-eye actions to set a selector or control lever operate slightly differently. These *purpose-driven* actions are examples of an explicit decision to take an immediate action. They are invoked at a priority higher than the background scan procedures. The elected action takes place immediately, interrupting the background scan pattern, and once the action has completed, the background scan pattern resumes.

Actions with a visual component that extend over time operate with another slight variation. Visually tracking the runway to support the side-step maneuver operates similarly with respect to priority, but to accommodate the extended time duration it allows intervals at which elements of the background visual scan can intervene. The interval between these purpose-driven scans varies depending on the particular circumstances. For example, as the aircraft closes on a target altitude, the interval between altimeter readings decreases until the scan is terminated as the target altitude is established. After the purpose-driven scan is terminated, the background scan pattern serves to monitor the maintenance of the current altitude.

The mix of a background scan pattern interleaved with purpose-driven scans is much like that found by Bellenkes, Wickens, and Kramer (1997). What we have established as the background scan pattern, they refer to as "minding the store." What we have called a purpose-driven scan, they have referred to as attending to the "action," the changing parameter(s) for the particular maneuver. Bellenkes et al. (1997) found that skilled pilots maintained their "minding the store" scans as they attended to the "action." It is this skilled behavior that we have modeled.

Information obtained by the pilot models in their scan of the HSI and the SVS is representative of the information derived from the major flight deck instruments. The HSI is a rich information source used by the pilots in tracking their progress along the flight path in each of the scenarios. Some information is immediately available in symbolic form: the aircraft's heading and the display scaling. Some information is readily determined in symbolic form once a little graphical interpretation has been accomplished: the name of the next waypoint and the distance to the next waypoint, Some is geometrically interpreted: that the changing heading is converging on the desired heading for the new flight plan leg. The pilots also recognize the current waypoint as the last one before landing and use this information in deciding when to terminate the their scan of the HSI. As the pilots track their progress along the flight path, they use the HSI range selector to appropriately adjust the display range. Figure 6 provides screen shot of the HSI display as NASA186 is traversing the leg to GOLET. Each of the basic information items can be seen in the figure. The form of the display is based on the HSI display described in the RNAV cognitive task analysis (Keller & Leiden, 2002a).

The scan of the SVS is particularly important to this research effort. The pilot models readily derive heading, speed, altitude, and altitude rate as numeric quantities from the SVS much as they do from the PFD. The SVS-equipped scenarios modeled to date are in IMC conditions with an 800-foot cloud ceiling. As the airport comes into the display's field of view, the captains use the SVS to acquire the airport and the runway. The pilots are able to track the runway using the SVS before the aircraft breaks out of the cloud cover. Once below the cloud cover, the modeled pilots rely primarily on the out-the-window view to track the runway.

5. Study-Induced Improvements the D-OMAR Model

An important goal of the NASA modeling program was to extend the state of the art in human performance modeling (Foyle et al., 2003). From our own perspective, we were anxious to see just how well our existing architecture and models would hold up when challenged by the baseline and SVS scenarios. The basic model architecture (see Section 4) provides an integration of perceptual, cognitive, and motor capabilities that produce human-like multi-tasking behaviors. The architecture that has supported modeling in the commercial aviation and related domains for a number of years held up quite well and was not changed.

Not surprisingly, the area of greatest change was in the specifics of the procedures for the baseline and SVS scenarios. The basic goals and procedures for the approach and landing scenarios were inherited from the earlier research on taxi error in the NASA O'Hare scenarios (Deutsch & Pew, 2001; 2002). The ILS approach and landing procedures used at O'Hare were extended to reflect the Keller and Leiden (2002a) cognitive task analysis for the RNAV approach as used at SBA for the NASA part-task scenarios. In a similar manner, the Keller and Leiden (2002b) addendum was used in developing the procedures for the pilot's use of the SVS. Goodman et al. (2003) provided detailed information on the specifics of the trial scenarios and additional information on the SVS. The addition of the RNAV approach and landing procedures and the refinements to the ILS approach and landing procedures are important improvements to the aircrew models that have also contributed to related NASA tasks (Deutsch & Pew, 2004b).

That said, there were two interesting points at which changes were required to meet the objectives of the modeling program. One of the more complex areas in the model addresses the arbitration necessary in managing the balance between intra-crew conversations and communication with the air traffic controllers. Section 5.1 outlines the changes that were required in the conversation model. The second point at which model changes were required was related to generating and gathering eye-tracking data that could be properly compared with the eye-tracking data from the NASA part-task scenarios. The changes related to capturing and processing eye-tracking data are described in Section 5.2.

5.1. Scenario Complexity Can Identify Model Shortcomings

Using a second aircraft in the late reassignment scenario was a small step in adding to the realism of the scenario. It was enough to trigger the need to refine model behaviors in dealing with the conflict that arises when there is air traffic controller communication at the same time that an aircrew has critical flight deck operations on which they need to communicate. The nominal behavior of the aircrew models is to suspend their intra-crew conversation when the air traffic controller interrupts a conversation in progress. This basic rule broke down first in the

O'Hare scenario when the first officer required an okay from the captain before accepting the exit taxiway offered by the approach controller. The captain had to inject an intra-crew statement within the on-going air traffic controller exchange to convey the acceptance of the exit taxiway to the first officer that was then relayed to the controller.

In the two-aircraft late reassignment scenario a related situation occurred. The lead aircraft was approaching decision height as the tower controller was providing the trailing aircraft with the clearance to land. The lead aircraft's first officer was monitoring the descent to decision height and found it necessary to "speak through" the air traffic controller conversation with the trailing aircraft to notify the captain in a timely manner that they were approaching decision height. On the speaker side, the priority associated with the spoken-message procedure was increased so that the statement was delivered immediately rather than postponed until the air traffic controller exchange with the trailing aircraft was completed. On the listener side, the priority was similarly increased so that attention was directed to the intra-crew message rather than the air traffic controller exchange with the trailing aircraft. The required changes resulted in important improvements in properly structuring intra-crew and ATC communications.

As scenarios become more complex, they stress our models and drive the development of model capabilities that provide better insight into pilot behaviors. Increased workload can negatively impact aircrew performance. When this increased workload is manifest in human performance models it may well promote the development of better models—models that help us to understand resulting changes in aircrew performance, enable us to identify the errors that might be the product of high workload situations, and provide a setting in which to develop and evaluate error mitigation strategies.

5.2. Generating and Gathering Model Eye-tracking Data

Section 3.1.3 included a comparison of human subject and model eye-tracking data. Gathering the data to support the comparison required substantive changes in the eye movement portions of the model. As originally developed, the implemented scan pattern was *sufficient* to gather the flight deck instrument data necessary to accomplish the required tasks as the aircrew preceded through the approach and landing. The tasks are under-constrained in the sense that for much of the approach and landing sequence there is appreciably more time available than needed to accomplish the required operations. Hence, in the original sufficiency model, there were numerous periods of idle eye time.

In order to generate eye movement data consistent with the part-task experiment data, the model had to be revised to consume the idle eye time. To accomplish this, we reduced the intervals between the cycles of the basic elements of the background scan pattern. These included the scans of the windscreen, PFD, the HSI, and the SVS when it was included on the flight deck. This change had the desired effect of eliminating the idle eye time, and interestingly, but not surprisingly, had negligible effect on pilot performance.

With the idle time now accounted for, minor software changes were required in collecting dwell time data for each targeted display item or flight deck control. Lastly, the data was averaged across flight deck instruments and controls after being partitioned by flight plan waypoints to correspond to approach phases used by NASA (Goodman et al., 2003) in preparing the human subject data. With these changes in place, the comparisons between human subject and model data as presented in Section 3.1.3 were readily completed. The collection and processing of model eye movement data is a valuable new capability that will certainly be well used in the future.

6. Conclusions

D-OMAR served very well to support the development of the SVS scenarios.

D-OMAR provided excellent support for developing the SBA approach and landing scenarios. The aircrew models that executed the ILS landings at O’Hare for the previous study (Deutsch & Pew, 2001; 2002) proved to be readily extendable. The RNAV approach and the use of the SVS, as detailed in the cognitive task analyses (Keller & Leiden, 2002a; 2002b), required the implementation of a broad range of new goals and procedures that, while complex, were reasonably straightforward to develop within the framework established for the human performance models for the O’Hare scenarios. As the RNAV procedures were developed, they also shared many of the goals and procedures for the ILS approach. As the changes were made, they did at times impact the ILS procedures and it was necessary to take additional steps to insure the continued operation of the O’Hare scenarios. As a result, we now have a working RNAV approach, and ILS approach and landing procedures that are more complete than they were at the end of the original O’Hare studies.

The modeling task drove model development relevant to NASA goals.

While most of the changes required to complete the SVS scenarios were in procedures related directly to the RNAV approach and the use of the SVS, there were two underlying areas in the models that required attention. The first of these, as discussed in Section 5, concerned arbitration in the aircrew model related to properly interleaving intra-crew and ATC communications. In this case, existing model capabilities were extended to address requirements that surfaced in developing the SVS scenarios. The second change addressed the requirement to compare model and human subject instrument dwell time data. The new ability to collect model dwell time data is an important addition to the D-OMAR modeling framework that was essential for the current task and will certainly be useful in the future. These improvements in the model framework and in the aircrew models themselves are very much in the spirit of the NASA goal for the overall project—the improvement of the human performance models that can then be used to support NASA aviation safety research studies (Foyle et al., 2003).

The NASA part-task experiment provided data essential to the modeling effort.

The NASA part-task simulation data (Goodman et al., 2003) has proven to be a very valuable resource from which much was learned and we believe might well be revisited in the future. In particular, the fixation and dwell sequence data files and the eye tracker video tapes provide a level of detail in instrument scanning essential to model development. For the present, the models look at an instrument and read the instrument’s data items in a single pass. The eye tracker videotapes suggest that the pilots selectively and repeatedly scan individual items within a display before moving on to the next display. Clark (1998) reports similar patterns of saccade sites being visited and revisited in the short term. While the eye fixations on particular fields within an instrument can be observed in the videotapes, the collected data was resolved at the instrument level, not at the individual field level (i.e., fixation time on the PFD rather than fixation time on the altitude tape).

Newly available data should be exploited to further improve eye movement models.

A number of recent studies (e.g., Bellenkes, et al, 1997; Wickens, Xu, Helleberg, Carbonari., & Marsh, 2000; Wickens et al., 2004; Diez, Boehm-Davis, Holt, Pinney, Hansberger, Schoppek, 2001; Hüttig, Anders, & Tautz, 1999; Anders, 2001) have also used eye tracking to examine pilot scan patterns. To the extent that these data are made available, there will be further information that tells us *what* pilots’ scan patterns actually are. For the purposes of the present study, our revised eye movement model was adequate—small changes in parameter values had little impact on overall performance. To the extent that additional effort is devoted to more carefully calibrating the eye movement model, issues related to particular differences in instrument design might be more accurately addressed. As the developers of aircrew models and the theory that underlies our models, we are obligated to seek to explain *how* basic human capabilities operate together to yield these scan patterns. Our models will better represent pilot performance to the extent that these behaviors are better explained. Results of the recent eye-tracking studies will be helpful in pursuing this goal.

The model-based “design” and evaluation of the Enhanced-SVS demonstrated the potential that human performance models have for forecasting potential system vulnerabilities and exploring mitigation strategies.

Our long-term goal is to continue to improve our understanding of pilot behaviors as represented in human performance models and use the models and that knowledge to explore the means to identify accident precursors and establish procedural and system changes to mitigate their onset. The response to the finding related to pilot scan patterns when the SVS was added to the flight deck as a second attitude instrument is representative of the how human performance modeling has been productively employed. In adjusting the model procedures to use the SVS, we uncovered an unanticipated effect that led us to revisit the human subject data. The subsequent analysis of the human subject data then supported the hypothesis that introducing the SVS to the flight deck impacted the time that subjects devoted to the navigation display. When the SVS was present, more time was devoted to the attitude displays (the PFD and the SVS) and less time was devoted to the navigation display (the HSI). While the particular scenarios employed in the part-task simulation did not produce the time-pressured situations in which this might have led to a problem, we felt the change in pilot scan patterns was an issue requiring attention.

The question addressed became: How might the advantages of the SVS be secured, while at the same time, not creating a flight deck environment that altered the balance in pilot scan patterns related to aircraft attitude and navigational tasks? Modeling provided a cost-effective means to pursue the initial answers to this question. Using the D-OMAR modeling environment, we were readily able to design and evaluate an Enhanced-SVS that combined the basic capabilities of the traditional PFD and the SVS. By executing the four part-task SVS scenario we obtained model data that suggested that the Enhanced-SVS design had the potential to restore the balance in time devoted to attitude and navigation evidenced in the baseline flight deck environment. Exercising the model produced the anticipated shift back to the more efficient scanning behavior observed in the baseline condition while preserving the advantages of SVS capabilities.

These results are important not only for the substantive predictions they generated, but also as a concrete example of the ways in which human performance models can contribute to design decisions early in the design process. Using the modeling framework, we were able to “design” an Enhanced-SVS and demonstrate the validity of the hypothesis that the Enhanced-SVS would restore the baseline pilot scanning behavior. The cost in time and effort of exercising the model condition to demonstrate the validity of the hypothesis was a small fraction of the cost of building a revised human-in-the-loop simulation, running a new set of subjects, and analyzing the resulting eye-movement data.

7. References

Anders, G. (2001). Pilot's attention allocation during approach and landing: Eye- and head-tracking research in an A330 full flight simulator. In *Proceedings of the 11th International Symposium on Aviation Psychology*. Columbus, OH.

Anderson, J. R., & Lebiere, C. J. (1998). *The atomic components of thought.* Mahwah, NJ: Lawrence Erlbaum Associates.

Bargh, J. A, & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist, 54,* 462-479.

Bellenkes, A. H., Wickens, C. D., & Kramer, A. F. (1997). Visual scanning and pilot expertise: The role of attentional flexibility and mental model development. *Aviation, Space, and Environmental Medicine, 68,* 569-579.

Brachman, R. J., & Schmolze, J. G. (1985). An overview of the KL-ONE knowledge representation system. *Cognitive Science, 9,* 171-216.

Clark, A. (1999). *Being there: Putting brain, body, and world together again.* Cambridge, MA: MIT Press.

Deutsch, S. E. (1998). Interdisciplinary foundations for multiple-task human performance modeling in OMAR. In *Proceedings of the Twentieth Annual Meeting of the Cognitive Science Society.* Madison, WI.

Deutsch, S. E., & Adams, M. J. (1995). The operator model architecture and its psychological framework. In *Proceedings of the 6th IFAC Symposium on Man-Machine Systems.* Cambridge, MA.

Deutsch, S. E., & Pew, R. W. (2001). *Modeling human error in D-OMAR.* BBN Report 8328. Cambridge, MA: BBN Technologies.

Deutsch, S. E., & Pew, R. W. (2002). Modeling human error in a real-world teamwork environment. In *Proceedings of the Twentieth-fourth Annual Meeting of the Cognitive Science Society (pp. 274-279).* Fairfax, VA.

Deutsch, S. E., & Pew, R. W. (2003). *Modeling the NASA baseline and SVS-equipped approach and landing scenarios in D-OMAR.* BBN Report 8364. Cambridge, MA: BBN Technologies.

Deutsch, S. E., & Pew, R. W. (2004a). Examining new flightdeck technology using human performance modeling. In *Proceedings of the 48th Meeting of the Human Factors and Ergonomic Society Meeting.* New Orleans, LA.

Deutsch, S. E., & Pew, R. W. (2004b). Modeling aircrew decision-making in D-OMAR. BBN Report 8387. Cambridge, MA: BBN Technologies.

Diez, M., Boehm-Davis, D. A., Holt, R. W., Pinney, M. E., Hansberger, J. T., & Schoppek, W. (2001). Tracking pilot interactions with flight management systems through eye movements. In: *Proceedings of the 11th International Symposium on Aviation Psychology.* Columbus, OH.

Edelman, G. M. (1987). *Neural Darwinism: The theory of neuronal group selection.* New York: Basic Books.

Edelman, G. M. (1989). *The remembered present: A biological theory of consciousness.* New York: Basic Books.

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 (HPM) element. In D. C. Foyle, A. Goodman, & B. L. Hooey (Eds.), *Proceedings of the 2003Conference on Human Performance Modeling of Approach and Landing with Augmented Displays*, (NASA/CP-2003-212267), 1-13. Moffett Field, CA: NASA.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences, 20,* 1-55.

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. In D. C. Foyle, A. Goodman, & B. L. Hooey (Eds.), *Proceedings of the 2003 Conference on Human Performance Modeling of Approach and Landing with Augmented Displays,* (NASA/CP-2003-212267), pp. 71-89. Moffett Field, CA: NASA.

Hüttig, G. Anders, G., & Tautz, A. (1999). Mode awareness in a modern glass cockpit: Attention allocation to mode information. In *Proceedings of the 10th International Symposium on Aviation Psychology*. Columbus, OH.

Keller, J., & Leiden, K. (2002a). *Information to support the human performance modeling of a B757 flight crew during approach and landing: RNAV.* Boulder, CO: Micro Analysis and Design.

Keller, J., & Leiden, K. (2002b). *Information to support the human performance modeling of a B757 flight crew during approach and landing: SVS addendum.* Boulder, CO: Micro Analysis and Design.

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). SOAR: An architecture for general intelligence. *Artificial Intelligence, 33,* 1-64.

Logan, G. D. (1988). Automaticity, resources, and memory: Theoretical controversies and practical implications. *Human Factors, 30,* 583-598.

Neumann, O. (1987). Beyond capacity: A functional view of attention. In H. Heuer & A. F. Sanders (Eds.), *Perspectives on perception and action*. London: Lawrence Erlbaum.

Norman, R. M. (2002). *Synthetic vision system concept assessment report.* Long Beach, CA: The Boeing Company.

Orasanu, J., Martin, L., & Davison, J. (1998). Errors in aviation decision making: Bad decisions or bad luck? In *Proceedings of the Fourth Conference on Naturalistic Decision Making.* Warrenton, VA.

Posner, M. I., Peterson, S. E., Fox, P. T., & Raichle, M. E. (1988). Localization of cognitive operations in the human brain. *Science, 240,* 1627-1631.

Schmolze, J. G., & Brachman, R. J. (1981). *Proceedings of the 1981 KL-One workshop.* BBN Report No. 4842. Cambridge, MA: BBN Technologies.

Shapiro, R. (1984). *FLEX: A tool for rule-based programming.* BBN Report No. 5843. Cambridge, MA: BBN Technologies.

Wickens, C. D., Alexander, A. L., Thomas, L. C., Horrey, W. J., Nunes, A., & Hardy, T. J. (2004). *Traffic and flight guidance depiction on a synthetic vision display: The effects of clutter on performance and visual attention allocation.* Technical Report AHFD-04-10/NASA(HPM)-04-1. Savoy, IL: University of Illinois.

Wickens, C. D., Xu, X., Helleberg, J. R., Carbonari, R., & Marsh, R. (2000). *The allocation of visual attention for aircraft monitoring and avoidance: Baseline measures and implications for freeflight.* Savoy, IL: University of Illinois.

Williams, D., Waller, M., Koelling, J., Burdette, D., Doyle, T., Capron, W., Barry, J., & Gifford, R. (2001). *Concept of operations for commercial and business aircraft synthetic vision systems, Version 1.0.*

Yarbus, A. F. (1967). *Eye movements and vision.* New York: Plenum Press.