Purtee, M. D. Krusmark, M. A., Gluck, K. A., Kotte, S. A., & Lefebvre, A. T. (2003). Verbal protocol analysis for validation of UAV operator model. Proceedings of the 25th Interservice/Industry Training, Simulation, and Education Conference, 1741-1750. Orlando, FL: National Defense Industrial Association.

Verbal Protocol Analysis for Validation of UAV Operator Model

Mathew D. Purtee, Kevin A. Gluck
Air Force Research Laboratory,
Warfighter Training Research Division
Mesa, Arizona
kevin.gluck@mesa.afmc.af.mil
mathew.purtee@mesa.afmc.af.mil

Michael A. Krusmark
L-3 Communications, Inc.
Mesa, Arizona
michael.krusmark@mesa.afmc.af.mil

Sarah A. Kotte, Austen T. Lefebvre
United States Air Force Academy
Colorado Springs, Colorado
c03sarah.kotte@usafa.edu
c04austen.lefebvre@usafa.edu

ABSTRACT

Scientists at the Air Force Research Laboratory’s Warfighter Training Research Division in Mesa, AZ are engaged in a basic research program to advance the state of the art in computational process models of human performance in complex, dynamic environments. Current modeling efforts are focused on developing and validating a fine-grained cognitive process model of the Uninhabited Aerial Vehicle (UAV) operator. The model is implemented in the ACT-R cognitive modeling architecture. The design of the model is inspired by the well-known “Control and Performance Concept” in aviation. The study described here was conducted in order to assess how accurately the model represents the information processing activities of expert pilots as they are flying basic maneuvers with a UAV simulation. The data suggest: (a) pilots verbalize attention to performance instruments more often than control instruments, despite the fact that they generally appear to be using the control and performance concept to fly these maneuvers, (b) the distribution of operator attention across instruments is influenced by the goals and requirements of the maneuver, and (c) although the model is an excellent approximation to the average proficiency level of expert aviators, for an even better match to the process data, the model should be extended to include the use of trim and a meta-cognitive awareness of the passage of time.

ABOUT THE AUTHORS

Mathew D. Purtee is a Warfighter Training Research Analyst for the Air Force Research Laboratory. His key contributions involve verbal protocol analysis and integrating virtual reality with maintenance training. Mr. Purtee has earned a B.S. (1999) in Psychology from Washington State University.

Michael A. Krusmark is a Research Psychologist working for L-3 Communications at the Air Force Research Laboratory’s Warfighter Training Research Division in Mesa, AZ. He earned a Masters degree in Cognitive Psychology from Arizona State University. His research interests include qualitative and quantitative methods for validating human behavior models.

Kevin A. Gluck is a Research Psychologist at the Air Force Research Laboratory’s Warfighter Training Research Division in Mesa, AZ. Dr. Gluck earned a PhD in Cognitive Psychology from Carnegie Mellon University in 1999. He is the Director of AFRL/HEA’s Performance and Learning Models Research Program and his research is in the area of basic and applied computational cognitive process modeling.

Sarah A. Kotte is a cadet at the United States Air Force Academy. She is pursuing a B.S. in Behavioral Sciences. Her research interests include cognition and applied research in human factors. Ms. Kotte has previously worked as a research assistant at the Air Force Research Laboratory, including work examining the effects of simulator training on pilot performance.

Austen T. Lefebvre is currently attending the United States Air Force Academy. He is studying for a B.S. in Human Factors Engineering. His areas of interest include applied research, cognition, and human performance. Previous research includes an internship at the Air Force Research Laboratory.

Report Documentation Page	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE **DEC 2003**	2. REPORT TYPE **Conference Proceedings**	3. DATES COVERED **30-01-2001 to 30-11-2003**

4. TITLE AND SUBTITLE **Verbal Protocol Analysis for Validation of UAV Subject Matter Experts**	5a. CONTRACT NUMBER **AFOSR 02HE01COR**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER **61102F**
6. AUTHOR(S) **Mathew Purtee; Kevin Gluck; Michael Krusmark; Sarah Kotte; Austen Lefebvre**	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Air Force Research Laboratory/RHA,Warfighter Training Research Division,6030 South Kent Street,Mesa,AZ,85212-6061**	8. PERFORMING ORGANIZATION REPORT NUMBER **AFRL; AFRL/HEA**
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) **Air Force Research Laboratory/RHA, Warfighter Readiness Research Division, 6030 South Kent Street, Mesa, AZ, 85212-6061**	10. SPONSOR/MONITOR’S ACRONYM(S) **AFRL; AFRL/RHA**
	11. SPONSOR/MONITOR’S REPORT NUMBER(S) **AFRL-RH-AZ-PR-2003-0004**

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

Published in Proceedings of the 25th Interservice/Industry Training, Simulation, and Education Conference (2003), 1741-1740, held in Orlando FL on 1-4 Dec 03

14. ABSTRACT

Scientists at the Air Force Research Laboratory?s Warfighter Training Research Division in Mesa, AZ are engaged in a basic research program to advance the state of the art in computational process models of human performance in complex, dynamic environments. Current modeling efforts are focused on developing and validating a fine-grained cognitive process model of the Uninhabited Aerial Vehicle (UAV) operator. The model is implemented in the ACT-R cognitive modeling architecture. The design of the model is inspired by the well-known ?Control and Performance Concept? in aviation. The study described here was conducted in order to assess how accurately the model represents the information processing activities of expert pilots as they are flying basic maneuvers with a UAV simulation. The data suggest: (a) pilots verbalize attention to performance instruments more often than control instruments, despite the fact that they generally appear to be using the control and performance concept to fly these maneuvers, (b) the distribution of operator attention across instruments is influenced by the goals and requirements of the maneuver, and (c) although the model is an excellent approximation to the average proficiency level of expert aviators, for an even better match to the process data, the model should be extended to include the use of trim and a meta-cognitive awareness of the passage of time.

15. SUBJECT TERMS

UAV; Uninhabited aerial vehicles; Subject matter experts; SMEs; Basic research; Human performance; Cognitive process model; Control and performance concept; Simulation; Training; ACT-R

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Public Release**	**10**	

Verbal Protocol Analysis for Validation of UAV Operator Model

Mathew D. Purtee, Kevin A. Gluck
Air Force Research Laboratory,
Warfighter Training Research Division
Mesa, Arizona
kevin.gluck@mesa.afmc.af.mil
mathew.purtee@mesa.afmc.af.mil

Michael A. Krusmark
L-3 Communications, Inc.
Mesa, Arizona
michael.krusmark@mesa.afmc.af.mil

Sarah A. Kotte, Austen T. Lefebvre
United States Air Force Academy
Colorado Springs, Colorado
C03sarah.kotte@usafa.edu
C04austen.lefebvre@usafa.edu

PREFACE

Scientists at the Air Force Research Laboratory's Warfighter Training Research Division in Mesa, AZ are engaged in a basic research program to advance the state of the art in computational process models of human performance in complex, dynamic environments. One of the current modeling efforts is focused on developing and validating a fine-grained cognitive process model of the Uninhabited Aerial Vehicle (UAV) Operator. The model interacts with a Synthetic Task Environment (STE) that provides researchers with a platform to conduct studies using an operationally-validated task without the logistical challenges typically encountered when working with the operational military community. This paper will begin by setting the context for the modeling through some background information on the STE. We then briefly describe the general design of the model and compare the model's performance to human performance. The remainder of the paper centers on the use of concurrent and retrospective verbal protocols as a source of validation data for the implementation of the model. The paper concludes with a description of the implications of the verbal protocol results for model development and future research.

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 show that experienced UAV pilots reach criterion levels of performance in the STE faster 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 Basic Maneuvering 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. Before each maneuver is a 10-second lead-in, during which the operator is supposed to fly straight and level. At the end of this lead-in, the timed maneuver (either 60 or 90 seconds) begins, and the operator maneuvers the aircraft at a constant rate of change with regard to one or more of the three flight performance parameters (airspeed, altitude, and/or heading). 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. Maneuvers progressively increase in complexity by requiring the operator to make constant rate changes along two and then three axes of flight. 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.

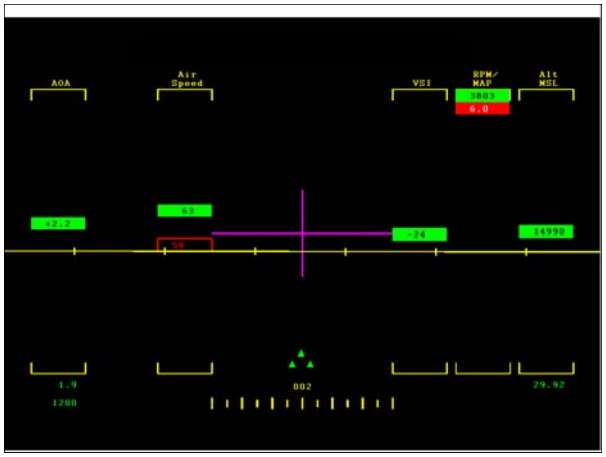
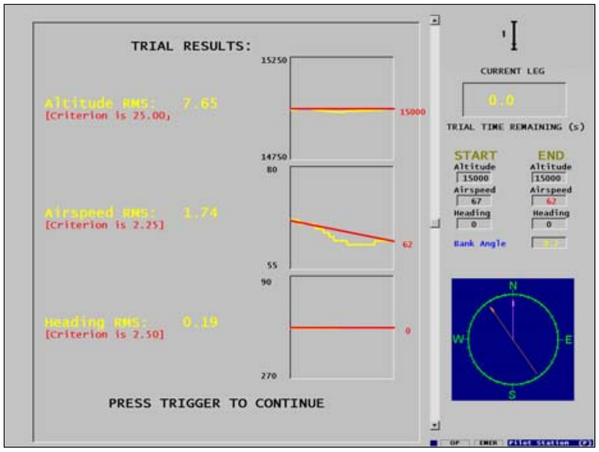


Figure 1. Predator UAV Heads-Up Display

During the basic maneuvering task the operator sees only the Heads-Up Display (HUD), which is presented on two computer monitors. Instruments displayed from left to right on the first monitor (see Figure 1) are Angle of Attack (AOA), Airspeed, Heading (bottom center), Vertical Speed, RPM's (indicating throttle setting), and Altitude. The digital display of each instrument moves up and down as values change. Also depicted at the center of the HUD are the reticle and horizon line, which together indicate the pitch and bank of the aircraft. On a second monitor there are a trial clock, a bank angle indicator, and a compass, which are presented from top to bottom on the far right column of Figure 2. During a trial, the left side of the second monitor is blank. At the end of a trial, presented on the left side of the second monitor is a feedback screen (see Figure 2), which depicts deviations between actual and desired performance on altitude, airspeed, and heading plotted across time, as well as quantitative feedback in the form of RMSD's.


TRIAL RESULTS:

Altitude RMS: 7.65
[Criterion is 25.00]

Airspeed RMS: 1.74
[Criterion is 2.25]

Heading RMS: 0.19
[Criterion is 2.50]

PRESS TRIGGER TO CONTINUE

CURRENT LEG
0.0
TRIAL TIME REMAINING (s)

START	END
Altitude 15000	Altitude 15000
Airspeed 67	Airspeed 62
Heading 0	Heading 0

Bank Angle

Figure 2. Feedback Screen at the End of Maneuver 1

THE UAV OPERATOR MODEL

The computational cognitive process model of the Air Vehicle Operator (AVO) was created using the Adaptive Control of Thought–Rational (ACT-R) cognitive architecture (Anderson, Bothell, Byrne, & Lebiere, 2003). ACT-R provides theoretically-motivated constraints on the representation, processing, learning, and forgetting of knowledge, which helps guide model development. The UAV Operator model was implemented using default ACT-R parameters. Due to space constraints, description of the model will emphasize the conceptual design. For additional model details regarding knowledge representation and architectural parameters, the interested reader is encouraged to see Gluck, Ball, Krusmark, Rodgers, and Purtee (2003), which includes such details, or contact the authors.

The Control and Performance Concept

The "Control and Performance Concept" is an aircraft control strategy that involves first establishing appropriate control settings (pitch, bank, power) for desired aircraft performance, and then crosschecking instruments to determine whether desired performance is actually being achieved (Air Force Manual on Instrument Flight, 2000). The rationale behind this strategy is that control instruments have an immediate first order effect on behavior of the aircraft which shows up as a delayed second order effect in performance instrument readings. Figure 3 is a graphical depiction of the "Control and Performance Concept," as implemented in the UAV Operator model.

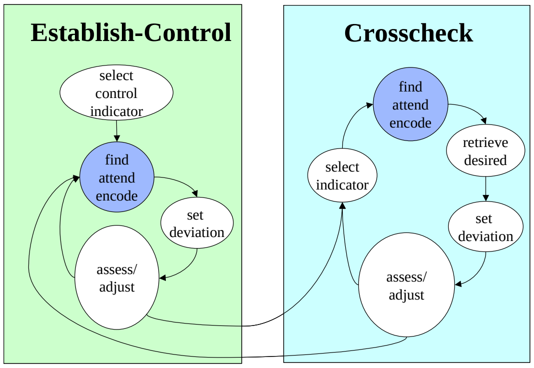


Figure 3. The Model's Conceptual Design

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.

Comparison With Human Data

Human data were collected from 7 aviation Subject Matter Experts (SMEs) at AFRL's Warfighter Training Research Division in Mesa, Arizona. Because recent world events have placed high operational demands on Predator AVOs, we were not able to recruit AVOs to participate in the current research. Therefore, participants were active duty or reserve Air Force pilots with extensive experience in a variety of aircraft, but none had actual Predator UAV flying experience or training. All were mission qualified in Air Force operational aircraft, and all had commercial rated certification. With the exception of one participant, all had airline transport certificates and instrument ratings. Five participants were instructor pilots that graduated from the USAF instructor school. The seven participants had an average of 3,818 hours flying operational aircraft. Prior to data collection, participants completed a tutorial on the Basic Maneuvering Task, during which they familiarized themselves with dynamics of UAV flight and the STE.

Participants completed the 7 basic maneuvers in order, starting with Maneuver 1 and ending with Maneuver 7. Each maneuver was flown for a fixed number of trials that ranged from 12 to 24, depending on the difficulty of the maneuver. SME data plotted in Figure 4 come from successful trials only, where success is defined as flying within performance deviation criteria used by Schreiber et al. (2002). We chose to use human data from successful trials only because (a) participants were not AVOs, and we could minimize and/or eliminate possible effects of learning in the SME's data by using successful trials only, and (b) the current modeling goal is to develop a performance model of skilled aircraft maneuvering, which is best achieved by comparing all model trials with human trials in which participants did well at executing the maneuver.

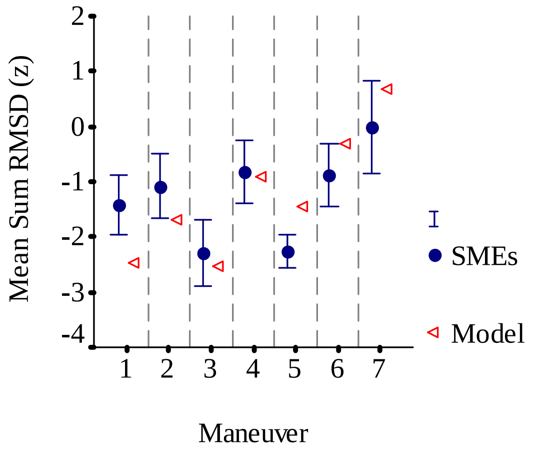


Figure 4. Comparison of SME and Model Performance by Maneuver

Figure 4 plots human and model data for each of the seven maneuvers. Airspeed, altitude, and heading RMSDs were combined to generate a composite measure of performance by first standardizing each performance parameter, because they are on different scales, and then adding the z-scores together. The resulting Sum RMSD (z) scores were then averaged across trials to provide a Mean Sum RMSD (z) score for each participant on each maneuver (49 scores total: 7 participants on each of 7 maneuvers), which were used to compute the means and 95% confidence intervals plotted in Figure 4.

The model data are an average of 20 model runs for each maneuver. The model data are converted to z scores by a linear transformation, using the means and standard deviations used to normalize airspeed, altitude, and heading 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 point predictions for each maneuver because we use exactly the same model for every trial run, without varying any of the knowledge or ACT-R 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 ran it 20 times to get an average. This is not the same as simulating 20 different people doing the task, rather 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.

Across maneuvers, the model corresponds to human performance with an r^2 = .64, indicating that the proportion of variance in the SMEs data accounted by the model is relatively high. In Figure 4 the strength of association between SME and model data can be seen by comparing mean trends, which show that the pattern of results across maneuvers is very similar. Even as the same general mean trend is observed in both the SME and model data, there is deviation between the two, with 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] Although this

---

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

may seem like a large deviation, in research presented elsewhere (Gluck et al., 2003), we have presented a bootstrapping analysis suggesting that deviation of this size is comparable to deviation observed when comparing any one SME's data to the other six SMEs' data. Moreover, given that we have not specifically tuned the model parameters to optimize its fit to the human data, we consider this fit to be fairly good.

Beyond merely examining the quantitative fit of model to human performance data, it is important to consider whether the model is producing desired performance in a way that bears close resemblance to the way human pilots actually do these maneuvering trials. We are interested in developing a model of an UAV operator that not only reaches a level of performance comparable to human operators, but also a model that uses the same *cognitive processes* involved in producing that level of performance. We propose that verbal protocols can be used to reveal valuable insights into these cognitive processes, and will devote the remainder of the paper to examples and discussion relevant to the use of verbal protocols for evaluating the similarity between model and human cognitive processing in complex, dynamic domains.

VERBAL PROTOCOL ANALYSIS

Verbal reports are a source of evidence about human cognition (Ericsson & Simon, 1993). Verbal reporting provides insight into experts' attention patterns and cognitive activity. Studying verbal reports of expert pilots provides information regarding their attention to instruments and mental processes while operating aircraft, which can provide a better understanding of pilots' strategies and goals. Such information subsequently can be used to improve computational cognitive process models of pilot behavior as well as pilot training. Verbal protocols provide a window into the mind of the participant, but do not impose a heavy cognitive or physical burden on the participant. In the aviation world this is especially beneficial because researchers want as much information as possible with as little interruption to the task as possible.

It is important to distinguish two types of protocol collection: concurrent and retrospective. Concurrent protocol collection takes place during an experiment as a participant performs a task. The resulting data is of high density, and provides a good view into the real-time cognitive activities of the participant, since forgetting over time is not a factor (Kuusela & Paul, 2000). Retrospective protocol collection requires that after the task is completed, participants think back about their processing and report what they think they were doing. Combining both concurrent and retrospective reporting is recommended (Ericsson & Simon, 1993; Kuusela & Paul, 2000), because it provides multiple sources of verbal evidence on which to base one's conclusions.

Ericsson and Simon (1993) proposed three criteria that must be satisfied in order to use verbal protocols to explain underlying cognitive processes. First, protocols must be relevant. The participant must be talking about the task at hand. It is important to keep participants on track. The second criterion is consistency. Protocols must flow from one to the other and be logically consistent with preceding statements. If protocols jump from topic to topic without any transitions, this could indicate that intermediate processing is occurring without representation in the protocols. In other words, there is information missing in the statements provided. Third, protocols must generate memories for the task just completed. A subset of the information given during the task should still be available after completion of the task. This ensures that the participants gave information that actually had meaning to them. Additionally, it indicates that the information provided was important to the participant at that time.

It is important to consider certain aspects of the task when deciding whether to collect verbal protocols (Svenson, 1989). One aspect is level of familiarity with the task. If the participant is unfamiliar with the task and must concentrate on learning it, protocols regarding strategy will not be provided. Participants must be very familiar with the task so that protocols will be meaningful and relevant to strategy. The participants in the study described here are expert aviators and were intimately familiar with basic aircraft maneuvering and instrument flight. Another relevant aspect is the complexity of the task. A simple task runs the risk of becoming automated, thus not eliciting rich protocols. Svenson recommends that a task have at least four separate categories of information that can be verbalized. In the task used, there are 10 instrument displays relevant to basic maneuvering and it was clear none of the participants believed that the task was simple or easy.

A shortcoming of concurrent verbal protocols is that it is virtually impossible to capture all cognitive events. However, we assume that, on the whole, participants verbalize most of the contents of their verbal working memory, and that verbalization patterns will reflect patterns of attention and/or cognitive processes.

Table 2. Code Definitions and the Overall Frequencies that they were Reported

Code	Definition	Frequency
Goals		
Altitude	Refers to altitude performance target(s)	112
Heading	Refers to heading performance target(s)	58
Airspeed	Refers to airspeed performance target(s)	40
General	Underspecified goal statement	14
Prospective	Future intention that includes explicit reference to future time	1
Control Instruments		
Bank Angle	Mentions bank angle	828
Pitch	Mentions pitch or reticle	316
RPM	Mentions RPMs	238
Trim	Mentions Trim	24
General	Mentions general control settings	12
Performance Instruments		
Altitude	Mentions altitude or altitude change	2428
Heading	Mentions heading or any of the heading indicators	1049
Airspeed	Mentions airspeed	2264
Time	Mentions time remaining, time passed, or current time	1316
General	Mentions general performance process or outcome.	791
Actions		
Throttle	Statements of action or *current* intent specific to throttle	1368
Stick Pitch	Statements of action or *current* intent specific to pitch	1298
Throttle or Stick Pitch	Statements of action or *current* intent that could be either throttle or pitch	1281
Stick Roll	Statements of action or *current* intent specific to roll	1422
Trim	Statements of action or current intent specific to trim	133
General	Unspecified or under-specified statement of *current* intent	423
Other		
Evaluative Exclamations	Vague, reactive expressions	132

METHOD

Participants were the 7 aviation SMEs that were previously described in the comparison between human and model data. While performing the Basic Maneuvering Task, participants verbalized on odd numbered trials. The recorded verbalizations were then transcribed, segmented, and coded. Following completion of all trials of each maneuver, SMEs were asked a series of questions to determine what strategies they believed they were using to complete each maneuver, which are the retrospective reports of strategy.

Concurrent Verbal Reports

Segmenting. The transcribed stream of continuous concurrent protocol data was segmented into distinct verbalizations. Table 1 lists the rules that guided segmentation of the transcribed data. One researcher segmented all of the verbalizations, while another segmented approximately one third of the data. The two agreed on 88.5% of segmentations. Disagreements were mutually resolved for the final data set, which contains 15,548 segments.

Coding. To quantify the content of the segmented verbalizations, a coding system was developed, which is presented in Table 2. The coding system has five general categories of verbalizations: Goal, Control, Performance, Action, and Other. Within each general category of verbalization are more specific codes that allow a more fine-grained analysis of the attentive and cognitive processes of the pilots in this study. One researcher coded all of the segmented verbal protocol data while another researcher coded a third of the data set. Agreement between the 2 coders was high, with Kappa = .875.

Table 1. Segmentation Rules

1. Periods, question marks, exclamation points, "…" and "(pause)" always indicate a break.
2. Segment breaks are optional at commas and semi-colons.
3. Conjunctions and disjunctions (and, or, so, but) typically indicate a segment break.
4. Judgment verbalizations should be kept in the same segment with the reference instrument ("airspeed is at 62, that's fine").
5. Exclamations (e.g., "Jeez", "Damn", "Whoa") are separate segments.
6. "OK …" and "Alright …", when followed by a comma are included in the same segment with the text that follows.
7. Repeated judgments separated by a comma (e.g., "bad heading, bad heading") are not segmented.
8. When separated by a period (e.g., "Bad heading. Bad heading.") They are separate segments.

Effect of concurrent verbal reports on performance. One might be concerned that providing concurrent verbal reports increased cognitive demands of the Basic Maneuvering Task and therefore degraded performance. Because participants provided concurrent verbal reports on odd trials only, we were able to assess whether performance was worse when participants provided verbalizations. Because performance on the first trial of each maneuver was dramatically worse than performance on the second and subsequent trials, the first two trials of each maneuver were eliminated from the comparison of verbal protocol condition. Across all trials but the first two trials of each maneuver, no effect of verbal protocol condition was found on altitude, airspeed, and heading RMSDs, suggesting that performance was not degraded when participants provided concurrent verbal reports.

Retrospective Reports

The retrospective reports were coded by two behavioral scientists for the presence of references to: (a) the use of a "control and performance" strategy, (b) reference to trim, and (c) reference to clock use. A response was coded as indicating use of the Control and Performance Concept if a participant mentioned setting one of the control instruments. Responses were coded further to include information about which control instruments were set (i.e., pitch, bank, or power): A response was coded as indicating use of trim if the participant mentioned using trim, no trim if the participants did not mention the use of trim, and abandon trim if the participant discussed or alluded to using trim and then discusses that trim use was discontinued. When the participant mentioned clock use in some form, either as a reference to the clock itself, discussing checkpoints or timing, or the use of seconds in their response, this was coded as a reference to clock use.

RESULTS AND DISCUSSION

Evidence That Participants Used the Control and Performance Concept

Concurrent verbal reports. The Control and Performance concept informed our expectations of how attention would be verbalized across coding categories. We expected that if participants were using the control and performance concept, then they would verbalize control statements just as frequently, or more so, than performance statements. Figure 5 displays the mean percentage of concurrent verbal reports that were coded as goal, control, performance, and action statements. The mean percentages of verbalizations within each code category were computed by first calculating the percentage of verbalizations of each code within each trial, and then averaging within-trial percentages of codes across trials and maneuvers. As you can see in Figure 5, the distribution of coded verbalizations across category code was relatively consistent among participants, and they tended to verbalize attention more to performance instruments than to control instruments. Goals were verbalized least frequently, possibly because when goals were verbalized, it was usually slightly before timing checkpoints at 15, 30, and 45 seconds into a trial, and those checkpoints only occur three or four times per trial.

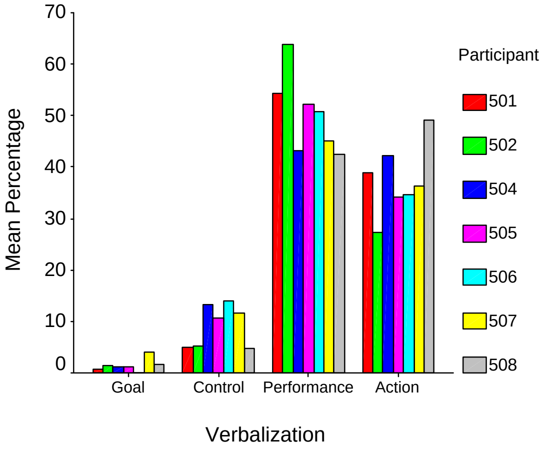


Figure 5. Percentage of Verbalizations Within Category for Each Participant

Figure 6 presents the mean percentage of specific control statements that were verbalized by maneuver. As can be seen, when participants verbalized their attention to control instruments, it was primarily to the bank indicator. Naturally, that almost always occurred on the trials that involved heading changes (2, 4, 6, and 7), but we will focus on effects of maneuver on verbalization patterns in the next section. **[**Rarely did participants verbalize that they were attending to pitch, which would have been represented in statements where they mentioned "pitch", "reticle", "ADI", and the like. Participants verbalized attention to RPM's even less frequently. With attention to performance instruments verbalized at 4-5 times the rate of attention to control instruments, the concurrent verbal protocols do not reveal the pattern predicted if the participants were using a Control and Performance strategy for their basic maneuvers. Based solely on results of concurrent verbal reports, there seems to be little evidence that participants used the Control and Performance concept as a strategy for maneuvering the simulated Predator UAV.

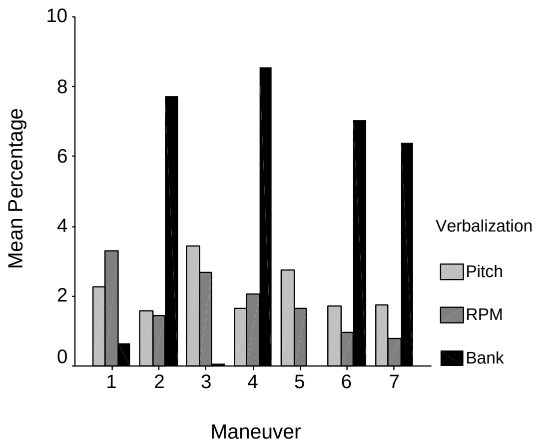


Figure 6. Percentage of Control Verbalizations Within Each Maneuver

Retrospective reports of strategy. If we consider the participants' retrospective reports of strategy, however, we find that all participants reported using the Control and Performance Concept on all maneuvers. Figure 7 depicts for each maneuver the number of participants that reported maneuvering the UAV by setting pitch, RPM, or bank values. As can be seen, on all maneuvers most participants reported that they were attending to at least one control instrument in an attempt to set values required for a given maneuver, and that is the essence of the Control and Performance Concept.

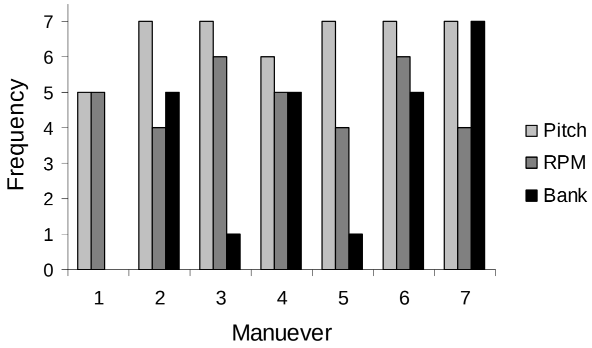


Figure 7. Frequency of Reports Indicating Setting Pitch, Bank, and RPM Values on Each Maneuver

Discussion and Implications for Modeling. How do we reconcile data from retrospective reports suggesting that participants were using the Control and Performance Concept with data from concurrent verbal reports suggesting that they were not? One possible explanation comes from how information is represented in different instruments on the HUD. Reports from participants suggest that on most maneuvers they were using the ADI to "set a pitch picture" to control the UAV simulator. The ADI represents graphically information about the pitch and roll of the UAV. Thus, before a participant can verbalize information from the HUD, it has to be encoded in its graphical representation, converted to a verbal representation, and then verbalized. With the exception of the compass and heading rate indicators, which depict heading information graphically, all other instruments on the HUD of the UAV represent information with digital values. Thus, because of the high demands of the task, it is entirely plausible that when participants are attending to the ADI they fail to verbalize it in concurrent reports because the cognitive effort in doing so would interrupt their natural stream of thought, and degrade their performance. Moreover, the fact that the ADI is not labeled on the HUD, whereas most other control and performance instruments are, further hinders the process of verbalizing attention to the ADI. In summary, the propensity for participants to verbalize attention to performance instruments and not control instruments is likely due to the relative ease with which performance instrument values are verbalized and the difficulty with which control instrument values are verbalized.

Regarding the computational cognitive process model, these results are encouraging. The paucity of evidence in the concurrent verbal protocol data for a maneuvering strategy based on the Control and Performance Concept is more than made up for by the overwhelming evidence for that strategy in the retrospective reports. It clearly is the case that the general maneuvering strategy around which the model was constructed is a realistic one, and we are satisfied that it is the right way to represent expert performance in the basic maneuvering tasks. Future analyses of eye tracking data (now underway) should further substantiate this conclusion.

Evidence That Participants Allocated Their Attention Differently Across Maneuvers

Concurrent verbal reports. Figure 8 displays performance verbalizations with respect to specific maneuvers. Similar to the "bank" verbalizations in Figure 6, there is a large effect of maneuvering goal on "heading" verbalizations. Participants verbalized attention to heading much less frequently on maneuvers where they did *not* change heading (1, 3, and 5) compared to maneuvers where they did change heading (2, 4, 6, and 7).

If we look at the goals that participants verbalized during concurrent reports, we find further evidence for task specific allocation of attention (See Figure 9).

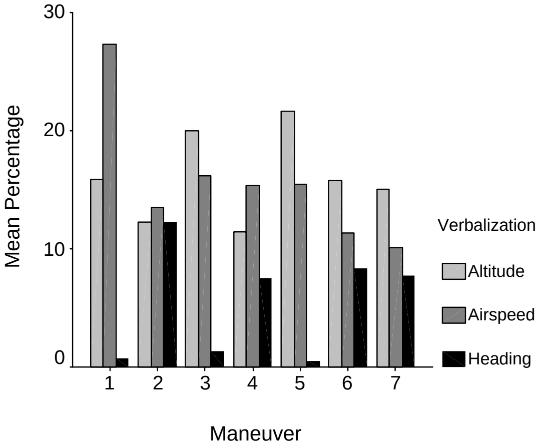


Figure 8. Percentage of Performance Verbalizations within each Maneuver

Heading goals were verbalized much less frequently, or not at all, on maneuvers that required no heading change (Maneuvers 1, 3, & 5). Likewise, altitude and airspeed goals (particularly altitude) were verbalized much more often on maneuvers that required altitude or airspeed changes (Maneuvers 3, 5, 6, & 7; and 1, 4, 5, & 7 respectively).

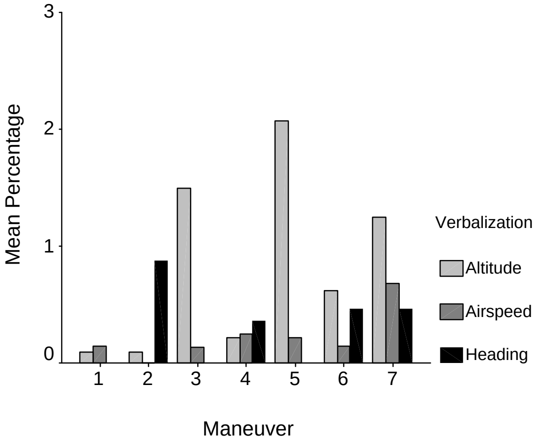


Figure 9. Percentage of Goal Verbalizations within Each Maneuver

Retrospective reports of strategy. Finally, participants' retrospective reports further corroborate the claim that the goal of the maneuver influences allocation of verbalized attention across instruments. If we look again at Figure 7, we see that most participants reported using a strategy of attending to the bank angle indicator to set desired roll primarily on maneuvers that require a heading change (2, 4, 6, & 7). Because proper pitch and power settings are required for all maneuvers, participants did not report strategies suggesting differential use of these indicators across maneuvers.

Discussion and Implications for Modeling. Evidence from both concurrent and retrospective reports are consistent in suggesting participants allocate their attention differently depending on the maneuver. Refreshingly, the model is already implemented in this way. The declarative memory structure in the model is designed such that the maneuvering goal spreads activation to declarative chunks representing instruments that are relevant to that particular goal, thereby increasing the probability of selecting a relevant instrument on the next shift of visual attention. So we do see a similar effect of maneuver on the distribution of the model's attention. The model does not actually *verbalize*, of course, so a more direct comparison is not possible.

Additional Evidence Informing Model Development

In addition to coding retrospective reports for evidence of Control and Performance strategies, we also coded these reports for use of trim and timing checkpoints. Information on use of the trim and the clock provides additional information regarding the strategies of participants when attempting to complete the maneuvers.

Two of the seven SMEs reported using trim on three maneuvers, including the most difficult maneuvers, 6 and 7. One other SME reported using the trim on earlier maneuvers, but abandoned its use on later maneuvers, as it failed to be an effective strategy. Although the sample size is small for such a comparison, the two pilots that reported success when using trim were not any better at successfully completing maneuvers than pilots that did not use trim. Currently, the model does not use trim at all when flying the basic maneuvers. This seems like a reasonable design decision, given that less than half of the human experts chose to use trim on these trials, and not all of those who did use trim thought it was effective. Admittedly, however, the model's generalizability and real-world utility would increase if we incorporated the knowledge necessary for trim use. This is an opportunity for future improvements to the model.

Retrospective strategies were also coded for use of the clock. Six of the seven pilots reported using the clock, or timing checkpoints, to successfully complete the task. It is hardly surprising that this strategy was used by most participants, since the instructions for each maneuver suggest specific timing checkpoints for monitoring progress toward the maneuvering goal.

However, that the clock was used consistently by participants suggests that it should be incorporated into our model of a UAV operator, and in fact it is. The checkpoints recommended in the maneuver instructions are represented as additional declarative chunks in the model. These are retrieved from memory whenever the model checks the clock, and then used to modify the desired aircraft performance goal, on the basis of how far the model is into the maneuver. Anecdotal evidence suggests there is a subtle difference between the way the model uses the clock and the way humans use it. The participants are slightly more likely to check the clock near the recommended timing checkpoints, presumably because they have a meta-cognitive awareness of the passage of time. The model has no such awareness of psychological time. Adding that capability in a psychologically plausible way would be a substantial architectural improvement, but is outside the scope of our current research effort.

CONCLUSION

This study assessed how accurately our UAV Operator model represents the information processing activities of expert pilots as they are flying basic maneuvers with a UAV simulator. A combination of concurrent and retrospective verbal protocols proved to be a useful source of data for this purpose. Results showed that (a) the general Performance and Control Concept strategy implemented in the model is consistent with that used by SME's, (b) the distribution of operator attention across instruments is influenced by the goals and requirements of the maneuver, and (c) although the model is an excellent approximation to the average proficiency level of expert aviators, for an even better match to the process data it should be extended to include the possible use of trim and a meta-cognitive awareness of the passage of time.

In future research verbal reports will be combined with eye-tracking data to provide the best possible understanding of the cognitive processes involved in flying basic maneuvers with the UAV STE. Even further down the road, we will be extending the basic maneuvering model to a model that flies reconnaissance missions (in the STE).

ACKNOWLEDGEMENTS

We thank the Air Force Office of Scientific Research for their support. This research is funded by AFOSR grant 02HE01COR. Our sincere thanks also to the SMEs at AFRL for their time and effort, without which this study would literally have been impossible.

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.

Ericsson, K. A & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87,* 215 – 251.

Gluck, K. A., Ball, J. T., Krusmark, M. A., Rodgers, S. M., & Purtee, M. D. (2003). A computational process model of basic aircraft maneuvering. *Proceedings of the 5th International Conference on Cognitive Modeling*, 117-122. Universitats-Verlag Bamberg.

Kuusela, H. & Paul, P. (2000). A comparison of concurrent and retrospective verbal protocol analysis. *American Journal of Psychology 113*(3), 387-404.

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.

Reisburg, D. (1999). Learning. In *The MIT Encyclopedia of the Cognitive Sciences* (Vol. 1, pp. 460-461. Cambridge: MIT Press.

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.

Svenson, O. (1989). Illustrating verbal protocol analysis: Individual decisions and dialogues preceding a joint decision. In H. Montgomery & O. Svenson (Eds.), *Process and structure in human decision making*. (pp. 83-98). Oxford, England: John Wiley & Sons.