UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

PREDICTING SUBJECTIVE WORKLOAD RATINGS: A COMPARISON AND SYNTHESIS OF THEORETICAL MODELS

A Dissertation

SUBMITTED TO THE GRADUATE FACULTY

In partial fulfillment of the requirements for the

Degree of

Doctor of Philosophy

By

JERRY M. CRUTCHFIELD
Norman, Oklahoma
2005

UMI Number: 3178305

UMI Microform 3178305

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

PREDICTING SUBJECTIVE WORKLOAD RATINGS: A COMPARISON AND SYNTHESIS OF THEORETICAL MODELS

A Dissertation APPROVED FOR THE DEPARTMENT OF PSYCHOLOGY

BY

____
Scott Gronlund

____
Francis Durso

____
Robert Terry

____
Randa Shehab

____
Carol Manning

Acknowledgments

I would like to express gratitude to my wife Martha and my parents, James and Judy Crutchfield for their patience and support throughout my lengthy matriculation. I would also like to thank Dr. Craig Rosenberg for designing, writing, and debugging the code that would make the Human Agent Module a working reality. Without any one of you my dissertation would have never happened. Many thanks.

Table of Contents

List of Tables

List of Illustrations

Abstract

Output data from a computer simulation of two air traffic control (ATC) scenarios were fit to workload ratings that ATC subject matter experts provided while observing each scenario in real time. Simulation output enabled regressions to test the assumptions of a variety of workload prediction models. The models included operational models that use observable situational and behavior variables (such as number of aircraft and communications by type) and theoretical models that use queuing and cognitive architecture variables (such as weightings of activities performed, amount of busy time, and sensory and cognitive resource usage) to predict workload. Regression results suggest models that include number of activities performed weighted by priority are best able to account for the highest amount of variance in subjective workload ratings.

PREDICTING SUBJECTIVE WORKLOAD RATINGS: A COMPARISON AND SYNTHESIS OF THEORETICAL MODELS

Concurrent with technological developments that have extended human physical and sensory capabilities has been an increase in the cognitive complexity of tasks humans are routinely asked to perform. Examples include the statistical analyst responsible for graphically representing the relationships among three different types of measures, the nuclear power plant operator responsible for monitoring numerous sensors for cues indicating the necessity of a certain switch action, and the air traffic controller responsible for efficiently directing growing levels of traffic while continuing to maintain safe aircraft separation. Tasks such as these require the performer to attend to vast amounts of incoming information and manage several ongoing activities in parallel.

Unfortunately, cognitively complex tasks tax the mental capabilities of human operators and give rise to situations of high workload. 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 (Manning, Mills, Fox, Pfeiderer, & Mogilka, 2001b; Xie & Salvendy, 2000). When complex tasks place high attentional demands on an operator, the result is an overall increase in workload.

Although there are variables shown to moderate the exact relationship between performance and workload for given situations (Hancock, Williams, Manning, & Miyake, 1995; Jex, 1988; Raby & Wickens, 1994;), in general, high workload levels tend to be associated with increases in operator error and decreases in overall performance (Lysaght et al., 1989; Morisson & Wright, 1989; Morrow, Lee, &

Rodvold, 1990). These findings, coupled with the growing amount of complex tasks found in today's work environments, have led to a growing interest in workload research. This is particularly true in the domain of air traffic control (ATC) where safety and operational efficiency often hinge upon human performance of highly complex tasks. Researchers recognize that workload levels inherent to cognitively complex ATC tasks may lead these tasks to be vulnerable to performance decrements.

Human factors researchers seek to understand workload and its characteristics so as to aid in the design of new systems that are hoped to reduce the occurrence of these performance decrements. If a thorough understanding of workload is attained the knowledge can be applied to the design of human-machine systems that reduce cognitive task complexity and foster appropriate workload levels. The pursuit of these goals requires identifying the factors that go into workload and developing tools that can measure and predict workload levels.

Unfortunately, the findings of workload research over the last three decades have revealed the construct to be a challenging one to characterize (Hendy, Liao, & Milgram, 1997; Meshkati, 1988; Xie & Salvendy, 2000). 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. The workload construct also seems to be multidimensional as well (Hart & Staveland, 1988; Hendy, Liao, & Milgram, 1997; Reid & Nygren, 1988). Hart and Staveland's National Aeronautics and Space Administration Task Load INDEX (NASA TLX) for example, measures workload along the dimensions of performance, frustration, effort, and mental, physical, and temporal demand.

Workload is also not a construct that can be directly measured (Xie & Salvendy, 2000). Rather it must be measured indirectly through related variables like primary and secondary task performance, physiological data, and subjective workload ratings. Numerous studies have been performed that explore the validity of these three measurement approaches (Gopher & Braune, 1984; Hart & Staveland, 1988; Lysaght et al., 1989; Reid & Nygren, 1988; Sarno & Wickens, 1995; Stein, 1985; Xie & Salvendy, 2000). The studies show that each of the three workload measurement approaches has both advantages and disadvantages depending upon the environment and domain from which the workload measures are to be taken and the purpose for which the measures are intended to be used.

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

Whereas many variables have the potential to moderate workload, there are a large number of variables that modelers can choose from to make workload predictions. Consequently, many different types of models have been developed to predict workload and workload related concepts such as dynamic density (Aldrich & Szabo,1986;

Kopardekar & Magyarits, 2003; North & Riley, 1989; Parks & Boucek, 1989; Rogers, Mogford & Mogford, 1998; Sarno & Wickens, 1995; Schmidt, 1978). These 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, the model approaches rely on entirely distinct sets of predictor variables.

One type of model applied to workload predictions is the queuing model. Queuing theorists model complex task performance by representing the process in terms of servers and clients (Schmidt, 1978; Tulga & Sheridan, 1980). Servers are processors capable of serving the clients who wait in queues to use them. Schmidt (1978) applied the queuing approach to the prediction of workload in the ATC domain. In his model the air traffic controller is represented as a server and the air traffic control 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 activities, and the relative priority of activities are used to predict the impact of workload (Schmidt, 1978; Tulga & Sheridan, 1980).

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 (Cardosi, 1993; Manning, Mills, Fox, Pfleiderer, & Mogilka, 2001a; Morrow & Rodvold, 1998; Porterfield, 1997). Variables such as these are often selected for analysis as they provide objective measures of workload that can be accessed without interfering with an air traffic controller's work. The discussion herein shall refer to models that use these types of variables as operational models due to the specificity of these types of variables to a given domain.

The identification of variables to be used 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 position of aircraft and deliver control commands to the aircraft vocally over a radio channel. Control commands, or clearances, include changes to aircraft altitude, heading, and speed. Clearances are given to direct aircraft to particular waypoints on the way to a destination, 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 the aircraft at different points in its 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, number of clearances by type, and number of handoffs performed as means to estimating workload (Cardosi, 1993; Manning et al., 2001a; Morrow & Rodvold, 1998; Porterfield, 1997).

One of the most comprehensive analyses of operational variables in the ATC domain to date was performed by Manning, et al. (2001a). In this study a wide range of operational variables were used in a regression analysis to predict workload. Twenty-three operational variables such as total aircraft controlled, total handoffs initiated, and number of altitude changes, were analyzed along with variables for number of communications and communication time. The operational variable values were derived from recordings of actual air traffic control. Manning, et al. first used a Principle Components Analysis on the values and reduced the variables into five sets. These sets were then used in multiple regression analyses to predict controller subjective 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. (2001a) study was also interesting in the way it collected the workload values that the operational variables were used to predict. In that study, workload was represented by subjective workload ratings. Although criticized due to findings that show dissociation between subjective workload ratings and performance (Gopher & Braune, 1984), 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 the observable and unobservable workload factors (Stein, 1998). 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. (2001a) study, controller subject matter experts (SMEs) observed recordings of air traffic control and indicated the amount of workload they believed the controller controlling the traffic was experiencing instead.

Although the Manning, et al. (2001a) 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 does not as 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 additional workload as a message delivered by a new technology. It may be the case 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 and cognitive 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 that all tasks can be broken down into, cognitive modeling may provide a way to account for the differences between any alternate systems that are modeled.

Early cognitive models included a variety of information processing models fashioned after the one described by Atkinson and Shiffrin (1968). These models were created to account for empirical data regarding the allocation of attention, multiple task performance, and the location of processing bottlenecks (Norman, 1968; Treisman, 1970). The information processing models were limited in that they implied the existence of only a single resource pool to be used by all types of cognitive processes.

Information processing models therefore failed to account for why tasks that shared certain characteristics in common seemed to interfere with each other more than others and why some tasks could be performed in parallel with little adverse impact. Multiple-resource theory was developed to explain data that the information processing models could not.

Although there are many types of cognitive models, most cognitive models applied to workload research are based on Wicken's Multiple-Resource Theory (1984). 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 different response resources (manual, vocal, etc.) for example. If two tasks require the 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 and cognitive resource usage and resource 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. A technique already seeing a great deal of application to the study of complex tasks was task analysis. 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 (Sanders & McCormick, 1993). 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 (1986) developed a process whereby the uses of theoretical cognitive, sensory, and motor resources were mapped on to 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 of an operator's 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 (1989). This model was developed at Boeing to predict pilot workload. Similar to the VACP approach, the approach created by Parks and Boucek used separate task networks for separate resource types. Task networks were created for cognitive, visual, auditory, manual hands, and manual feet resource usage. By enhancing their 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.

Theorizing that the earlier cognitive approaches to workload prediction could be improved by accounting for cognitive resource interference, North and Riley (1989) 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 were estimated to represent how much different parallel resource usages would interfere with performance. Workload predictions were found similar to the VACP approach except that the amount of relative task interference was included 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. Cognitive models are based on cognitive theories that may be controversial or otherwise difficult to confirm.

Sarno and Wickens (1995) tested and compared the assumptions of Parks and Boucek's (1989) TLAP, Aldrich and Szabo's (1986) VACP, and North and Riley's (1989) 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 between 56 and 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 features were associated with increases in prediction performance. Results showed that prediction performance was best for models that represented the use of multiple resources. The results also showed that workload prediction was not improved for models when the degree of resource usage interference was included in the calculations.

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, in the same domain, and 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.

Fortunately, Boeing Air Traffic Management's Regional Traffic Model (RTM) has made it possible to compare the predictive value of operational, cognitive, and queuing variables simultaneously. This model was developed as an analysis tool for the ATC domain. The RTM output includes variables such as number of aircraft under control, and number of communications given by type. Furthermore, the cognitive architecture found within the Human Agent Module (HAM) of the RTM models the use of cognitive, sensory, and motor resources and records when tasks requesting those resources are in conflict.

This dissertation used output from the RTM and HAM to test and compare the assumptions of both the operational and theoretical models. Two air traffic scenarios were run using the RTM and the output was used to derive queuing, operational, and cognitive model variables. The variables used are shown in Table 1. 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.

Table 1

Operational and Theoretical Variables Used to Predict Workload

Operational Variables	Theoretical Variables
Number of Communications	Task Load
Communication Congestions	Number of Activities
Number of Aircraft Controlled	Number of Activities by Priority
Number of Handoffs	Number of Activities by Difficulty
Number of Altitude Changes	Number of Tasks Performed
Number of Heading Changes	Visual Resource Use
Number of Speed Changes	Spatial Cognition Resource Use
	Verbal Cognition Resource Use
	Resource Usage Conflicts

In this dissertation the validity of each of the three model types was first tested individually. The test of operational variables performed for this dissertation was comparable to that performed by Manning, et al. (2001a). Similarly, the test of cognitive variables performed was comparable to that performed by Sarno and Wickens (1995) except that in this dissertation the cognitive variables were used to inform the relationship between the usage of different cognitive resources and subjective workload instead of task performance. A comparison of the predictive value of each of the different model types resulting from the analyses may serve to narrow the range of plausible theories used to characterize the workload construct.

After testing the models individually, regressions were performed to predict workload using combinations of variables from across the operational and theoretical model types. The test of these combinations was performed to identify the overall model that best predicted workload. It was proposed that the additional detail provided by theoretical models could combine with the predictive value of operational variables to provide a wide-ranging picture of the workload experience. This analysis consequently serves the more practical goal of this dissertation: to provide ATC researchers with a quantitative tool that can be used to predict the workload of alternate ATC systems. The results of this analysis furthers the continuing effort to find a workload prediction equation that can be used in lieu of actual workload ratings, prior to human-in-the-loop simulations or when physical prototypes are not yet available.

Method

Participants

Two ATC SMEs were compensated for their participation in this study. Both of these participants were ex-air traffic controllers currently employed as ATC training consultants. One participant's area of specialization was in control for TRACON environments and the other participant's area of specialization was in control for Center environments.

Materials

The Regional Traffic Model

Boeing Air Traffic Management's Regional Traffic Model is a fast time 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. The model was designed to provide trade studies with higher fidelity analyses than are provided through the use of economic analysis tools but in a shorter time span and at a lesser cost than analyses provided through the use of human-in-the-loop simulations. The model was also designed to be flexible so as to model efficiently a variety of alternate air traffic management operational concepts prior to making investments in the development of physical system prototypes. Through the use of models like the RTM, analysts can predict to some degree how effective a system will be and where failures or reductions in performance are likely to occur. Analysts can make changes to the system as it is represented in the model and collect further 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. These modules include Aircraft, Airspace, Communication, Surveillance, Traffic Generation, and Human Agent modules among others. In the Traffic Generation Module, for example:

> [stochastic traffic generation] 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 (Haraldsdottir, Schoemig, Warren, Tong, & Crutchfield, 2004, p. 2).

The Surveillance module represents the accuracy and delay associated with Radar or other technological sensor systems.

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 the possible impact different operational concepts will have on the performance of human operators. Increase and reduction of human operator workload is one type of impact with which ATC analysts are often concerned. 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 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 and can be used to predict the impact a given system has on the mental workload of the controller.

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 in 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. 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 controller HAM is capable of representing both of these types of controllers.

The HAM was designed and coded by Craig Rosenberg to meet the requirements and specifications provided by the author of this dissertation. The specifications included the processes and resources that would make up the model's cognitive architecture. Figure 1 presents a diagram of the theoretical cognitive architecture the model represents. The HAM and RTM were developed for use by Boeing Air Traffic Management and the details of the model are proprietary to Boeing. However, accompanying Figure 1 is a general description of how the HAM works.

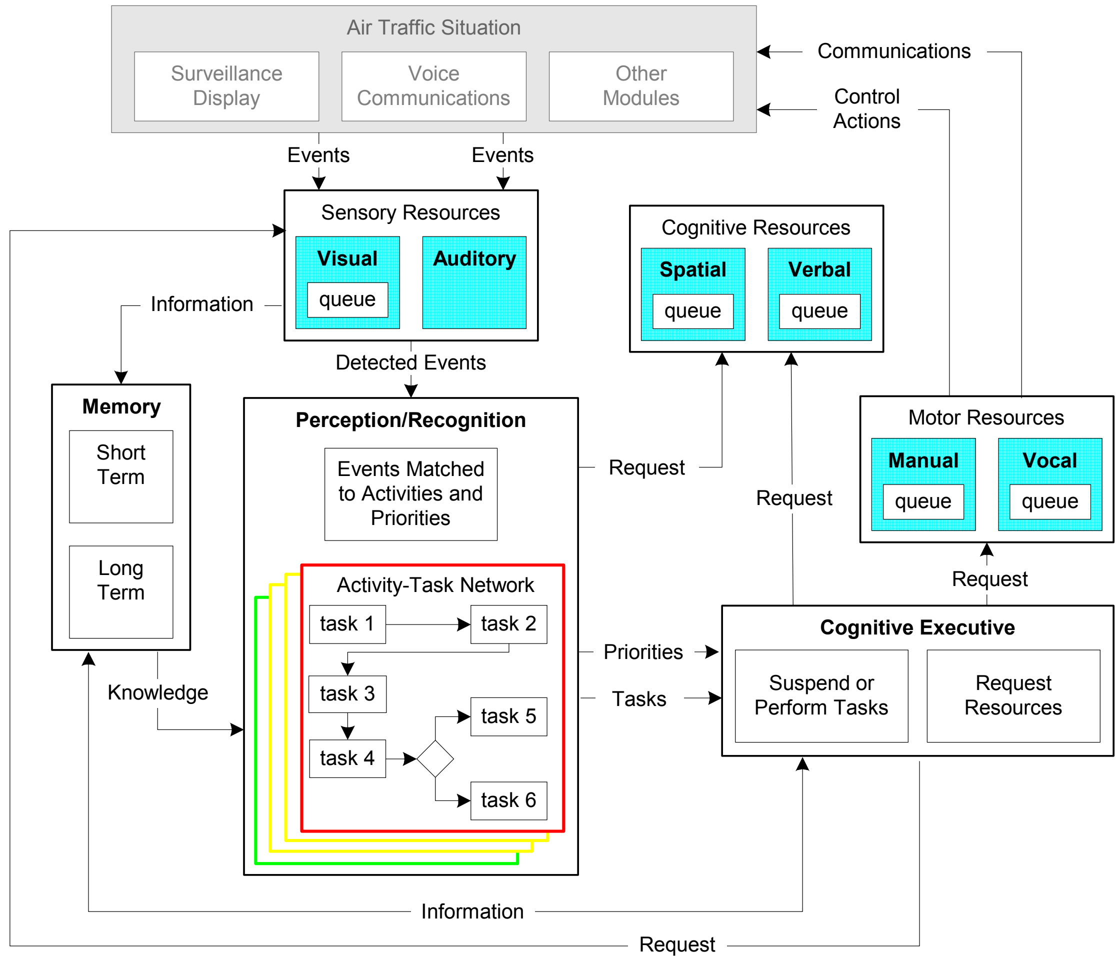

Figure 1. Cognitive Architecture represented in the Human Agent Module.

The controller HAM accomplishes ATC by first receiving traffic related events from other RTM modules. Events include notification that an aircraft has passed a waypoint or deviated from assigned speed 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, the event must be recognized. The HAM recognizes events by associating them with programmed activities and tasks. In the HAM, activities are made up of two or more tasks. The representative activities performed in response to the events were obtained from previously performed task analyses (Human Technology Inc., 1991; Rodgers and Drechsler, 1995) and through knowledge elicitation from controller SMEs. A relative priority ranking for each of the activities was also elicited. Table 2 provides a listing of many of the activities and tasks this version of the HAM simulated.

Table 2

HAM Activities and Tasks

Activities	Tasks
Conflict Detection and Resolution	Review radar display for potential violation of aircraft separation standard
	Mentally project aircraft future position/altitude/path
	Determine if aircraft are separated by less then prescribed minima
	Review potential conflict situation for resolution
	Determine appropriate action to resolve aircraft conflict situation
	Determine if conflict is resolved
Receive Handoff	Determine that aircraft is entering the sector
	Determine response to handoff request
	Coordinate with adjacent controller
	Receive handoff request
	Accept the Handoff

Activities	Tasks
	Controllers response to the request
	Review if the restrictions have been met
	Wait for pilot to check in and issue instructions
Establish Radio Communication	Receive initial radio contact from pilot
	Verify completeness of message
	Verify aircraft has proper ATIS code
	Provide correct ATIS code to pilot
Initiate Handoff	Determine that aircraft is leaving the sector
	Discuss transfer of control
	Determine adjacent controller request
	Initiate handoff
	Receive handoff acceptance
	Issue change of frequency to the pilot
Control Instructions	Formulate clearance with appropriate instruction
	Issue clearance and instruction to the pilot
	Detect acknowledgment/readback of issued clearance
	Verify aircraft compliance with the clearance
	Query pilot regarding compliance with clearance

Activities	Tasks
Conformance Violation	Review radar display for potential violation of conformance criteria
	Determine maneuver to restore flight
	Determine if aircraft is in conformance

When the controller HAM performs tasks associated with traffic events it calls upon representations of sensory, cognitive, and vocal resources. These resources make up the HAM's cognitive architecture. Tasks are theorized to require certain resources be available before it can be successfully completed. 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 two tasks require the same 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, represent human performance accuracy and delay in the implementation of the solution.

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 that show aircraft speed and altitude. Sector boundaries and the airway routes on which the aircraft flew were 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 (as a part of a "14s a Pair" runway configuration). 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 use.

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 along the trombone to ensure spacing at the Final Approach Fix (FAF) before handing the aircraft off to the Tower controller.

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

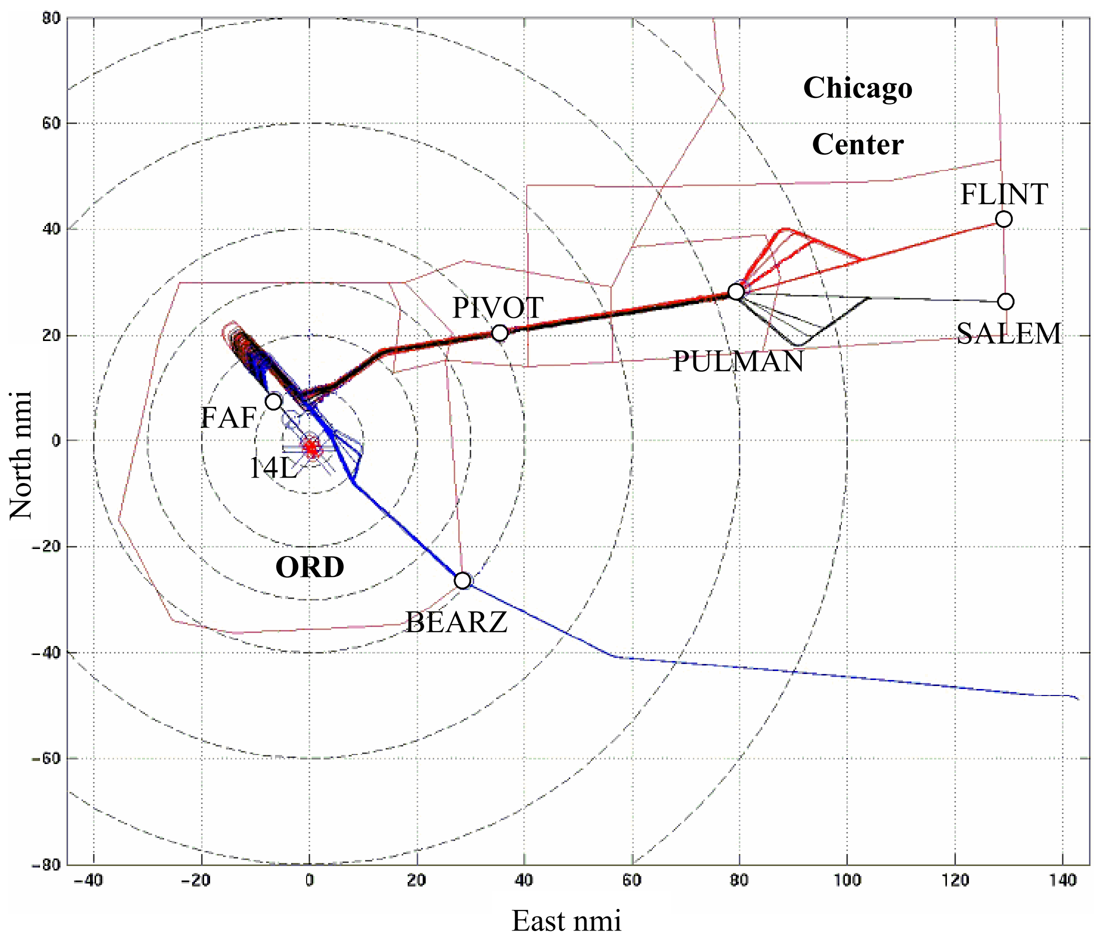

Figure 2. Illustration of lateral flight paths modeled within the RTM.

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. Parameters associated with controller and pilot performance can be found in Appendix A.

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 participant that specialized in TRACON control. 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 (Stein, 1985). 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 (instructions used are provided in Appendix B). Upon completion of both scenarios the observers were asked to indicate in writing any criteria they purposely tried to use when deciding upon what workload level to indicate.

Results

Several variables were selected from the scenario output for regression analysis to predict the workload ratings. These variables are related to three different theoretical model types that have been used to predict workload in previous studies and are further defined below. Descriptive statistics and correlations for these variables and the workload ratings are also provided. The correlations are followed by the model performance results for each of the three different theoretical model types. In the last part of this section all types of predictor variables are combined to identify the models that predict the greatest variance in participant workload ratings.

Variables Selected

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. These variables were derived from scenario output for each 4-minute period that a workload rating was collected. The way in which each of these variables is derived from the model is described below. The descriptions are organized according to the theoretical model type with which the variable is most associated.

Queuing Models

The variables in this category represent those used by Parks and Boucek in their TLAP approach (1989) and Schmidt in his queuing analysis approach (1978) to predict workload. Although ultimately derived from data at the task level, the variables used in these approaches are aggregate measures such as frequency, difficulty, and relative priority of activities, and overall task load. The following variables were chosen to represent queuing variables from the RTM output:

Number of Activities Performed. This is the number of Activities that were processed, to the point of using at least one sensory, cognitive, or motor resource, during a 4-minute time period.

Number of Activities Performed Weighted by Difficulty. This is the number of Activities that were processed, to the point of using at least one sensory, cognitive, or motor resource, during a 4-minute time period multiplied by a difficulty rating for the activity provided by a controller SME.

Number of Activities Performed Weighted by Priority. This is the number of Activities that were processed, to the point of using at least one sensory, cognitive, or motor resource, during a 4-minute time period multiplied by a priority rating for the activity provided by a controller SME.

Task Load. Taskload is the ratio of time on tasks to total time available. It is calculated by finding the amount of time in every 4-minute period that the controller is not performing any tasks (singly or in parallel) in seconds, subtracting that from 240 seconds, and then dividing the difference by 240 to obtain the ratio. The tasks considered in this analysis are only the tasks that use cognitive, sensory, or motor resource representations. Other tasks, such as those that involve waiting for a given Event to occur, are not included. When the controller is not using the Visual Resource in the completion of a particular task, it defaults to the use of the Visual Resource to perform scans of the radarscope for new events. This time spent using the Visual Resource performing the general monitoring task is not included in the total time on tasks calculation as it is assumed to occur when the controller has nothing else with which to occupy its time.

Operational Models

This category represents variables related to observable behaviors, specific to the ATC domain. Variables such as communication channel congestion, number of aircraft under control, and number of handoffs performed are studied to provide a non-intrusive and reliable means to predict controller workload and airspace complexity (Cardosi, 1993; Manning, et al., 2001a; Morrow & Rodvold, 1998; Porterfield, 1997). The following variables were chosen to represent operational models from the RTM output:

Number of Communications. This is the number of communications sent by the controller during a 4-minute time period. Both the total number of communications and the number of each particular type of communication (heading, speed, and altitude clearances and frequency changes) will be examined in the analysis.

Communication Channel Congestion. This is the sum of all the time durations that the communication channel was in use during a 4-minute time period.

Number of Aircraft. This is the number of aircraft in a controller's airspace during a 4-minute time period.

Number of Handoffs Completed. This is the number of frequency changes issued to aircraft in a particular sector during a 4-minute time period.

Cognitive Models

This category represents variables used by cognitive models based on multiple resource theory. These models include VACP (Aldrich & Szabo, 1986) and the W/INDEX (North & Riley, 1989). The following variables were chosen to represent cognitive variables from the RTM output:

Number of Tasks Performed. This is the number of tasks that utilize cognitive, sensory, or motor resources that were initiated during each 4-minute time period. The general monitoring task is not included in this calculation.

Resource Usage Conflicts. A resource request conflict occurs when a task requests the use of a sensory, motor, or cognitive resource that is already currently being used by another task. The request conflicts are a theoretical representation of resource interference experienced by individuals engaged in multi-tasking. These request conflicts will be summed for every 4-minute time period.

Descriptive Statistics

The average values for the derived variables and the participant provided workload ratings are shown in Table 3. At the time the scenarios were run, the version of the RTM used was experiencing bugs causing some anomalous aircraft and controller behaviors. For this reason, workload ratings and model output that were collected for 4-minute time periods in which the anomalous behavior occurred were excluded from the analysis. Fifteen TRACON and fifteen Center workload ratings remained and were combined for the analysis.

All derived variable values showed an increase from the Low Traffic condition to the High Traffic condition. The increase of these variables across traffic levels would suggest that all of these variables share a positive relationship. Number of aircraft controlled was greater for the TRACON controller, as was the proportionate increase in number of aircraft from Low Traffic condition to High Traffic condition. This was due to the fact that the TRACON sector was being fed by more than one Center sector. An increase in runway arrival rate 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 those related to pilot requests and overflights. Neither sector under Low or High Traffic conditions were 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.

Table 3.

Means of the Workload Ratings and Derived Predictor Variables.

	TRACON		Center	
Variable per 4 minute Time Segment	Low Traffic	High Traffic	Low Traffic	High Traffic
Workload Ratings	3.57	5.25	2.00	2.57
Task Load (percentage of time performing tasks)	56.79	78.96	38.33	42.50
Number of Activities	9.43	17.25	4.38	7.71
Number of Activities Weighted by Priority	39.86	71.63	16.38	28.14
Number of Activities Weighted by Difficulty	37.71	66.00	14.13	24.86

	TRACON		Center	
Variable per 4 minute Time Segment	Low Traffic	High Traffic	Low Traffic	High Traffic
Number of Communications	10.29	19.88	4.50	19.00
Communication Congestion (seconds)	50.15	92.37	22.32	91.07
Number of Aircraft	3.00	5.38	2.00	2.71
Number of Handoffs	1.00	1.63	0.63	1.71
Number of Altitude Changes	1.71	3.50	1.88	3.14
Number of Heading Changes	5.14	9.50	0.63	9.00
Number of Speed Changes	1.71	3.38	0.63	3.43
Number of Tasks	28.57	54.63	13.25	53.29
Task Specific Visual Resource Use	10.29	20.00	5.13	19.71
Task Specific Spatial Cognition Resource Use	2.57	4.75	0.63	4.86
Task Specific Verbal Cognition Resource Use	15.00	28.25	7.50	27.00
Resource Usage Conflicts	2.57	7.88	0.00	6.43

Participant Reported Criteria

After observing both scenarios, each participant provided in writing the criteria they personally chose to use when estimating workload. Together the participants reported using number of aircraft, potential number of conflicts, controller/pilot communication load, and handoff requirements as mental criteria for judging workload levels for the two scenarios. If the participants successfully adhered to the reported criteria, variables related to the criteria would be highly predictive of the workload ratings.

Correlations

Pearson correlations were found for all pairs of output variables. These correlations indicate the strength and direction of the relationship between each variable pair. As these variables are derived from the same model output, it is expected that a higher than average degree of correlation be found for some variables. Extraneous factors present in the real world or even a lab environment will attenuate the visible relationships between variables. However, this analysis is interesting because it allows us to quantify the relationships between variables from different theoretical model types.

Tables 4 and 5 show the Pearson r for each of the comparisons. Correlation values shown in the table with an asterisk are significant to the .05 level. As indicated by the descriptive statistics, all of the variable pairs were shown to have positive relationships.

Overall there was a high degree of correlation among the variables. Some correlations found were greater than .9. The relationship between Number of Tasks performed and Number of Communications for example, was .97. This is indicative of the fact that most Tasks were performed for the purpose of either creating or delivering clearances. Other variables such as Number of Activities, Number of Activities Weighted by Difficulty and Number of Activities Weighted by Priority are highly correlated because they share the same root output.

Table 4

Correlations between Derived Queuing and Operational Variables, TRACON and Center Sector Combined, for the Two Model Runs

	nact	prior	diff	ncom	cong	nac	nho	nalt	nhdg	nspd	tld
Activities Weighted by Priority	.99*	1.0									
Activities Weighted by Difficulty	1.0*	.99*	1.0								
Number of Commun-ications	.72*	.68*	.69*	1.0							
Commun-ication Congestion	.71*	.67*	.67*	.99*	1.0						
Number of Aircraft	.84*	.86*	.85*	.62*	.62*	1.0					
Handoffs	.54*	.51*	.50*	.68*	.73*	.33	1.0				
Altitude Changes	.55*	.50*	.52*	.70*	.65*	.35	.27	1.0			
Heading Changes	.70*	.68*	.68*	.97*	.95*	.65*	.61*	.57*	1.0		
Speed Changes	.57*	.55*	.53*	.88*	.89*	.58*	.71*	.43*	.84*	1.0	
Task Load	.65*	.69*	.67*	.36	.36	.73*	.21	.26	.38*	.33	1.0

Strong relationships were also found among variables used in different theoretical model types. For example, Number of Communications was highly related to Verbal Cognition Resource Use. This relationship results from the reliance communication tasks have on the use of the verbal cognitive resource for the formulation of message content. The relationships between the Visual Resource Use and both the number of aircraft Heading Changes and the number of aircraft Speed Changes was also high, indicative of a reliance on the visual resource for the completion of these two tasks as well.

Table 5

Correlations between Derived Cognitive Variables and Queuing and Operational Variables, TRACON and Center Sector Combined, for the Two Model Runs

	ntask	nvis	nspat	nverb	confli
Number of Activities	.65*	.60*	.46*	.70*	.60*
Activities Weighted by Priority	.62*	.57*	.45*	.66*	.59*
Activities Weighted by Difficulty	.62*	.57*	.44*	.67*	.59*
Number of Communications	.97*	.93*	.82*	.99*	.83*
Communication Congestion	.96*	.93*	.83*	.96*	.81*
Number of Aircraft	.63*	.63*	.55*	.61*	.60*
Handoffs	.63*	.63*	.54*	.60*	.39*
Altitude Changes	.63*	.52*	.40*	.74*	.68*
Heading Changes	.97*	.93*	.86*	.96*	.82*
Speed Changes	.90*	.91*	.82*	.85*	.71*
Task Load	.37*	.37*	.34	.36	.33

Model Performance by Type

Each of the variables recorded was used in a regression analysis to predict workload ratings. In some cases, pairs of variables from a model type were analyzed together. The results of these analyses are provided in the tables below. The tables provide both the R and the R^2 value indicating the amount of variance accounted for by a model. The tables also provide the F and the p values indicating the level of significance the model reached. These results indicate the ability to predict subjective workload for each of the independent theoretical model types as represented by the HAM and the RTM. Successful models identify candidates for variables that could be used in place of subjective workload ratings when it comes to predicting workload for new ATC systems.

Queuing Models

Activity Variables. The variables tested here include Number of Activities Performed and Number of Activities Performed weighted by either difficulty or priority. These variables represent aggregates of tasks performed to complete activities. Results are presented in Table 6. All three models did well at predicting workload 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 been suggested in the literature before (Hendy, Liao, & Milgram, 1997), time pressure may play an important role in the subjective experience of workload.

Table 6

Performance of the Activity Models

Variables	R	R^2	F	p
Number of Activities Performed	.849	.721	72.335	.000
Number of Activities Weighted by Difficulty	.857	.735	77.714	.000

Variables	R	R^2	F	p
Number of Activities Weighted by Priority	.876	.767	92.346	.000

Task Load. The amount of time the HAM was engaged in using sensory or cognitive resources was derived and the ratio of busy time to free time was calculated for each 4-minute time period. This Task Load ratio was analyzed in a linear regression to predict the subjective workload ratings. Table 7 provides the regression analysis results. The Task Load model was successful at predicting 45% of the variance in workload ratings.

Table 7

Performance of the Task Load Model

Variables	R	R^2	F	p
Task Load	.671	.450	22.880	.000

The Task Load Model uses an aggregate of the data provided by the Activity Models in its prediction of workload. The Activity Models use an aggregate of the tasks completed in its prediction of workload. The results of this analysis suggest that the Activity Models are better predictors of workload than the Task Load models. Neither of these Queuing Models represent tasks at the individual resource level and may therefore be relatively insensitive to what occurs inside of activities. However, they require less advanced work to prepare than cognitive models do. These results suggest that Activity Models may provide a powerful tool to predict workload at an appropriate stage of the design process before many of the task performance details associated with a system are known.

Operational Models

Communications. As communication variables have been used to estimate workload level in the ATC domain, number of communications and amount of time the radio communication channel was busy were analyzed separately using linear regression to predict the subjective workload ratings. Table 8 provides the regression analysis results. Neither model predicted more than 38% of the variance in subjective workload ratings. This performance result falls below the 49% found by Manning et al. (2001a).

Table 8

Performance of the Communication Models

Variables	R	R^2	F	p
Number of Communications	.610	.372	16.579	.000
Communication Congestion	.596	.355	15.406	.001

Other Operational Models. Other operational variables identified as possible predictors of controller workload were analyzed. Combinations of these variables and communication model variables were also tested. Table 9 provides the regression analysis results. Numbers of individual clearance types were not good predictors of workload. Of all the clearances, Number of Heading Clearances was the best predictor accounting for only 38% of the variance.

Number of Aircraft by itself performed well predicting 69%. In fact all six of the best performing models (predicting between 69 and 70% of the variance) included Number of Aircraft as one of the variables. These results paired with the participant reported criteria suggest that a measure as simple as Number of Aircraft under control can be a fairly accurate representative of subjective workload.

Table 9

Performance of the Other Operational Models

Variables	R	R^2	F	p
Number of Altitude Clearances	.390	.152	5.035	.033
Number of Heading Clearances	.619	.384	17.425	.000
Number of Speed Clearances	.541	.292	11.565	.002
Number of Handoffs Performed	.376	.141	4.610	.041
Number of Aircraft	.828	.686	61.161	.000
Number of Altitude Clearances and Number of Heading Clearances	.621	.385	8.463	.001
Number of Altitude Clearances and Number of Speed Clearances	.569	.323	6.450	.005
Number of Altitude Clearances and Number of Handoffs Performed	.480	.231	4.046	.029
Number of Altitude Clearances and Number of Aircraft	.835	.697	31.041	.000
Number of Altitude Clearances and Communication Congestion	.596	.355	7.430	.003
Number of Heading Clearances and Number of Speed Clearances	.620	.385	8.447	.001
Number of Heading Clearances and Number of Handoffs Performed	.619	.384	8.401	.001
Number of Heading Clearances and Number of Aircraft	.835	.698	31.186	.000
Number of Heading Clearances and Communication Congestion	.620	.384	8.414	.001
Number of Speed Clearances and Number of Handoffs Performed	.541	.292	5.580	.009
Number of Speed Clearances and Number of Aircraft	.832	.692	31.306	.000
Number of Speed Clearances and Communication Congestion	.596	.355	7.442	.003

Variables	R	R^2	F	p
Number of Handoffs Performed and Number of Aircraft	.836	.698	31.275	.000
Number of Handoffs Performed and Communication Congestion	.602	.362	7.659	.002
Number of Aircraft and Communication Congestions	.835	.697	31.044	.000

Operational Models of workload were as good predictors for the model as they were for the data collected by Manning et al. (2001a). Models that include one of the easiest types of variables to obtain for ATC, Number of Aircraft, did the best at predicting the subjective workload ratings of our participants. Unfortunately, the Number of Aircraft found in a scenario tells us very little about how one system contributes to workload levels versus another.

Cognitive Models

Resource Usage. The total number of Tasks completed by the HAM as well as the number of calls to the Visual Processor and Verbal and Spatial Cognition Resources for each 4-minute segment was recorded. These variables were analyzed separately using linear regression to predict the subjective workload ratings. Table 10 provides the regression analysis results. The highest performing variable from this list, Number of Tasks Performed, was successful at accounting for 36% of the variance in subjective workload ratings. These results suggest that an aggregate of cognitive and sensory resource usage may be as valuable for the prediction of workload as the usage of any particular resource by itself.

Table 10

Performance of the Resource Usage Models

Variables	R	R^2	F	p
Number of Tasks Performed	.599	.358	15.643	.000
Use of the Spatial Cognition Resource	.521	.271	10.432	.003
Use of the Verbal Cognition Resource	.592	.350	15.106	.001
Use of the Visual Processor Resource	.584	.341	14.502	.001

Resource Usage Conflicts. When a task requires the usage of a sensory, motor, or cognitive resource that is already being used by another task, the second task must wait until the first task stops using the resource before it can proceed. The number of these resource usage conflicts that occurred during each 4-minute time segment was used in a linear regression analysis to predict subjective workload ratings. Another model that combined number of resource conflicts with total number of tasks performed in each 4-minute time period was also analyzed. Due to the previous prediction performance given by the W/INDEX model used in Sarno and Wickens (1995) study it was theorized that these two variables might perform well together as subjective workload predictors. Table 11 provides the regression analysis results. The Resource Usage Conflicts model predicted roughly 41% of the variance while the combined model predicted roughly 42%. It is interesting to note that Sarno and Wicken's version of the W/INDEX model predicted a similar amount (42%) of the variance in TASKILLAN performance data.

Table 11

Performance of the Resource Usage Conflicts Models

Variables	R	R^2	F	p
Resource Usage Conflicts	.639	.408	19.279	.000
Resource Usage Conflicts and Number of Tasks Performed	.645	.416	9.628	.001

The Cognitive Models did only a moderate job at predicting workload. It is interesting to note that 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 Models may not fare well by themselves, they can potentially provide designers with useful information regarding resource usage.

Combining Variables

As the RTM and HAM produce variables for all three theoretical model types from the same scenario, it was theorized that the model types could be directly compared by using the different variables in a regression analysis. In this study, however, there was not a sufficient amount of workload ratings to perform any regression procedures using more than two variables at a time. Therefore the analysis was conducted by testing all variables (except where prohibited by co-linearity) in pairs.

The regressions identified seventeen variable pairs that produced models accounting for over 75% of the variance in workload ratings. A Bootstrap analysis procedure was applied to the predicted workload values of each of these models. Results of this analysis showed that none of the predicted values of any of the models were 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 collected does not allow for statistically reliable comparisons to be made.

All seventeen 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. Table 12 describes the top five workload predicting models.

Table 12

Model Performance for the Top Five Predicting Models

Model	R	R^2	F	p
Activity by Priority and Resource Conflicts	0.889	0.791	50.950	0.000
Activity by Priority and Number of Aircraft	0.889	0.790	50.689	0.000
Activity by Priority and Use of Spatial Cognition	0.887	0.787	49.869	0.000
Activity by Priority and Taskload	.881	.775	46.578	.000
Activity by Difficulty and Number of Aircraft	.878	.770	45.298	.000

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 + .067(Resource Conflicts) + .049(Activities Weighted by Priority). Figures 3 and 4 present line graphs allowing a visual comparison of workload predicted using this model equation with the actual workload ratings collected. 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 for use to represent workload levels in design situations where actual subjective workload ratings cannot be assessed.

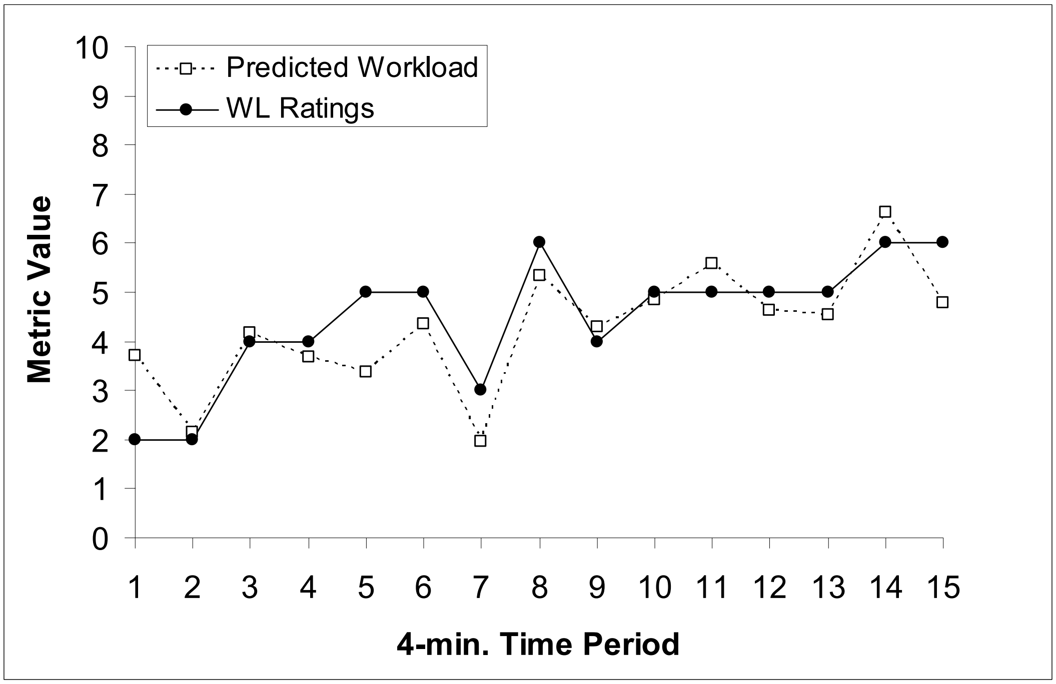

Figure 3. Workload predictions compared with ratings from the TRACON observer

Discussion

Results of this dissertation suggest that number of activities performed per 4-minute time period is a good predictor of workload. By itself this variable predicted almost 77% of 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 resources 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.

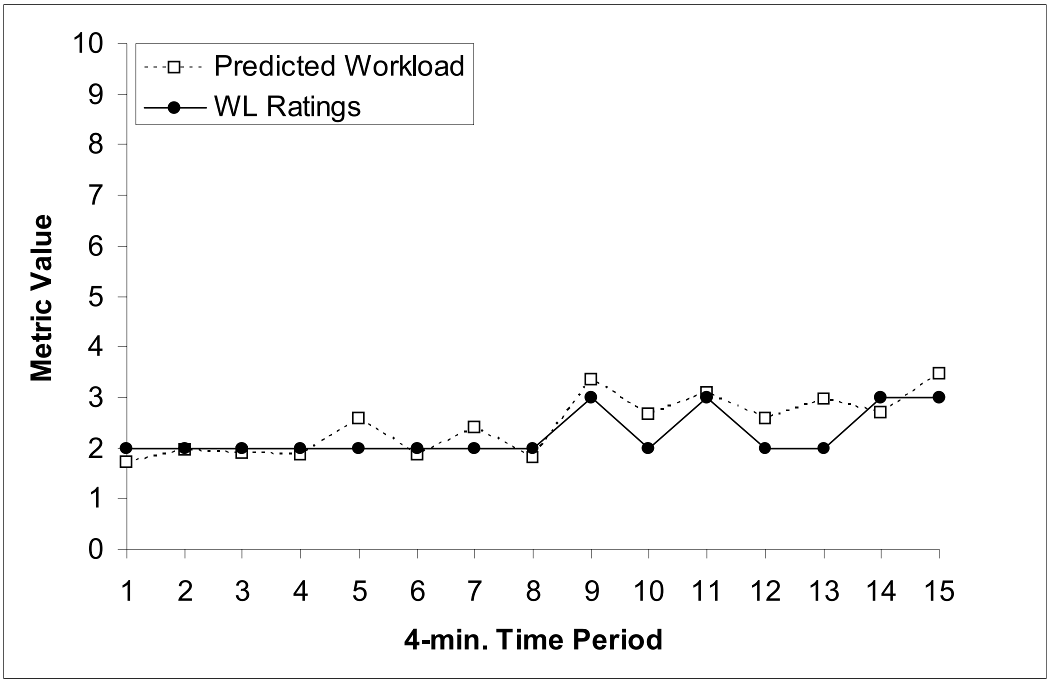

Figure 4. Workload predictions compared with ratings from the Center observer

The R^2 value of number of activities increased when this variable was weighted either by 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 to resulting workload (Hendy, Liao, & Milgram, 1997). 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 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 the workload associated with new systems.

One reason to include cognitive modeling is that a descriptive analysis of resource usage provides designers with a fairly comprehensive picture of 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 what 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 occurrences 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. 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.

One example of how inclusion of resource conflicts provides important information to the representation of workload can be found in the means through which clearances are delivered to pilots in the ATC domain. Voice communications can cause additional workload through the occurrence of readback errors and step-ons whereby communications from one pilot may occlude part of a communication from another pilot. In this mode of communication, the controller uses auditory and vocal resources to communicate. The use of datalink to send clearances by computer using visually displayed messages at first suggests a means of alleviating errors and workload associated with voice communications. However, traditional datalink clearances rely heavily on the controller's visual resources to both select and send a message and to watch for the response to that message. Considering the controller is already using her visual resources to scan the radar for potential conflicts etc., this type of datalink usage may instead increase associated workload over the use of voice communications. As the results of this analysis suggest a predictive value to resource usage conflicts, the author suggests that a cognitive architecture model, such as portrayed in the HAM, can be a valuable tool for systems designers concerned with the prediction of human workload.

It is important that work continue to be done to validate models such as the ones discussed here. For example, the reliability of the workload analysis results in this dissertation could be improved by either using a larger number of participants or developing the model to represent more of the factors that may impact workload. Additional participant resources would allow for the use of a third level of scenario and also increase the number of workload ratings. Either of these improvements would allow for less bounded workload ratings. The validity of model output to predict the impact of future technologies on workload could be further explored through the use of scenarios that include models of alternate operational concepts (data link, automated traffic advisories etc.) and subjective workload ratings by SMEs who have had experience with these types of prototype systems.

Valid workload prediction models provide potential benefits to the design of systems that are compatible with human capabilities. Regardless of model validity it is not the purpose of modeling to replace the use of higher fidelity analysis such as human-in-the-loop evaluations. Rather the value of modeling comes from narrowing the focus of the subsequent analysis. The approach allows analysts to make changes to a proposed system, as it is represented in a model, and collect data in a relatively quick and efficient manner. Models that can predict workload are valuable because they offer insight as to where human-system limitations lie. Perhaps more importantly, however, is that models such as these provide needed guidance early in the design life cycle prior to the development of costly physical prototypes for use with human-in-the-loop evaluations.

References

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: The Human Factors Society.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), The Psychology of Learning and Motivation (Vol. 2, pp. 89-195). Orlando, FL: Academic Press.

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.

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.

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.

Hart, S. G., & Staveland, L. E. (1988). Development of a NASA TLX (task load index): Results of empirical and theoretical research. In P.S. Hancock & N. Meshkati (Eds.), Human Mental Workload (pp. 139-183). Amsterdam: Elsevier.

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.

Human Technology, Inc. (1991). Cognitive task analysis of en route air traffic control: Model extension and validation (Report No. OPM-87-9041). McLean, VA: Author.

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.

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.

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, US Army Research Institute for the Behavioral and Social Sciences, Willow Grove.

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: Office of Aerospace Medicine.

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.

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.

Norman, D. A. (1968). Toward a theory of memory and attention. Psychological Review, 75, 522-536.

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.

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.

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

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

Reid, G. B. & Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P. A. Hancock & N. Meshkati (Eds.), Human Mental Workload (pp. 185-218). Amsterdam: Elsevier.

Rodgers, M.D. & Drechsler, G.K. (1995). Conversion of the TRACON operations concept database into a formal sentence outline job task taxonomy (DOT/FAA/AM-95/16). Oklahoma City, OK: Human Factors Research Laboratory, Civil Aeromedical Institute, Federal Aviation Administration.

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.

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.

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.

Stein, E. S. (1985). Air Traffic Controller Workload: An Examination of Workload Probe (DOT/FAA/CT-TN84/24). Atlantic City Airport, NJ: Federal Aviation Administration Technical Center.

Treisman, A. M. (1970). Contextual cues in selective listening. Quarterly Journal of Experimental Psychology, 12, 242-248.

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.

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.

Appendix A

Air Traffic Controller Parameter	Description	Value	Units
activityLimit	Represents the maximum number of Activities for which a controller can work on tasks simultaneously.	7	none
advisoryScTime	This is the duration of time for which the controller HAM uses the Spatial Cognition resource when a Task indicates the use of the Spatial Cognition resource.	1.42	sec
altitudeMonitoringCorrectionTime	Indicates the time the HAM Controller waits before restarting to monitor for altitude violations, for a given aircraft after an altitude correction has been issued for the aircraft.	60	sec
altitudeMonitoringLowerBound	Altitude that an aircraft can go lower than its commanded altitude before monitoring will give the HAM Controller an altitude violation Event.	60.96	meters
altitudeMonitoringUpperBound	Altitude that an aircraft can go higher than its commanded altitude before monitoring will give the HAM Controller an altitude violation Event	60.96	meters
auditoryProcessorTime	This is the duration of time for which the controller HAM uses the Auditory Processor when a Task indicates the use of the Auditory Processor. The controller HAM's auditory processor resource represents what is referred to in the domain of Cognitive Psychology as the verbal	0.117	sec

Air Traffic Controller Parameter	Description	Value	Units
	cognition resource.		
cognitiveCycleTime	Indicates the increment of time before which the tolerance processor checks the HAM Controller memory for the presence of memory codes. Also influences how often ongoing Activities are reprioritized and how quickly new Activities can be initiated.	0.117	sec
enrouteMinimumSpeed	This represents the minimum speed that a controller is realistically likely to direct an aircraft to fly to, in the en route environment.	128.61	meters /sec
etaStaTolerance	The tolerance (in seconds) of the difference between the ETA and the STA. If the difference is less than this tolerance, the controller will not do anything special as far as trying to get an aircraft to meet its STA.	5	sec
lateOnBaseTurnDelay	The amount of time that the Controller HAM will wait before it turns an aircraft onto base during the error condition of "late on base turn".	20	sec
lateOnBaseTurnPercentage	This represents the percentage of time that controllers are late in turning an aircraft onto the base.	0	%
maxTraconSpeed	This represents the fastest speed that aircraft can legally fly in the TRACON.	128.61	meters /sec

Air Traffic Controller Parameter	Description	Value	Units
maxTromboneController	The furthest beyond the nominal trombone that the HAM Controller will allow an aircraft to go before turning it onto base.	5556	meters
minTraconSpeed	This represents the minimum speed that a controller is realistically likely to direct an aircraft to fly to, in the TRACON environment.	87.455	meters /sec
minTromboneController	The furthest before the nominal trombone that the HAM Controller will allow turning an aircraft to base.	-5556	meters
motorProcessorTime	This is the duration of time for which the controller HAM uses the Motor Processor when a Task indicates the use of the Motor Processor. More specifically, this refers to the manual motor processor and not the vocal motor processor.	1	sec
msgLenChangeFrequencyMean	Average amount of time that a change frequency communication will take.	9.46	sec
msgLenClearanceMean	Average amount of time that a clearance communication will take.	9.46	sec
msgLenConfirmMean	Least amount of time the communication from the controller to the Pilot will take when the controller is responding to a Pilot check in.	9.46	sec
msgLenILSMean	Average amount of time the communication from the controller to the Pilot will take when the controller is directing the Pilot to intercept ILS.	9.46	sec

Air Traffic Controller Parameter	Description	Value	Units
responseSelectionProcessorTime	This is the duration of time for which the controller HAM uses the Response Selection Processor when a Task indicates the use of the Response Selection Processor.	0	sec
spatialCognitionProcessorTime	This is the duration of time for which the controller HAM uses the Spatial Cognition resource when a Task indicates the use of the Spatial Cognition resource.	1.42	sec
speedMonitoringCorrectionTime	Indicates the time the HAM Controller waits before restarting to monitor for speed violations, for a given aircraft after a speed correction has been issued for the aircraft.	30	sec
speedMonitoringLowerBound	Speed an aircraft can travel less than that issued by a controller before monitoring will send a speed violation Event to the Controller HAM.	55.144	meters /sec
speedMonitoringTime	Indicates the time the HAM Controller waits before restarting to monitor for speed violations, for a given aircraft after a speed clearance has been issued for the aircraft.	30	sec
speedMonitoringUpperBound	Speed an aircraft can travel greater than that issued by a controller before monitoring will send a speed violation Event to the Controller HAM.	27.572	meters /sec
traconSpeedIncrement	Represents the speed increment in which controllers issue speed clearances to a Pilot in the	5.144	meters /sec

Air Traffic Controller Parameter	Description	Value	Units
	TRACON.		
TromboneStepController	Distance increment the controller checks between the min and max trombone for options to meet the STA.	1609.34	meters
useIMC	Denotes whether the HAM Controller uses IMC control procedures or VMC control procedures.	True	True/ False
visualDetectionTime	Represents the mean time that it takes a controller to detect a visual event on the radar screen.	7	sec
visualPerceptualWorkload	Indicates resource interference ratio used in some workload prediction calculations.	1	ratio
visualProcessorTime	This is the duration of time for which the controller HAM uses the Visual Processor resource when a Task indicates the use of the Visual Processor resource.	0.66	sec

Pilot Performance Parameter	Description	Value	Unit
confirmWaitAdditional	A constant amount of time added to the time that a Pilot would wait after the communications channel is free before talking.	0	sec
confirmWaitRange	Upper bound of uniform distribution starting at 0, from which a time representing how long a Pilot waits after the comminication channel is free before talking is selected.	0	sec
lnavDelay	Amount of time between when an evNewHeading command is received by the Pilot from the Controller HAM and an evSetLateralControlModeLNAV event is sent from the Pilot to the aircraft.	5	sec
msgLenLowerBound	Shortest length of a Pilot communication.	0	sec
msgLenMean	Average length of a Pilot communication.	0	sec
msgLenUpperBound	Longest length of a Pilot communication.	0	sec
msgLenVar	Variance assocated with the length of a Pilot communication.	0	none
PilotSlowingMean	Percentage of time that no early pilot slowing occurs.	100%	%
PilotSlowingVar	Variance of distribution of percentage of time that no early pilot slowing occurs.	0	none
reactionTimeLowerBound	Shortest reaction time for the Pilot.	0	sec
reactionTimeMean	Average reaction time for the Pilot.	0	sec
reactionTimeUpperBound	Longest reaction time for the Pilot.	0	sec
reactionTimeVar	Variance assocated with the reaction time of the Pilot.	1	none
readbackErrorPercentage	Represents the percentage chance of a controller clearance being read back incorrectly by a pilot.	0.80%	%
timeTillCheckin	Amount of time that it takes until the Pilot checks in with a new sector.	1.67	sec

Appendix B

ATWIT Instructions

The purpose of this research is to obtain an accurate evaluation of controller workload. By workload, I mean all the physical and mental effort that you must exert to do your job. This includes maintaining the "picture," planning, coordinating, decision-making, communicating, and whatever else is required to maintain a safe and expeditious traffic flow.

In the next 2 hours you will watch two approximately 40 minute air traffic scenarios. At 4-minute increments during the scenarios I will ask you to provide your estimate of how hard a specialist controlling the traffic scenario would be working at that moment. Although a computerized model is actually controlling the scenarios, try to evaluate them as if an experienced controller was controlling the traffic observed instead.

Please provide your workload estimate using a scale from 1 to 10. I will review what these rating numbers mean in terms of workload. At the low end of the scale (1 or 2), workload is low – a controller can accomplish everything easily. As the numbers increase, workload is getting higher. The numbers 3, 4, and 5 represent increasing levels of moderate workload where the chance of error is still low but steadily increasing. The numbers 6, 7, and 8 reflect relatively high workload where there is some chance of making errors. At the high end of the scale are the numbers 9 and 10, which represent a very high workload, where it is likely that a controller will have to leave some tasks unfinished. Please provide your workload estimate as quickly as possible.