A Computational Process Model of Basic Aircraft Maneuvering

Kevin A. Gluck[1]
Jerry T. Ball[2]
Michael A. Krusmark[2]
Stuart M. Rodgers[1]
Mathew D. Purtee[1]

[1]Air Force Research Laboratory [2]L3 Communications
Warfighter Training Research Division, 6030 S. Kent St., Mesa, AZ 85212-6061 USA
(all email addresses are *first.last@williams.af.mil*)

Abstract

This paper describes a computational process model of basic aircraft maneuvering. It is an embodied performance model, implemented in ACT-R, that operates a Predator UAV synthetic task environment. The design of the model is borrowed from the Control and Performance Concept, a widely taught technique for instrument flight, and from discussions with subject matter experts. Comparisons with human data show the model to be a good approximation to expert human performance, although the model shows more intra-maneuver variability. The paper concludes with a description of methodological and implementation details that make this cognitive modeling effort distinctive.

Introduction

There is a long and rich history of human performance modeling in aviation psychology, extending back to the creation of the Psychology Branch of the Aero Medical Laboratory at Wright Field in 1945, with Paul Fitts as its first Director (Pew, 2001). Over the subsequent decades, psychologists, engineers, and computer scientists have investigated a wide variety of phenomena associated with situation awareness, aircraft control, attention, and task management. Wickens (2002) notes that a great deal of laboratory research has taken place to isolate and understand these complex perceptual, cognitive, and psychomotor processes. He goes on to suggest that "modeling the complex interactions among these phenomena remains a critical challenge posed by aviation to psychological researchers who are interested in 'scaling up' their theories to real-world problems" (p. 132).

We have taken on precisely this challenge in using an integrated cognitive architecture to develop a computational cognitive process model of basic aircraft maneuvering. Specifically, it is a model of an air vehicle operator (AVO) for a Predator Uninhabited Air Vehicle (UAV). The model interacts with a Synthetic Task Environment (STE) created for use by cognitive scientists who are interested in conducting their research in the context of an operationally-validated task, without the logistical challenges of working with the real operational military community. This paper will begin by setting the context for the modeling through some background information on the STE. We then describe the representations and processes built into the model and compare the model's performance to human performance. The paper concludes with a description of methodological and implementation details that make this cognitive modeling effort distinctive.

Background on UAV STE

The core of the STE is a realistic simulation of the flight dynamics of the Predator RQ-1A System 4 UAV. This core aerodynamics model has been used to train Air Force Predator operators at Indian Springs Air Field in Nevada. Built on top of the core Predator model are three synthetic tasks: the *Basic Maneuvering Task,* in which a pilot must make very precise, constant-rate changes in UAV airspeed, altitude and/or heading; the *Landing Task* in which the UAV must be guided through a standard approach and landing; and the *Reconnaissance Task* in which the goal is to obtain simulated video of a ground target through a small break in cloud cover. The design philosophy and methodology for the STE are described in Martin, Lyon, and Schreiber (1998). Tests using military and civilian pilots showed that experienced UAV pilots perform better in the STE than pilots who are highly experienced in other aircraft but have no Predator experience, indicating that the STE is realistic enough to tap UAV-specific pilot skill (Schreiber, Lyon, Martin, & Confer, 2002).

Basic maneuvering is the focus of the current modeling effort. The structure of the task was adapted from an instrument flight task designed at the University of Illinois to study expertise-related effects on pilots' visual scan patterns (Bellenkes, Wickens, & Kramer, 1997). The task requires the operator to fly seven distinct maneuvers while trying to minimize root-mean-squared deviation (RMSD) from ideal performance on altitude, airspeed, and heading. Each maneuver starts with a 10-second

straight and level lead-in section as the participant prepares to execute the maneuver. At the end of this lead-in, the timed maneuver segment (either 60 or 90 seconds) begins and the operator is required to maneuver the aircraft at a constant rate of change with regard to one or more of the three flight performance parameters. The initial three maneuvers require the operator to change one parameter while holding the other two constant. For example, in Maneuver 1 the goal is to reduce airspeed from 67 knots to 62 knots at a constant rate of change, while maintaining altitude and heading, over a 60-second trial. They increase in complexity by requiring the operator to fly maneuvers that change in combinations of two parameters. Maneuver 4, for instance, is a constant-rate 180° left turn, while simultaneously increasing airspeed from 62 to 67 knots. The final maneuver requires changing all three parameters simultaneously: decrease altitude, increase airspeed, and change heading 270° over a 90-second trial.

During the basic maneuvering task the operator sees only the Heads-Up Display (HUD). The HUD includes various digital and analog instruments, such as Angle of Attack (AOA), Airspeed, Heading (bottom center of display), Vertical Speed Indicator, RPM's (indicating the throttle setting), and Altitude. The digital displays move up and down as the value of the instrument changes. There is also a reticle and horizon line, which together indicate the pitch and bank of the aircraft.

At the end of a trial, the results for the altitude, airspeed and heading deviations are displayed graphically, with actual and desired values on each performance parameter plotted across time. Quantitative RMSD's provide numerical feedback for tracking performance.

Gray (2002) noted that one of the challenges involved in using existing simulation environments in research on computational human behavior representation is that typically those environments were not designed for interaction with a cognitive architecture and are implemented in a different programming language than is the modeling architecture. An attractive and common solution to this challenge is to reimplement or backwards engineer the simulation into a form amenable to cognitive modeling. In the case of the current project, however, reimplementation of the aerodynamics model and real-time simulation of aircraft handling in Lisp was not a reasonable option. It was imperative that a way be found to get the model to interact directly with the existing STE. This was somewhat of a challenge, because the model is implemented in ACT-R 5.0 (Anderson, Bothell, Byrne, & Lebiere, 2002), running on top of Allegro Common Lisp (ACL) 6.2, while the UAV STE is coded in C. The solution has been to run the cognitive model on a separate hardware platform and give it some Lisp code that communicates with the STE through a non-blocking socket "polling" mechanism. The current interface to the STE relies on a reimplementation of the control inputs process to receive input from the cognitive model instead of the stick and throttle. Another process running on the cognitive model platform receives data from the STE, and makes that data available to the cognitive model via a Lisp-based "mock HUD", which is where the model actually gets its instrument readings while it is flying.

The Model

Description of the model will begin with an explanation of the general task management structure, continue with the representation of declarative and procedural knowledge for flying a UAV, and finish with a section on architectural parameters used in the model.

The Control and Performance Concept

There is an instrument flight strategy called the "Control and Performance Concept" (Air Force Manual on Instrument Flight, 2000). This aircraft control process involves first establishing appropriate control settings (pitch, bank, power) for the desired aircraft performance, and then crosschecking the instruments to determine whether the desired performance is actually being achieved. The rationale behind this strategy is that control instruments have an immediate first order effect on the behavior of the aircraft which shows up as a delayed second order effect in the performance instrument readings.

At the beginning of a trial, the model first uses the stick and throttle to establish appropriate control settings (pitch, bank, power), then it initiates a crosscheck of the instruments to assess performance and to insure that control settings are maintained. In the process of executing the crosscheck, if the model determines that an instrument value is out of tolerance, it will adjust the controls appropriately. A subtle implication is that, in order to effectively use the Control and Performance Concept, it is necessary for a pilot (or a model) to know what the appropriate control settings are for various types of desired aircraft performance. That brings us to the next section of the paper, on knowledge representation for the UAV Operator Model.

Declarative Knowledge

Declarative knowledge is represented in four critical ways in this model: the goal chunk, crosscheck intent chunks, instrument chunks, and knowledge of appropriate control settings. Each of these is discussed below.

The goal chunk contains the knowledge, or links to the knowledge, needed to fly the Predator. It serves the purpose of representing the operator's situation awareness. The goal chunk is organized hierarchically into three categories: maneuver knowledge (e.g., intent of the maneuver, how long into the trial it is), control

knowledge (e.g., current, desired, and deviation values for the control instruments), and performance knowledge (e.g., current, desired, and deviation values for the performance instruments). Clearly this is a lot of information, all of which is important to instrument flight. A common modeling practice in ACT-R models is to restrict the size of declarative memory chunks to 3-5 slots. In the case of the goal chunk for the UAV Operator Model, however, we found this to be unmanageably restrictive. There is just too much information about the pilot's cognitive state and the aircraft's physical state that needs to be available for decision making. On the other hand, having all aircraft state data available to the model at all times would be too powerful. Therefore, the productions are designed in such a way that, at any one time, only a few slots in the goal chunk are actually used. For example, if the model has just attended to airspeed, then the current-airspeed slot is available to the model. Slots with values from previous attention-decision cycles are not assumed to be available, and new values must be encoded from the instruments or retrieved from memory. Thus, although the goal chunk has a sizeable number of slots, only a few of them have available values at any one time.

Movement of attention from one instrument to the next is decided via retrieval of a crosscheck-intent chunk, based on the current instrument, the maneuver intent, and the time-segment. The retrieval of a crosscheck-intent chunk also sets the context for the current attention-decision cycle (i.e., standard crosscheck or control focus).

The model assumes the operator has declarative representations of the instruments on the HUD. Instrument chunks contain a slot for the location of the instrument and a slot for encoding the current value of that instrument.

Finally, the model represents knowledge of the control settings that are appropriate for executing the required maneuvers. This knowledge is crucial for establishing the correct settings at the start of a trial, following the lead-in period. Knowledge of the desired control instrument settings at given points in a scenario (e.g. 15 seconds, 30 seconds, 45 seconds) is important for insuring that the control instrument settings are being maintained and that performance objectives are being achieved.

Procedural Knowledge

In order to do well on the basic maneuvering trials in the STE, the moment the trial starts the pilot must initiate a maneuver that results in approximately the right rate of change in the performance instruments. Therefore, there is a set of productions that are specific to the maneuver being executed and that represent learned behavior about how to initiate that maneuver. The execution of these productions is triggered by an auditory beep which occurs at the start of a trial, via ACT-R's audition module, or by recognition that the lead-in period is nearing completion.

The model has separate productions for establishing control and crosschecking, since the behavior of the model is different in these two cases. Establishing control begins with the selection of an instrument for which control needs to be established. This happens at two key points: 1) at the beginning of a trial when the values of control instruments are first set, and 2) whenever the assessment of a control instrument shows a large enough deviation to cause the model to focus on a control instrument. Through a series of production firings, (*Find*, *Attend*, *Encode*), attention shifts to the control instrument and its current value is encoded. If the desired-value is not already available in the goal chunk from a previous attention-decision cycle, it is retrieved from memory (*Retrieve-Desired*). The current-value and desired-value are compared and a numeric deviation is computed which is converted into a qualitative value (e.g. very-small, small, medium, large, very-large) during *Set-Deviation*. Then the qualitative size of the deviation is considered and, if necessary, an adjustment is made to the stick or throttle (*Assess-Adjust*). If an adjustment is required because a control instrument is off, the model sets its state to continue focusing on the control instrument on the next production cycle. Otherwise, the model sets its state to begin, or return to, a normal crosscheck.

The process of crosschecking is largely identical to the process of establishing control. The major difference is that both the control and performance instruments are candidates for attention. If the model attends to an instrument that deviates significantly from the desired value, it returns to the control loop. Moderate deviations result in adjustments to the stick and/or throttle without leaving the crosscheck loop.

Parameter Settings

A variety of parameters in ACT-R can be modified to influence the behavior of a model. One of the long-term architectural goals in the ACT-R community is to settle on default, or at least "commonly accepted," values for all of these parameters, in order to further guide the process of developing a model. The parameters that are relevant to the UAV Operator Model (with their values in parens) are: Production Utility Noise (1), Goal Weight (1), Latency Factor (1), Decay Rate (.5), and Activation Noise (.25). These are all values that are considered to be architectural defaults, or values that have been commonly used in other models.

It is important to emphasize that this is an ACT-R model with default values for the parameters mentioned above, and the design of the model is a direct translation of a well-known instrument flight technique. Nothing about the design of the model or the global parameters

above is tuned or optimized to any specific dataset. The question remaining to be addressed is ... when ACT-R uses the Control and Performance Concept to operate the Predator UAV STE, how does its performance compare with human pilot performance?

Comparison with Human Data

Human data were collected from seven aviation Subject Matter Experts (SMEs) at our lab in Mesa, Arizona. These are experienced Air Force pilots with an average of more than 3000 hours of flight time in different aircraft, but who had no prior Predator UAV training. Participants completed each maneuver for a fixed number of trials that ranged from 12-24, depending on the difficulty of the maneuver. Each participant completed the maneuvers in order, starting with Maneuver 1 and ending with Maneuver 7. The SME data plotted in the figures below come from successful trials only, where success is defined as flying within the performance deviation criteria used by Schreiber, et al. (2002). The important thing to understand is that the human data come from trials in which the SMEs flew well, relative to the performance goals. We use these data for the comparison because the current modeling goal is to develop a performance model of skilled aircraft maneuvering. Therefore, the appropriate comparison is between all model trials and human trials in which the participants did well at executing the maneuver.

Performance

At the highest level of analysis, we are interested in how closely the model approximates expert pilot performance *on the whole*. When UAV operators fly a mission, they typically are responsible for executing hundreds or thousands of maneuvers over many hours. We would hope that *on the whole* the model's performance is at a level of proficiency that reasonably approximates the proficiency of our experts.

Aggregating up to the level of average task performance for flying the UAV STE requires averaging over the airspeed, altitude, and heading deviation performance measures. Those measures are on different scales. Therefore, the RMSD data within each performance measure are converted to z scores. These normalized values are then summed for each trial, resulting in a Sum RMSD (z) score. Those scores are averaged to provide a Mean Sum RMSD (z) score for each participant in each maneuver (49 scores total – 7 participants on each of 7 maneuvers). Those scores are then averaged across maneuvers, to get an average RMSD (z) for each participant. Those seven averages are used to compute a Grand Mean RMSD (z) score and a 95% Confidence Interval for participant performance. The Grand Mean and 95% CI are plotted in Figure 1.

The model data are an average of 20 model runs in each maneuver. The model data are converted to z scores by a linear transformation, using the mean and standard deviation from the normalization of the RMSD's in the SME data. Model data are aggregated up in the same manner as the human data. The model data are plotted as a point prediction because we use exactly the same model for every run, without varying any of the knowledge or parameters that might be varied in order to account for individual differences. The model is a baseline representation of the performance of a single, highly competent UAV operator. There are stochastic characteristics (noise parameters) in ACT-R that result in variability in the model's performance, so we run it 20 times to get an average. This is not the same as simulating 20 different people doing the task. It is a simulation of the same person doing the task 20 times (without learning from one run to the next). The confidence intervals in the human data capture between-subjects variability. Since we just have one model subject, it would be inappropriate to plot confidence intervals. Therefore, it is a point prediction.

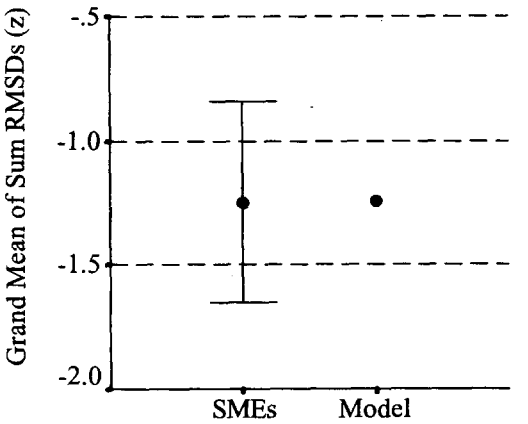

Figure 1: Aggregate comparison of SME performance and model performance

It is reassuring that, at least on the whole, the model flies the UAV STE at a level of proficiency equivalent to that of expert pilots. If we de-aggregate down to the level of average performance on each maneuver, we see that the fit of the model to pilot performance does vary by maneuver. Those data are available in Figure 2.

Across maneuvers, the model corresponds to human performance with an $r^2 = .64$ and a root mean squared scaled deviation (RMSSD) of 3.45, meaning that on

average the model data deviate 3.45 standard errors from the SME data.[1]

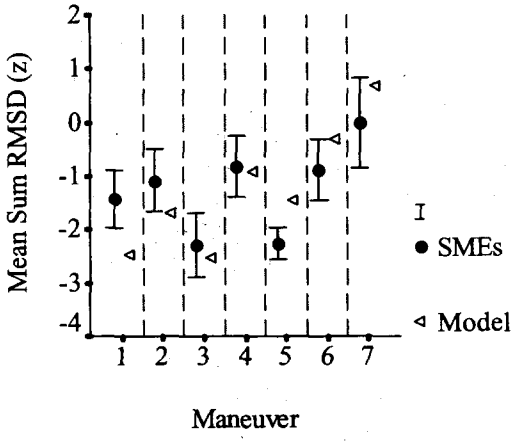

Figure 2: Comparison of SME and model performance by maneuver

It is hard to know whether to be pleased with these fits, since there are still no commonly accepted standards for assessment (Estes, 2002) in the cognitive modeling community. To get a better sense for how we should interpret these results, we ran the same goodness of fit measures for each of the human participants, pulling them one at a time, without replacement, from the sample. We tested the fit of P1 to the data from P2-P7, then the fit of P2 to the data from P1, P3-P7, and so on. The average human fit is $r^2 = .75$ and RMSSD = 2.95. So the model's fit to overall human performance is only a little worse than the average individual human pilot's fit to overall human performance. In fact, it turns out P5's fit to the other participants is $r^2 = .63$ and RMSSD = 4.92, which is actually worse than the model's fit. We interpret this as evidence that the model is a good approximation to expert performance on this task.

There are two things worth noting about the model data. First, the fact that it is a performance model, and not a learning model, does play a role in decreasing the fit to the human data. Since the SMEs progressed through the seven basic maneuvers in sequence, it would be reasonable to assume that more learning occurred during Maneuver 1 relative to Maneuver's 2 through 7. This would explain the relatively large performance difference between SMEs and the model on Maneuver 1. In fact, if we compute the fit using only data from Maneuver's 2 through 7, r^2 increases to .74 and RMSSD drops to 3.20.

Second, it is noteworthy that the model is sensitive to maneuver complexity. Significant main effects of the number of axes maneuvered were observed for both the Model, $F(2,137) = 59.02$, $p < .001$, and SMEs, $F(2,449) = 37.05$, $p < .001$. For both the Model and SMEs, performance was significantly better on one-axis maneuvers compared to two-axes maneuvers, $t(137) = 6.77$, $p < .001$ and $t(449) = 2.95$, $p < .01$, and on two-axes compared to three-axes maneuvers, $t(137) = 5.56$, $p < .001$, and $t(449) = 6.82$, $p < .001$, respectively. Thus, the model captures these difficulty effects, even though it was not intentionally engineered to do so. These effects emerge naturally from the general design of the model.

Variability

There is variability in the model's behavior, but that variability is not represented in Figures 1 and 2. The appropriate comparison for assessing the extent to which the variability in the model's behavior is a good approximation to human variability is a within-subjects comparison. The standard deviation (SD) of the RMSD (z) scores was computed for each participant and the model, separately by maneuver. These SD's were then aggregated up to the task level, in a manner identical to that used for the performance data. The resulting Grand Mean and 95%CI, along with the point prediction for the model's variability, are plotted in Figure 3.

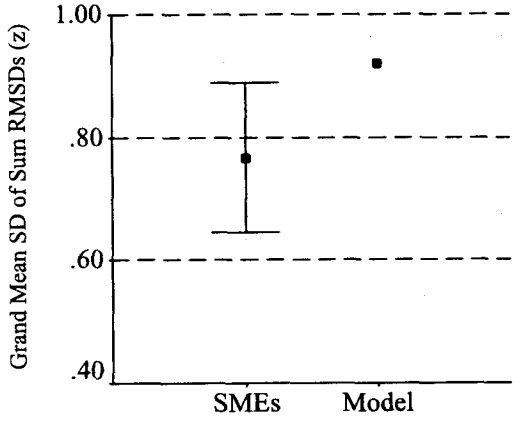

Figure 3: Aggregate comparison of SME variability and model variability

Here we see that variability in the model's performance actually exceeds the human pilot variability. Compared to SMEs, the model was much more variable in performance across trials within each maneuver. The variability in model performance is partially due to the noise parameters in ACT-R, which influence chunk activations and production selection, but also is due to the shifting, dynamic environment in which the model is operating.

Figure 4 plots the SME and model variability comparison by maneuver. Plotting the data by maneuver

[1] See http://www.lrdc.pitt.edu/schunn/gof/index.html for a discussion of RMSSD as a measure of goodness of fit.

reveals that the model was within the 95%CI for human variability on 3 of the maneuvers. However, the model was so much more variable from trial to trial in Maneuver 7 that it's average variability ends up being greater than that seen in the human data. The fit on average within-subject variability between model and SME data was r^2 = .25, and RMSSD = 4.85.

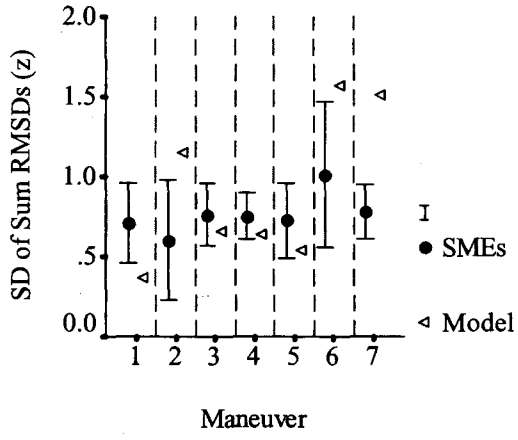

Figure 4: Comparison of within-subject variability

Currently, we are exploring the data logs in more detail to understand why we have more variable performance in the model than we would like, and particularly to understand what is happening in Maneuver 7.

Discussion

Our approach so far in this model development effort has been to use the ACT-R architecture in its current state, and see how far it gets us. We will close by discussing two characteristics of this effort that distinguish it from "typical" or "classical" cognitive modeling efforts.

Perhaps the most important distinction is that, rather than fine-tuning the knowledge and parameters of the model to some specific data set in a post-hoc data fitting exercise, we have used default parameters, and implemented a very general control strategy. This is a modeling approach that can only be attempted with a realistic expectation of success after a user community has had an opportunity to apply an architecture to a sufficiently broad set of empirical results that default, or commonly accepted, parameter settings begin to emerge. It is exciting to see that ACT-R has begun to reach that stage, as evidenced by our results. This does not preclude some possible future attempts at optimizing parameter values or conducting sensitivity analyses, but the default parameters have served us well.

A second point of distinction is that we have implemented this model without modifying or circumventing the architecture. There are no ad hoc modules or buffers. We are using the default 50 msec cognitive cycle time, and all perceptual inputs and motor movements are implemented using class definitions that are consistent with the design of the perceptual and motor modules. It is not unusual, when taking an architecture into unexplored territory, to have to modify or circumvent it in some way, even if temporarily, in order to get the desired behaviors or effects. We have not had to do that.

Acknowledgments

We thank the Air Force Office of Scientific Research for their support. This research is funded by AFOSR grant 02HE01COR.

References

Air Force Manual on Instrument Flight. (2000).

Anderson, J. R., Bothell, D., Byrne, M. D., & Lebiere, C. (2002). An integrated theory of the mind. Submitted to *Psychological Review.*

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*(7), 569-579.

Estes, W. K. (2002). Traps in the route to models of memory and decision. *Psychonomic Bulletin & Review, 9(1)*, 3-25.

Gray, W. D. (2002). Simulated task environments: The role of high-fidelity simulations, scaled worlds, synthetic environments, and microworlds in basic and applied cognitive research. *Cognitive Science Quarterly* 2(2), 205-227.

Martin, E., Lyon, D. R. & Schreiber, B. T. (1998). *Designing Synthetic Tasks for Human Factors Research: An Application to Uninhabited Air Vehicles.* Proceedings of the Human Factors and Ergonomic Society.

Pew, R. W. (2001). Biosketch: Paul Morris Fitts. In W. Karwowski (Ed.), *International Encyclopedia of Ergonomics and Human Factors* (Vol. 1, pp. 17-18). New York: Taylor & Francis.

Schreiber, B. T., Lyon, D. R., Martin, E. L., & Confer, H. A. (2002). *Impact of prior flight experience on learning Predator UAV operator skills* (AFRL-HE-AZ-TR-2002-0026). Mesa, AZ: Air Force Research Laboratory, Warfighter Training Research Division.

Wickens, C. D. (2002). Situation awareness and workload in aviation. *Current Directions in Psychological Science, 11(4)*, 128-133.