Federal Aviation Administration

DOT/FAA/AM-07/6
Office of Aerospace Medicine
Washington, DC 20591

Predicting Subjective Workload Ratings: A Comparison and Synthesis of Operational and Theoretical Models

Jerry Crutchfield
Federal Aviation Administration
Civil Aerospace Medical Institute
Oklahoma City, OK 73125

Craig Rosenberg
The Boeing Company
Tukwila, WA 98108

March 2007

Final Report

NOTICE

Technical Report Documentation Page

1. Report No. DOT/FAA/AM-07/6	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Predicting Subjective Workload Ratings: A Comparison and Synthesis of Operational and Theoretical Models		5. Report Date March 2007 6. Performing Organization Code
7. Author(s) Crutchfield J,[1] Rosenberg C[2]		8. Performing Organization Report No.
9. Performing Organization Name and Address [1]FAA Civil Aerospace Medical Institute P.O. Box 25082 Oklahoma City, OK 73125	[2]The Boeing Company 9725 East Marginal Way S. Tukwila, WA 98108	10. Work Unit No. (TRAIS) 11. Contract or Grant No.
12. Sponsoring Agency name and Address Office of Aerospace Medicine Federal Aviation Administration 800 Independence Ave., S.W. Washington, DC 20591		13. Type of Report and Period Covered 14. Sponsoring Agency Code
15. Supplemental Notes Work was accomplished under approved task AM-BHRR-522		

16. Abstract

Output from a computer simulation of two air traffic control (ATC) scenarios was fit to workload ratings that ATC subject-matter experts provided while observing each scenario in real time. Simulation output enabled regression analyses that tested the assumptions of a variety of workload prediction models. These included both operational models that use observable situational and behavioral variables (e.g., number of aircraft and communications by type) and theoretical models that use queuing and cognitive architecture variables (e.g., activities performed, amount of busy time, and sensory and cognitive resource usage). Results suggested the models that included number of activities performed weighted by priority accounted for the highest amount of variance in subjective workload ratings.

17. Key Words Air Traffic Control, Modeling, Workload		18. Distribution Statement Document is available to the public through the Defense Technical Information Center, Ft. Belvior, VA 22060; and the National Technical Information Service, Springfield, VA 22161	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 14	22. Price

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Predicting Subjective Workload Ratings: A Comparison and Synthesis of Operational and Theoretical Models

INTRODUCTION

Workload is a term often used to refer to the amount of work or effort required to perform an activity over a given time period [1, 2]. Although certain variables have been shown to moderate the exact relationship between performance and workload for given situations [3, 4, 5], high levels of workload generally tend to be associated with increases in operator error and decreases in overall performance [6, 7]. These findings have led to an enduring interest in workload research. This is particularly true in the domain of air traffic control (ATC) where safety and operational efficiency often hinge upon performance of highly complex tasks. Researchers recognize that high workload inherent in cognitively complex ATC tasks may lead these tasks to be vulnerable to performance decrements.

Unfortunately, research findings of over the last three decades have also revealed the workload construct to be a challenging one to characterize [2, 8, 9]. Workload seems to result from several different contributing factors. These factors include operator individual differences, fatigue, expertise, environment, time pressure, number of tasks, task modality, and task difficulty.

Despite obstacles, advancement in workload research has enabled the development of mathematical models used to support the analysis of operator workload. Many of these models have been developed for use in the ATC domain. Computer models provide predictions of workload that approximate those that would otherwise have to be gained from the use of system prototypes and subject-matter expert (SME) interactions. Through the use of valid workload models, analysts can predict how effective a system will be and where failures or reductions in performance are likely to occur.

There are a large number of variables that modelers can choose to make workload predictions. Consequently, many different types of models have been developed to predict workload and workload-related concepts such as dynamic density [10, 11, 12, 13, 14, 15, 16]. Workload models vary in the domains to which they have been applied and in the amount and method of validation they have received. These models often differ in their approaches as well. Some approaches rely on objective variables observable in the environment or situation, while other models rely on variables derived from theoretical constructs or processes. Even though these models were created to predict the same general theoretical concept, their approaches rely on entirely distinct sets of predictor variables.

One type of model applied to workload prediction is the queuing model. Queuing theorists model complex task performance by representing the process in terms of servers and clients [15, 16]. In Schmidt's model, the ATC specialist was represented as a server, and the ATC tasks to be completed were represented as the customers of the server. In this type of theoretical model, number of activities, the difficulty associated with performing the activities, and the relative priority of activities are all used to predict workload [15, 16].

Researchers in the ATC domain have used the occurrence of certain quantifiable situational factors and observable air traffic controller behaviors as variables to predict workload as well [7, 17, 18, 19]. Variables such as these are often selected for analysis as they provide objective measures of workload that can be accessed without interfering with a controller's work. The discussion herein shall refer to models that use these variables as operational models due to the association they have with a specific domain.

The identification of variables for use in operational models requires an understanding of the domain under consideration. In the ATC domain, for example, controllers typically monitor a radar scope showing the positions of aircraft and deliver control commands to the aircraft verbally over a radio channel. Control commands, or clearances, include changes to aircraft altitudes, headings, and speeds. Clearances are given to direct aircraft to particular waypoints, increase or assure a safe distance between all aircraft, or slow and descend an aircraft so as to land on a runway. Furthermore, different types of controllers control aircraft at different points in their journey. In our example, an Air Route Traffic Control Center (or simply Center) controller may hand off an aircraft to a Terminal Radar Approach Control (TRACON) controller who slows and descends the aircraft, and hands the aircraft off to a Tower controller for landing. From these ATC activities, researchers have identified variables such as number of aircraft under control, altitude changes, and handoffs performed as means of estimating workload [7, 17, 18, 19].

In Manning et al. [17], a wide range of operational variables was used in a regression analysis to predict workload. Twenty-three operational variables were analyzed along with variables for number of communications and communication time. The operational variable values were derived from video and audio data recordings of air traffic control. Manning et al. first used a Principle Component Analysis on the values and reduced the variables into five sets. These sets were then used in multiple regression analyses to predict controller workload ratings. In this way the authors were able to identify a model that could predict 72% of the variance in workload.

In addition to the number of variables used, the Manning et al. [17] study was also interesting because of the way the authors collected the workload values that the operational variables were used to predict. In this study, workload was represented by subjective workload ratings. Although criticized due to findings that show dissociation between subjective workload ratings and performance [20], subjective ratings are among the most popular workload measurement techniques. The subjective technique has a great deal of face validity and theoretically allows the researcher to tap personal perceptions of workload that result from the interactions of both observable and unobservable workload factors [21]. Subjective workload ratings are usually collected from operators as they perform their tasks or shortly afterward. Operators report the amount of workload personally experienced. However, in the Manning et al. study, controller SMEs instead observed recordings of air traffic control and indicated the amount of workload they believed the controller controlling the traffic was experiencing.

Although the Manning et al. study showed that operational variables provided promising results for predicting controller workload in a known ATC system, the ability of operational models to predict workload for a system that currently does not yet exist remains to be determined. Take, for example, the application of the operational modeling approach to the prediction of workload associated with an ATC operational concept that includes the use of new technology (e.g. datalink) to deliver aircraft clearances. The operational approach would seem to assume that a message delivered by voice would result in the same amount of workload as a message delivered by a new technology. It may be the case, however, that the weighting of workload predictive variables is different for a system that uses a different mode of communication, supports the controller with automated decision aids, or relies on a different set of procedures.

Cognitive models are a type of theoretical model that may be useful for the prediction of workload with proposed new systems. Cognitive models allow for a representation of performance at the sensory, cognitive, and motor resource level. Although this level of representation requires an additional investment in time and effort, it provides a theoretical way to account for the unobservable aspects of workload that operational models do not. By modeling the cognitive aspects of workload, cognitive modeling can provide a way to account for the differences between any alternate systems that are modeled.

Although there are many types of cognitive models, most cognitive models applied to workload research are based on Wicken's Multiple-Resource Theory [22]. Multiple-Resource Theory posits that there are separate and independent pools of resources for separate types of processing. There are different sensory resources (audio, visual, etc.) and response resources (manual, vocal, etc.) for example. If two tasks require simultaneous use of the same resource, interference will occur and task performance will suffer. As the concept of workload assumes that human performance is limited by finite resources, Multiple-Resource models rely on sensory, cognitive, and motor resource usage and interference to predict workload.

Models such as those based on Multiple-Resource Theory were developed to describe cognitive processes at a minute level. Before these models could be applied to the prediction of workload, a method of extrapolating the models to represent the processing involved in complicated real world tasks was needed. Task analysis is a means of describing all the steps that must be carried out to perform a function and the sequence with which those steps must be taken. In task analysis, activities such as knowledge elicitation and role-playing exercises are used to identify functions and then break those functions down into activities. Many types of task analyses produce task networks. In task networks, activities are further broken down into tasks, and the information requirements for each task are defined. Task analysis provides a means to extrapolate cognitive models for efficient application to complex, real-world situations.

Aldrich and Szabo [10] developed a process whereby they mapped uses of theoretical cognitive, sensory, and motor resources onto a task network. Their model became known as the VACP model because separate task networks were created for Visual, Auditory, Cognitive, and Psychomotor resource usage. Tasks along these networks were also rated for difficulty. Workload predictions were calculated for any given moment by adding up the difficulty ratings for all tasks being performed at that moment. The VACP model was capable of providing additional information regarding which resources were being utilized when and with what frequency.

Another early workload prediction model utilizing Multiple-Resource Theory was Parks and Boucek's Time-Line Analysis and Prediction (TLAP) model [13], developed to predict pilot workload. Similar to the VACP approach, the approach created by Parks and Boucek used separate task networks for each different resource type. Networks were created for cognitive, visual, auditory, manual hands, and manual feet resources. By enhancing the task analysis with a cognitive architecture, Parks and Boucek were able to provide a theory-based prediction of when tasks could be performed in parallel. The aggregate ratio of overall operator busy time to time available that emerged from these theoretical task networks was used to predict level of workload.

North and Riley [12] extended the above approaches by incorporating an interference matrix into their Workload Index (W/INDEX) model. The interference matrix indicated the degree to which tasks interfere with each other at the resource level. Values from 0 to 1.0 were estimated to represent how much different parallel resource usages would interfere with performance. Workload predictions were found to be similar to the VACP approach except that the value for relative task interference was also a factor in the calculations.

Without validation it would be impossible to know whether models such as W/INDEX perform better at workload prediction than models such as VACP or TLAP. Although it is important that any model type be validated, validation is particularly important for cognitive models because they are based on cognitive theories that may be controversial or otherwise difficult to confirm.

Sarno and Wickens [1] tested and compared the assumptions of Parks and Boucek's [13] TLAP, Aldrich and Szabo's [10] VACP, and North and Riley's [12] W/INDEX. These models were tested against two types of performance data: data recorded from participants as they attempted a combination of derived tracking, monitoring, and decision making tasks, and data collected from participants as they took part in a TASKILLAN helicopter simulation. All models tested accounted for 56 to 84% of the performance variation in the derived tasks but accounted for only 12 to 42% of the variance in TASKILLAN performance. By removing and combining model features, Sarno and Wickens were able to narrow down which model variables were associated with improvements in prediction. Results showed that prediction was best for models that represented the use of multiple resources. The results also showed that workload prediction was not improved when the degree of resource usage interference was included in the analysis.

Although Sarno and Wicken's study was useful for a comparison among subtypes of cognitive models, for designers and researchers to answer the broader question of whether workload can be better characterized by queuing, operational, or cognitive model variables, requires that the model types be tested together, against the same data. All three types of models have been employed with some degree of success to the analysis of real-world problems. However, even when differing model types have been applied to the same domain, they were not validated against the same data set.

The current paper used output from Boeing Air Traffic Management's Regional Traffic Model (RTM) and its Human Agent Module (HAM) to test and compare the assumptions of both the operational and theoretical models. The RTM output includes variables such as number of aircraft under control and number of communications by type. Furthermore, the cognitive architecture found within the HAM models the use of cognitive, sensory, and motor resources and records when tasks requesting those resources are in conflict.

Two air traffic scenarios were run using the RTM, and the output was used to derive queuing, operational, and cognitive model variables. These variables were used in regression analysis to predict subjective workload ratings. The workload ratings were provided by ATC SME's who observed the two scenarios as they were being run by the model in real time.

METHOD

Participants

Two ATC subject matter experts were compensated for their participation in this study. Both of these participants were former air traffic controllers employed as training consultants. One participant's specialization was in TRACON environments and the other participant's specialization was in Center environments.

Materials

The Regional Traffic Model. Boeing Air Traffic Management's RTM is a fast-time, discrete event-modeling tool developed to allow engineers and decision makers to compare and assess the impact of theoretical new technologies and procedures on air traffic management performance. Through the use of models like the RTM, analysts can predict to some degree how effective a system will be and where failures or enhancements in performance are likely to occur. Analysts can make changes to the system as it is represented in the model and collect data in a relatively quick and cost-efficient fashion.

The RTM is made up of a number of modules that represent the generic functionalities inherent in the air traffic management system: Aircraft, Airspace, Communication, Surveillance, Traffic Generation, and Human Agent Modules among others. In the Traffic Generation

Module, stochastic traffic generation, for example: "can be configured in terms of inter-arrival times to specify various demand scenarios as well as in terms of traffic type and wake vortex class composition. This provides the ability to represent aircraft arrivals into Center airspace at appropriate miles-in-trail" [23]. The HAM was developed as part of the RTM to represent the behavior and performance of human air traffic controllers and pilots. It was also developed to enable the prediction of human operator workload. The HAM is a part task network model and part cognitive architecture model. Whereas there are modules in the RTM that produce data regarding traffic generation, aircraft performance, aircraft spacing, surveillance, and communication channel performance, the HAM produces data regarding the time of occurrence, duration, and frequency of controller activities and tasks, and the usage of motor, sensory, and cognitive resources in the completion of those tasks. These data are used to derive human task performance delay, error rate, and communication channel congestion metrics.

The controller HAM controls air traffic in a way that is representative of how traffic is controlled today or in a way that we expect it to be controlled under alternate operational concepts. It accepts control of an aircraft and guides it along its course by issuing altitude, heading, or speed clearances through the communication channels. The controller HAM also uses these clearances to maintain safe distances between the aircraft and provide collision avoidance maneuvers. In today's air traffic environment, controllers are differentiated by the type of airspace they control. TRACON controllers control the airspace immediately around airports and deal with the arrival and departure phases of flight. Center controllers typically deal with aircraft undergoing the en route phase of flight often associated with higher altitudes. The HAM is capable of representing both types of controllers.

The controller HAM accomplishes ATC as described in Figure 1. First, the HAM receives traffic-related events from other RTM modules. Events include notification that an aircraft has passed a waypoint or deviated from assigned altitude, among others. The processing of these events may be delayed, depending upon the availability of the sensory resources represented within the HAM. Once the existence of an event is known, it must be recognized. The HAM recognizes events by associating them with programmed activities and tasks. In the HAM, activities are operational goals (e.g. Accept Handoff, Resolve Conflict). Activities are achieved through the performance of two or more tasks (e.g. Issue Altitude Clearance, Determine if Aircraft is in Conformance). The representative tasks performed in response to the events were obtained from previously performed task analyses [24] and through knowledge elicitation from controller SMEs. Relative priority rankings and difficulty rankings for all of the activities were also elicited, and the priority rankings were used in the model.

When the controller HAM performs tasks associated with traffic events, it calls upon representations of sensory, cognitive, and motor resources. These resources make up the HAM's cognitive architecture. Tasks are theorized to require the use of certain resources before they can be successfully completed. Two tasks that require the use of two different resources can be performed in parallel. However, if a task requires a resource that is currently in use, a resource conflict is logged, and the subsequent task is placed in a model queue until the other task is completed. If the two tasks require the resource simultaneously, the task associated with the higher priority activity will gain access to the resource first. In this way, controller activities can be interrupted by higher priority activities but tasks cannot.

Finally, the performance of the HAM is set through parameters associated with each task. Therefore, not only is the HAM able to represent the way in which a human solves given air traffic control problems but also, through instantiation of these parameters, is able to represent differing amounts of human performance accuracy and delay in the implementation of the solution.

The Total Airport and Airspace Modeler (TAAM). The TAAM tool, from Preston Aviation Solutions, provides a viewer functionality that enables visualization of model results using a perspective similar to ATC radar displays. This tool allows RTM data to be replayed at a rate representative of real time. Aircraft are depicted as radar targets accompanied by data blocks showing aircraft speed and altitude. Sector boundaries and the airway routes on which the aircraft flew are also depicted.

Procedure

The RTM was used to run two 150 minute air traffic scenarios. These scenarios depicted a representation of westbound arrivals from three Chicago Center sectors into Chicago O'Hare's (ORD) TRACON and runway 14L. One of the scenarios modeled a Low traffic-level condition and the other modeled a High traffic-level condition. The RTM output from these runs included a record of human-controller task completions, air-ground communications, and sensory and cognitive resource uses.

The RTM Traffic Generator parameters were populated to provide aircraft that differed in equipage (weight and performance classes). The ratio of aircraft equipage types used was representative of traffic into ORD during a typical day from August 2000. The scenario that depicted the Low traffic condition was populated such that approximately 15 aircraft would land on runway 14L per hour. The scenario that depicted the High

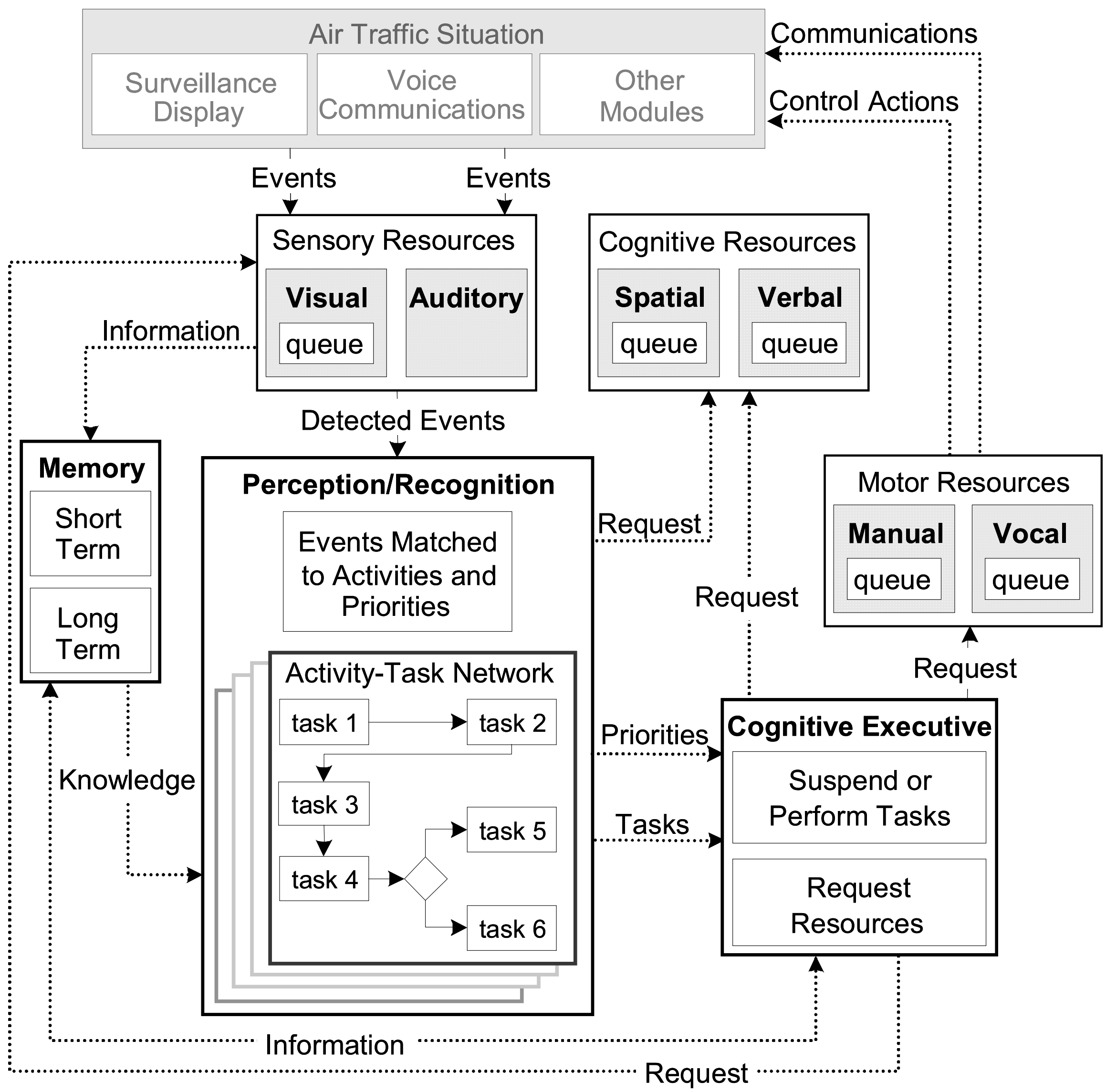

Figure 1. Cognitive Architecture Represented in the HAM.

traffic condition was populated such that approximately 24 aircraft would land per hour.

An illustration of the approximate lateral profile followed by the simulated aircraft can be seen in Figure 2. Aircraft enter the Center sector at the FLINT and SALEM waypoints and travel westward. The Center sector controller merges the two traffic streams at PULMAN before handing the aircraft off to the next controller. Aircraft enter the TRACON just after PIVOT in the Northeast and after BEARZ in the south. The TRACON Final controller takes control of the air traffic from the south just after the northward vector, merges the two traffic streams, and vectors the aircraft on the downwind to ensure spacing at the Final Approach Fix (FAF) before handing the aircraft off to the Tower controller.

The RTM input parameters that represented the behavior and performance of both the humans and the technological systems in these scenarios were chosen and instantiated to model the way traffic is controlled today with today's technology. Air routes used in the model of Center airspace and vectors used in the model of TRACON airspace matched those used in current Chicago operations. Communication system performance matched that of today's analog voice systems.

The output from the two model runs was loaded into the TAAM viewer and replayed in real-time for the participants to observe. The participants each viewed 80 minutes of the output, 40 minutes from the Low traffic-level condition and 40 minutes from the High traffic-level condition. Each time segment observed started with a representative number of aircraft already in its respective airspace. The TAAM depicted display was limited to the Pullman sector for the participant that specialized in Center control and the ORD sector for the TRACON control specialist. Prior to viewing, both participants were briefed as to the nominal flight profiles used in the respective scenarios. It is also worthy of note that, as the RTM produces no audible output, participants viewing the scenarios had to infer communication messages by observing changes to aircraft heading, speed, and altitude visible in the aircraft data blocks.

Workload ratings were elicited from the participants as they observed the scenarios. The workload rating collection procedure was a modification of the Air Traffic Workload Input Technique [21]. The participants were informed that at 4-minute intervals during the scenarios they would be asked to estimate the level of workload they believed someone controlling the current traffic situation would be experiencing. The participants provided their answers, in pencil and paper format, on a scale from 1 to 10 with 1 being extremely low workload and 10 being extremely high workload.

RESULTS

Descriptive Statistics

Several RTM output variables were selected for analysis to predict the workload ratings provided by the participants. These variables were selected based on their theoretical ability to predict workload as suggested in previous studies. The variables are listed on Table 1, in the first column. These variables were derived from scenario output for each 4-minute period that a workload rating was collected. Means for the variables and the workload ratings are shown in columns 2-5.

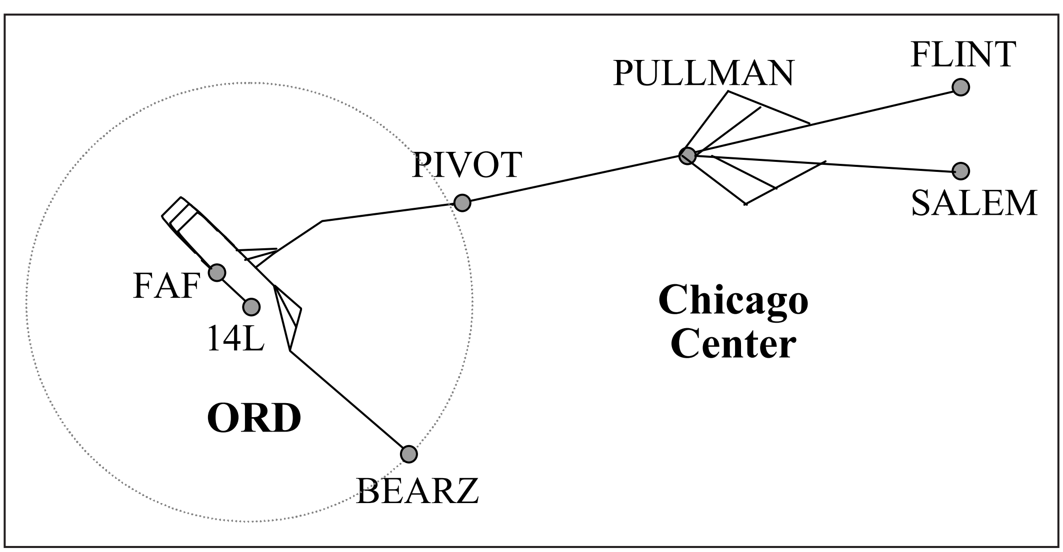

Figure 2. Lateral Flight Paths Modeled Within the RTM.

Table 1. Variable Types, Means From the Scenarios, and Results of the Regression Analyses.

	Means				Model Performance			
	TRACON Traffic		Center Traffic					
Variable per 4 Minute Time Segment	**Low**	**High**	**Low**	**High**	**R**	**R^2**	**F**	**p**
Subjective Workload Ratings (1-10)	3.57	5.25	2.00	2.57				
Operational Variables								
Number of Aircraft	3.00	5.38	2.00	2.71	0.828	0.686	61.161	0.000
Number of Heading Changes	5.14	9.50	0.63	9.00	0.619	0.384	17.425	0.000
Number of Communications	10.29	19.88	4.50	19.00	0.610	0.372	16.579	0.000
Communication Congestion (in seconds)	50.15	92.37	22.32	91.07	0.596	0.355	15.406	0.001
Number of Speed Changes	1.71	3.38	0.63	3.43	0.541	0.292	11.565	0.002
Number of Altitude Changes	1.71	3.50	1.88	3.14	0.390	0.152	5.035	0.033
Number of Handoffs (accepted & initiated)	1.00	1.63	0.63	1.71	0.376	0.141	4.610	0.041
Theoretical Variables - Queuing								
Number of Activities Completed Weighted by Priority	39.86	71.63	16.38	28.14	0.876	0.767	92.346	0.000
Number of Activities Completed Weighted by Difficulty	37.71	66.00	14.13	24.86	0.857	0.735	77.714	0.000
Number of Activities Completed	9.43	17.25	4.38	7.71	0.849	0.721	72.335	0.000
Task Load (time on tasks/240 seconds)	56.79	78.96	38.33	42.50	0.671	0.450	22.880	0.000
Number of Tasks Performed	28.57	54.63	13.25	53.29	0.599	0.358	15.643	0.000
Theoretical Variables - Cognitive								
Resource Usage Conflicts	2.57	7.88	0.00	6.43	0.639	0.408	19.279	0.000
Verbal Cognition Resource Requests	15.00	28.25	7.50	27.00	0.592	0.350	15.106	0.001
Visual Processor Resource Requests (task specific only)	10.29	20.00	5.13	19.71	0.584	0.341	14.502	0.001
Spatial Cognition Resource Requests	2.57	4.75	0.63	4.86	0.521	0.271	10.432	0.003
Top 5 Combination Models								
Activities by Priority & Resource Usage Conflicts					0.889	0.791	50.950	0.000
Activities by Priority & Number of Aircraft					0.889	0.790	50.689	0.000
Activities by Priority & Spatial Cognition Requests					0.887	0.787	49.869	0.000
Activities by Priority & Task Load					0.881	0.775	46.578	0.000
Activities by Difficulty & Number of Aircraft					0.878	0.770	45.298	0.000

All derived variable values showed an increase from the Low traffic condition to the High traffic condition. Number of aircraft controlled was greater for the TRACON controller. This was because the TRACON sector was being fed by more than one Center sector. An increase in runway arrival rate, for the scenarios, was attained by proportionately increasing air traffic frequencies at each of the Center airspace entry points.

Workload levels were rated higher for the TRACON sector than for the Center sector. The Center sector controller for this model did not have to perform some of the common tasks that many real Center controllers would have to perform, including responding to pilot requests and overflights. Neither sector under Low or High traffic conditions was rated as presenting the simulated controller with more than a moderate level of workload. These low ratings may have given rise to a floor effect for some variables.

Model Performance

Each of the variables recorded was used in a regression analysis to predict workload ratings. The results of these analyses are provided in columns 6-9 of Table 1. The table provides both the R and the R^2 value indicating the amount of variance accounted for by each model. The table also provides the F and p values indicating the level of significance the model reached. These results indicate the ability to predict subjective workload for each of the model types as represented by the HAM and the RTM. Successful models identify candidates from among the operational and theoretical variable types that could be used in place of subjective workload ratings when it comes to predicting workload for new ATC systems.

The operational variables analyzed for this study included number of communications, communication channel congestion, number of clearances by type (altitude, heading, and speed changes), number of aircraft being controlled, and handoffs. Unfortunately, neither of the communication variables predicted more than 38% of the variance in subjective workload ratings in this study (as compared to 49% found by Manning et al. [17]). Even the best predictor among the clearance types (number of heading changes) did not predict more than 39% of the variance.

Number of aircraft, however, performed well, predicting almost 69% of the variance. These results suggest that a measure as simple as number of aircraft can be a relatively strong representative of subjective workload by itself. Its usefulness is limited by the fact that the number of aircraft found in a scenario tells us very little about how one system contributes to workload levels versus another.

The theoretical queuing variables tested were task load, number of activities completed, and number of activities completed weighted by either difficulty or priority. These variables represent aggregates of tasks performed to complete activities. The task load model was successful at predicting 45% of the variance in workload ratings. However, three activity level models performed better, accounting for between 72 and 77% of variance. It is interesting to note that the best-predicting model of the three used priority, a relative measure of time criticality, to weight the number of activities. As has previously been suggested in the literature [8], time pressure is an important contributor to the subjective workload experience.

The theoretical cognitive variables analyzed included total number of tasks performed by the HAM, as well as the number of calls to the verbal, spatial, and visual processor resources, and resource usage conflicts for each 4-minute segment. The highest performing variable from this list, resource usage conflicts, predicted roughly 41% of the variance in subjective workload ratings. Resource usage conflicts predicted relatively well, considering that this variable requires the most detail about how tasks are being carried out and relies heavily on cognitive theory. Although the cognitive variable models may not fare well by themselves, they can potentially provide designers with useful information regarding resource usage.

Further regression analyses were conducted by testing pairs of variables together. In this study, there was an insufficient amount of workload ratings to perform any regression procedures using more than two variables at a time. Operational variables, theoretical queuing variables, and theoretical cognitive variable pairs were all tested, except where prohibited by co-linearity. Additionally, as the RTM and HAM output includes both operational and theoretical variables from the same scenario, it was theorized that the operational and theoretical model types could be directly compared. Toward this end, regressions were also performed on pairs that included one variable from both the theoretical and operational variable types.

The regressions identified 17 variable pairs that produced models accounting for more than 75% of the variance in workload ratings. Model performance for the top five predicting models is shown in Table 1. A Bootstrap analysis was applied to the predicted workload values for each of these models. Results of this analysis showed that none of the predicted values for any of the models was significantly different from any of the others. Although comparing the amount of variance accounted for across the various models may provide hints as to trends in model performance, the small number of workload ratings does not allow for statistically reliable comparisons to be made.

All 17 top predicting pairs included either number of activities weighted by priority or number of activities weighted by difficulty as a variable. Number of activities weighted by priority in combination with any one of either taskload, number of aircraft, spatial cognitive resource use, or number of resource conflicts produced the four best prediction models.

The model pairing number of activities weighted by priority and number of resource conflicts produced the highest R^2 value. The coefficients and constant for this model make up the following workload prediction equation: Workload = 1.328 + 0.067 (resource conflicts) + 0.049 (activities weighted by priority). The results of this analysis suggest the model equation for number of activities weighted by priority and number of resource conflicts is the most suitable to represent workload levels in design situations where actual subjective workload ratings cannot be assessed.

DISCUSSION

Results of this study suggest that number of activities completed per 4-minute time period is a good predictor of workload. By itself, this variable predicted 72% of the variance in workload ratings. As derivation of this variable requires only a minimal task analysis, this is potentially good news for designers who lack in-depth knowledge about new task procedures or who lack the time and budget to perform in-depth cognitive analyses. In this study, number of activities was a better workload predictor than the domain specific operational variables such as frequencies of clearances by type, number of handoffs, average number of aircraft under control, and those related to communications.

The predictability of number of activities increased when this variable was weighted either by activity priority or difficulty. Priority is an indicator of the time criticality of an activity. The finding that the priority weighting improved this model tends to corroborate workload theories that have identified time pressure as a major influence on resulting workload [8]. As the relative priority rankings of activities is not likely to change across systems, the "number of activities weighted by priority" model will be insensitive to comparisons of systems that change the amount of workload contributed by activities without changing the number of activities that need to be performed. This limitation would not exist for the "number of activities weighted by difficulty" model, should it be possible to estimate a different set of difficulty weightings for activities performed using the new technology.

The R^2 value of activities weighted by priority was further improved when paired with the variable representing the number of resource conflicts that occurred during the 4-minute time period. Based on the results of the regression analysis alone, the model using activities weighted by priority and number of resource conflicts is the preferred model to use to predict workload. However, taken at face value these results only show a 2% increase in prediction associated with the cognitive component of the equation.

Gaining this extra prediction accuracy required the development of a cognitive architecture and the assignment of cognitive resource usage estimates to tasks in a task network. The cost in budget and schedule needed to perform this cognitive modeling may not seem worth the extra 2% gain. However, there are other important reasons to consider using cognitive modeling to predict workload associated with new systems.

One reason to include cognitive modeling is that a descriptive analysis of resource usage provides guidance to designers regarding factors that are likely to impact the workload of a new system. The model using number of activities weighted by priority can be used to predict when a system is likely to foster a high level of workload, but it is unlikely, by itself, to say much about which elements may be causing the workload increase. Descriptive statistics such as number of uses of the visual processing resource or number of uses of the communication channel can suggest to a designer where the problem areas are likely to occur, should suboptimal workload levels be predicted.

A second reason is that the inclusion of the variable representing number of resource conflicts into the equation with number of activities weighted by priority brings the model a much-needed consideration for behaviors that take place within the activities. A workload model that uses number of activities weighted by priority, assuming the priorities of activities do not change between systems, will not distinguish between systems that require similar numbers of activities. Even workload models that predict and record cognitive resource usage at the task level will not distinguish between two systems that simply shift the resource usage modality without changing the number of tasks being performed. However, measures such as resource usage conflicts provide information as to how the system and procedures integrate with human limitations and therefore increase the sensitivity of the model. As the results of this analysis suggest a predictive value to resource usage conflicts, the authors suggest that a cognitive architecture model, such as that portrayed in the HAM, can be a valuable tool for systems designers concerned with the prediction of human workload.

REFERENCES

1. Sarno, K.J. & Wickens, C.D. (1995). Role of multiple resources in predicting time-sharing efficiency: Evaluation of three workload models in a multiple-task setting. *The International Journal of Aviation Psychology, 5 (1),* 107-130.

2. Xie, B. & Salvendy, G. (2000). Review and reappraisal of modelling and predicting mental workload in single- and multi-task environments. *Work & Stress, 14 (1)*, 74-99.

3. Hancock, P.A., Williams, G., Manning, C.M., & Miyake, S. (1995). Influence of task demand characteristics on workload and performance. *The International Journal of Aviation Psychology, 5* (1), 63-86.

4. Jex, H.R. (1988). Measuring mental workload: Problems, progress, and promises. In P. Hancock and N. Meshkati's (Eds.) *Human mental workload,* (pp. 5-39). New York: Elsevier Science Publishers.

5. Raby, M. & Wickens, C.D. (1994). Strategic workload management and decision biases in aviation. *The International Journal of Aviation Psychology, 3 (3),* 211-240.

6. Lysaght, R.J., Hill, S.G., Dick, A.O., Plamondon, B.D., Linton, P.M., Wierwille, W. W., Zaklad, A.L., Bittner, A.C., Jr., & Wherry, R. J., Jr. (1989). *Operator workload: Comprehensive review and evaluation of workload methodologies*. Report No. 2075-3, Willow Grove, PA: US Army Research Institute for the Behavioral and Social Sciences.

7. Morrow, D. & Rodvold, M. (1998). Communication issues in air traffic control. In M.W. Smolensky & E.S. Stein (Eds.) *Human factors in air traffic control* (pp. 421-456). San Deigo, CA: Academic Press.

8. Hendy, K.C., Liao, J., & Milgram, P. (1997). Combining time and intensity effects in assessing operator information-processing load. *Human Factors, 39 (1)*, 30-47.

9. Meshkati, N. (1988). Toward development of a cohesive model of workload. In P. Hancock and N. Meshkati's (Eds.) *Human mental workload,* (pp. 305-314). New York: Elsevier Science Publishers.

10. Aldrich, T.B. & Szabo, S.M. (1986). A methodology for predicting crew workload in new weapon systems. In *Proceedings of the Human Factors Society 30th Annual Meeting*. Santa Monica, CA: The Human Factors Society.

11. Kopardekar, P. & Magyarits, S. (2003). Measurement and prediction of dynamic density. In *Proceedings of the 5th USA/Europe air traffic management research and development seminar,* Budapest, Hungary.

12. North, R.A., & Riley, V.A. (1989). W/Index: A predictive model of operator workload. In G. McMillan, D. Beevis, E. Salas, M. Strub, R. Sutton, and L. Van Breda's (Eds.), *Applications of human performance models to system design* (pp. 81-89). New York: Plenum Press.

13. Parks, D.L., & Boucek, G.P. (1989). Workload prediction, diagnosis, and continuing challenges. In G. McMillan, D. Beevis, E. Salas, M. Strub, R. Sutton, and L. Van Breda's (Eds.) *Applications of human performance models to system design* (pp. 47-63). New York: Plenum Press.

14. Rodgers, M.D., Mogford, R.H., & Mogford, L. S. (1998). *The relationship of sector characteristics to operational errors* (DOT/FAA/AM-98/14). Washington DC: FAA Office of Aviation Medicine.

15. Schmidt, D.K. (1978). A queuing analysis of the air traffic controller's work load. *IEEE Transactions on Systems, Man, and Cybernetics 8 (6),* 492-498.

16. Tulga, M.K. & Sheridan, T.B. (1980). Dynamic decisions and work load in multitask supervisory control. *IEEE Transactions on Systems, Man, and Cybernetics, 10 (5),* 217-232.

17. Manning, C.A., Mills, S.H., Fox, C., Pfleiderer, E., & Mogilka, H.J. (2002). *Using air traffic control taskload measures and communication events to predict subjective workload* (DOT/FAA/AM-02/4). Washington, DC: FAA Office of Aerospace Medicine.

18. Cardosi, K.M. (1993). Time required for transmission of time-critical air traffic control messages in an en route environment. *The International Journal of Aviation Psychology, 3 (4),* 303-313.

19. Porterfield, D.H. (1997). Evaluating controller communication time as a measure of workload. *The International Journal of Aviation Psychology, 7 (2),* 171-182.

20. Gopher, D. & Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures? *Human Factors, 26 (5),* 519-532.

21. Stein, E.S. (1985). *Air traffic controller workload: An examination of workload probe* (DOT/FAA/CT-TN-84/24). Atlantic City Airport, NJ: FAA Technical Center.

22. Wickens, C. (1984). *Engineering psychology and human performance.* Columbus, OH: Merrill.

23. Haraldsdottir, A., Schoemig, E.G., Warren, A.W., Kwok-On, T., & Crutchfield, J.M. (2004). Analysis of arrival management performance with continuous descent trajectories using the Regional Traffic Model. *Proceedings of the 23rd digital avionics systems conference*, Salt Lake City, UT.

24. Human Technology, Inc. (1991). Cognitive task analysis of en route air traffic control: Model extension and validation (OPM-87-9041). McLean, VA: Author.