Modeling Pilot Cognitive Behavior for Predicting Performance and Workload Effects of Cockpit Automation

Guk-ho Gil

David Kaber

Sang-Hwan Kim

Karl Kaufmann

Theo Veil

MODELING PILOT COGNITIVE BEHAVIOR FOR PREDICTING PERFORMANCE AND WORKLOAD EFFECTS OF COCKPIT AUTOMATION

Guk-ho Gil, David Kaber, Sang-Hwan Kim, Karl Kaufmann & Theo Veil
Industrial & Systems Engineering, North Carolina State University
Raleigh, NC
Paul Picciano
Aptima Corporation
Boston, MA

The objective of this research was to develop a model of pilot cognitive behavior to predict performance and workload while using varying degrees of cockpit automation to serve as a basis for future systems design. A cognitive task analysis (CTA) was conducted on expert pilot performance a flight control panel (FCP), control-display unit (CDU) and flight management system, and an enhanced CDU (CDU+) providing pre-programmed arrivals from air traffic control in a simulated landing and approach task. Cognitive models were developed from the CTA using an enhanced form of the GOMS language, including a set of additional task operators, to represent pilot actions on cockpit interfaces. Pilot performance and workload data from a parallel empirical study of the same flight tasks were used as a basis for validating the cognitive model output. Indices of automation complexity were formulated based on counts of task methods and steps, required chunks of information, and information transactions coded in the enhanced GOMS models. These indices revealed high complexity for the FCP mode and low complexity for the prototype CDU+ mode. The automation index values were positively and significantly correlated with pilot heart rate (as an objective measure of workload) and vertical path deviation error from the experimental data set. The computational cognitive models of pilot behavior in using forms of cockpit automation were demonstrated to be a viable tool for predicting pilot workload and flight performance under high workload flight conditions.

Early research on cockpit automation (e.g., Wiener & Curry, 1980) identified potential human performance consequences resulting from a technology-centered approach to automation design implementing automation wherever and whenever possible, while leaving unanticipated and unstructured tasks to the pilot. These consequences include pilot complacency, vigilance decrements, loss of situation awareness and decision making problems. A number of empirical studies subsequently demonstrated such negative effects of technology centered automation design (e.g., Parasuraman et al., 1992; Endsley & Kiris, 1995) both in the aviation context and other domains. On this basis, human-centered approaches to cockpit automation (e.g., Billings, 1997) were proposed. This includes considering the information processing and performance capabilities of pilots as well as how pilots interact with cockpit interfaces. Empirical studies were conducted to determine the impact of various levels of automation on human performance, workload and situation awareness in aviation-related tasks (e.g., Endsley & Kaber, 1999), which led to guidelines for the use of intermediate modes of automation (between manual control and full automation). Beyond this, qualitative models for selecting the types and levels of automation applicable to human-machine systems (Parasuraman et al., 2000) were developed.

The main issue with the existing approaches to cockpit automation design is that they require empirical data as a basis for design alternative selection or they are based on collections of design guidelines with limited theoretical explanation of why such guidelines might be effective. Experimental studies to obtain necessary data are time consuming and costly. Also, the lack of a cognitive explanation for why certain design principles may be useful limits understanding of when and how guidelines should be applied. With this in mind, there is a need to develop computational models of pilot behavior in interacting with cockpit automation as a basis for reducing experimentation to assess or validate specific forms of automation. Such models can also provide a basis for explaining the effects of automation design guidelines in terms of perceptual processing, memory transactions, decision rule use, and response execution; thereby providing a more theoretical foundation of human-centered design of automation. Based on the prior research, the objective for the present study was to develop a computational (computer-based) model of pilot cognition interacting with various forms of cockpit automation as a basis for future system design.

Method

Flight Simulator and Flight Scenario

A PC-based flight simulator was setup for cockpit automation prototyping and to collect data on pilot performance for use in the cognitive model validation step. The simulator setup consisted of two PCs and flight deck controls, including a yoke, a throttle quadrant, and rudder pedals (see Figure 1 (a) for the simulator setup and displays) integrated with the X-Plane simulator software. Two LCD monitors were arranged vertically with the lower display presenting the instrument panel of the Boeing 767-300, including the primary flight display (PFD), flight control panel (FCP), and control display unit (CDU) (or flight management system (FMS)) interface. The upper display showed an out-of-cockpit view of the dynamic flight situation rendered by X-Plane. The display contents of the two monitors were synchronized using a TCP/IP network supported by the X-plane software.

(a) (b)

Figure 1. *Simulator setup (a) and image of X-Plane displays (b).*

A realistic arrival and landing scenario was created to support the objectives of conducting a CTA on pilot interaction with cockpit automation and the experimental study of the performance effects of automation in addressing normal events during a high workload phase of flight. Reno-Tahoe International Airport (KRNO) was chosen for its proximity to significant terrain and selection of instrument approaches and arrivals. There were three critical events pilots encountered in the flight scenario. The first critical event was a re-clearance from the northern standard terminal arrival (STAR) to the southern STAR due to a runway changing. This occurred 5 NM from the first waypoint, which served both STARs, and pilots had a very short period of time to interpret the clearance and command the aircraft to turn onto the new STAR. The second critical event was a northbound leg of the STAR to align the aircraft with the ILS final approach. This leg was defined as the backcourse of the ILS serving the opposite runway. Backcourse procedures are familiar to all instrument rated pilots, but they are not often encountered in normal service. This required extra effort from pilots to recall and carryout the correct procedures at the proper times. The last critical event was a clearance to descend from the initial altitude. If there was any delay in beginning the descent or if the rate of descent was too low, intercepting the glideslope became very difficult.

Three Interfaces Representing Different Forms of Cockpit Automation

There were three different modes of cockpit automation (MOAs) that were simulated through the X-Plane software. Each MOA had four types of information processing functions (TOF) including perception of flight status (TPF-P), flight information analysis (TOF-IA), decision making on flight path (TOF-DM), and pilot action implementation (TOF-AI). In the FCP mode, X-Plane presented the B-757/767 flight control panel. Pilots used the FCP display for tracking altitude and speed (TOF-P) and they dialed-in flight path targets (TOF-AI) during the experiment. Because, X-Plane does not provide the B-757/767 CDU, a new realistic CDU interface was developed using the X-Plane SDK. This was then employed for the CTA and pilot performance study. With respect to the CDU+ mode, the main difference from the CDU mode was that the system was capable of presenting to the pilot (TOF-P) ATC suggested routes including vertical path, when changing or deciding on other routes (TOF-DM) under inclement weather conditions, etc. With these pre-planned routes, pilot control actions (TOF-AI) were dramatically reduced, as the CDU+ required no pilot interaction during the STAR, once the desired runway for landing was selected.

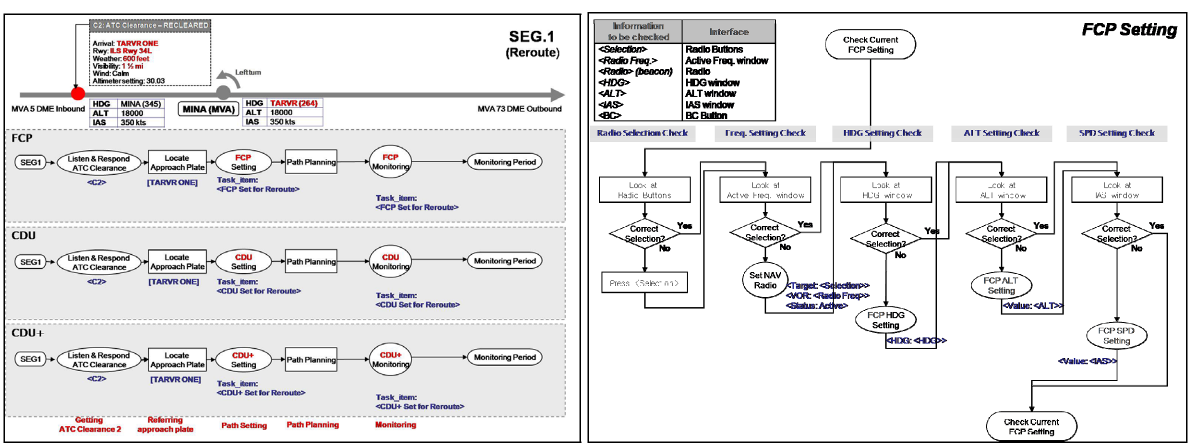
Cognitive Task Analysis

There was a need to develop an understanding of the commercial transport pilot's working context as a starting point for the cognitive modeling effort. Kieras (1997) suggested that cognitive modeling starts with a CTA. The purpose of this step in the research was to identify expert pilot behaviors in flying the high workload landing approach scenario using the different forms of cockpit automation simulated through the enhanced X-Plane setup. Specifically, the CTA was expected to reveal pilot goals, decisions, information requirements, and tasks in achieving goals at various stages in the approach. Information from verbal protocols and goal-directed task analyses (Endsley, 1993) was used to develop the computational cognitive models of pilot behavior with the FCP, CDU and CDU+ modes of control.

The CTA required several steps, including: (1) videotaping expert pilot performance with the X-Plane simulation in the test flight scenario; (2) recording pilot verbal protocols and transcribing them; (3) formulating pilot task lists for each MOA. Table 1 shows an example of task items for the FCP mode at a specific location (73 DME from the MINA VOR (MVA) outbound) after receiving a clearance from ATC according to the flight scenario; (4) developing pilot action flow diagrams (AFDs) of overt and cognitive behaviors as the basis for cognitive model coding. Figure 2 shows example AFDs for the use of the three different MOAs in the rerouting task (Figure 2(a)) and a sub-task flow to check FCP settings and the required information for the task (Figure 2 (b)); and (5) expert pilot verification of the AFDs for accuracy in describing behaviors with the automation in the various phases of the approach. For the first, second and fifth steps of this procedure, a highly experienced former USAF transport pilot (C-130) with ATP certification served as the expert pilot.

Table 1. *Example of task items for FCP use.*

Location	Current Status (Expected)		ATC Clearance		FCP Tasks	FCP Objects
MVA 73 DME outbound	NAV1	MVA/I-RNO			Descending (to 16000)	V/S mode
	NAV2	FMG			Speed down	IAS knob
	Source	NAV1	Altitude	16000	Switch Radio	NAV1 Radio
	Altitude	18000	Speed	250	HDG Setting (344)	HDG knob
	IAS	350	Altimeter	30.03	BC toggle on	BC button
	HDG	283			Altimeter Setting	Altimeter knob



(a) (b)

Figure 2. *Example of AFDs for general flow of rerouting task (a) and sub-task for checking FCP setting (b).*

Development of GOMS Models

Enhanced GOMS (E-GOMS) models were created based on the results of the CTA, specifically the AFDs. The general structure and flow of the models was similar to NGOMSL (Kireas, 1997); however, E-GOMS included an expansion on theGOMSL available operator set to more accurately represent pilot actions on cockpit interfaces (e.g., dialing knobs). The E-GOMS models included a main (task) goal, sub-methods and operators for each sub-method, as well as task item representation. Two major features of each model were the description of the action flow and the information object set. Models not only represented pilot behaviors, but also the information to be manipulated during flight tasks (e.g., from external ATC clearances or internal path planning). All information objects were coded as audio objects with their own variables and values. For example, the information object for the CDU SPD/ALT setting had two variables, a SPD value and an ALT value. Internal path plans were represented as task-items.

Empirical Study

A lab experiment was conducted to assess the effects of the FCP, CDU and CDU+ modes of automation on pilot performance, and subjective and objective workload responses (NASA-TLX and heart rate, respectively). The experiment used the same scenario as used for the cognitive model development (high workload landing approach with a "last minute" reroute, steep descent and speed reduction). The main objective was to test hypotheses on the potential for pilot flight control errors in response to critical events based on the nature of the automation interfaces and functionality (e.g., the CDU MOA was expected to produce greater waypoint over shoot errors upon the reroute due to the complexity of flight path reprogramming). The experiment also served to generate a data set for preliminary validation of cognitive model output.

Results

Experiment

Pilot performance results revealed highly significant effects of MOA among data segments including vertical and lateral path deviations ($p<0.0001$). Pilot objective workload (heart rate) revealed significant effects of MOA and there was an interaction of MOA and flight segment across test trials ($p=0.0487$) when trial order was considered in the statistical model. Pilot subjective workload ratings (NASA-TLX) revealed a marginally significant effect of MOA ($p=0.0949$) when trial order was considered in the model. In general, these results indicated an influence of the FCP, CDU, and CDU+ modes of control on pilot behavior and motivated the cognitive model development effort.

Cognitive Model Outputs

As previously mentioned, the cognitive models were analyzed manually for pilot performance predictions with the various forms of cockpit automation. Since the flight scenario was divided into three segments for analyzing the actual pilot performance data from the lab experiment, the cognitive model outputs were also determined and analyzed according to the same three segments (rerouting, turning, and final approach). In general, the outputs from the E-GOMS models can be characterized as task complexity indices for each MOA and flight segment. Four indices were determined for this study, including: (1) the number of sub-methods to perform tasks during a flight segment; (2) the total number of steps in the model, including those as part of required sub-methods; (3) the required number of information elements to complete a task during a segment (including the sub-methods); and (4) the number of information transactions between WM and LTM or external memory (e.g., pilot notes on an approach plate).

Table 2 shows the values for the task complexity indices for each MOA and flight segment, as determined from the E-GOMS models. It should be noted that the indices for the final flight segment are the same across MOAs because only the FCP mode was used in this segment. In general, the FCP mode produced larger index values than the CDU and CDU+ modes. The CDU+ mode generated the smallest index values among all modes. Therefore, the CDU+ mode was considered to pose the lowest level of task complexity and use of the FCP mode yielded the highest level of task complexity.

Table 2. *Calculated task complexity indices for each MOA and flight segment.*

		Mode of Automation		
		FCP	**CDU**	**CDU+**
Seg. 1	# of Sub-methods	7	6	6
	# of Steps	74	77	69
	# of Information	32	24	24
	# of Transactions	4	3	3
Seg. 2	# of Sub-methods	16	12	10
	# of Steps	169	151	124
	# of Information	63	46	42
	# of Transactions	12	11	11
Seg. 3	# of Sub-methods	7	7	7
	# of Steps	75	75	75
	# of Information	31	31	31
	# of Transactions	7	7	7

On the basis of these index values, the potential for flight errors can be predicted. Kieras (1997) noted that, if more than five (5) chucks of information must be maintained in WM at any given time, this lead to cognitive overload and, consequently, induce errors in performance. Figure 3 shows a plot of the number of chunks of information required by a pilot during the second flight segment (turning) under each MOA. It can be seen from the plot that the number of chunks for setting the FCP control to turn the aircraft at TARVR is 16, while the other modes of control (CDU and CDU+) required less than two (2) chunks of information. Even though the task of setting the FCP for turning can be further decomposed into heading setting, altitude setting, radio setting and air speed setting, the amount of information that must be manipulated by a pilot at a given time exceeds the criteria suggested by Kieras (1997) and the "magic number" of working memory capacity identified by Miller (1956). Thus, it can be predicted based on the cognitive model output that a pilot may make flight errors in setting the FCP for turning descent of the aircraft under high workload conditions. Based on the results of the experiment with actual pilots, it was observed that some participants did not set the FCP appropriately at this point in the flight and this produced greater path deviation than for the CDU or CDU+ modes.

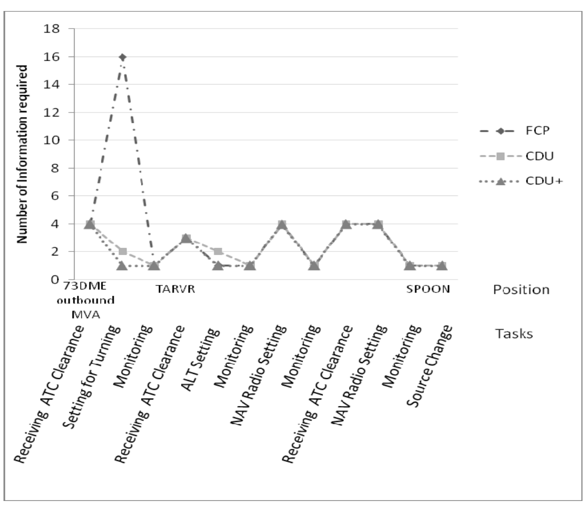


Figure 3. *Number of chunks of information required during the second flight segment.*

Comparison of Model Outcomes with Experiment Data

Non-parametric correlation (Spearman) analyses were conducted on the task complexity index data and observations on the workload and performance response measures from the experiment. Since only the FCP mode of control was used in the final segment of the flight scenario, data for the first and second segments were used for comparison of model outputs with the pilot HR and path deviation responses. In addition, a composite task difficulty index was determined based on the E-GOMS models for all three segments of flight (across all pilots) for correlation with the NASA-TLX scores, determined at the close of trials.

Results revealed the pilot HR responses were highly correlated with all four model-based task complexity indices (r= 0.928, 0.829, 0.928, 0.883; number of sub-methods, number of steps, required chunks of information, and information transactions, accordingly) with a significance level of p=0.05. Additional correlation results revealed NASA-TLX scores to be positively correlated with the number of sub-methods, number of method steps, and number of required chunks of information. Unfortunately, there were too few data points for the significance levels to be considered reliable. Related to this, the number of information transactions was not significantly correlated with the subjective workload data. In addition, there were positive linear relations between vertical path deviation and model outcomes including: number of sub-methods (r=0.978, p=0.008); number of steps (r=0.886, p=0.019); number of required information elements (r=0.978, p=0.008); and number of information transactions (r=0.971, p=0.001). However, there was no significant correlation between the lateral path deviation data and model outcomes. These results suggested that for the specific flight scenario, vertical path control performance may be most sensitive for revealing differences in cognitive processing due to modes of cockpit automation.

Conclusion

The computational cognitive models of pilot behavior in using the various forms of cockpit automation were demonstrated to be a viable tool for predicting pilot workload and flight performance under high workload flight conditions. The new cognitive modeling approach may support the development of a general models of pilot cognition, which may facilitate future automated cockpit design.

Acknowledgement

This research was supported by NASA Ames Research Center under Grant No. NNH06ZNH001. Mike Feary was the technical monitor. The opinions and conclusions expressed in this paper are those of the authors and do not necessarily reflect the views of NASA.

References

Billings, C. E. (1997). *Aviation automation: The search for a human-centered approach*. Mahwah, NJ: Lawrence Erlbaum Assoc.

Endsley, M. R. and Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, 42(3), 462-492.

Endsley, M. R. and Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors*, 37, 381–394.

Endsley, M. R. (1993). A survey of situation awareness requirements in air-to-air combat fighters. The International Journal of Aviation Psychology, 3(2), 157-168.

Kieras, D. E. (1997). A guide to GOMS model usability evaluation using NGOMSL. In M. Helander, T. Landauer & P. Prabhu. (Eds.), *Handbook of Human-Computer Interaction*, 2nd Ed. completely revision ed., 733-766.

Miller, G.A. (1956). *The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information*, V.63, N.2, pp.81-97.

Parasuraman, R. , Bahri , T. , Deaton, J. , Morri son, J. & Barnes, M. (1992). *Theory and design of adaptive automation in aviation systems* (Progress Rep. No. NAWCADWAR-92033-60). Warmister, PA: Naval Air Warfare Center, Aircraft Division.

Parasuraman, R. , Sheridan, T. B. and Wickens, C. D. (2000). A model of types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics*, 30, 286–297.

Wiener, E. L. & Curry, R. E. (1980). Flight-deck automation: Promises and problems. *Ergonomics*, 23, 95–1011.