EFFECT OF AIR TRAFFIC CONTROLLER TASKLOAD AND TEMPORAL AWARENESS ON TASK PRIORITIZATION

Esa M. Rantanen and Brian R. Levinthal
University of Illinois at Urbana-Champaign
Savoy, IL 61874

This paper describes an experiment that was conducted to provide an empirical foundation for estimation of parameters for air traffic controller performance modeling efforts presently pursued within the NASA DAG-TM CE-6 model development. The focus of the work was the task prioritization scheme used in these models. A total of 11 retired FAA controllers and supervisors assigned to the FAA Technical Center volunteered to participate in the experiment. A part-task experimental simulation that presented the participating controllers with several simultaneous tasks in four quadrants, or panes, on a single display was used. Only one pane and typically one task could be viewed at a time. This allowed for measurement of controllers' attention to each task. All events unfolding in the experimental scenarios and controllers' actions were recorded and timed as well. From these data, several dependent variables were derived, focusing on the temporal aspects of controllers' performance and their prioritization of simultaneously available tasks. The results indicate that taskload was manipulated successfully and resulted in measurable differences between experimental conditions in both taskload and performance, the latter evinced by the time elapsed in a window of opportunity for a given task before action was taken on it as well as time remaining in the window of opportunity when action was completed. However, it appears that either the controllers were not aware of these temporal features of their tasks or that other factors dominated their prioritization decisions. Task prioritization may hence be driven by task characteristics that are categorical rather than continuous and quantifiable.

Introduction

The intricacies of control of complex and dynamic systems are particularly well illustrated in the nation's air traffic control (ATC) system. The importance of an up-to-date mental model of the traffic situation to the controller is self-evident, as are the temporal demands of the controllers' task. Anticipatory behavior of air traffic controllers, however, is not overt: Anticipation is not an end in itself, it is seldom expressed in verbal communications, and may not result in directly observable behavior (Boudes & Cellier, 2000). Yet, accurate anticipation lies in the core of successful control of air traffic and the performance of a controller by allowing early detection of conflicts (i.e., two aircraft coming closer to each other than a minimum separation required) and formulation of conflict-free traffic flows. A controller who fails to anticipate the development of traffic situation has already 'lost the picture' and is forced from proactive into a reactive mode of behavior, rapidly increasing his or her stress, workload, and propensity for unrecoverable errors.

The temporal dimensions in controllers' tasks are also likely to gain in significance with the introduction of automation applications in ATC. One large-scale effort to increase the National Airspace System (NAS) capacity is the NASA Distributed Air/Ground (DAG) Traffic Management (TM) concept of distributed decision-making. The goal of DAG CE-6 (Concept Element) is to integrate the controller DST with data link to minimize lags/delays while providing controllers with as much flexibility in options as they have today. However, substantial qualitative differences in the working methods and practices of the controllers are to be expected (c.f., Hopkin, 1995; Wickens et al., 1997; Wickens et al., 1998). These differences may in turn have important impact on the controllers' performance and workload, potentially quantifiable by the temporal characteristics of their tasks and the way they are carried out.

In addition to the importance of temporal performance of air traffic controllers, time may offer a useful domain for research of a multitude of human factors aspects. All scientific research of mental models and subsequent engineering applications are dependent on methods of measurement (Chapanis, 1959). Apart from the relevance of time to anticipatory behavior in control of dynamic systems, it offers attractive methods for the measurement of covert mental models. Time has a long history as a means to investigate cognitive processes, manifested by extensive reaction time research. Timing data (e.g., response times) are relatively easy to obtain under both experimental and naturalistic conditions, and time is a variable that is common to the human, the task, and the environment. Time offers thus a common unit of measurement of human performance in the context of the task, and can be used to infer the goodness of the temporal dimension of the operator's mental model of the task or system being controlled. Grosjean and Terrier (1999) defined temporal awareness as a "representation of the situation including the recent past and the near future," (p. 1443) echoing definitions of mental models (e.g., Rouse & Morris, 1986) and situation awareness (e.g., Endsley, 1995). In an experiment mimicking control of three simultaneous processes (simulated production lines) Grosjean and Terrier (1999) discovered that subjects who had developed good temporal awareness made fewer errors, prioritized their work more effectively, and managed their rest periods better than those with poorer temporal awareness. Temporal awareness was thus found to be a good predictor of performance.

Task network models use human/system task sequence as the primary organizing structure and hence appear as particularly suitable approach to modeling air traffic controllers' jobs, which consist of many tasks with varying degrees of dependency. As all tasks and subtasks unfold in time, it may be hypothesized that their successful management is primarily a temporal task and the controller's performance is predominantly determined by his or her time management skills and the goodness of his or her temporal awareness of the situation. Time is hence an attractive variable for investigating the interactions of ATC task load and controller performance as well as a congenital parameter in task network models. The purpose of this research was to provide an empirical foundation for estimation of parameters for air traffic controller performance modeling.

Method

Participants

Participants for this study were recruited among retired FAA controllers and supervisors assigned to the FAA Technical Center's Human Factors Research and Development Laboratory (HFRDL) at Atlantic City International airport, NJ. A total of 11 volunteers participated in the experiment. All participants were male, with a mean age of 55.64 years (SD = 9.1), ranging from 38 to 66 years. All were also very experiences in a variety of ATC facility types with a mean experience as a controller of 23.45 years (SD = 6.67), ranging from 11 to 33 years.

Apparatus

The experimental apparatus was a custom-built ATC simulator. The simulation program was written in C++ and ran on two laptop computers with 14-inch TFT displays and 1024 x 768 –pixel resolution. The simulator mimicked the display system replacement (DSR), including data link (DL) capability, allowing for accurate timing of participant interactions with the DL interface. A regular mouse was provided for moving between tasks (as described below) and control inputs.

Experimental Task

The experimental task mimicked the job of air traffic controllers. The participants viewed air traffic scenarios on four separate quadrants, or panes, on a single computer display. The scenarios could be viewed only one at a time by moving a cursor to the desired pane. This task balanced the requirements of realism and experimental control and it allowed for accurate measurement of times of the different events unfolding in the experimental scenarios as well as timing of the participants' actions in response to them. Six subtasks modeled in the NASA CE-6 modeling effort were selected for the experiment: (1) receive handoff, (2) initiate handoff, (3) transfer communications, (4) respond to DL request to change altitude, (5) perform conflict resolution, and (6) perform metering.

Design

Independent variables. The primary independent variable was taskload, which was manipulated through several other variables over which the experimenter has complete control. It should be noted, however, that control over these variables was constrained by the participants' actions after the onset of the experiment, that is, the eventual sequence and timing of the tasks depended on individual participants' different time management skills and strategies as well as other individual performance differences. Time required (TR) was manipulated primarily by differential difficulty of conflict situations, based on findings of Rantanen and Nunes (in press). Pilot testing revealed a mean time required for participants to use the datalink system's flyout menus to communicate altitude, speed, or heading clearances to pilots, respond to downlink requests, and initiate and receive handoffs. Time available (TA) consisted of the individual windows of opportunity (WO) for each task encountered per trial. In each trial, certain windows of opportunity overlapped reliably, regardless of individual difference in performance, as a result of the discrete trial onset times. For example, the WO for receiving flights from handoff would not vary across participants because it was related directly to trial onset time and the initial speed of the flights, while the extent of the WO for conflict resolution, DL responses, or resolving metering violations would be subject to individual differences. The ratio of time required and time available was the basis of the definition and computation of nominal taskload. By nominal we mean that it was calculated a priori, at the outset of the scenarios. A total of 31 scenarios were created and used to form a total of 20 experimental scenarios: 8 high taskload conditions, 8 low taskload conditions, and 4 transitions.

Design. The basic design of the experiment was a 3 (taskload, Low, Transition, High) x 2 (order Lo-Tr-Hi, Hi-Tr-Lo) x 2 (replicates) factorial design. In the subsequent analyses, however, only low and high taskload conditions were considered, the transitional scenarios split between the two conditions. Four scenario files, one file per quadrant (pane) on the display, started the experimental blocks. At the end of each scenario, a new scenario files filled the pane. An experimental block was comprised of 4 (panes) x 5 scenarios, which followed each other in a seamless sequence. Three levels of taskload were included in each block: two scenarios per pane of high taskload, two scenarios per pane of low taskload, and one transition scenario per pane. The order the scenarios were presented was balanced so that each participant encountered a block that started with low taskload (first two scenarios per pane) and ended with high taskload (the last two scenarios per pane) as well as a block in which the scenarios with different taskload were presented in an opposite order (starting high and ending low).

Dependent variables. To derive the objective performance metrics, a number of actions were timed and recorded for each task. These timed actions were used to derive a number of dependent variables for the purposes of this research. The elapsed time from opening of WO at the time the task was performed (time to first action, or TFA) was calculated by subtracting the time the WO opened for the task from the first action on the task (e.g., mouse click on a flyout menu). Note that this value may also be negative if the task was performed before the WO opened. The time remaining in the WO after completion of the task (TRm) was calculated by subtracting the time of the last action on the task from the time the WO for that task closed. It was hypothesized that good performance would be manifested by prompt actions in tasks (small TFA values) and ample time remaining after completing the task (large TRm values).

Results

The experimental simulation program recorded all events and actions taken by the participants into a text file in a form of a time line. The data were then processed by another program for reduction. This program read the timeline and reorganized the data into an output file so that tasks were entered in rows every time the participant did something about them (including looking at a task, or 'dwelling' in it), plus other measures pertinent to that particular task.

We wanted to determine the actual taskload as influenced by the participants' control actions and strategies, as we anticipated the actual taskload to be different from the nominal one determined from the outset of the experiment. An index of taskload (actual taskload, TL_A) was provided by the following formula:

$$TL_A = \frac{n(TR_{avg})}{TE} \qquad (1)$$

where n is the total number of tasks present in an epoch and TR_{avg} the average time required to perform these tasks. The TE is the duration of the epoch, in this case 300 s (5 min).

It is acknowledged that many tasks had zero time required to perform them, for example, acceptance and initiation of handoffs and transfer of communication only required a single mouse-click. Furthermore, it is clear that physically performing the task, by keyboard entries or clicking through menus with a mouse, only constitutes a small fraction of the total time required to perform the task, that is, the overt actions do not reveal planning and decision-making processes, which almost certainly require most of the controller's time. Nevertheless, multiplication of the time required by the number of tasks compensate to some degree the very short (i.e., 0) performance times in an epoch, and indeed this index showed clear differences between the different taskload conditions (Figure 1). The differences between taskload were significant (two-sample t-test, $p < .05$) for all but the transition epoch.

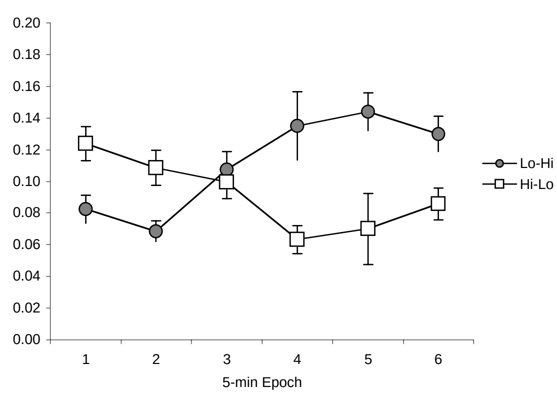

Figure 3.1. Mean actual taskload index values (by eq. 1) by taskload condition and 5-minute epochs; the switch from low to high and high to low taskload condition about 10 minutes into the block is apparent.

Task prioritization was analyzed by observing which tasks were performed before others when all were 'available' simultaneously, that is, the WO for performing the tasks were open at a same time. Specifically, the probabilities a given task was chosen to be performed first among a number of simultaneously available tasks were derived by the following method:

(1) Divide the experimental block in time 1-minute epochs. This epoch duration was somewhat arbitrary, but its minimum was determined by the necessity to have more than one task 'available' within it and its maximum by the notion of simultaneity. The average number of tasks in the 1-minute epochs was 6.45, with a range from 2 to 11.

(2) For each epoch and tasks in the epoch, the actions taken by the controller were recorded (essentially, the first action a controller took on a task).

(3) Based on the first-action times, tasks within an epoch were ranked (1st, 2nd, etc.)

(4) These data were then sorted by task and each task pair was analyzed separately, counting the times a task in a pair was acted on before the other task(s) in the pair within an epoch.

(5) These counts were then summed across participants and experimental blocks, and the proportion of times one task in a pair was acted on before the other was calculated.

(6) The above procedure was repeated for another set of 1-minute epochs, offset by 30 s from the above, to maximize the number of task pairings. This could be seen as a resampling technique, and the combined results improve the reliability and accuracy of the probability estimates.

The results of this analysis are depicted in Table 1. Taskload condition appears to have had only minimal impact on prioritization between tasks in pairs. However, before interpreting these results particular limitation of this analysis must be observed: this method considered the tasks separately, that is, whether a particular task was performed first or not within the 1-minute epoch, and assigned the task into a group based on the given variable value (TFA or TRm) for that task only. In reality, however, the tasks were not independent but considered by participants relative to each other. To determine whether TFA of TRm of each task in a pair was a factor in the participant's choice of task to be performed first, other methods of analysis must be employed.

Table 1.
Probabilities (proportions) a given task was performed before another task when both were available (i.e., their WOs were 'open') simultaneously. Key to the task acronyms: CR = Conflict Resolution, DL = Downlink request (climb/descent), FR = Frequency Change, IH = Initiate Handoff, MV = Metering Violation, RH = Receive Handoff.

Taskload	Proportion	
Task Pair	CR/DL	DL/CR
Low	0.592	0.408
High	0.471	0.529
	CR/MV	MV/CR
Low	1.000	0.000
High	0.714	0.286
	CR/RH	RH/CR
Low	0.306	0.694
High	0.331	0.669
	DL/FR	FR/DL
High	1.000	0.000
	DL/RH	RH/DL
Low	0.275	0.725
High	0.334	0.666
	FR/IH	IH/FR
Low	0.400	0.600
High	0.125	0.875
	RH/FR	FR/RH
Low	1.000	0.000
High	0.833	0.167
	RH/MV	MV/RH
Low	0.875	0.125
High	0.839	0.161
	DL/MV	MV/DL
Low	0.750	0.250
High	0.500	0.500
	MV/RH	RH/MV
Low	0.188	0.813
High	0.250	0.750

As was discussed above, it was of interest to examine whether the participants' temporal awareness, that is, awareness of the TFA or TRm of each task at hand (i.e., tasks with simultaneously open WOs) played a role in their decisions to prioritize one task over another. To do this, we considered tasks in pairs, as was done in previous task prioritization analyses. In this analysis, however, we contrasted the TFA and TRm values of each task in a pair according to the eventual priority given to a task. Specifically, we hypothesized that it the controller was aware of the time elapsed in the WO (TFA), he or she might perform the task with longer TFA first and the task with a more recently opened WO second. Hence, the hypothesis may be operationalized as

$$TFA(1) > TFA(2) \Leftrightarrow TFA(1) - TFA(2) > 0 \qquad \text{(H1)}$$

Another hypothesis was that if the participants were aware of the impending *closing* of a WO, that might have contributed to a sense of urgency and a task with a shorter TRm value would be performed first. Specifically,

$$TRm(1) < TRm(2) \Leftrightarrow TRm(1) - TRm(2) < 0 \qquad \text{(H2)}$$

We tested these hypotheses for task pairs of conflict resolution (CR) and datalink request (DL) as well as CR and receiving handoff (RH) by calculating the proportions of positive and negative outcomes of the above hypotheses (see Table 2 below).

Table 3.15.
Proportions of the positive and negative outcomes from the hypotheses H1 an H2 as stated above. However, the results are not only mixed (i.e., split between positive and negative) but actually opposite to the hypotheses. In both task pairs, TFA had a higher proportion of negative values than positive, and the TRm a higher proportion of positive values than negative

Task Pair	Variable	N Pos	N Neg	% Pos	% Neg.
CR/DL	TFA(1)–TFA(2)	81	166	32.8	67.2
CR/DL	TRm(1)–TRm(2)	125	77	61.9	38.1
CR/RH	TFA(1)–TFA(2)	98	352	21.8	78.2
CR/RH	TRm(1)–TRm(2)	186	81	69.7	30.3

The results did not confirm the hypotheses. As a matter of fact, they were opposite to what was hypothesized in that the participants seemed to perform tasks with more recently opened WOs before tasks that have been available longer, and tasks with more time before closing of their WOs before more urgent tasks by the same measure.

We also performed an ANOVA to see whether the above hypotheses differed between task priority and taskload, with taskload level and task priority as factors, plus their interaction in the model. For CR/DL and TFA, neither of the main effects was significant, that is, the difference in TFA values between the tasks in the pair did not differ significantly between taskload conditions or between task priorities. The interaction between these factors was significant, however, $F(1, 243) = 14.87$, $p < .001$. For TRm, task priority was significant, $F(1, 198) = 638.55$, $p < .05$, but no other factors or interactions. Analysis of the CR/RH task pair yielded similar results; for TFA, only the interaction between task priority and taskload was significant, $F(1, 446) = 4.84$, $p < .05$, and for TRm, there were no significant results.

Discussion

The combined results from this study suggest that task prioritization may be driven by task characteristics that are categorical rather than continuous and quantifiable. Support for this conclusion is provided by the very different trends in TFA for the three different tasks analyzed, conflict resolution, receiving handoffs, and responding to downlinked requests. Of these, conflict resolution was clearly the most difficult task, as well as the most important. The difficulty of detecting conflicts as well as the time required to construct and implement resolutions to them probably made this task more vulnerable to influences of workload and time pressure than simpler tasks. It must also be remembered that accepting handoffs is, in addition to being a quick and easy task to perform, a prerequisite to subsequent control of the flight (e.g., to implement conflict resolution) and hence the average prioritization between conflict resolution and receiving handoffs is inherently biased towards the latter.

Another aspect worth considering is the nature of the analyses and differences between experimental simulations and realistic situation in operational ATC. Statistics (i.e., minimization of probabilities of both Type I and II errors) is dependent on sufficiently large number of observations, which necessitates aggregation of observations across individual participants and experimental blocks. Yet, even in relatively constrained task environments such as our experiment these observations exhibit substantial variability. For example, aggregation of conflict resolution tasks and receiving handoffs as was done here did not consider the often unique characteristics of each of these instances. Parsing the data according to such characteristics, however, would severely limit the number of observations available for analysis and undermine the reliability of the results. This is a classical ‘Catch-22’ situation for which the only remedy is to collect much more data over extended periods of time.

Finally, large differences in performance of individual participants should not be overlooked. These differences were statistically highly significant in almost all analyses we performed and bespeak of inherent variability in working techniques, strategies, and performance of individual controllers working on the same tasks.

Acknowledgments

This work was supported under National Aeronautics and Space Administration and Micro Analysis and Design Contract NASA-MAD-7400.005.01. Parimal Kopardekar from NASA and Ken Leiden from MAAD served as technical monitors. Views expressed herein are those of the authors and do not necessarily represent official NASA, FAA, or MAAD positions. We express our appreciation to Ben Willems and Pamela Della-Rocco at the FAA William J. Hughes Technical Center for their invaluable assistance in securing participants and venue for running the experiment and the individual participants for volunteering their time and expertise for this study. Special thanks are due to Sharon Yeakel for programming the experimental simulator and data processing tools.

References

Boudes, N., & Cellier, J.-M. (2000). Accuracy of estimations made by air traffic controllers. *International Journal of Aviation Psychology, 10*(2), 207-225.

Chapanis, A. (1959). *Research techniques in human factors*. Baltimore, MD: Johns Hopkins University Press.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*(1), 32-64.

Grosjean, V., & Terrier, P. (1999). Temporal awareness: Pivotal in performance. *Ergonomics, 42*(11), 1443-1456.

Hopkin, V. D. (1995). *Human factors in air traffic control*. Bristol, PA: Taylor & Francis.

Rantanen, E. M., & Nunes, A. (In press). Hierarchical conflict detection in air traffic control. *International Journal of Aviation Psychology*.

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin, 100*(3),349-363.

Wickens, C. D., Mavor, M., & McGee, J. (1997). Flight to the future: *Human factors of air traffic control.* Washington, DC: National Academy of Science.

Wickens, C. D., Mavor, M., Parasuraman, R., & McGee, J. (1998). *The future of air traffic control*. Washington, DC: National Academy of Science.