DEVELOPMENT OF TIME AND WORKLOAD METHODOLOGIES FOR MICRO SAINT MODELS OF VISUAL DISPLAY AND CONTROL SYSTEMS

by

Sandra A. Moscovic

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Industrial and Systems Engineering

APPROVED:

Robert J. Beaton, Chair

Gregory J. Buhyoff

Robert D. Dryden

Paul T. Kemmerling

Harry L. Snyder

May, 1992

Blacksburg, Virginia

DEVELOPMENT OF TIME AND WORKLOAD METHODOLOGIES FOR MICRO SAINT MODELS OF VISUAL DISPLAY AND CONTROL SYSTEMS

by

Sandra A. Moscovic

Committee Chair: Robert J. Beaton, Ph.D.

Industrial and Systems Engineering

(ABSTRACT)

The Navy, through it's Total Quality Leadership (TQL) program, has emphasized the need for objective criteria in making design decisions. There are numerous tools available to aid human factors engineers meet the Navy's need. For example, simulation modeling provides objective design decisions without incurring the high costs associated with prototype building and testing. Unfortunately, simulation modeling of human–machine systems is limited by the lack of task completion time and variance data for various objectives. Moreover, no study has explored the use of a simulation model with a Predetermined Time System (PTS) as a valid method for making design decisions for display interactive consoles.

This dissertation concerns the development and validation of a methodology to incorporate a PTS known as Modapts into a simulation modeling tool known as Micro Saint. The operator task context for the model was an interactive displays and controls console known as the AN/SLQ-32(V). In addition, the dissertation examined the incorporation of

a cognitive workload metric known as the Subjective Workload Assessment Technique (SWAT) into the Micro Saint model.

The dissertation was conducted in three phases. In the first phase, a task analysis was performed to identify operator task and hardware interface redesign options. In the second phase data were collected from two groups of six participants who performed an operationally realistic task on 24 different configurations of a Macintosh AN/SLQ-32(V) simulator. Configurations of the simulated AN/SLQ-32(V) were defined by combinations of two display formats, two color conditions, and two emitter symbol sets, presented under three emitter density conditions. Data from Group 1 were used to assign standard deviations, probability distributions and Modapts times to a Micro Saint model of the task. The third phase of the study consisted of (1) verifying the model-generated performance scores and workload scores by comparison against scores obtained from Group 1 using regression analyses, and (2) validation of the model by comparison against Group 2.

The results indicate that the Modapts/Micro Saint methodology was a valid way to predict performance scores obtained from the 24 simulated AN/SLQ-32(V) prototypes ($R^2 = 0.78$). The workload metric used in the task network model accounted for 76 percent of the variance in Group 2 mean workload scores, but the slope of the regression was different from unity ($p = 0.05$). The statistical finding suggests that the model does not provide an exact prediction of workload scores. Further regression analysis of Group 1 and Group 2 workload scores indicates that the two groups were not homogenous with respect to workload ratings.

ACKNOWLEDGMENTS

This research would not have been possible without the assistance of my graduate studies committee: Dr. Robert Beaton, Dr. Gregory Buhyoff, Dr. Robert Dryden, Prof. Paul Kemmerling, and Dr. Harry Snyder. All committee members were a wealth of knowledge and support. As committee chairman, Dr. Beaton worked closest to this research effort and his insights, expertise, and encouragement were a constant source of motivation.

The other students working on the redesign project for the Navy – Bo Dyess and Richard Miller – were instrumental to this work and their support and friendship are greatly appreciated. Richard is credited with the programming of the Macintosh prototypes and making his computer equipment available for the rest of the project team.

Numerous people have supported my academic endeavors. Despite being many states away, my mom has been always at my side with love, support, and her unwavering confidence in me. The loving memory of my father keeps me inspired to make something of myself. My sister Anita has also been there for me in rough times, as have been my other siblings Dianne, Paul, and Doug. Thanks go to my friends, especially Chris, Linda, Deborah, General Doppelt, and the rest for all their encouragement and compassion.

A final word of appreciation goes to NAVSEA for funding this project and to the Air Force for giving me this opportunity to learn.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

The complexities and costs of modern technological systems have produced a need for computer simulation tools in human factors engineering. Computer modeling enables human factors engineers to support design decisions and recommendations with quantitative data prior to prototyping actual systems. The rapidly increasing availability of computer modeling packages which aid design engineers is evidence of the escalating interest in computer modeling. Given the intricacy of human-computer systems, computer modeling in human factors engineering is expected to become an important part of normal design activities.

One currently popular tool in human-machine modeling is the program Micro Saint (Chubb, Laughery, and Pritsker, 1987). Micro Saint, developed for simulation of operator tasks through task state networks, is a microcomputer version of the System Analysis of Integrated Networks of Tasks (SAINT). Operator tasks are modeled by a network of nodes representing constituent elements or subtasks. The ordering of the tasks is indicated by branches among the nodes, where several branches emanating from one node represent available choices leading to an operator decision. Micro Saint simulates operator task activity by generating successive runs through the network using subtask completion times, variances, and probability distributions.

The utility of Micro Saint in the visual displays and controls realm has not been fully explored by human factors engineers. There is, however, a reason which impedes the use of all task networking models—acquiring appropriate data on task completion times (means, standard deviations, and probability distributions) to use in the models. A logical progression in the development of human factors engineering tools is to explore methods to create time parameter taxonomies or methodologies. A secondary means of enhancing modeling tools consists of extending the models to include indices of cognitive workload.

The design of hardware and software systems is a non-trivial problem for human factors engineers. In the past, the human factors approach consisted of developing prototypes and subsequently attaining data by empirical testing. Prototypes are expensive, time-consuming, and offer little flexibility with respect to different design configurations. Contrary to good design practice, prototypes often are not developed until most design decisions are made (Chubb et al., 1987).

Inadequacies of the prototyping method have been overcome somewhat by the engineer's ability to predict operator performance in different system designs through computer modeling. Computer simulation provides the predictive capability in a manner that is compatible with evaluation techniques (e.g., task analysis). In the last decade, some attempts have been made to merge predictions of operator workload with other analytical methods prior to system development (Laughery, 1989).

Micro Saint was developed expressly with the human factors engineer in mind (Laughery and Drews, 1984). That is, the program was designed for modeling human-operator systems. Unlike general purpose modeling languages, Micro Saint does not concentrate on micro-level activities. Rather, Micro Saint addresses issues of strategic and tactical decision making by the operator (Laughery and Drews, 1984). Application of Micro Saint begins with an analysis of the system in terms of constituent operator subtasks. The subtasks are modeled in terms of a network. The computer simulation imitates system operation through algorithms that replicate interrelationships among the subtasks. Overall system performance is simulated by Monte Carlo selections of subtasks and times.

Micro Saint has been used to model complex human-operator systems such as tank crew performance (Laughery and Hegge, 1984), Close-In Weapon System (CIWS) loading operations (Tijerina and Treaster, 1990), performance on Naval Surface Ship Combat Direction Systems (CDS) (Osga, 1989), and the effects of medical chemical defenses on military performance (Laughery and Hegge, 1984). The predecessor of Micro Saint (titled SAINT) has been used to model operator performance in air defense systems (Chubb et al., 1987). In general, Micro Saint has been valuable in evaluating different design alternatives and different manning configurations.

Although there has been relatively widespread use of the modeling tool, there are still some significant concerns associated with implementing Micro Saint (B. Bachert, personal communication, June 17, 1991). One difficulty resides in the assignment of time quantities to the task network. The assignment of task completion times is particularly difficult when modeling a system that does not exist. Two critical pieces of information – the mean time and standard deviation for each sub-element – are tedious and expensive to obtain. Typically, time data are collected by averaging precise time measurements made on numerous iterations of the tasks of interest. This process is laborious and time intensive.

In general, there are three sources of time data for task network models. First, these data may be extracted from laboratory studies, trainers, and simulators. Second, operational data from war-gaming or routine operations may be used. Third, subjective opinions provide a source of information of task completion time values. Laboratory sources are used most often but some data have been collected from simulators. Reportedly, almost no models have been built from operational sources (Boff and Lincoln, p. 1365).

Laboratory data are difficult to use in real-world applications due to the controlled situations from which they are obtained. Although data from simulators are representative of actual tasks and less contrived than laboratory tasks, rarely are they collected. Operational data are subject to contamination due to a lack of control and often they are not collected systematically. Many of the past efforts to derive time parameters consisted of subjective methods, using expert opinions to estimate time values (Boff and Lincoln, 1988).

Boff and Lincoln (1988) defined a data bank as a "...systematically organized and formatted compilation of data arranged according to a special taxonomic scheme to answer specific questions" (p. 1373). There are no existing time data banks for network modeling of human operator systems.

One goal of the present effort is to explore the use of task network modeling in human factors research by creating an effective method of approximating subtask performance times and deviations. Establishing a valid and reliable methodology to collect and assign time parameters to subtasks should facilitate the use of Micro Saint in human factors engineering. Micro Saint can be enhanced further by a methodology to include indices of mental workload in the task network.

A possible starting point to establish a methodology for incorporating time data into Micro Saint draws on another tool commonly used in a different field of industrial and systems engineering; that is, Predetermined Time Systems (PTSs). Predetermined times are used for creating work standards for jobs different from those where the time data were measured originally. Values of similar elements from previously measured jobs are used to predict the unmeasured elements of some different job (Turner, Mize, and Case, 1987).

Although PTSs were constructed as databases for determining work standards, they are premised on the same ideas as task network modeling. Modeling and PTSs require that performance be defined in terms of component tasks which are assigned individual time values. Both tools assume that a skilled operator performs the sub-elements of a task without delays.

It may be argued that the use of predetermined time data for network modeling is impractical for portions of task models that deal with mental processes (such as reading or decision making) because PTSs estimate the time to perform a series of known *actions*. There are two reasons that the utility of exploring PTSs in the network modeling application is warranted. First, time systems such as Modapts incorporate methods to estimate times to read text (Shinnick, 1987). Second, decision times can be modeled either as separate factors (nodes) or can be included indirectly with the operator's overt body movement times.

In short, a taxonomy for assigning Micro Saint time values has not been developed and PTSs present a reasonable approach to this end. Additionally, incorporation of cognitive workload indices in Micro Saint would be useful. With these objectives, a concurrent effort to perform a human factors evaluation of a display and control console for a shipboard anti-missile early warning system provides the context in which to investigate the integration of completion time and mental workload methodologies into task network models.

BACKGROUND

The relevant background for the current research efforts is arranged in four separate sections. The first section describes the Micro Saint modeling tool. The second section describes some of the more popular PTSs with an emphasis on Modapts. The third section discusses cognitive workload in general and the Subjective Workload Assessment Technique (SWAT) specifically. The fourth section describes the AN/SLQ-32(V) Display Control Console (DCC), which provides the operator task context of the present research.

Micro Saint

System Analysis of Integrated Networks of Tasks (SAINT) was developed in the mid-1970s by the Air Force Aerospace Medical Research Laboratory in response to the need for a tool to evaluate complex human operator-system interactions (Laughery and Drews, 1984). The SAINT methodology simulates operator tasks and activities as a network of nodes and branches. The network portrays the sequential relationships of the subtasks, where the various branches reflect decision points for the operator.

SAINT was used infrequently because it was difficult to learn and intricate to use (Laughery and Drews, 1984). In January 1985, the *Journal for the Society for Computer Simulation* presented a seven-item list of features that a panel of simulation experts felt were necessary for effective modeling programs. The list included: (1) no coding, (2) no manuals, (3) user-friendly language, (4) greater software availability, (5) windows, (6) integrated applications, and (7) a conceptual framework for the expression of problems (Laughery and Drews, 1984). This prompted Micro Analysis and Design Corporation to develop the microcomputer version of SAINT, appropriately named Micro Saint.

The primary development goal of Micro Saint was to allow models to be built easily by non-simulation experts. In 1986, an IBM PC version of the Micro Saint software was developed. And, in 1991, Micro Analysis and Design produced an Apple Macintosh version of Micro Saint. The Macintosh version of Micro Saint is more powerful than the PC version for two reasons: (1) task network models require less time to create and (2) time to learn the software was minimal (Laughery and Drews, 1984).

The Micro Saint technology is a network modeling approach involving the decomposition of a system or operation into a series of subactivities or tasks. This decomposition uses the same principles of task and function analyses.

The Micro Saint model user must decide which level of system decomposition will meet the desired objectives. A model can be a single network or a hierarchy of several subnetworks. A model can be composed of relatively separate subnetworks which act autonomously.

After the task networks have been defined, the model user determines the variables that are relevant to the research effort. The model user must be able to describe the way variables are affected by the tasks or nodes in the network and, then, construct these relationships within the model. The relationships among tasks and the interrelationships among networks incorporated through the variables are crucial for valid modeling. For each node, the user can input this information either as a "beginning effect" which is a change in the variable as a function of the current node beginning or an "ending effect" which is a change in the variables as a function of node completion.

Beside the variables that the model user defines, Micro Saint maintains four system variables for each network:

1. Clock - keeps track of the elapsed time since the beginning of the model execution.
2. Duration - keeps track of the time that each entity passing through the model spends in each node or queue.
3. Run - keeps track of the number of times the model is run.
4. Tag - keeps track of the entities within an executing model.

To construct the model, the model user is required to provide five pieces of information for each subtask as listed below:

- release condition
- mean time
- underlying distribution
- follow-on tasks
- decision criteria for the follow-on tasks.

The release condition refers to the state of the system that must exist before the current node or task can execute. For example, before the task "launching chaff" can be performed (the task at hand), the launcher must be loaded and aimed (release conditions).

Micro Saint computes the execution time for each task using the probability distribution that the modeler specifies. The underlying time distribution of each task (e.g., rectangular, exponential, gamma, normal, lognormal, or some alternate distribution derived by the modeler) allows simulation of the task execution time as realistically as possible.

Follow-on tasks (e.g., what happens next) of each task are listed, and the modeler specifies the logic that Micro Saint will use when selecting from alternate follow-on tasks during each run. The decision types available to the modeler include "single" (only one follow-on task possible), "multiple" (all follow-on tasks execute simultaneously), "tactical" (the follow-on job with the greatest of some modeler-defined expression), and "probabilistic" (the follow-on jobs occur randomly but with unequal probabilities as defined by the modeler).

Although construction of the task network logic can be performed readily, specifying subtask times is an obstacle that discourages the use of Micro Saint models. In relatively small models, it is not prohibitive to obtain time data using video cameras equipped for frame-by-frame analysis. But in large models, the time and cost to gather task times under all conditions can be a formidable undertaking.

One route around the limitation imposed by the acquisition of time parameters is to use a general purpose methodology that will provide good estimates of time. An initial possibility for developing a methodology for appropriate subtask time parameters draws upon established PTSs for first–order approximations. A brief description of PTSs is given below.

Predetermined Time Systems

Time and motion methods are not new to the field of industrial engineering or the discipline of human factors engineering. Karger and Bayha (1987) define PTSs as:

> ...an organized body of information, procedures, and techniques employed in the study and evaluation of work elements performed by human power in terms of the method or motions used, their general and specific nature, the conditions under which they occur, and the application of prestandardized or predetermined times which their performance requires (p. 42).

The origin of work–time studies dates to the early 1900 "scientific management revolution" when systematic techniques were applied to management of people in the workforce (Tripartite Working Group, 1976; Turner, et al., 1987). The first task analytic methods incorporating time and motion concepts trace to two concurrent efforts: Frederick Taylor's stopwatch time study procedures and Frank and Lillian Gilbreth's PTS procedures in 1911. Predetermined time standards refer to the collection and cataloging of time data that are applied subsequently to tasks or jobs other than the ones from which they were collected originally. Rather than determine the standard times for each job through an individual study, standard databases were organized from a number of related studies (Mundel, 1985).

From this starting point, a number of techniques were formulated that embodied the idea of breaking a manual operation into its component parts and assigning appropriate times to each. Early industrial engineers used elemental approaches, where component times were determined through observation of body parts used to perform a motion, the distance the body part had to travel, the manual control required, and the force (weight or resistance) involved (Meister, 1985).

Mundel (1985) reported that time standards derived from synthesized databases, assuming that errors in the data are random, yield more reliable time data due to the greater number of data points.

Three of the most prominent PTSs are Methods Time Measurement (MTM), Computerized Maynard Operation Sequence Technique (MOST), and the Modular Arrangement of Predetermined Time Standards (Modapts). Each of these systems will be discussed individually.

Methods Time Measurement. Of all the PTSs, the MTM is the most well-known (Zandin, 1980). MTM was developed and published in 1948 by Harold B. Maynard, G.J. Stegemerten, and J.L. Schwab. MTM is well established, very detailed in its procedural requirements, and generally considered the most accurate and accepted system to date (Zandin, 1980).

MTM is based on film-time studies of various industrial jobs. Use of the MTM system requires a trained analyst to describe the work of interest in terms of manual activities that correspond to basic elemental motions (Chaffin, 1987). To use the MTM system with accuracy and consistency, an analyst must attend a 24-80 hour training course (Chaffin, 1987; Zandin, 1980).

The time values for each elemental motion are given in one hundred-thousandths of an hour (0.00001 hour), referred to as one Time-Measurement Unit or TMU (1 TMU = .036 second). The primary assumption behind MTM is that times are determined for each basic element (motion) as a function of three variables: (1) type of terminal condition, (2) the load being transported, and (3) the distance of the motion. A detailed data card exists for each basic motion (reach, move, grasp, position, release, body, leg and foot motions, etc.) Once each basic motion and corresponding its variables have been identified, times are chosen from the appropriate data card. An example of the MTM data card for the motion "Move" is illustrated in Table 1.

TABLE 1
Partial MTM Card for "Move" (Mundel, 1985, p. 401)

Distance Moved Inches	Time TMU				Wt. Allowance			CASE AND DESCRIPTION
	A	B	C	Moving B	Wt.(lb.) Up to	Factor	Constant TMU	
3/4 or less	2.0	2.0	2.0	1.7	2.5	0	0	A Move object to other hand or against stop.
1	2.5	2.9	3.4	2.3				
2	3.6	4.6	5.2	2.9	7.5	1.06	2.2	
3	4.9	5.7	6.7	3.6				
4	6.1	6.9	8.0	9.2	12.5	1.11	3.9	
5	7.3	8.0	9.2	5.0	12.5	1.11	3.9	
6	8.1	8.9	10.3	5.7	17.5	1.17	5.6	
7	8.9	9.7	11.1	6.5	17.5	1.17	5.6	B Move object to approximate or indefinite location. ...
8	9.7	10.6	11.8	7.2	22.5	1.22	7.4	
9	10.5	11.5	12.7	7.9	22.5	1.22	7.4	
10	11.3	12.2	13.5	8.6	22.5	1.22	7.4	
...	...	...	...	...	...	...	...	

For the "Move" element, there are three terminal conditions or cases/descriptions of the motion (only two are shown in Table 1):

- Move the object to the other hand or against a stop.
- Move the object to an approximate or indefinite location.
- Move the object to an exact location.

The load that is moved is specified in pounds (lbs). The effect of the load is not as straightforward as the distance of the movement or the terminal condition. The load, depending on the number of hands used to lift it, is assigned a constant time delay and a proportional time delay.

Basic motion distances must be measured accurately and classified correctly. Because of the detail required in MTM analysis, applicator error can be a problem (Zandin, 1980).

The nine additional types of motions are listed below (note that each has a unique data card):

- Reach - a motion made with unloaded hands or fingers.
- Position - fine motions to align an object at the end of a larger motion.
- Release - either a distinct motion by the fingers or a release of an object without overt motions.
- Disengage - an involuntary "rebound" motion after two objects suddenly separate under exertion.
- Grasp - an overt motion to gain control of an object.

- Eye focus/travel - the time required for the eyes to move to and fixate on a specific object.
- Turn/apply pressure - a rotation of the hand about the long axis of the arm in order to manipulate controls, tools, or other objects.
- Body, leg/foot motion - transporting the body with given step values.
- Simultaneous motions - rules are provided for motions that concurrently occur, only the longest time is recorded.

Another underlying assumption of MTM is that the analyses reflect 100% performance levels. Also, MTM does not consider posture of the worker. For example, a biomechanically correct push is not any different from a push using incorrect posture.

Additional versions of the MTM, namely MTM-2 and MTM-3, were developed to reduce the applicator error and the time consumed in application of the method. Versions 2 and 3 average or group together certain basic motions and/or variables to decrease the amount of applicator error. Unfortunately, versions 2 and 3 were not as accurate as MTM-1 (Zandin, 1980).

Computerized Maynard Operation Sequence Technique. MOST was developed by one of the men who developed MTM. MOST was the result of efforts to simplify the task of the analyst, since the MTM procedure was complicated and arduous.

Unlike most PTSs that seek to set work standards based on the time to perform elemental motions, MOST is founded on the idea that work can be measured by the movement of objects. The theoretical underpinnings are related directly to the physics definition of work; that is, work equals force multiplied by distance. MOST assumes that the movements of objects follow certain patterns such as reach, grasp, move, and positioning the object (Zandin, 1980). These patterns are arranged in a sequence of activities that are followed in the movement of an object. Subsequently, the pattern sequences are modeled and used as a standard guide for assessing the movement of the object. Another assumption of MOST is that the motion contents of the activities are independent from one another.

MOST is built on the idea that there are two ways an object can be moved. Either the object is moved freely through space (e.g., a clean lift) or it is moved while in contact with another surface (e.g., pushed along the floor). Each movement type is modeled by a different sequence. The use of tools or items of equipment also is represented by a unique sequence model. While the sequence of elemental subactivities is invariant, the time associated with each subactivity changes. In short, MOST assumes that all manual work can be described by three basic modeling sequences:

- General Move - object is moved freely through space.
- Controlled Move - object is moved in contact with another surface.
- Tool/Equipment Use - the use of common hand tools.

Each of the three modeling sequences is composed of different subactivities, as listed in Table 2.

TABLE 2
MOST Sequence Models and Subactivities (Zandin, 1980, p. 6)

MANUAL HANDLING		
ACTIVITY	SEQUENCE MODEL	SUB ACTIVITIES
General Move	**ABGABPA**	A - Action B - Body Motion G - Gain Control P - Place
Controlled Move	**ABGMXIA**	M - Move Controlled X - Process Time I - Align
Tool/Equipment	**ABGABP ABPA**	F - Fasten L - Loosen C - Cut S - Surface Treat R - Record T - Think M - Measure

Each subactivity is assigned a time-related index number based on the "motion content" of that subactivity (Zandin, 1980). This process is referred to as parameter indexing. The appropriate index number is looked up on a data card that corresponds to the activity. The General Move data card is shown in Table 3.

Given the information in Tables 2 and 3, a fully indexed General Move sequence may look something like the following example:

$$A_6\ B_6\ G_1\ A_1\ B_0\ P_3\ A_0.$$

The time units in MOST are identical to those in MTM, where 1 TMU equals 0.00001 hour, or 0.036 second. Times are calculated by adding the index numbers and multiplying the sum by 10. Therefore the time to perform the General Move sequence above is (6 + 6 + 1 + 1 + 0 + 3 + 0) x 10 = 170 TMU, or 1 minute.

MOST times represent the time that a skilled person working at normal speed would require to perform a task. (This is the equivalent of the MTM 100% performance levels.) In the 1970s, MOST incorporated a new sequence model and index card to represent clerical movements.

TABLE 3
MOST General Move Data Card (Zandin, 1980, p. 19)

	ABGABPA	GENERAL	MOVE		
INDEX	A Action Distance	B Body Motion	G Gain Control	P Place	INDEX
0	≤ 2 in ≤ 5 cm			Hold Toss	0
1	Within Reach		Light Object Light Objects Simo	Lay Aside Loose Fit	1
3	1-2 Steps	Bend and Arise 50% OCC	Non Simo Heavy or Bulky Blind or Obstructed Disengage Interlock Collect	Adjustments Light Pressure Double	3
6	3-4 Steps	Bend and Arise		Care or Precision Heavy Pressure Blind or Obstructed Intermediate Moves	6
10	5-7 Steps	Sit or Stand			10
16	8-10 Steps	Through Door Climb On or Off			16

MOST was computerized so that a computer could perform the work measurement and relieve the analyst of paper work (Karger and Bayha, 1987). The core computer program is an interactive dialog comprised of operator inputs on element variables and the computer calculation of the relevant times. Computerized MOST contains certain modules that allow the analyst to calculate the process times for various operations (e.g., welding, machining, etc.). In addition, the computerized version provides simulation capabilities. Thus, the analyst has the ability to display different configurations for comparison (Karger and Bayha, 1987). The computerized version of MOST makes the application of the system more uniform and expedient as well as offering the convenience of a database to store the time information (Zandin, 1980).

MODular Arrangement of Predetermined Time Standards. Under the direction of G. C. Heyde, Modapts was developed in 1965 by an informal working group of 17 individuals representing 14 different companies (Karger and Bayha, 1987; Shinnick, 1987). Consistent with predetermined time systems, Modapts was introduced as a work measurement technique. The early version was known as Concise Modapts. In 1969, a clerical database was collected and it was called Office Modapts. Additional data from warehouse activities were compiled and used to create Transit Modapts in 1974. In 1981, these three Modapts databases were combined into Modapts PLUS. This current database embodies 44 elements of human body movements in addition to numerous activities.

One MOD is the amount of time required to perform one finger movement and has the value of 0.129 second. The value of one MOD emerged from time studies over a wide range of jobs at more than 20 different industrial organizations (Shinnick, 1987).

To use the system, every motion is assigned a two-part code. The first part of the code is an alphabetic character indicating the body member used in the motion. The second part of the code is a number used to calculate the time required to complete the motion by multiplying it by 0.129 second. For example, a hand move is encoded M2 and is assigned a time value of 0.258 second (2 x 0.129). The Modapts values for all movements are included in Appendix A.

In the basic coding algorithms, Modapts employs several rules for determining times of simultaneous movements and combined movements. Two categories of movement control (low and high) are recognized in Modapts. A low conscious control activity is an automatic response requiring little muscular control, no visual control, and no mental control. Low conscious control activities are considered to be free of hesitation because the activity is predetermined. In contrast, a high conscious control involves muscular control and either visual or other sensory modality feedback. Some movements represented in the Modapts system are diagrammed in Figure 1.

Modapts mean time data assume that movements are produced by a skilled operator performing a familiar task.

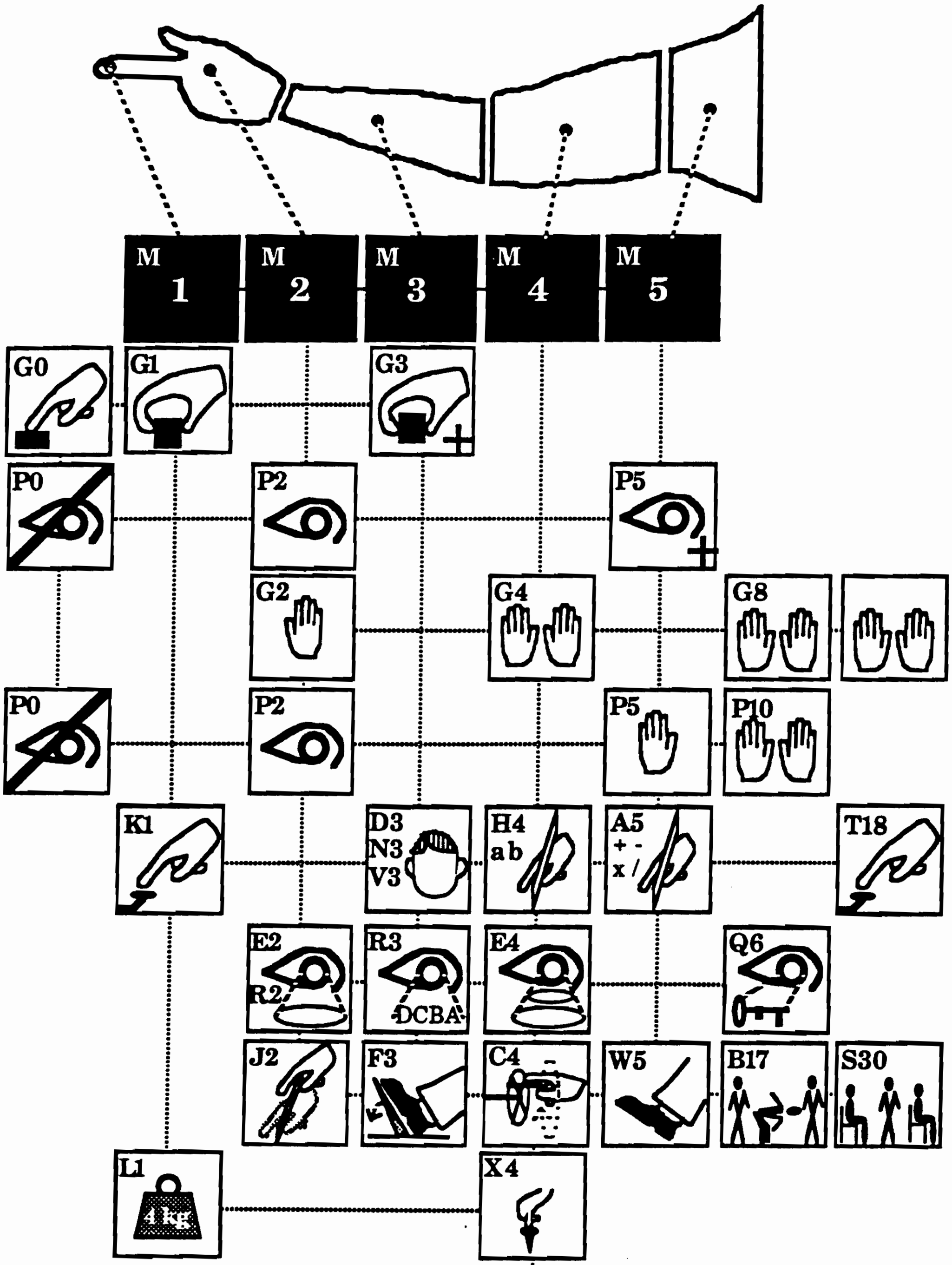


Figure 1. Diagram of selected Modapts movements (adapted from Shinnick, 1987).

Discussion. Comparing MTM, Computerized MOST, and Modapts, all three predetermined time systems are based on the idea of breaking a task into elemental motions or subactivities and they use multiples of a basic time unit to calculate total task times. For MTM and MOST, the time unit is a TMU, or 0.036 sec. Modapts uses the time unit entitled a MOD, or 0.129 sec.

There are certain theoretical issues that need to be considered when choosing a PTS. MTM, though a more precise and accurate system than MOST (Zandin, 1980), is far more detailed due to the assumption that an operator will perform a task *exactly and precisely* every time. This assumption of MTM is impractical given the variabilities of human performance that exist in the workplace. This is not to say, however, that MTM does not fulfill a need for data required on short, highly repetitive motions in small work spaces where it is important to minimize even small variances in operator activity. MOST, on the other hand, has a "built-in" balancing adjustment whereby a 2-minute padding is added into every sequence to offset variances due to individual differences as well as individual variability. Because of this balancing, MOST would be inappropriate to use when studying sequences of actions that are less than two minutes in duration.

MTM is an "exact" measurement system, in which it is necessary to record precise types of actions, distances, weights, and other variables. MOST is a less complex system which describes the *work* of a sequence of activities based on a physical definition of work (i.e., force x distance). Modapts, unlike MTM and MOST is based on the idea that distances do not need to be measured due to the fact that the body parts used for each motion reflect the distance (and, therefore, time).

Modapts, unlike MTM and MOST, recognizes that operators do not always perform at particular levels. While Modapts yields (like MTM and MOST) times for an average trained operator working at a typical pace, a simple adjustment to the MOD can be made subjectively to account for extremely hard (fast) work or extremely slow work. This adjustment is a constant for all actions, as the MOD value is simply increased or decreased.

Clearly, all three PTSs are meant to generalize over a range of different jobs. However, MTM emphasizes repetitive, cyclic actions, as evident from the precise measurements and number of variables. Computerized MOST focuses on specific materials handling type jobs (although a clerical database has been added). Modapts is the most general, PTS and has been used in a variety of applications including design for handicapped people (Shinnick, 1985).

Modapts was tested by Price Waterhouse in comparison to other predetermined time systems (MTM, MSD, and Work Factor) as well as to stopwatch time studies (Shinnick 1987). They found all systems to have comparable accuracy (Gerber and Heyde, 1985; Shinnick, 1985; Shinnick and Erwin, 1989). Computerized MOST was comparable to MTM for all body motions, with the exception of short, cyclic motions (Zandin, 1980). Modapts was approximately twice as fast to apply as MTM-1, and the application time of Computerized MOST falls somewhere in between these two systems. Shinnick (1985) also reports that Modapts was recommended by the International Labor Office as a "...valid and useful work measurement system" and that Modapts complies with MIL-STD 1567A specifications developed by the Department of Defense for work measurement programs of defense contractors. Given the apparent comparability of predetermined time values of various systems, Modapts is appealing because of its ease of application (Shinnick and Gerber, 1985).

Cognitive Workload Measures

With respect to the current research effort, one purpose is to investigate the feasibility of modeling cognitive workload as a variable in Micro Saint task networks. Laughery (1989) has examined task network modeling as a basis for analyzing operator workload. An extension to the Micro Saint simulation package for the analysis of operator workload was used by Drew, Laughery, Kramme, and Archer based on a technique done by McCracken and Aldrich in 1984 (Laughery, 1989). The theoretical basis for determining operator workload was the multiple resource theory proposed by Wickens, Sandry, and Vidulich in 1983.

The McCracken and Aldrich technique requires that for each node of the network model, the workload demand in each of four processing channels be assigned a value. The identified channels are auditory, visual, cognitive, and psychomotor output. A benchmark scale was created by McCracken and Aldrich for each of the channels to help the modeler decide a value that best characterized the subactivity. By using Micro Saint to simulate the task network and applying the workload values, the operator's attentional requirements are obtained.

Although valuable as a step toward analyzing workload, the McCracken and Aldrich extension poses several difficulties. For instance, their approach assumes several channels of workload and that workload can be summed across tasks. There is no consensus on what values constitute "excessive" workload and the model does not address the issue of task dumping—a strategy often performed by overloaded operators (Laughery, 1989). Because the research on this extension is ongoing, and not much has been published on the technique, it will not be addressed further at this time. Instead, the current research attempts to integrate an existing workload index into Micro Saint modeling.

The construct of cognitive workload represents the attentional effort or cost to the operator of performing at a certain level (Gopher and Braune, 1984; Hart, 1987). Evidence of the cognitive workload issues caused by the growing complexity of modern systems is found in the large number of reviews of cognitive workload measures (e.g.,Gartner and Murphy, 1976; Hart, 1987; Williges and Wierwille, 1979). There are three main categories of mental workload measures: psychophysiological measures, performance based measures, and subjective measures (Eggemeier and O'Donnell, 1982). Mutual agreement exists regarding the importance of mental workload, but there exists substantial controversy over the best type of workload measurement method (Nygren, 1991; Reinhart, Glynn, Dye, Takahama, and Snyder, 1988). Eggemeier and O'Donnell (1982) argue that all three forms of measures should be a part of any comprehensive cognitive workload study, but in practice this is not feasible. Each type of measure assesses unique workload components and has inherent advantages and disadvantages.

Psychophysiological measures. The various psychophysiological measures are founded on the premise that states of cognitive workload can be inferred from physiological conditions (Eggemeier and O'Donnell, 1982). Physiological measures assess the psychosomatic effects of stress (McCloy, 1987). The most commonly used psychophysiological measure is heart rate, or sinus arrhythmia (SA). Typically, heart rate variability decreases as the cognitive workload increases. Other common psychophysiological measurements of cognitive workload are the electromyogram (EMG), the galvanic skin response (GSR), measures of cardiovascular function (blood pressure, blood oxygen level, etc.), measures of brain activity (electroencephalogram), and measures of sensory functions (such as eye functions.)

The primary advantage of physiological measures is objectivity. Other advantages of this type of measure are minimal interference with the primary task and the use of concrete units. Disadvantages of physiological measures include extensive instrumentation, susceptibility to unrelated disturbances and individual differences, and non-specificity to a particular task. Non-specificity means that a person's cognitive workload status may be due to a number of different factors or combination of factors that are unrelated to the cognitive processes under study. Hartman (1980) warned that cognitive workload data must be interpreted with caution.

Performance measures. Performance measurements of cognitive workload assess impairments to performance levels (McCloy, 1987). Performance measures are grouped into broad classes based on the mission: weapon systems performance, primary task performance, secondary task performance, and laboratory task performance (Eggemeier and O'Donnell, 1982).

Performance measures provide immediate and direct responses that are objective. More importantly, performance measures reflect how well the objectives of a task were satisfied. Performance measures are limited in that they do not reflect the hypothesized association between cognitive workload and time pressure. Different performance decrements reflect different tasks, and some tasks demand greater attentional allocation by the operator than others. The handicap of performance measures in assessing cognitive workload is an insensitivity to levels below the operator's maximum capacity. While an operator may have to put forth considerable effort to accomplish the objective, the measure tells the researcher only that the cognitive workload level was not impossible.

Subjective measures. Subjective cognitive workload measures assess the conscious experience of the operator through self-report estimates of the operator's ability to cope with and achieve desired objectives. Nygren (1991) reported that subjective measures have become the most practical and popular cognitive workload assessment technique. Many researchers believe that subjective ratings are the most valid and sensitive way to tap cognitive workload (Johanssen, Moray, Pew, Rasmussen, Sanders, and Wickens, 1979).

Disadvantages of subjective workload measures stem from scaling problems and limitations in diagnostic and predictive value. Some subjective measurements do not have equal scale increments since the underlying psychometrics are expressed in psychological, verbal, or spatial terms. Ordinal data are the most common but are task specific and preclude comparisons across tasks.

Vidulich and Wickens (1986) reported that cognitive processes responsible for the subjective assessments may differ from the processes responsible for performance. Dissociations between subjective measures and performance are related to factors such as automaticity, presentation rate, and motivation (Vidulich and Wickens, 1986).

In total, the advantages of subjective measures far outweigh the disadvantages (Gopher and Braune, 1984). Although subjective measures have a high propensity for between-subject diversity, they are reliable. Subjective measures are an integral part of most aircrew workload assessments (Hart, 1987; Reid, Shingledecker, Eggemeier, and Nygren, 1981). For both practical and theoretical reasons, subjective measures offer valuable information.

Reid et al. (1981) reported that in such instances as operational flight tests, the subjective measure is the only feasible index of cognitive workload. It is not intrusive and does not require excessive instrumentation. Theoretically, the level of cognitive workload may not always be evident in performance (Hart, 1978; Hartman, 1980). Johannsen et al. (1979) pointed out that regardless of what performance and behavioral measures may indicate, if a person feels effortful and loaded then he is indeed effortful and loaded.

Subjective measures are easy to use, score, and analyze. Hart (1987) reported that the construct validity of subjective ratings has been demonstrated in tests against known cognitive workload levels, resulting in converging evidence as well as accumulated evidence.

Different types of subjective measurement techniques include: (1) unidimensional numerical ratings given with or without behavioral anchors, (2) multidimensional evaluations, (3) rank ordering of tasks with respect to cognitive workload, (4) task specific protocols, checklists, or questionnaires, and (5) stereo-tape recorder techniques (Long, 1974).

Some of the most popular subjective measures are Cooper-Harper Scales, the NASA-TLX (a multidimensional rating procedure that uses six subscales), and the Subjective Workload Assessment Technique (SWAT).

This present study will use the SWAT for several reasons. First, workload is a multidimensional construct and, therefore, a multidimensional measurement tool is believed to be most appropriate. Second, although the NASA-TLX technique may be more accurate than the SWAT (Nygren, 1991), SWAT requires less time and effort for the operators (Tan, 1990). Third, Cooper-Harper Scales reportedly are more sensitive to low and medium workloads, while the SWAT appears to be more sensitive to workload differences under heavy loadings (Reinhart et al., 1988.) Fourth, SWAT is particularly suited for empirical testing (Nygren, 1991).

Subjective Workload Assessment Technique. Reid, Shingledecker, Eggemeier, and Nygren (1981) and Reid, Shingledecker, and Eggemeier (1981) developed the Subjective Workload Assessment Technique (SWAT) expressly for aircrew applications. SWAT has been used in a variety of flight and simulated flight studies (Biferno, 1985; Bortolussi, Kantowitz, and Hart, 1986; Damos, 1984; Hart, Battiste, and Lester, 1984; Vidulich and Tsang, 1985).

Few researchers disagree that cognitive workload is multidimensional, but there is no consensus on the number of dimensions or levels. Driven by cost and other practical considerations, Reid et al. (1981) examined a three-dimensional regression approach based on work by Sheridan and Simpson (1979), who identified three major factors which jointly affect cognitive workload: information processing, mental effort, and psychological stress resulting from task demands. Reid and his colleagues referred to each factor as a dimension and labeled them time load, mental effort, and psychological stress, respectively. As the workload on each dimension increases, the overall cognitive workload also increases.

Time load refers to the total time that the operator is busy. A person can complete all of the cognitive work with some time to spare when time load is low. As time load increases, spare time decreases.

Mental effort refers to the amount of attention required for a task regardless of the number of subtasks or time limitation. The concentration or attention required for a task is minimal or automatic when the mental effort is low. As the mental demand increases, the degree of concentration or attention increases.

Psychological stress load refers to the contribution to total workload from any condition that produces anxiety, frustration, and confusion while performing a task or tasks. One feels relaxed at a low level of stress. As the stress load increases, confusion, anxiety, and frustration increase.

Each SWAT dimension has three levels, as described below:

Time Load

1. Often have spare time. Interruptions or overlap among activities occur infrequently or not at all.
2. Occasionally have spare time. Interruptions or overlap among activities occur frequently.
3. Almost never have spare time. Interruptions or overlap among activities are very frequent or occur all the time.

Mental Effort

1. Very little conscious effort or concentration is required. Activity is almost automatic, requiring little or no attention.
2. Moderate conscience mental effort or concentration is required. Complexity of activity is moderately high due to uncertainty, unpredictability, or unfamiliarity. Considerable attention is required.
3. Extensive mental effort and concentration are necessary. Very complex activity requiring total attention.

Stress Load

1. Little confusion, risk, frustration, or anxiety exists and can be easily accommodated.
2. Moderate stress due to confusion, frustration, or anxiety noticeably adds to workload. Significant compensation is required to maintain adequate performance.

3. High to very intense stress due to confusion, frustration, or anxiety. High to extreme determination and self-control required.

Multidimensional scaling techniques are a fundamental building block of SWAT. The mathematical procedures of conjoint analysis for SWAT can be found in Nygren (1982).

Before the interval scale is constructed for a group, axioms of SWAT are first tested to ensure that the group is homogeneous (Reid, Eggemeier, and Nygren, 1982). An individual can be evaluated against his or her own scale, the group scale, or against one of the six prototype scales that emerge from the accumulation of data.

The SWAT is a two-step technique. Each person completes a scale development phase and, subsequently, an event rating phase. During the scale development phase, necessary data are collected to develop a cognitive workload scale for the group (Reid, Shingledecker, and Eggemeier, 1981; Reid, Shingledecker, Eggemeier, and Nygren, 1981). Tests on the scales are performed to ensure that the assumptions of the SWAT were not violated. The second phase, event rating, is conducted after some specific task of interest or segment of time.

Scale development. The three dimensions of the SWAT each contain three levels. The 27 unique combinations are placed on separate cards. The cards are shuffled, stacked, and given to the person to rank order.

The rank orders are used in the conjoint measurement analysis. Nygren developed a computer program, SWAT1, that will perform the necessary analyses once the rank orders have been input (Nygren, 1982).

The first step in the analysis determines for each case whether a group or individual scale is necessary. This manipulation essentially tests the agreement of the ordering among people using Kendall's coefficient of concordance. A high coefficient (> 0.75) represents a high level of agreement among individuals involved in the test.

The conjoint program derives a workload scale with interval properties based on orderings of the scores. The program outputs a cognitive workload score of 0 to 100 for each of the 27 scale values.

Reliability tests revealed that the subjects agreed 87 percent of the time with their original sorts after two weeks, and 75 percent after three months (Reid et al., 1981).

Event scoring. The simplicity of the event rating phase has been recognized as the most advantageous characteristic of SWAT. After the task or segment of interest has been performed, the three dimensions (T, E, and S) are ranked by a person involved in the test as 1 (low), 2 (medium), or 3 (high). The combination of the three ratings is matched to the position of the corresponding card according to the scale developed in the first phase. The interval number assigned to that position is the index of cognitive workload associated with the task or segment. For example, say a person rated a task as a "two" in time load, a "three" in effort, and a "one" in stress. This "2-3-1" combination is located in the previously developed rank order. The number between 1 and 100 that this position has been assigned by the conjoint analysis then becomes the cognitive workload score for the task.

The AN/SLQ-32(V) DCC

The AN/SLQ-32(V) (referred to as the "Slick") is a naval electronic warfare equipment built by Raytheon Company, Electro-Magnetic Systems Division (Blake, 1991). The system was developed to provide Anti-Ship Missile Defense (ASMD). A concurrent human factors evaluation and redesign of the AN/SLQ-32(V) Display and Control Console (DCC) provided the mechanism for investigation of the present dissertation objectives. The AN/SLQ-32(V) is illustrated in Figure 2.

The DCC provides the communications and control interface between the operator and the electronic digital processing unit. The monochrome 12-inch diagonal screen displays the electronic emitter environment on a polar display along with all related menus for operating the system. Information on the functional status of the system is displayed by a series of indicator lights located below the screen.

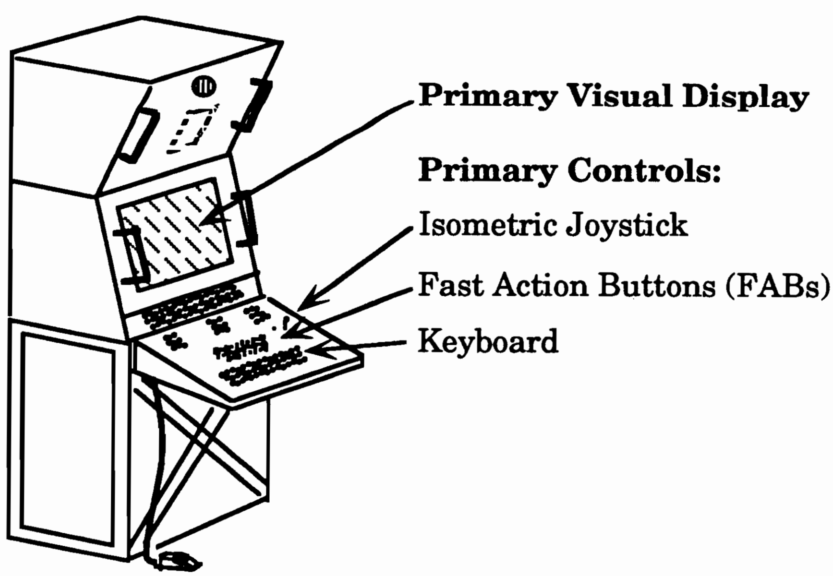


Figure 2. The AN/SLQ-32(V) display and control console.

The primary control panel is located below the screen and consists of a QWERTY keyboard control group, an isometric joystick, 24 Fast Action Buttons (FABs), an audio control group, and an Active Electronic Counter-Measures (AECM) control group. Ancillary control groups for battle shorts are located above the display screen along with the speaker for functional recognition (audio identification of emitter scan types). Chaff launcher controls are positioned above the keyboard. Electronic warfare (EWs) operators also attend to a 5-inch ULQ-16 visual monitor used for frequency spectrum analysis.

The AN/SLQ-32(V) requires continuous around-the-clock monitoring by at least one EW. In addition to surveillance of the radar screen, the EW is required to perform a multitude of tasks including various identity checks, categorization of threats and non-threats at various levels, maintenance of a library of encountered craft and missiles (airborne, surface, and subsurface), record keeping of hostile munitions capabilities, keeping the ship commanders and the commanders of other own-allegiance craft apprised of all threats, deployment of chaff, deployment of decoys, and management of anti-electronic countermeasures. Shipboard computers detect the electronic emissions of vessels and attempt classification, but the EW makes the definitive identifications. EWs assimilate information received by various computers with publications to formulate emitter identifications.

Operation of the AN/SLQ-32(V) has been defined in terms of a task analysis. Three representative tasks were selected and modeled in Micro Saint. These three models were checked for completeness and accuracy in terms of the sequence of steps by expert review of several operators. The tasks are tracking an emitter, identifying and classifying an unknown emitter, and performing a library update.

OBJECTIVES

There are two goals of this research effort. The first goal is to investigate the development of a method for assigning time parameters (means and standard deviations) to Micro Saint model networks for specific application in the visual displays and controls arena using Modapts. The second goal is to investigate the integration of a cognitive workload index into Micro Saint based on SWAT. The research was couched within the context of a model of one operational task on the AN/SLQ-32(V) DCC. The data were collected in conjunction with other on-going research intended to establish the best DCC design. The stated goals were accomplished by three objectives as follows:

- Definition of the operator's task using the AN/SLQ-32(V), development and validation of the task network model.
- Development of the Modapts parameters (times and distributions), standard deviations, and SWAT scores for the AN/SLQ-32(V) Macintosh prototype and subsequent refinement of the Micro Saint model.
- Assessment of the predictability of the Modapts and SWAT methodologies.

METHODS

Overview

The objectives were accomplished in three phases (Figure 3). The first phase consisted of a task analysis of the SLQ-32 operator's work, development of the Integrated Task (IT) for the AN/SLQ-32(V) and validation of the task network model using subject matter experts (EWs).

The second phase involved construction of the Macintosh prototypes of the SLQ-32, and included experimentation to collect mean times, standard deviations, and SWAT scores for both the baseline (or existing AN/SLQ-32(V) configuration) as well as seven alternate configurations for which various aspects of the DCC display had been redesigned. During the collection of these data, the necessary refinements to the Micro Saint model were made to reflect the performance of the participants using the Macintosh prototypes.

Phase 2 consisted of three distinct efforts: (1) describing the IT in procedural terms for the prototypes, generating the appropriate mean time values for the Micro Saint model using the Modapts system, examining the distribution of the time values, and identifying suitable standard deviations; (2) examining the SWAT scores to incorporate the workload measure into the models; and finally, (3) incorporating the times and workload parameters in the final Micro Saint model. The Modapts times were adjusted to fit the collected baseline data. The SWAT scores from all experimental trials were used to explore different factors which correlated to the workload scores for subsequent incorporation into the task network model.

The third phase consisted of running the model in simulations of 24 display configurations and verification of the model's fit to Group 1. Finally, the model was validated by assessing its predictive ability. The validation assessment was done by comparing performance and workload data obtained from the model to the same data collected from six additional participants (Group 2).

The methods of the study are outlined by steps in Figure 4. Methods for each step are presented in the order indicated in Figure 4.

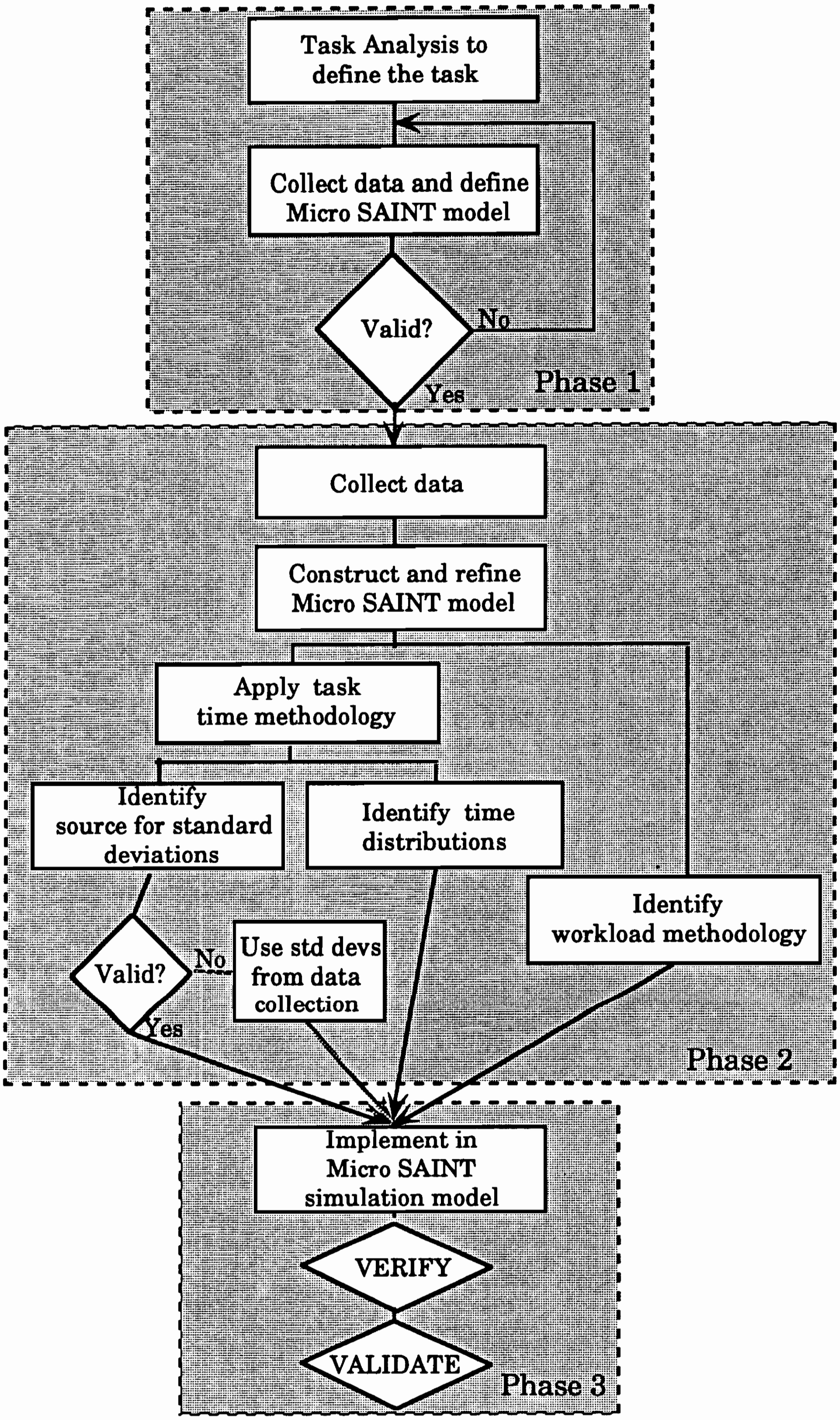


Figure 3. Diagram of methodology for the study.

METHODS		
Step/Objectives	Procedure	Analysis
Task Definition Task Selection & Validation	Task Analysis	Expert Review
Micro Saint Model Define Task Network	Human Experimentation, Video-Tape data extraction	Detail task network structure, specify likelihoods Multiple regressions (Errors, Error St Dev)
Model Times AssignTimes & Verify	Application of Modapts System	Multiple Regression
	Identify Mod Value	t-test on slope of Regression Equation
Model Distributions Assign Distributions of Times and Errors	Examine Distributions	Kolmogorov-Smirnov Tests
Standard Deviations Assign Standard Deviations	Research Modapts database, similar studies, and data obtained in Experiment	Opinion based on research findings
Workload Develop workload methodology	Subjective Workload Assessment Technique: Scaling Solution	Kendall's Coefficient of Concordance
	Workload Scores	Multiple Regressions
Model Verification	Comparison of Group 1 and Model	Simple Regressions: Performance & Workload
Model Validation	Comparison of Group 2 and Model	Simple Regressions: Performance & Workload

Figure 4. Breakdown of Methods into steps: Outline for Methods section.

Task Selection

As part of the AN/SLQ-32(V) redesign effort, a task analysis was performed. The AN/SLQ-32(V) operator is laden with tasks, as shown in the list of primary and secondary tasks outlined in Table 4 (Beaton, Dyess, Miller, and Moscovic, 1991).

Based on the analysis and scope of the project, one particular operational task was selected for the purposes of the redesign effort for the visual displays and controls console. In the IT, several functional subtasks are performed. The IT task was selected as the test-bed for display redesigns for several reasons. First, as indicated in Table 4, the primary responsibilities of the EW fall into three general activities as follows: (1) monitoring the electronic environment; (2) detecting a change in the electronic environment; and (3) evaluating emitters.

The IT task encompasses all three of these activities. Moreover, the IT task is the basis for all secondary tasks that the EW must perform. Thus, it is important for the EW to perform the task as efficiently as possible. The steps required for the integrated task are listed in Table 5.

Expert EWs at Norfolk Naval Base reviewed the basic task networks constructed for several of the steps detailed above for validation.

TABLE 4

Primary and Secondary AN/SLQ-32(V) Operator Tasks

TASKS	ASSIGNMENT
1. SLQ Operation and Emitter Management	
• initiation of system	Primary
• prepare electronic order of battle	Primary
• analyze available intelligence data	Primary
• surveillance	Primary
-detect emitters	Primary
-locate emitters	Primary
-tracking emitters	Primary
• emitter evaluation	Primary
-instantaneous id of high-threats	Primary
-auditory recognition of emitters	Primary
-match emitters to EOB and pubs	Primary
-manage emitter alerts	Primary
-designating emitter identifications	Primary
-correlating emitters to platforms	Primary
• emitter logging	Secondary
-written log	Secondary
-updating library	Secondary
2. ECM and AECM	
• employ ECM and AECM	Secondary
• missile counter-targeting	Secondary
• employ MK Decoy Launching System	Secondary
-initialize BDA	Secondary
-initiate/end decoy engagement	Secondary
-select decoy type, salvo size	Secondary
-select launcher and launcher tube	Secondary
-arm and fire each tube	Secondary
-verify launch	Secondary
-determine reseed interval	Secondary
-coordinate launcher reloading	Secondary
-conduct an IR engagement	Secondary
-conduct a distraction engagement	Secondary
-conduct a chaff engagement	Secondary
3. *Communication / Coordination*	
• coordinate with own force (& LAMPS)	Secondary
• provide early warning to own-ship	Secondary
• respond to CIC queries	Secondary
• coordinate with other CIC operators	Secondary
• brief watch relief (tactical environment)	Secondary

TABLE 5

Integrated Task (IT) Activities for the AN/SLQ-32(V) Operator

1	Operator hears an alert and new emitters appear.
2	Operator "hooks" (selects) emitter with his isometric joystick.
3	Operator assimilates the following emitter parameters: a. Close Control parameters (bearing, threat level, bias factor frequency, PRF, scan, scan type, assigned EFX). b. Auditory display of scan. c. Visual representation of scan (oscilloscope, or frequency spectrum analyzer–ULQ-16).
4	Operator searches through written publications and/or on-line libraries for an identification/confirmation.
5	Operator agrees or disagrees with the system designation and identification.
6a	Operator records the emitter into a log. Returns to step 2.
or	
6b	Operator searches again through library and/or publications for the correct designation.
7	Operator designates the ID; otherwise he adds the new emitter to the library.
8	Operator records the emitter into a log. Returns to step 2.

Micro Saint Model

An experimental study examining redesigned components of the DCC provided the data used to assess the fit of the Modapts time scores. This same study was used to collect workload scores for the different configurations tested, probability distributions, detailed procedures, and probabilities of different branches within the model. As mentioned in the Background section, the Modapts times provided the starting point for developing a time methodology. Those times were adjusted to fit the times collected. SWAT scores were collected at the end of each trial. A multiple regression analysis was used to explore different factors and weightings for the factors that correlate with overall workload level for each trial.

The following section describes the study that examined the redesigned components of the DCC. The data and research pertinent to the current efforts will not be the outcome of the design comparisons, per se, as those evaluations will be addressed elsewhere (Beaton, Dyess, Miller, and Moscovic, 1992). Instead, the present research involves the modeling of the task with the redesigned components and the *manipulations* of the time and workload data to fit collected data.

Apparatus and equipment. The test-bed for the prototypes of the AN/SLQ-32(V) display configurations was a Macintosh FX equipped with an extended keyboard and a mouse. The software for the interactive prototypes as developed in HyperCard and SuperCard (Beaton et al., 1992). Procedure and time data were collected with two standard video camera–recorders, accurate to 1/30 sec. The video tapes were dubbed with the Society of Motion Pictures and Television Engineers (SMPTE) standard time-stamp. One video camera recorded the screen displays, while the other monitored the operator using the prototype displays and controls console. Hooking error data were collected during the study by the computer. SWAT workload scores were elicited by the computer at the end of each test trial.

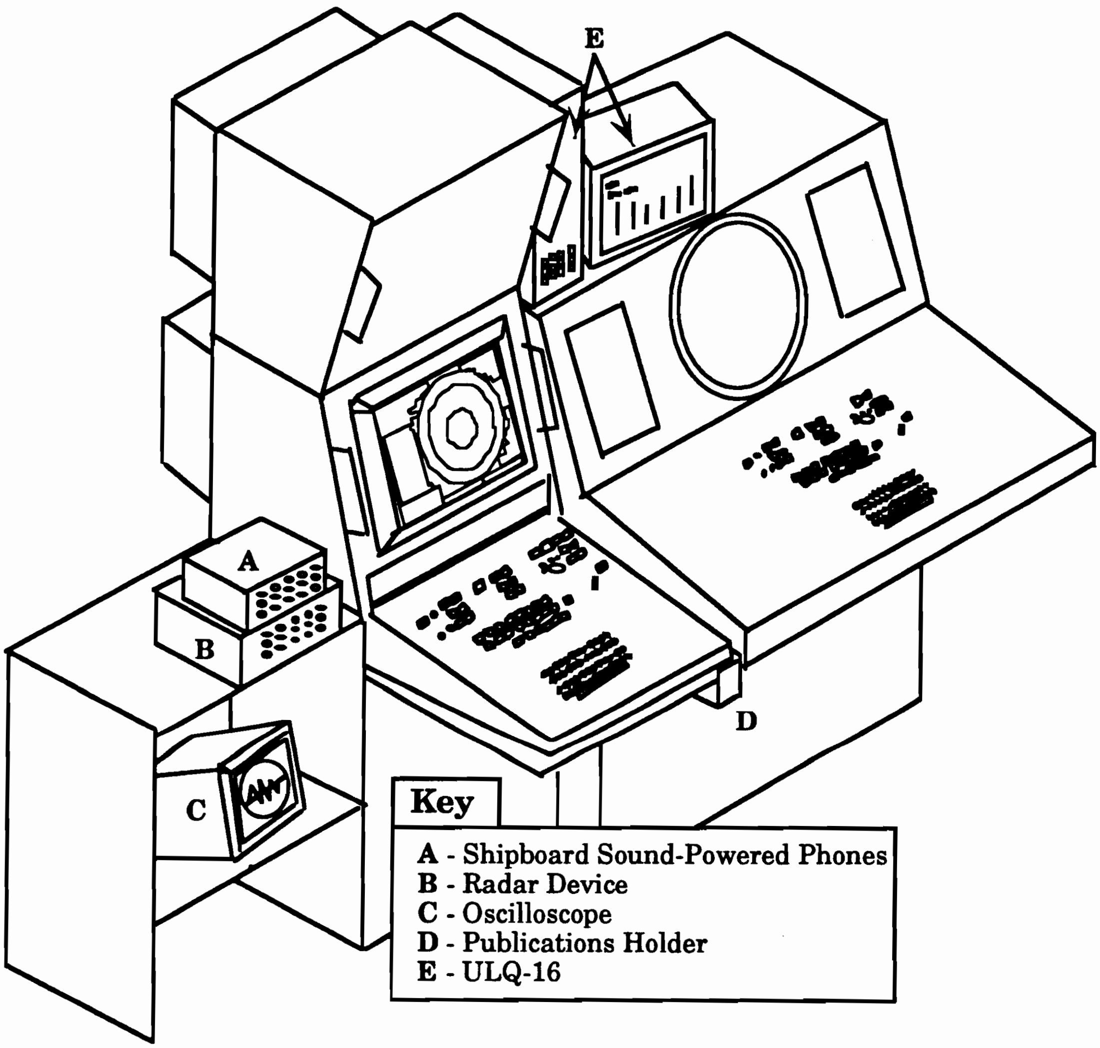
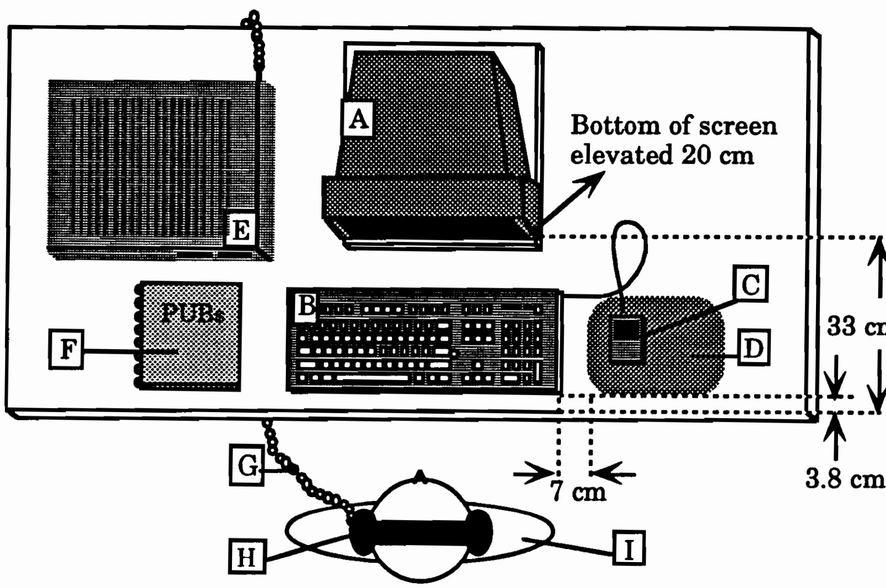


Figure 5. A typical AN/SLQ-32(V) layout onboard ship.

Configurations of the AN/SLQ-32(V) and auxiliary equipment vary from ship to ship. A common layout is depicted in Figure 5. The Naval Tactical Data System (NTDS) console is located to the right of the AN/SLQ-32(V). The ULQ-16 and oscilloscope are optional equipment. AN/SLQ-32(V) operators try to obtain a ULQ-16 and oscilloscope from other ship stores. This experiment assumed that the ULQ-16 is available to the operator.

There were two data collection stations set up with prototypes of the AN/SLQ-32(V). An illustration of a data collection station is shown in Figure 6.



A CRT Monitor	**D** Mouse Pad	**G** Headset Cord
B Extended Keyboard	**E** Macintosh FX	**H** Headset
C Mouse	**F** Publications	**I** Operator

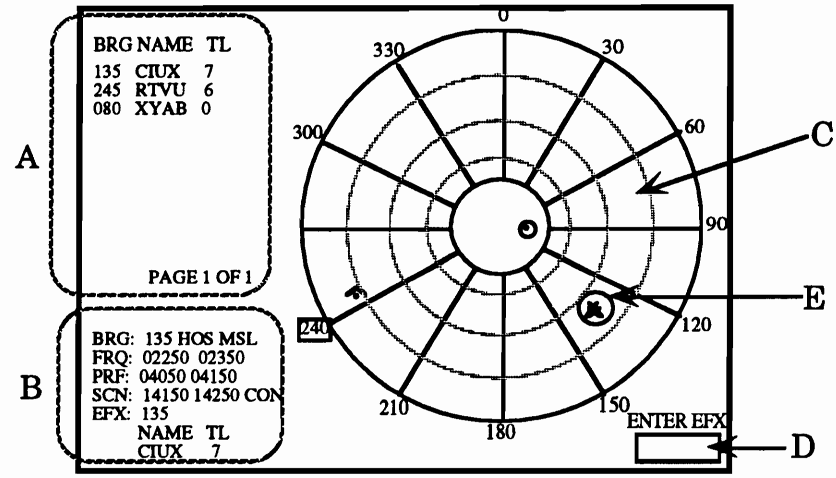
Figure 6. Layout of the experimental prototype AN/SLQ-32(V) station.

The training software (Beaton et al., 1992) was created using SuperCard, a HyperCard extension. An Apple Color 13-inch 8-bit color monitor (640 x 480 addressability), System 7.1, and 20 megabytes of RAM were used. The required Fast Action Buttons (FABs) and keyboard were simulated with a Macintosh Extended keyboard (the second highest keyboard repeat setting and the second fastest delay setting were used), and the Apple Mouse (using the second highest gain setting). Participants were permitted to have the mouse configured for either left- or right-handed use; however, all participants were right-handed.

An illustration of the screen display is given in Figure 7. During the data collection sessions, operators were prevented from switching to the operating system by an inadvertent mouse activation by AutoLock software. A ball cursor similar to that found on the AN/SLQ-32(V) (a circle with diameter approximately equal to 2 mm) was used.

The participants wore Realistic Nova 40 headsets that allowed the operator to listen to the scan types of emitters.

One of the workstations was in a small room with gray walls. The second station was set up in a larger room, but was made to resemble the first by enclosing the station with two black plastic curtains as the back and side wall. Black cloth was hung over the walls to darken the room. Both rooms were dimly lit with blue-filtered overhead fluorescent lights. The ambient illuminance was matched between the rooms (approximately 46 lux). The screen luminance also was matched between the stations (approximately 21.8 cd/m^2).



DISPLAY ITEM		FUNCTION
A	Threat Summary List	Lists all detected emitters
B	Close Control Information	Shows parameters for hooked emitter
C	Polar (or Range) Emitter Display	Primary display; shows location of emitters in relation to the ship
D	Designate Identification Box	Displays typed EFX numbers
E	Close Control Indicator	The small ball tab cursor enlarges and surrounds a selected emitter

Figure 7. The screen display of the Prototype AN/SLQ-32(V).

A library of 427 emitters was created and compiled into the Publications or "pubs." The pubs consist of two sections. The main body of the pubs contains a listing of the emitters sorted by frequency (FRQ). The rest of the information for each emitter is listed in columns of threat level (TL), bias factor (BIAS), category (CAT), pulse repetition frequency (PRF), scan, and cross-reference number (X-REF). After the operator matches FRQ, TL, BIAS, CAT, PRF, and scan, the corresponding X-REF number is used to look up the Emitter File Index (EFX) number from one of four pages in the front of the pubs. In this section of the pubs, all X-REF numbers are listed in ascending order. There are two identical X-REF numbers, followed by name of emitter, scan type, and EFX. The correct EFX is obtained by looking up the correct X-REF number, choosing the correct scan type from the two identical X-REF's , and selecting the corresponding EFX. An EFX is a three-digit number that uniquely identifies every known radar-emitting vehicle.

The extended keyboard in this experiment was coded so that the operator could use the number keys in the QWERTY section just as EW's use the AN/SLQ-32(V) FABs. Instead of having FABs to (1) designate the identification of the emitter and (2) listen to the emitters' scan types, the F5 and F8 special function keys were used, respectively. The F5 key was covered with a key cap labeled "Desig ID". The F8 key was covered with a key cap labeled "Sig Sel" for Signal Select.

Instead of using a separate display to simulate the ULQ-16 frequency spectrum analyzer, the operator was presented with the ULQ-16 information in the center of the Apple monitor when the Sig Sel key was pressed. The operator turned off this ULQ-16 by pressing the Shift key.

The task. As explained in *Task Selection,* the main function of the participant (operator) in this study is to perform the Integrated Task (IT). In the experiment, this task was performed in the following sequence:

1. Operator hears an alert and new emitters appear.
2. Operator selects an emitter with the input device.

3. Operators uses close control parameters, the audio scan of the emitter, and the information presented on the ULQ-16 to identify the emitter.
4. Operator searches through publications for a cross-reference (number that agrees with the Frequency and PRF.
5. Operator searches through the cross-reference index to find the cross-reference number. Based on the type of the sound the operator finds the correct EFX number to identify the emitter file index number (EFX).
6. The operator designates that emitter with the correct EFX number by selecting the DESIG ID button, entering the three-digit number, and pressing the RETURN key.
7. The operator processes the next emitter on the threat summary list (Step 2).

Independent variables. Four factors were manipulated in the experiment. The factors, identified as priority considerations for redesigning the AN/SLQ-32(V) display (Beaton et al., 1991), are described below.

Format (polar, range). The primary information on the AN/SLQ-32(V) screen is the Polar display. The Polar display consists of a circle separated into three concentric rings. Each area formed by the rings holds specific emitters. The innermost ring contains all friendly emitters, the middle ring contains all hostile missiles, and the outermost ring contains all others (e.g., hostile aircraft, hostile submarines, and unknowns). The Polar display aligns detected emitters along bearing lines; however, due to the position-coding scheme used in this format, there is no range information. The Polar display is illustrated in Figure 8.

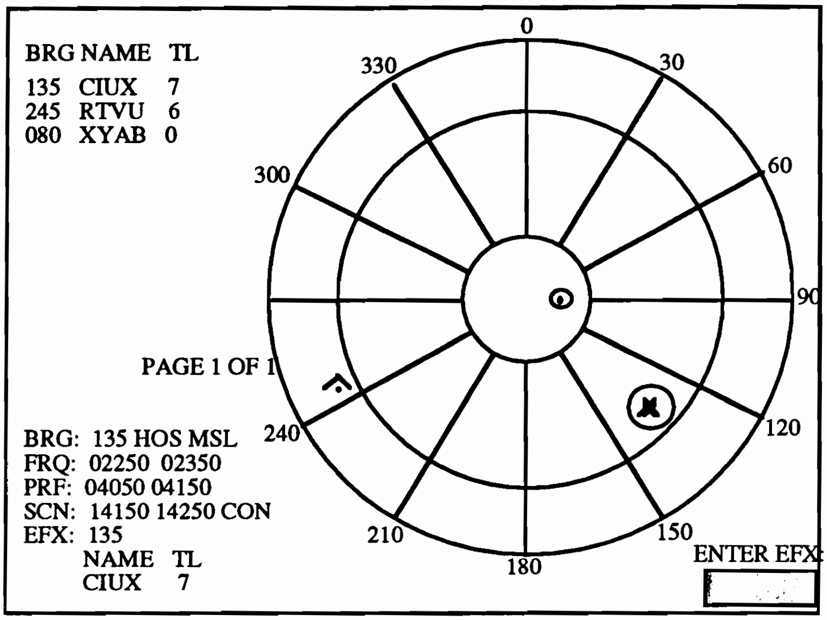


Figure 8. Screen display in the polar format.

The GeoSit display resembles the true-position display used in radar systems. It presents emitters in relation to the ship in terms of both bearing and range. There are concentric rings, but the areas indicated by the rings indicate different distances from the ship to the emitter. While it is known that for some tasks the EW must have range information, he currently obtains that information from other systems in the ship command information center. An example of the Range display format is shown in Figure 9.

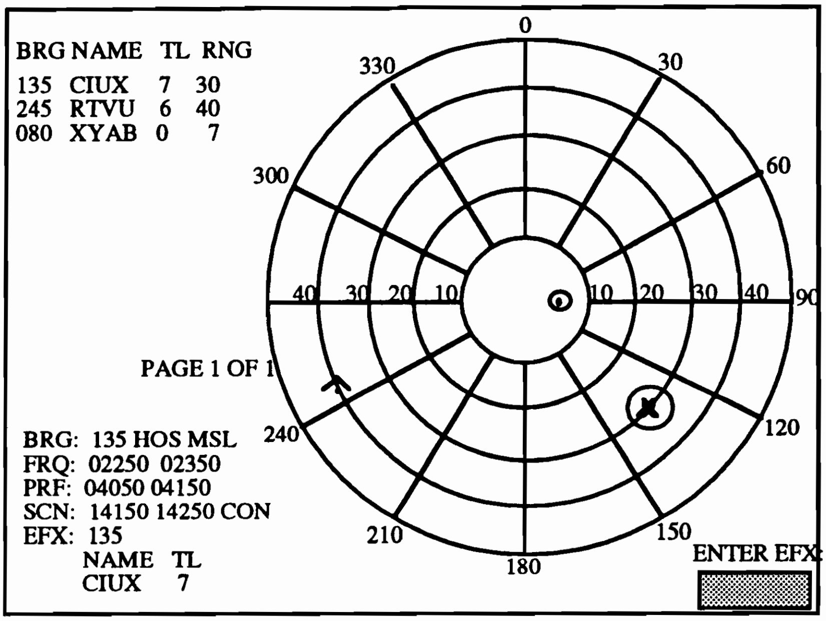


Figure 9. Screen display in the range format.

The original intent of the Polar display was to use location-coding so that the missiles would be salient and, therefore, quickly processed by the EW. However, the effects of using a Range display were examined to accommodate three needs: (1) the EW needs to have distance information for a number of tasks, (2) the EW needs to build and maintain a cognitive picture of the tactical environment, and (3) a Range display would provide consistency with other system displays aboard ship.

Color condition (black and white, color). The current AN/SLQ-32(V) screen display is a green phosphor with two levels of intensity coding – normal and "double-bright." The Macintosh prototypes simulated the existing display by using a black-and-white display.

Due to the enormous amount of information presented to the EW, the use of color has the potential to effectively manage some of the information processing required by the EW. The color-coded display consists of five different colors. Color was used to code emitter symbols and corresponding text for the purposes of distinguishing emitters of different threat levels (e.g., all hostile missiles were coded red, hostile non-missiles were amber, unknown were yellow, and friendlies were green). The Polar/Range lines were white.

Symbol set (old, new). The current symbol set consists of eight geometric symbols that represent emitter categories, as shown in Figure 10. The revised symbol set was based on newly developed NATO symbology (iconic symbols) and is shown in Figure 11.

Density (Caribbean, Gulf, Armageddon). The redesign configurations were tested in different emitter density situations. Three levels of emitter density were used: Caribbean, Gulf, and Armageddon. The number of emitters for each of the densities varied in the number of emitters present at the start of the trial (initial density) and the amount of time before new emitters appeared (four new emitters appeared at each interval). The characteristics for the different density levels are presented in Table 6.

OLD SYMBOL SET
(GEOMETRIC SYMBOLS)

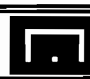
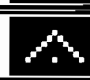
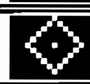
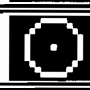
BIAS FACTOR	SYMBOL	THREAT LEVEL
MISSILE		7
UNKNOWN UNKNOWN		3
HOSTILE AIRCRAFT		6
HOSTILE SURFACE/LAND		5/4
HOSTILE SUBMARINE		5
FRIENDLY AIRCRAFT		0
FRIENDLY SURFACE/LAND		0
FRIENDLY SUBMARINE		0

Figure 10. Old (Geometric) symbol set and associated threat levels.

NEW SYMBOL SET
(ICONIC SYMBOLS)

BIAS FACTOR*	**SYMBOL**	**THREAT LEVEL (Hostile, Friendly)**
MISSILE		7,n/a
UNKNOWN UNKNOWN		3
AIRCRAFT		6,0
SURFACE SHIP		5,0
SUBMARINE		5,0
LAND-BASED		4,0

* Bias (Hostile, Friendly) was indicated by threat level and/or color, depending on the configuration.

Figure 11. New (Iconic) symbol set and associated threat levels.

TABLE 6
Emitter Density Levels

Level	Initial Density	Time between new emitters
Caribbean	10	70 seconds
Gulf	40	40 seconds
Armageddon	70	10 seconds

Dependent variables. There were two levels of dependent variables in this study. The first level was a macro-level for the processing of the emitter (IT). The macro-level dependent variables were men performance and workload scores. These data were available immediately from the computer–recorded times, errors, and workload scores obtained from the participants' responses. The second level was a micro-level of variables used to supplement the task network models. These micro-level variables were obtained from video-tapes. The micro-variables included subtask completion times and standard deviations for each node of the task network model. Each of the dependent variables is described below.

Performance. Performance of the participant was defined as a unitless metric encompassing both mean time to process an emitter and the mean number of hook errors. The performance score for completing each test trial was determined by the following function:

$$\text{Performance} = \frac{1}{b_1 \mathrm{T} + b_2 \mathrm{E}}, \tag{1}$$

where:

T = normalized mean time to process an emitter, in seconds,

E = normalized mean number of hooking errors, and

$b_1 = b_2 = 0.5$ (weighting coefficients for T and E, respectively).

The rationale behind the use of this performance measure is that EW performance is related inversely to the mean time to process an emitter and also is related inversely to the mean number of hooking errors. While it is desirable for the operator to work as quickly as possible, it is true that the fewer errors that are made, the fewer number of re-attempts must be made. Therefore, fewer errors are integral to less time and increased performance. Equal beta weights (e.g., b_1 and b_2) were given to time and errors.

Workload scores. A SWAT score was obtained for each configuration and density level. At the end of each trial, the participant gave a time load rating, then a mental effort load rating, and finally, a psychological stress load rating. The three ratings were used to derive a workload score using the rating scale developed on Day 1.

Subtask completion times and standard deviations. The times to complete the subtasks of the IT were extracted from the video tape recordings using frame counts. For each subactivity, 120 data points were collected.

Procedure. The participants were required to attend a one hour session prior to the study for general instructions, signing the informed consent form (an example is included in Appendix B), and testing for color vision deficiencies using Dvorine Color Plates. Participants also completed the SWAT rating scale development (card sort), and finally were scheduled for eight consecutive days to participate in the study. Six participants were chosen randomly and assigned to Group 1, the remaining six participants were assigned to Group 2. The instructions for the SWAT are included in Appendix C. Video tape data were collected on the first four participants of Group 1.

For each of the subsequent eight experimental sessions, a different design configuration was presented. A design configuration was a combination of the format (polar, range), symbol set (old, new), and color condition (black-and-white, color).

To bring the performance of the participant to a stabilized level for each configuration, a series of training blocks was given before the test trials. A sequence diagram of the events performed by participants in the experiment is shown in Figure 12.

Each day (Days 2-9) is referred to as a single session. For each of the eight sessions, a different design configuration was presented. The order for the eight different display configurations was randomized across participants. During a session, participants were allowed to take breaks at any juncture between training sets. The following sections describe the experimental sessions.

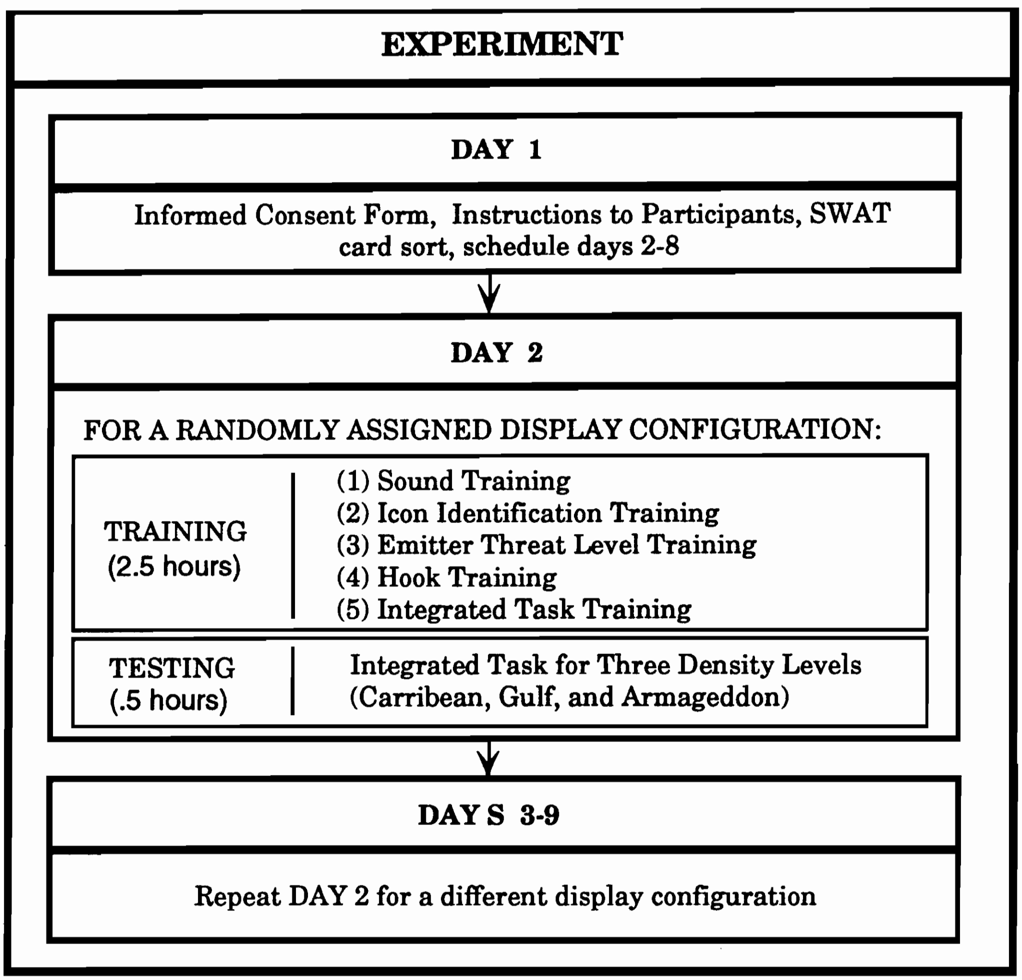


Figure 12. Sequence of events for participants.

Training. There were five blocks of training that each participant was required to complete per session. The training ensured that the participant attained a level of performance that was stabilized at his or her best under the task conditions. Intensive training was necessary to collect data that would capture the effects on "expert" operator performance associated with the different display configurations. Also, one of the assumptions of using the Modapts time method is that the operator is skilled and performing without hesitation. The five blocks of training are described below.

(1) Sounds. Participants were trained to recognize four sounds (scan types) and the associated waveform graphics that depict the frequency spectrum of the signal. The four sounds are Conical (CON), Standard (STD), Circular (CIR), and Sectional (SEC). These four sounds represent the signals generated by the emitter. Training included the simultaneous presentation of a sound and its corresponding graphic representation in a window on the computer screen. Each sound and graphic was presented by the experimenter and their attributes were explained. The instructions for the sound training are included in Appendix D.

An auditory probe, with the associated visual graphic on the computer screen, then was presented to the participant. Above the graphic was a set of buttons, each labeled with one of the scan types. The participant was instructed to use the mouse to select the button that represented the correct scan. Figure 13 gives an example in which the sound being presented is "STD" and the operator is about to select the correct choice.

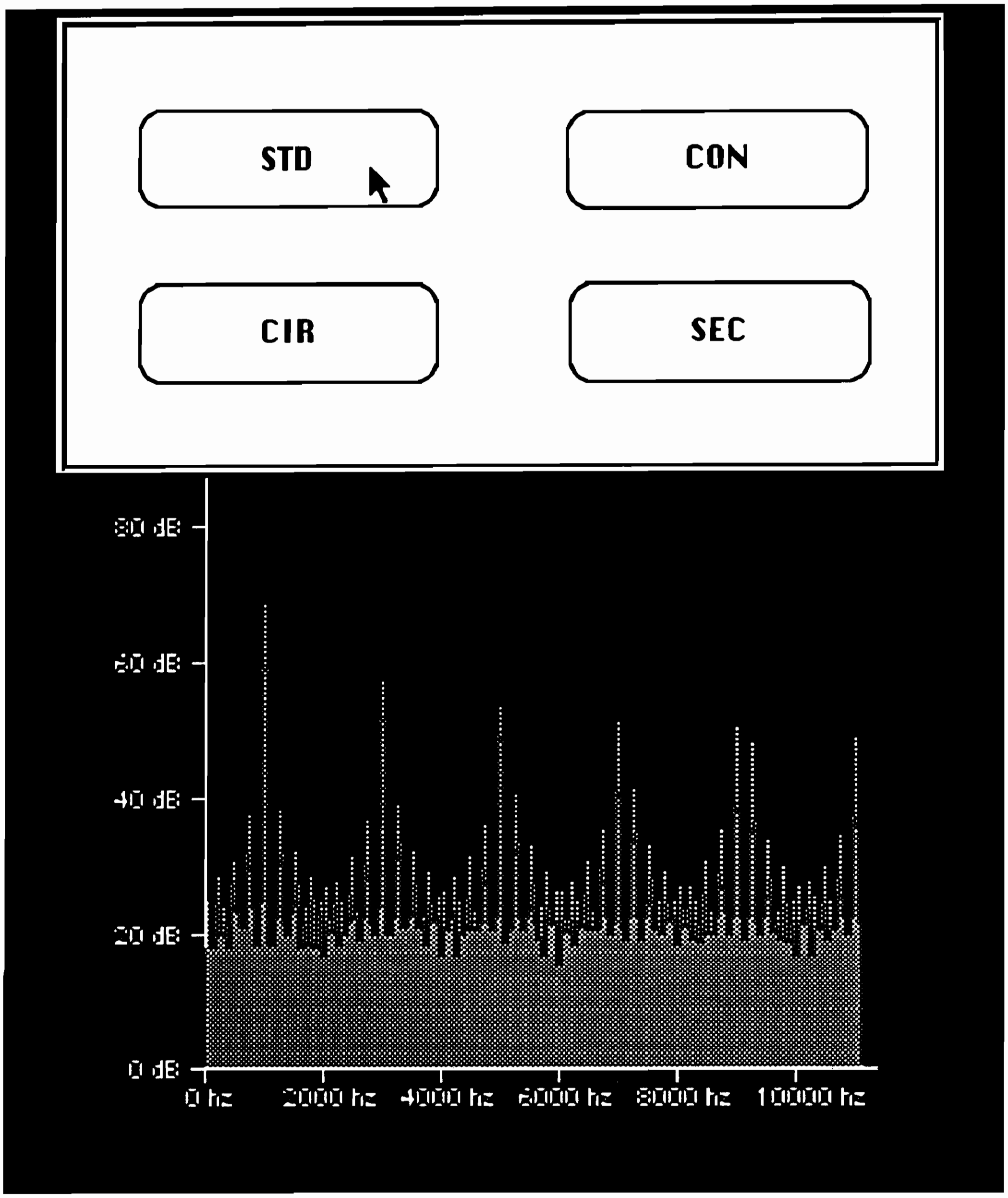


Figure 13. Screen format during sound training.

Participants were required to score at least 92 percent correct on the last set in order to continue to the next block of training. The 92 percent criterion was chosen *a priori*, allowing the participant to make a mouse selection error on no more than 2 of the 25 probes, as errors occurred due to the speeded nature of the training and not as a function of the operator's identification ability. Participants who failed to meet this criterion repeated the entire block. On Days 3-9, the participants were given only one set of 25 probes, with the same criterion.

(2) Icon Identification Training. The training for icon identification was identical to the training for sounds, with the exception that the icons set varied for each day. The instructions for icon training are included in Appendix D. Based on the particular configuration tested during the session, the participant was given a card illustrating the symbols (icon set and color condition). Participants were allowed to study the cards as long as they desired. Subsequently, five sets of 25 probes were presented on the screen. Again, the participants selected the correct identification using labeled buttons above the stimulus probe. An example of a screen display for the icon training is shown in Figure 14. The number of correct identifications was displayed at the end of each set.

Participants were required to score 92 percent correct on the fifth set in order to proceed to the next block of training. The 92 percent criterion was chosen *a priori*, allowing the participant to make a mouse selection error on no more than two of the 25 probes, as errors occurred due to the speeded nature of the training, and not as a function of operator's identification ability.

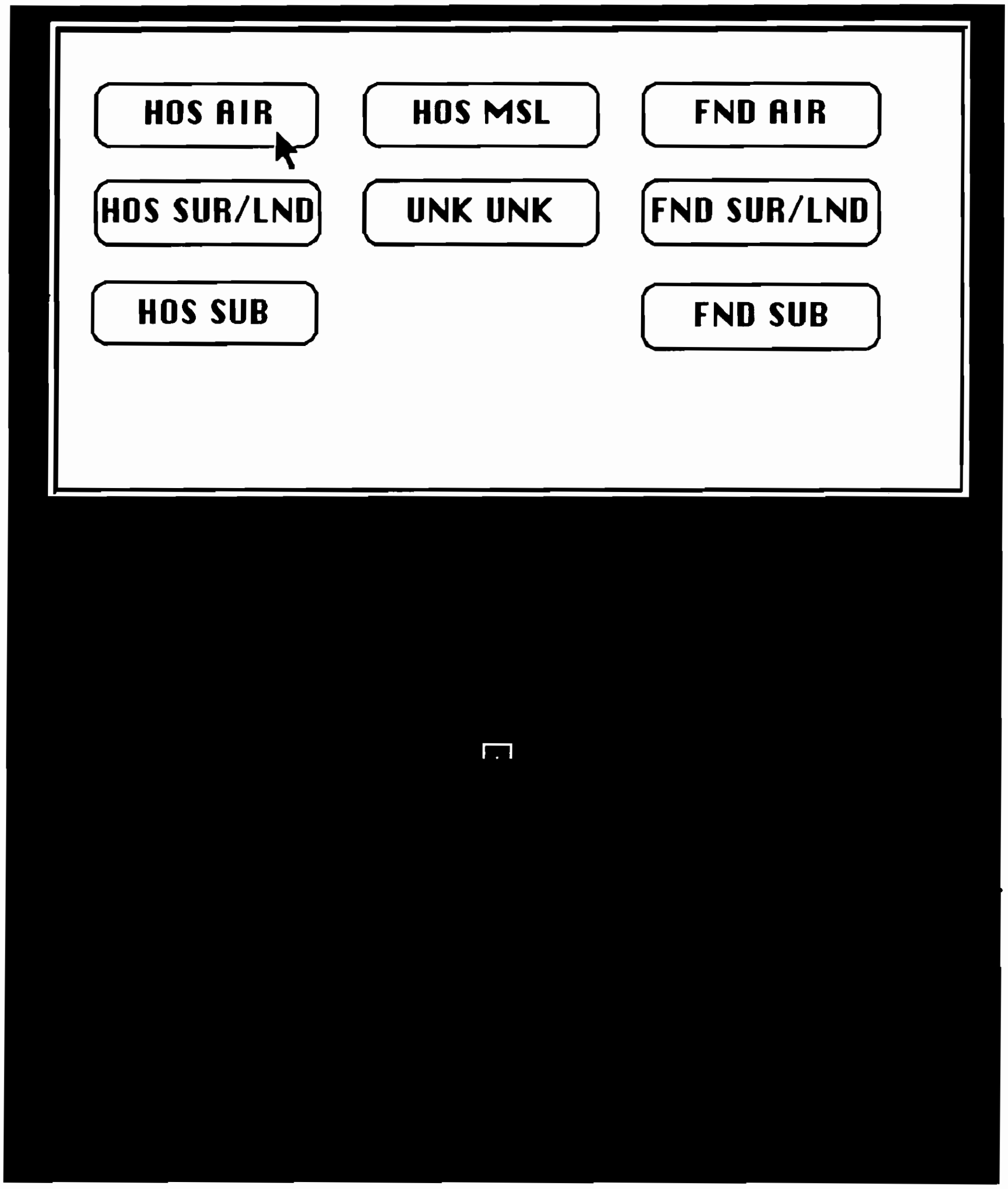


Figure 14. Example of screen format for icon identification training using the black and white geometric symbol set.

(3) Threat Level Training. The training for the threat levels was identical to the training for icon identification. Participants were given cards to study that contained the threat level for each icon. The buttons were labeled with threat levels (0-7) instead of icon names. The instructions given during threat level training are included in Appendix D.

Again, participants were required to score 92 percent correct on the fifth and final set to proceed to the next block of training. The 92 percent criterion was chosen *a priori*, allowing the participant to make a mouse selection error on no more than 2 of the 25 probes, as errors occurred due to the speeded nature of the training, and not as a function of operator's identification ability.

(4) Hook Training. During hook training, the participants were presented with the primary display (polar or range) and a page of emitters on the threat summary list. (Each page holds 15 emitters.) The beginning of each trial was marked with an auditory tone. As soon as the tone sounded, the participants were required to read the first emitter on the threat summary list and *as quickly and as accurately as possible*, select that emitter on the Polar/Range display using the mouse. When the correct emitter was selected (e.g., hooked), the participant continued to the next emitter. The set was complete when all 15 emitters were hooked. At the end of each set, the mean time to hook an emitter was displayed. The experimenter constantly urged the participant to perform as quickly and as accurately as possible. Fifty sets were completed per session. (Analysis of pretest subject data revealed that subjects reached a level of asymptotic performance at around 40 sets.) During the final five sets of the block, the

participant was required to reach a speed of hooking that did not exceed six seconds. The criterion of six seconds was established by approximation based on the rates of three pilot-participants and on knowledge obtained by experimenter use of the system. A failure to meet this level resulted in dismissing the participant from the study. Appendix D details the instructions for the hook training.

(5) Integrated Task Training. Finally, the participants practiced performing the entire IT on as many emitters as possible in a five-minute trial. There were 10 trails on Day 2 (50 minutes total) and 8 trials on Days 3-9 (40 minutes total). During the IT training, the emitter density level was set at 55 emitters – the 15 real emitters to be hooked and 40 dummy emitters. At the end of the trial, the mean times to process an emitter and the number of errors were given as a means of feedback and motivation. In addition, while performing the IT, the operator was given an audio "chime" after each correct designation.

Participants had to process at least eight emitters during the final trial. The criterion of eight emitters was based on pilot participants and the operator's knowledge obtained by experience using the system. If the trial data showed excessive dummy error rates during the last trial, the data from trial seven were used to determine if the participant met the criteria. In other words, a high number of dummy errors indicates that there were many overlapping emitters in the scenario. In the event that this occurs, the score at the end of the scenario is not a true reflection of the operator's ability. In case the scenario caused a large number of errors, then trial seven was used for determining if the participant performed at the criterion level. Between trials, the experimenter encouraged the operator to work as quick and as accurate as possible.

Testing. There were three test trials at the end of each session. Except for two differences, the test trials were identical to the trails during the IT training. First, the test trials used one of the three density levels, in randomized order, described under the Independent variables section. Second, each of the three test trials was followed by a computer elicited SWAT rating instead of performance feedback.

Model Times

The Modapts system was used to generate times suitable for the actions embodied by the Micro Saint model. For activities that were not covered by the Modapts system (e.g., using the mouse, searching the display, etc.) or for activities that the video tape methodology was not sensitive to (e.g., eye travel times), alternate sources and/or other assumptions were made.

The principles/rules and values for the Modapts elements that were relevant when applying the system to the model will be described below. These descriptions were taken directly for the *Modapts Plus Resource Manual* (Shinnick, 1987).

Principles/Rules. Several of the appropriate Modapts elements for the IT were values that originated in the Modapts Clerical Database. The clerical elements are more detailed than the elements for the manufacturing database. The Modapts authors developed certain rules to make the application of the database simple and rational. Fortunately, the rules were consistent with making a meaningful simulation model of a displays and controls application. These rules and principles are discussed below.

Rule #1: The Predictability Rule. The creators of Modapts tried to make each element as large as possible because the larger the element, the more quickly the Modapts methodology could be applied. The intent was to include only chains of events that were reasonably predictable.

Rule#2: The "Use Area" Rule. All the activities for the clerical database occur within an area that requires no more than a movement of M4 to reach.

Rule #3: The Reasonable Capability Rule. The general application of the clerical database assumes that the time to complete a physical task does not vary significantly for different people. It is implicit that the operator does not have physical disabilities and that all operators have comparable motivation levels.

In contrast, the time to perform mental work varies for reading, writing, and mathematical functions. Modapts assumes a "reasonable capacity" for mental operations.

Modapts lists three different controlling factors, as follows:

(1) Muscular control (muscular precision required),

(2) Visual control (eye action required), and

(3) Mental control (conscious decisions).

Modapts recognizes two levels of control. In low conscious control, the operator requires little control in terms of the controlling factors and hesitations do not occur. In high conscious control, some muscular control is required and assisted by sensual feedback.

In the Modapts system, a person is not capable of performing two high-conscious control activities simultaneously without increasing the time to perform the activity. To apply the system properly, times are awarded as simultaneous activities or activities occurring in series, depending on the level of control.

Simultaneous motions – motions performed at the same time by different body members – may or may not be similar. In the Modapts system, when simultaneous motions occur, the time awarded is the time for the most difficult activity with some adjustment to account for both activities. If both motions are low conscious control activities, then the time is awarded for one of the activities. On the other hand, if both activities are high conscious control, then time is awarded to accomplish both individual motions. Modapts assumes that the high control activities require concentration and feedback, which the operator must process individually.

In Modapts, move elements (M) are not actions by themselves. The Move is usually joined by a Get or a Put. Long moves which require the body to bend slightly increases the distance of the move but not the time awarded to make the move. Body assistance is considered a simultaneous movement.

The award of the element and associated Mod value is always the lowest possible class of the motion. That is, although the body member used was the elbow, if the movement only required the wrist to move, then the value of the wrist movement is awarded.

Elements / Values. The specific elements that were relevant to the IT task are briefly described below.

Decide. The decide element deals only with binary decisions.

- A D3 is awarded for the unusual, not the usual.

Read. Read elements are used for instances when the operator reads to either get the overall message or when operator must actually register the word and its meaning.

- R2 : When the operator must read one word in a group of words and the purpose is to get the overall message. The words are familiar and the reading is done silently.
- R3: Awarded when the operator reads one word in a group of words where each word is registered. For example, an R3 would be awarded per word when proofreading or verifying.
- R3: One word is awarded for up to three digits. For example, the digit 7 required one R3 to read, as do the numbers 76 and 765. However, the digits 7654, 76543, and 765432 must receive two R3s to read.

Eye control. An eye control element is awarded when an activity is dependent on the eyes finishing some activity. In many of the move elements, the eye control times are incorporated into the work of the hand, such as high control Gets and Puts.

- E2: Eye fixations are awarded two Mods. An eye fixation is a mental recognition, such as check-reading a gauge.
- E2: Eye travel, or simply, the movement of the eye up and down, left or right. In the Modapts system, an E2 is awarded for the eye travel of 30 degrees (approximately 20.32 cm across, 38.1 cm from the eye).
- E4: Eye focus, or changing shape of the lens. One eye focus time is awarded when the distance of the next object is substantially different from the last.

Move. The move element is an action of the finger, hand, or arm. The action is a specific movement from and to articles or locations. Moves are often a part of action chains that include getting or putting objects.

- M1: A finger move (one Mod) is a movement performed with any finger. In the Modapts system, the movement often is a distance of 2.54 cm or less.
- M2: Hand moves are considered to cover a distance of 5.08 cm or more. The M2 must consist of the palm moving, such as in handling small objects.
- M3: When the forearm is used in the movement, usually a distance of at least 15.24 cm, an M3 is awarded. The M3 is only awarded if the wrist moved with the movement.
- M4: The wholearm move is performed with full arm outward, a distance of 45.72 cm. To award the M4, the shoulder must move. The Modapts definition of the M4 specifies that the

move is inside an imaginary cone with the point at the shoulder.

- M5: When the arms and the body trunk are involved in the movement (a usual distance of 76.2 cm), the M5 is awarded. The extended arm movement is like the M4, except that the distance requires that the body will move.

Gets. The element Get is an action to gain control of an object using the fingers.

- G0: Get by contacting; the simplest form of obtaining control of the object is awarded a G0. A contact Get is achieved by touching the object with the finger(s) or hand. G0 also is internal to the Move element.
- G1: Get by simple closing; this type of grasp is a simple closing of the fingers around an object. The G1 is a terminal activity; a move must precede it.

Juggle. A change in the position of the grasp accomplished without giving up control of the object is a Juggle element. If a Juggle occurs internal to the move, it does not have to be awarded.

- J2: A change of grasp.

Data Processing. Data processing elements include all necessary elements associated with the keystroking of computers (including eye fixations and eye travel).

- DP3.3: Keystroking using a video display.

Counting. The element Number deals with counting numbers and mentally retaining the information.

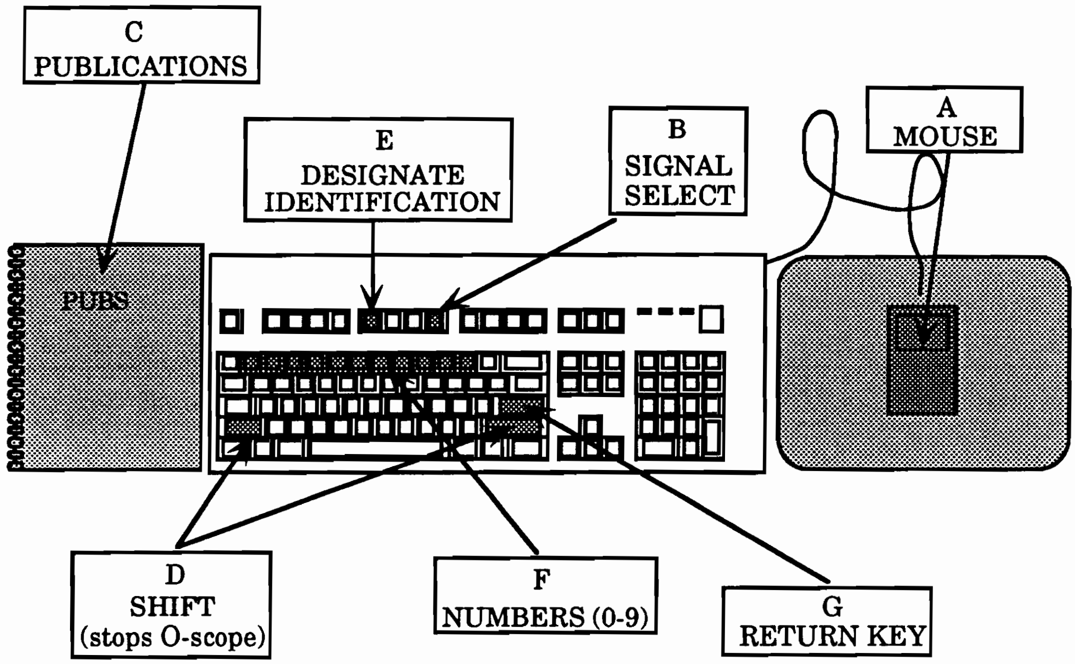
- N3: Awarded per item when the items are arranged.
- N6: Awarded per item when the items are not arranged.

Seek. The Seek element deals with locating things. There is a small amount of reading together with simple movements, and the actions overlap. A Seek element requires that a movement precedes the reading. The time for a Seek includes the time to read one key word per item. If more than one word must be read, then additional R2s must be added.

- SS6: Seek information in books, per page, with the pages not fully turned.
- SS10: Seek information in books, per page, with the pages fully turned.

After the elements relevant to the IT were identified, specific Modapts values were assigned to the activities of the IT defined by the Micro Saint model. For the physical actions, the assigned Modapts values are illustrated and listed in Figure 15.

Mod Value. The mean times to perform the activities listed in Figure 15 were determined from the video tapes of the experimental sessions. For each activity, 120 cases were measured. The Modapts times were compared to the actual times to assess the fit and to identify the proper Mod allowance factor. The activities of typing in the second EFX number (F1-F2), third EFX number (F2-F3), and pressing the Return key (F3-G) were combined into one activity category. The Modapts times for the activities were used as regressors on the actual performance times.



MOVEMENT	DISTANCE (cm)	MODAPTS
A–B	≈440	G1M5G0
B–C	200 (lap)/ 280 (desk)	M4G0
C–D	≈65(left)/305(right)/110 (lap)	M4G0
D–E	150	DP3.3
E–F1	65–110	DP3.3
F1–F2	0–172	DP3.3
F2–F3	0–172	DP3.3
F3–G	75–240	DP3.3
G–A	370	M5G0
Grasp (Mouse/Pubs)		J2

Figure 15. Diagram and list of IT actions and Modapts assignment.

A *t*-test was performed to determine if the slope was significantly different from 1. A slope significantly different from unity warrants further investigation of the Mod value to identify an allowance factor. A slope that approached unity indicates that the actual and Modapts times are comparable, given a high R^2 value.

Frequency Distributions

Micro Saint is a network simulation software package that enables the modeler to simulate real-life processes. The random variation of a process is simulated by including the standard deviations for each activity and specifying the appropriate frequency distribution of the activities. The distributions of the movement times, errors, and search times through the publications were analyzed using nonparametric statistical tests; specifically, Kolmogorov-Smirnov (K-S) tests were used to determine the appropriate distribution of times or occurrences for the IT activities. In other Micro Saint modeling systems similar to the abstraction level of the IT, the gamma distribution was used to model movement times (Osga, 1989 Tijerina and Treaster, 1987). Law and Kelton (1991) cited both lognormal and gamma distributions as likely candidates for the movement distributions.

For the movement and search times extracted from the video tapes, outliers were eliminated from the data sets. Outliers were defined as two data points exceeding two standard deviations from the mean and were not included to eliminate times which included spurious movements (e.g. pushing up eyeglasses, adjusting eyeglasses, etc.). The actual times were converted to lognormal times, and a K-S one-sample test on the lognormal conversions was performed. Gamma distributions were generated using alpha and beta values computed from the original data sets. Two-sample K-S tests compared the gamma distributions to the actual data.

Model Standard Deviations

The Modapts methodology does not specify time standard deviations for any activity. Three sources for specifying standard deviation information were examined. The sources were:

(1) standard deviations from relevant data sets used in constructing the Modapts database,

(2) standard deviations cited in other Aegis class studies, and

(3) the actual standard deviations collected from the participants.

Workload Methodology

Scaling solution. After all participants completed the SWAT card sort as described in a previous section, the rank orders were analyzed using the SWAT software. A Kendall's coefficient of concordance was computed to determine if a single group scaling solution can be used for the all participants. The SWAT software also produces the scaling solution for workload event rating scores.

Workload scores. SWAT scores obtained by the participants after each test trial were converted to workload scores based on the group scale. Workload scores from Group 1 were analyzed with multiple regressions. The predictors Format, Color condition, Symbol set, Density, mean time to process an emitter, and mean number of hook errors on workload scores were regressed on mean workload scores obtained on all 24 configurations by Group 1.

Model Verification

The model times defined by the Modapts methodology using Group 1 data, actual standard deviations from Group 1, and the workload regression equations from Group 1 were built into the Micro Saint IT model. The Modapts time assignments, actual standard deviations, gamma frequency distributions, and workload regression equation were integrated into the Micro Saint model for the IT. During final refinements to the model, the estimated times based on the Mod value of 0.146 were consistently two or more seconds longer than the actual times. By inspection, no single aspect of the model accounted for the difference. Therefore, the standard value of the Mod (0.129 sec) was used instead of 0.146 sec. This point is discussed further in the Discussion and Conclusions section.

Verification of the proposed methodologies was accomplished by comparing Group 1 mean performance scores (defined in Equation 1) to mean performance scores produced by the Micro Saint Model, and mean workload scores obtained from Group 1 to mean workload scores generated by the Micro Saint Model.

During the test trials, the six participants in Group 1 were able to process between 8 and 16 emitters on each configuration. Using 12 as a typical number of processed emitters, the model was run a corresponding number of times; that is, the model was run 72 times (6 x 12) for each configuration. A random number generator was used to assign random number seeds for each model configuration.

Mean performance scores (defined in Equation 1) were calculated for the modeled configurations and compared to the Group 1 mean performance scores using simple regression.

Mean workload scores obtained from the model for each configuration were compared to Group 1 mean workload scores using simple regression.

Sensitivity analysis. A sensitivity analysis was performed to check the effect of varying the weights (b_1 and b_2) used in defining the performance variable shown in Equation 1. Three different weights were chosen to examine $b_1 > b_2$, $b_1 = b_2$, and $b_1 < b_2$. The three weighting schemes were: (1) $b_1 = 0.75$, $b_2 = 0.25$, (2) $b_1 = 0.50$, $b_2 = 0.50$, and (3) $b_1 = 0.25$, $b_2 = 0.75$.

Using the three different weighting schemes, mean performance scores were computed for the model and Group 1. Performance scores for the model were regressed on performance scores for Group 1 across all 24 configurations. The R^2 values from each regression (one for each scheme) was compared to the R^2 value for every other scheme using a z-test, where:

$$z = \frac{R_a^{\ 2} + R_b^{\ 2}}{\sqrt{\dfrac{MS_{E(a)}}{df_a} + \dfrac{MS_{E(b)}}{df_b}}} \qquad (2)$$

$R^2{}_a$ and $R^2{}_b$ represent the R^2 values produced by regression of model performance scores on Group 1 performance scores using a particular weighting scheme. $MS_{E(a)}$ and $MS_{E(b)}$ are the regression residual mean squares associated with the regression using the specified weighting schemes, and the *df*s for all regressions was 22. The corresponding *z* value for rejecting the null hypothesis that the two R^2s are equal at $p = .05/2$ is 1.69.

Model Validation

The final step of the research effort was to validate the Micro Saint model against a second group of participants. Group 2 data were not used in the model-building phases of the study. Mean performance scores were calculated for Group 2 participants using Equation 1 and compared to the model-generated scores with a simple regression. Likewise, mean workload scores generated from the model were regressed on mean workload scores obtained from Group 2 across all 24 configurations using a regression.

Cross validation. Correlation coefficients were calculated between Group 1 mean performance scores and Group 2 mean performance scores, and Group 1 mean workload scores and Group 2 mean workload scores across all 24 configurations. Since the groups were small (n = 6) cross-validation was done to determine that Groups 1 and 2 were not different in terms of performance scores and workload scores on the 24 configurations.

RESULTS

Task Definition and Validation

Expert EWs agreed that the task networks were an appropriate representation of the steps required to perform a functional analysis of emitters, that the alternate paths taken by the operator were accounted for, and that there were no omissions. Moreover, certain paths were deleted because they were, for all practical purposes, never used by the operator (e.g., the on–line library is used infrequently due to the inadequate interface and non-existent manipulation power.) After expert review of the procedural components of the task, the IT was checked for accuracy against video taped demonstrations performed aboard the U.S.S. Virginia.

The IT detailed in Table 5 represents the functional analysis of an emitter that an EW performs using the AN/SLQ-32(V). To examine the task under different display conditions using Macintosh prototypes, the task was modified for the prototype environment. These steps were the basis of the final Micro Saint model. The steps for the prototype SLQ-32 IT are listed in Table 7.

The procedural components of the Macintosh IT network model were verified by video taped sessions using the prototypes. The refined Micro Saint model will be discussed at length in the next section.

TABLE 7

The Integrated Task (IT) Activities for the Macintosh Operator

1	Operator hears an alert and new emitters appear.
2	Operator "hooks" (selects) an emitter with a mouse.
3	Operator assimilates the following emitter parameters: a. Close Control parameters (bearing, threat level, frequency, PRF, scan, scan type, assigned EFX). b. Auditory display of scan. c. Visual representation of scan (simulated frequency spectrum analyzer–ULQ-16).
4	Operator searches through written publications and/or on-line libraries for an identification/confirmation.
5	Operator agrees or disagrees with the system designation and identification.
6a	Operator records the emitter into a log. Returns to step 2.
or	
6b	Operator searches publications for the correct designation.
7	Operator designates the ID. Returns to Step 2.

Micro Saint Model

The IT is a serial task in terms of the motion/steps that the EW must take to process emitters. There is little variation regarding the procedural aspects of the model. The IT assumes that the operator is skilled and knows exactly what actions need to be taken using the information available regarding the parameters of the emitter.

The key to modeling is selecting the appropriate level of abstraction to represent the system under study (Law and Kelton, 1991). For this dissertation effort, there were two factors that affected the level of the network model; logically, both factors stem from the purpose of the project. First, the level of the model had to be general enough to allow for the different configurations of the study without requiring separate models to be built for each configuration change. Second, the model had to be specific enough to allow for examining different configurations of the DCC through movements that were reasonable to measure. After iterative refinement of the task network, the Micro Saint model diagrammed in Figure 16 was produced.

Each node of the model will be explained, as will the structure of the branches. In Micro Saint symbology, a circle represents a single node, while a square represents a subnetwork of nodes.

The top level of the IT model shows that there are three main functions – indicated here by subnetworks – for processing an emitter; an emitter is processed by hooking it (Network 2), evaluating it (Network 3), and, then, designating the emitter with the correct EFX (Network 4).

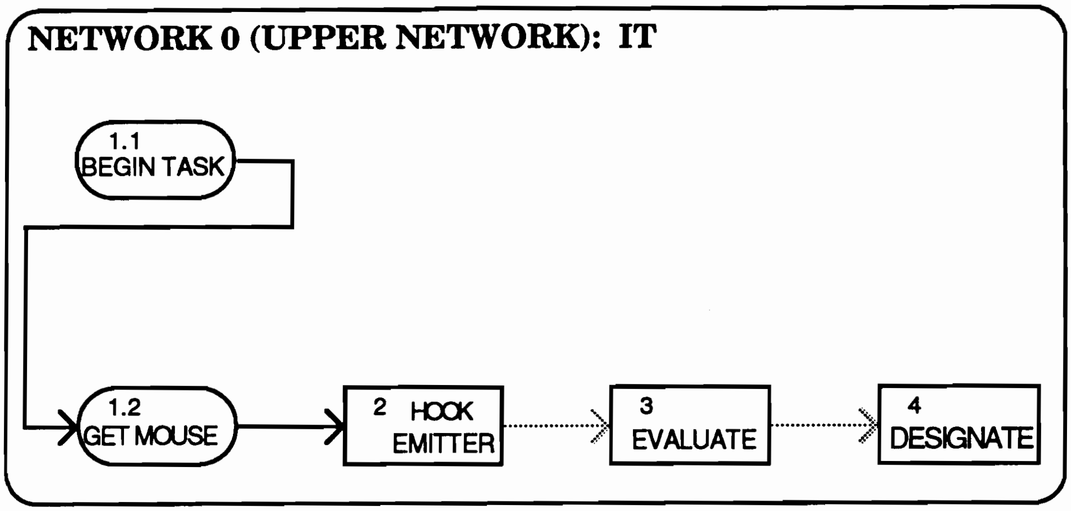


Figure 16. Network 0 (top level) of the Micro Saint model for the IT.

The first node of the model, identified as 1.1 and titled “Begin Task,” is a dummy variable which allows the experimenter to specify a display configuration. Each node has an accompanying window to specify the parameters for that node. In Micro Saint terminology, this window is titled “Task Description.” For clarity, the task description window is referred to as the node parameters. The node parameters for 1.1 are listed in Table 8. All other nodes are described briefly in subsequent paragraphs. While all node parameters are included in Appendix E, the node-by-node descriptions include only times, distributions, and standard deviations for those activities that were not determined by using Modapts. The parameters that were defined using the Modapts methodology are described under the next section of this document.

TABLE 8

Node 1.1 Parameters

Task Number: 1.1	Name: BEGIN TASK	
Upper Network: 0		
Upper Name: IT		
Time Distribution:	Normal	
Expressions:		
{Release Conditions}	1;	
{Beginning Effect}		
	FORMAT:=POLAR,	{POLAR,RANGE}
	SYMBOL:=OLD,	{OLD, NEW}
	COLOR:=BW,	{BW, C}
	DENSITY:=1,	{1-CARIBBEAN, 2-GULF 3-
	ARMAGEDDON}	
	MOD:=0.129	{MOD value];
{Mean Time}	0 ;{Dummy Variable}	
{Standard Deviation}	0;	
{Launch Effect}	n/a	
Decision Type:	Single	
What Happens Next:	Probability:	
1.2 Get Mouse	Always	

Several functions were created in the model to invoke a value at a specified time. The functions with their values and descriptions are listed in Table 9. (The equations for the mean and standard deviation for errors are discussed in a later section).

The variables that were defined in the IT model are listed in Table 10. In task network modeling, the variables are essential to characterizing the system. Variables are used to modify or describe the relationships between different factors affecting the model.

TABLE 9

Functions for the IT Model

VARIABLE NAME	VALUE	DESCRIPTION
ERRORS	Equation	mean for gamma distribution
ERRSTDEV	Equation	standard deviation for gamma distribution
POLAR	0	indicator variable for Polar Format
RANGE	1	indicator variable for Range Format

TABLE 10

Variables for the IT Model

VARIABLE	INITIAL VALUE	DESCRIPTION
BW	0	indicator variable for black/white condition
C	1	indicator variable for color condition
clock	0	timer
COLOR	0	assigned a 0 or 1 depending on color condition
DENSITY	0	reflects density level (1, 2 , or 3)
DUM_ERROR	0	counts the number of hook errors
duration	0	duration of each node
ERROR	0	evokes ERRORS function
ERRORSTDEV	0	evokes ERRSTDEV function
FORMAT	0	assigned 0 or 1 depending on format condition
GRAB	0	indicates that pubs were grabbed
MOD	0	specifies input MOD value
NEW	1	indicator variable for new symbol set
NO	0	repeats hooking until correct
OLD	0	indicator variable for old symbol set
READ	0	indicates operator has read ULQ-16
RETRYA	0	probability of re-hooking from beginning
RETRYB	0	probability of hooking a close emitter
RETRY C	0	probability of hooking overlapped emitters
run	0	specifies pass through model
SYMBOL	0	assigned 0 or 1 depending on symbol set
WKLD	0	workload score

Using Micro Saint conventions, system defined variables – those variables built-in to the simulation software – are in lower case letters. User-defined variables are capitalized to distinguish them from system variables.

The basic steps for the Hook Network are shown in Figure 17. For the IT, variations due to the combination of independent variables are manifested in the Hook Network. Changing the display format, color condition, symbol set, and emitter densities affect search and locate times, specifically the speed at which the operator can locate an emitter and the number of errors the operator makes in attempting to hook the proper emitter. It is assumed that the remainder of the IT is strictly procedural in nature, such that the activities are not be influenced by the change in the display. Hook errors, also referred to as dummy errors, are modeled as a function of the different display configurations.

In Network 3 (Figure 18), the operator has hooked the correct emitter, and performs the steps required to evaluate the parameters associated with the emitter. Network 3 incorporates the only difference in IT strategies detected in the data collection. In most cases (five out of six), operators used the publications in one step. That is, most operators turned on the ULQ-16 until the proper EFX was determined from the pubs and the close control information. Once the X-Ref number had been determined, the operator immediately turned to the EFX section of the pubs to locate the correct EFX (this time is represented in node Pubs A). Subsequently, the operator moved to the activities of designating the identification in Network 4. However, in

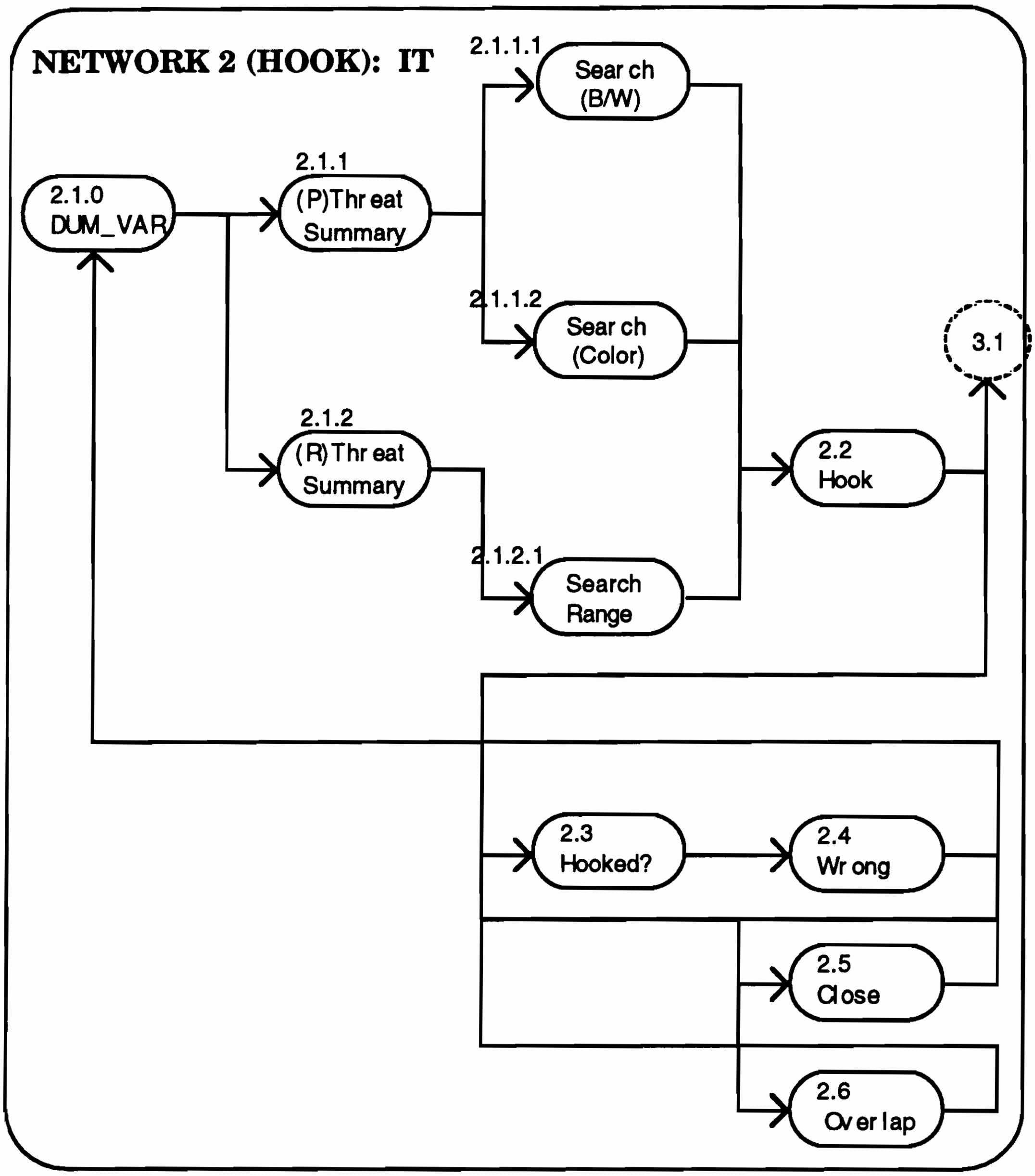


Figure 17. Network 2 (Hook) of the Micro Saint model for the IT.

one out of six cases, the operator's strategy was segmenting the pubs search into two separate searches. First the operator examined the close control parameters and ULQ-16 and, then searched the pubs for the correct X-Ref number (Pubs B node). Once the X-Ref number was determined, the operator turned off the ULQ-16 (Shift key), press the Desig ID FAB, and returned to the pubs to look up the correct EFX (Pubs C node). Finally, the operator turned his or her attention back to the keyboard to enter the appropriate EFX numbers. At node 4.4, differences in operator steps due to the different search strategies cease. Network 4 is illustrated in Figure 19.

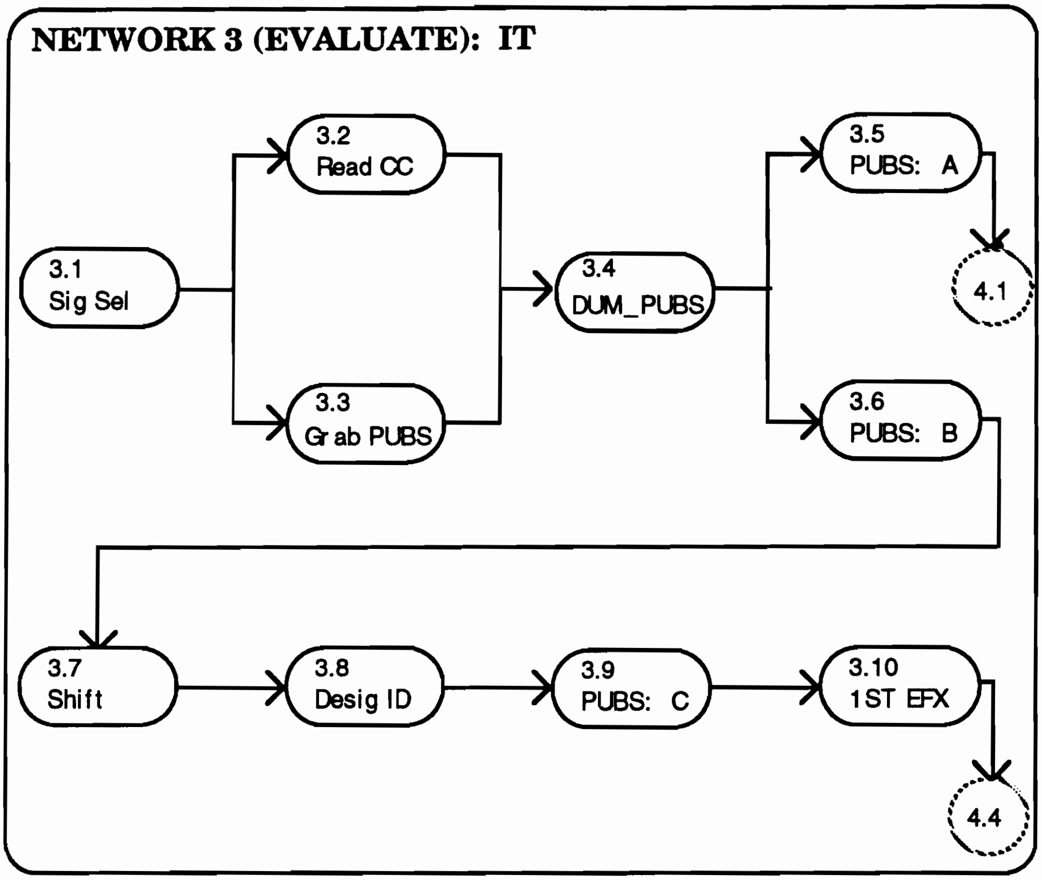


Figure 18. Network 3 (Evaluate) of the Micro Saint model for the IT.

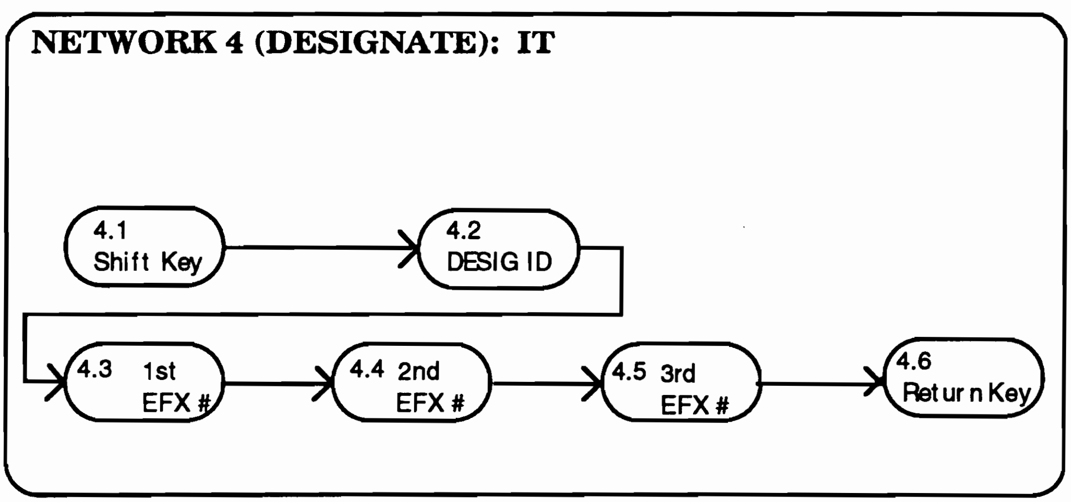


Figure 19. Network 4 (Designate) of the Micro Saint model for the IT.

1.1 BEGIN TASK. As mentioned above, dummy variable 1.1 is used to define the display configuration. There is no operator activity associated with node 1.1; therefore, no time (or standard deviation) is assigned. In node 1.1, the experimenter defines each of the four independent variables of the redesign experiment; thus, 24 different configurations are possible. The value of the MOD also can be adjusted in this first node. Selection of the MOD value is discussed later in the section on movement times. The performance and workload scores of each of the 24 configurations were the basis of verifying the model's fit to the six participants (Group 1) from whom the model was developed, as well as validating the model against six additional participants (Group 2). The follow-on node (1.2) represents the actual start of the simulation of the operator performing the IT.

1.2 Get Mouse. The start of the prototype-based IT begins with the operator moving his hand and grasping the mouse. Because the operator is processing emitters in rapid succession, the start point of the movement is leaving the Return Key from a previously processed emitter. The stop point of the node is when the mouse has been grasped.

2 Hook. Node 2 represents the network for the function of hooking an emitter.

2.1.0 DUMVAR. DUM VAR is used to separate Polar and Range Hook strategies. Based on the value of the variable FORMAT, the model simulates either the Polar configuration (FORMAT = 0) or the Range configuration (FORMAT = 1). Node 2.1.0 also contains the information for determining the number of hooking errors that occur in the current model run. SYSTAT Version 4, the System for Statistics for the PC (1988), software was used to determine the frequency distribution of errors for each of the 24 possible configurations (over all 12 participants) using Kolmogorov-Smirnov (K-S) tests. In all configurations, the results indicate that the hypothesis that the data were obtained from a gamma distribution was not rejected ($p = 0.05$). Table 11 provides a summary of the K-S tests.

TABLE 11
Kolmogorov-Smirnov Test Results of Gamma Distributions for Errors

FORMAT	SYMBOL	COLOR	DENSITY	p
Polar	Old	B/W	Caribbean	0.985
Polar	Old	B/W	Gulf	0.985
Polar	Old	B/W	Armageddon	0.985
Polar	Old	C	Caribbean	0.769
Polar	Old	C	Gulf	0.985
Polar	Old	C	Armageddon	0.985
Polar	New	B/W	Caribbean	0.768
Polar	New	B/W	Gulf	0.986
Polar	New	B/W	Armageddon	0.985
Polar	New	C	Caribbean	0.994
Polar	New	C	Gulf	0.986
Polar	New	C	Armageddon	0.985
Range	Old	B/W	Caribbean	1.000
Range	Old	B/W	Gulf	0.768
Range	Old	B/W	Armageddon	0.769
Range	Old	C	Caribbean	0.538
Range	Old	C	Gulf	0.986
Range	Old	C	Armageddon	0.421
Range	New	B/W	Caribbean	1.000
Range	New	B/W	Gulf	1.000
Range	New	B/W	Armageddon	0.769
Range	New	C	Caribbean	1.000
Range	New	C	Gulf	0.985
Range	New	C	Armageddon	0.769

StatView II (1987) software was used to perform multiple regressions of the configuration variables across all 24 configurations on (1) mean errors and (2) standard deviations and the configuration variables (main effects). The results of the multiple regression, outlined in Table 12, showed that a regression equation accounted for approximately 72 percent of the variance in mean error times ($p < 0.0001$). The reported R^2 value is adjusted for the number of variables in the regression.

Similarly, the results from the multiple regression for error standard deviations are outlined in Table 13. It was found that a multiple regression equation accounted for approximately 71 percent of the variance in standard deviations ($p < 0.0001$).

A Micro Saint model function, entitled ERRORS, was created to define the multiple regression equation for hooking errors. Another function, entitled ERRSTDEV, was created to define the multiple regression equation for hooking error standard deviations. Both functions were implemented with regression equations using the beta coefficients and intercepts in Tables 12 and 13, respectively. In any particular model configuration, the number of hooking errors was chosen from a gamma distribution with mean equal to ERRORS and standard deviation equal to ERRSTDEV.

There are two possible follow-on nodes to Node 2.1.0. The Micro Saint decision type of "multiple" was implemented so that for a Polar display configuration the follow-on node was 2.1.1 and for a Range display configuration the follow-on node was 2.1.2.

TABLE 12

Multiple Regression: Errors

Count	r	R^2_{adj}	
24	0.85	0.71	$p < 0.0001$

Variable	Beta (obtained)	Beta (Std)	p
FORMAT	-3.982	-0.582	<0.0001
SYMBOL	-2.073	-0.303	0.0232
COLOR	-1.897	-0.277	0.0358
DENSITY	1.900	0.454	0.0015
INTERCEPT	3.787		

TABLE 13

Multiple Regression: Error Standard Deviations

Count	r	R^2_{adj}	
24	0.84	0.71	$p < 0.0001$

Variable	Beta (obtained)	Beta (Std)	p
FORMAT	-1.454	-0.593	0.0014
SYMBOL	-0.870	-0.237	0.0372
COLOR	-1.188	-0.046	0.0064
DENSITY	1.014	0.500	0.0004
INTERCEPT	1.916		

2.1.2 (P) Threat Summary. Because the video tape measurement system was not sensitive to specific reading times, the time to read the information from the threat summary list was approximated by Modapts terms. In the threat summary list of a Polar configuration, operators need to read bearing and threat level. Using Modapts definitions, the time to read the bearing and the threat level is equivalent to reading one word.

2.1.1.1 Search (B / W) and 2.1.1.2(Color). Node 2.1.1.1 represents the time required for an operator to search the Polar display for an emitter when the color configuration is black and white. Node 2.1.1.2 represents the time for the operator to search the Polar display for an emitter when the configuration includes color. The subtask begins after the bearing and threat level have been read from the threat summary list and ends when the operator begins to move the mouse to the emitter symbol on the Polar display. The video tapes were not sensitive to eye fixations of the operator, so the time to search the primary display for the correct emitter was interpolated. Specifically, time data were taken from the video tapes that began with the operator looking at the threat summary list until the operator clicked on the emitter with the mouse. The time used in node 2.1.1 – time to read the threat summary list – plus a Modapts value for an eye travel between the threat summary list and the main display were subtracted from the measured time. Next, the time in node 2.2 (time to move the mouse and click) was subtracted. The resultant times were used as the independent variable in a multiple regression equation with Symbol, Color Configuration, and Density as the predictors of Polar search times. The results are shown in Table 14. The mean times for the Polar search nodes were assigned by the multiple regression equation using the obtained beta coefficients shown in Table 14.

TABLE 14
Multiple Regression: Polar Search Times

Count	r	R^2_{adj}	
12	0.88	0.77	$p < 0.0065$

Variable	Beta (obtained)	Beta (Std)	p
SYMBOL	-0.492	-0.495	0.1395
COLOR	-0.871	-0.280	0.0198
DENSITY	0.717	0.666	0.0045
INTERCEPT	1.964		

A regression of the variables Symbol, Color, and Density on standard deviations for Polar search times indicated that the standard deviations did not vary systematically with the variables ($R^2 = 0.03$), as shown in Table 15. Thus, the standard deviation of the grand mean was used as the standard deviation for each configuration in nodes 2.1.1.1 and 2.1.1.2.

TABLE 15
Multiple Regression: Polar Search Time Standard Deviations

Count	r	R^2_{adj}	
12	0.17	0.03	$p = 0.9705$

Variable	Beta (obtained)	Beta (Std)	p
SYMBOL	-0.196	0.020	0.7509
COLOR	0.035	-0.115	0.9550
DENSITY	0.126	0.121	0.7376
INTERCEPT	2.391		

2.1.2 (R) Threat Summary. The time to read the information from the threat summary list when the configuration included a Range format was approximated by Modapts terms. In the threat summary list of a Range configuration, operators need to read bearing, range, and a symbol icon. Using Modapts definitions, the time to read the bearing, range, and symbol icon each are equivalent to reading one Modapts word. In Modapts, a single word which requires careful reading is assigned a value of three Mods. However, only two Mods are assigned to a word when the operator simply glances at a word to get an idea of the meaning. It was assumed that the time to read bearing and range is three Mods and the time to look at the symbol icon is only 2 Mods.

2.1.2.1 Search (Range). Range information provided textually in the threat summary list and graphically in the primary display decreased the operator's search for an emitter from an entire bearing line – as done in the Polar format – to a pinpoint location. After the operator read the emitter parameters from the threat summary list, the operator looked at the primary display and performed a three Mod read for the bearing and two counting actions (each worth 2 Mods) for counting range lines.

A similar method was used to interpolate the times to search using a Range configuration as was used to determine the Polar search times (nodes 2.1.1.1 and 2.1.1.2). Specifically, the time the operator looked up at the threat summary list until the time the operator clicked the mouse on the correct emitter was extracted from the video tapes. From this time, the following three times were subtracted: (1) the time to read the threat summary list (node 2.1.2), (2) the time to read a bearing (three Mods) and time to count two range lines (two Mods apiece), and (3) the time to move the mouse to an emitter and click the button (node 2.2). This interpolation produced the search time for the node (a time that could not be extracted from the video tapes) from a measurable activity. The results of the multiple regression of Symbol, Color, and Density on search time using the Range configurations are listed in Table 16.

16
Multiple Regression: Range Search Times

Count	r	R^2_{adj}	
12	0.88	0.77	$p = 0.0058$

Variable	Beta (obtained)	Beta (Std)	p
SYMBOL	-0.441	-0.495	0.0911
COLOR	-0.104	-0.28	0.6620
DENSITY	0.682	0.666	0.0013
INTERCEPT	0.713		

Using the Symbol, Color, and Density as predictors of actual times, the multiple regression equation accounts for approximately 77 percent of the variance in Range search times (p = 0.0058).

Thus, the total time for the range search node was defined as the time to read on bearing line, count two bearing lines, and the search time calculated with a multiple regression equation using the beta coefficients listed in Table 16.

A shown in Table 17, multiple regression of Symbol, Color, and Density on standard deviations for range search times resulted in a low correlation (r = 0.17) and all beta coefficients were not significant, thereby indicating that the standard deviations did not vary as a function of these factors. Thus, the standard deviation for the grand mean of all range configurations (2.564) was used in node 2.1.

TABLE 17
Multiple Regression: Range Search Time Standard Deviations

Count	**r**	R^2_{adj}	
12	0.17	0.03	p = 0.9705

Variable	**Beta (obtained)**	**Beta (Std)**	p
SYMBOL	0.035	0.020	0.9550
COLOR	-0.196	-0.115	0.7509
DENSITY	0.126	0.121	0.7376
INTERCEPT	2.391		

2.2 Hook Emitter. The node "Hook Emitter" represents the time it takes an operator to move the mouse to a desired emitter on the primary display and to press the mouse button. The video tapes lacked the sensitivity to allow distinction of the small movements involved in positioning and clicking a mouse, so an existing model for defining the subtask time was adapted for this node from Card, Moran, and Newell (1983).

Fitt's Law states that movement time is a function of the size of the target and the distance to the target. Based on Fitt's Law, Card et al. (1983)

found that the time to move the hand (T_p) a distance, D, to a target of size S is given by:

$$T_p = C\log_2(D/S + 0.500) \tag{3}$$

where:

C = 0.100 [0.070-0.120] sec/bits.

Moreover, Card et al. (1983) showed that text selection using different input devices could be analyzed in terms of moving a hand to a target. Specifically, movement time using a mouse is given by:

$$T_p = K + C\log_2(D/S + 0.500), \tag{4}$$

where:

T_p = Positioning time,
D = Distance to target
S = Size of the target,
C = 0.100 [0.070-0.120] sec/bit, and
K = a constant.

Borrowing from Card et al. (1983), the equation above was used to determine the time for node 2.2, where:

D = 8.890 cm (the radius of the primary display),
S = 0.635 cm (symbol width), and
K = 1.030 (data from Card et al., 1983).

In this study, the average distance to the target was defined as half the diameter of the display; that is, D was assigned the radius of the primary display.

Card et al. (1983) cited a standard error of 0.070 for movement times in a text selection task using a mouse. This standard error value was used in the Micro Saint model.

2.3 Hooked? The activity entitled "Hooked?" is a decision node that determines if the correct emitter has been hooked. If the variable NO (i.e., error counter) is not zero when the simulation run passes through node 2.3, the model simulates a hook error. A hook error is followed by node 2.4. If the variable NO is zero, indicating that a correct emitter was hooked, the simulation proceeds to the next step of processing the emitter (i.e., Network 3–Evaluating the emitter).

2.4 Wrong. As mentioned above, when a hook error occurs, node 2.4 is evoked. A hook error is succeeded by one of the following activities:

2.1.0 - Rehook starting from reading the threat summary list,
2.5 - Moving the mouse to a close-by emitter and clicking, or
2.6 - Slightly moving the mouse on the present icon and clicking.

The video tapes and informal comments from the participants revealed that when hook errors occur, the likelihood for each of the three following steps (essentially, reasons for the error) differed depending on the display format. In other words, differences in likelihood of the three follow–on activities are linked to the attributes of the format. When a hook error occurred, the probabilities of performing node 2.1.0 (i.e., re-reading the threat summary list), 2.5 (i.e., trying a close-by emitter) and 2.6 (i.e., trying to hook an over-lapped emitter) were defined with variables RetryA, RetryB, and RetryC, respectively. Values of the Retry variables were estimated by random observations of 120 cases of hook errors for each of the format conditions. The initial assigned values for the variables are listed in Table 18.

In the Polar format, the most prevalent cause for hooking an incorrect emitter occurred because several emitters were overlapped (95% of hook errors). Due to the constraints of the system software (both AN/SLQ-32(V) and Macintosh prototypes), clicking on a target emitter overlapped with a dummy emitter often "hooked" the dummy emitter. The operator, then, attempts repeatedly to place the cursor over a small, non-over-lapping edge of the desired symbol. In the Range format, emitters did not overlap often because real vessels cannot occupy the same space in the environment. Range format hook errors attributed to over-lapping emitters was 33.3%.

TABLE 18
Assigned values for Retry Variables

Variable	Follow-on Node	Variable value (likelihood)	
		POLAR	RANGE
RETRYA	2.1.0 (Dum_Var)	3 (0.025)	77 (0.642)
RETRYB	2.5 Close emitter	27 (0.225)	3 (0.025)
RETRYC	2.6 Over-lap	90 (0.750)	40 (0.333)

In the Polar format, operators frequently hooked an incorrect emitter that was along the same bearing line as the correct emitter (22.5%). This problem is inherent to the display because of the lack of distance relationships along the bearing lines. The problem occurred most often for missiles. Missiles were presented in the largest region of the display, and the number of missiles was larger than the other emitter categories. Consequently, when a number of emitters were placed randomly on the display for each trial, there was greater likelihood that two or more missiles would occur on or about the same bearing, thus making it necessary to use the "trial and error" method of figuring out which missile was the correct emitter.

The likelihood of the operator going back to the threat summary list to read the emitter's parameters was small (2.5%) in the Polar format. In the Range format, however, the likelihood of going back to read the threat summary list was 64.2%. An operator equipped with both range and bearing information was fairly certain to hook the correct emitter the first time. When the hook attempt was incorrect, the most frequent reason was an error in reading the threat summary list.

The decision type for selecting the follow-on node was specified as probabilistic. Micro Saint software reads the values for the Retry variables which were dependent on Format and converts the numbers into probabilities by summing the value of Retry A-C and dividing each variable by the sum.

The Modapts methodology assigns a decision time of 2 Mods when a yes/no comparison or any binary decision is made and the outcome was the “unusual” case (i.e., failing to pass an inspection). In this network, a hook error is an unusual event; thus, a 2-Mod value is assigned to node 2.4. No standard deviation was assigned due to lack of appropriate data.

2.5 Close. As described above, node 2.5 accounts for the time to move the cursor from an incorrect emitter to a close-by emitter and to click the mouse on it. The time for this node was determined using the same equation as was used for the original hook, except that D was changed to reflect the average distance of a close-by emitter (1.27 cm). The equation used to determine the time was:

$$T_p = K + C\log_2(D/S + 0.500) \text{ sec,} \tag{5}$$

where:

T_p = Positioning time,
D = 1.27 cm,
S = 0.64 cm (symbol width),
C = 0.100 [0.070-0.120] sec/bit, and
K = 1.03 (data from Card et al., 1983).

The standard deviation was defined with the standard error found in Card et al. (1983). The follow-on task is 2.3, which determines whether the hook was correct or an error.

2.6 Overlap. The time to retry a hook when the emitters are overlapped was determined by Equation 5 (adapted from Card et al., 1983).

$$Tp = K + C\log_2(D/S + 0.500), \tag{6}$$

where:

D = 0.080 cm,
S = 0.635 cm (symbol width),
C = 0.100 [0.070-0.120] sec/bit, and
K = 1.030 (data from Card et al., 1983).

A distance of 0.08 cm was an approximate distance of moving the mouse from one edge of a dummy emitter to the edge of the overlapped emitter. The standard deviation was defined as the standard error found in Card et al. (1983). The follow-on task is 2.3, which determines whether the hook was correct or an error.

3.0 Evaluate. Subnetwork 3 models the steps for evaluating the emitter's parameters and identifying a correct EFX number.

3.1 Sig Sel. Pressing the Signal Select FAB is a movement starting from the time the hand leaves the mouse to the time the Signal Select button is pressed. Immediately following the Signal Select key press, the operator performs two activities simultaneously: (1) reads the close control parameters and the ULQ-16 information (node 3.2), and (2) reaches for the pubs (node 3.3).

3.2 Read CC Parameters. The Modapts convention for reading words was used to define the number of Mods for reading the close control parameters. The two parameters that the operator needs are the FRQ (two words) and the PRF (two words).

3.4 Dum Pubs. Node 3.4 is a dummy variable which starts with the pubs search only after the parameters have been read *and* the pubs are in hand (simultaneous actions noted in 3.1).

3.5 PUBS: A. Operators used two different approaches for identifying an emitter's EFX. In the first approach, the operator refers to the pubs for the appropriate X-Ref number and turns immediately to the EFX section of the pubs to lookup the EFX. In the second approach, the operator does not turn immediately to the EFX section of the pubs after having located the correct X-Ref number. Instead, the operator turns off the ULQ-16 and presses the Design ID key, then returns to the pubs to lookup the correct EFX number.

Pubs A represents the activity of a single pubs search strategy in which time to search includes obtaining an X-Ref number and, subsequently, the correct EFX. Node 3.5 starts when the operator looks at the pubs and ends when the operator begins reaching for the keyboard. Modapts-defined seek times are used in the node (discussed in detail in the next section).

3.6 PUBS: B. PUBS: B represents the first of a two-part pubs search strategy. PUBS: B represents the search in the pubs to find the correct X-Ref number. Node 3.6 begins when the operator looks at the pubs and ends when the operator begins reaching for the keyboard. Modapts-defined seek/search times are used in the node (discussed in detail in the next section). In the two-part pubs search strategy, the ULQ-16 (node 3.7) is turned off after the operator has found the X-Ref number.

3.7 Shift Key. For the Macintosh prototype, the operator uses the Shift key to turn off the ULQ-16. In the Micro Saint simulation, the movement starts when the operator's hand leaves the pubs and ends when the Shift key is pressed.

3.8 Desig ID. Operators press the Desig ID button after turning off the ULQ-16. Node 3.8 represents the activity of pressing the Desig ID key. It starts when the operator's hand leaves the Shift key and ends when the operator presses the Desig ID key. In all cases, the Shift key press precedes the Desig ID key press.

3.9 PUBS: C. PUBS: C represents the second half of the two-part pubs search strategy, in which the operator searches to find the correct EFX after the X-Ref is known. Node 3.9 starts when the operator looks at the pubs and ends when the operator begins reaching for the keyboard. Modapts-defined seek/search times are used in the node (discussed in detail in the next section).

3.10 1st EFX. Node 3.10 models the operator's hand leaving the pubs after a PUBS: C search and, subsequently, pressing the first digit of the EFX. The movement ends when a numeral (0-9) key is pressed. After the first digit of the EFX is entered, the steps become the same for operators who used either the single pubs search strategy or the two-part search strategy.

4 Designate. Subnetwork 4 represents the steps for designating the emitter with an EFX.

4.1 Shift Key. For the Macintosh prototype, the operator uses the Shift key to turn off the ULQ-16. In the Micro Saint simulation, the movement starts when the operator's hand leaves the pubs and ends when the Shift key is pressed.

4.2 DESIG ID. The movement to press the Designate ID FAB begins when the operator's finger leaves the Shift key and ends when the operator presses the Desig ID key. In all cases, the Shift key immediately precedes the Desig ID key.

4.3 EFX-1. Node 4.3 models the action of entering the first digit of the EFX after performing the PUBS: A search. In contrast to node 3.10, the action begins when the operator's finger leaves the Designate ID FAB. Like node 3.10, the action ends when the number key is pressed.

4.4 EFX-2. The action of entering the second digit of the EFX begins when the operator's finger leaves the key of the first digit and presses the second.

4.5 EFX-3. The action of entering the third and final digit of the EFX number begins when the operator's finger leaves the key of the second digit and presses the third.

4.6 Return Key. In the IT, the final press of the Return key enters the EFX into the system and assigns it to the emitter that was hooked. The motion begins when the operator's finger leaves the third digit of the EFX and ends when the Return key is pressed.

Model Times

The regression of predicted Modapts times for the nine selected activities on actual times showed that the Modapts system accounting for 96 percent of the variance in actual activity times ($p < 0.0001$). The regression is illustrated in Figure 20.

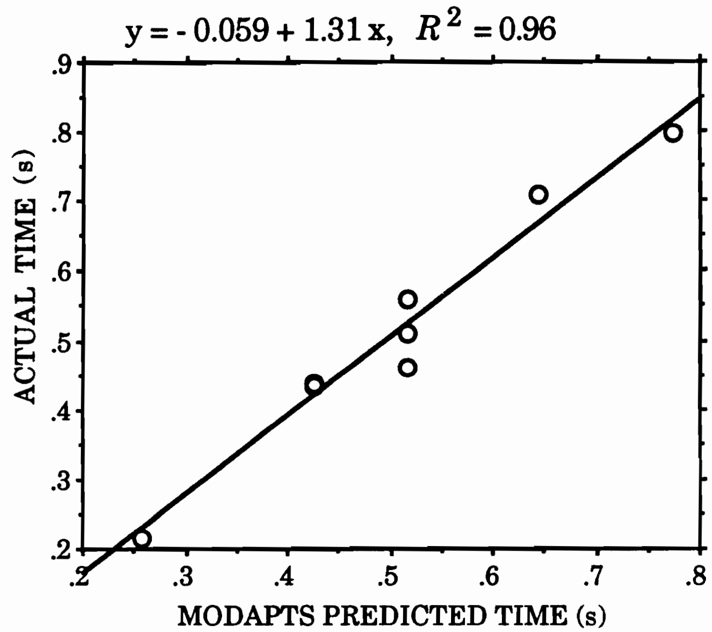


Figure 20. Regression of Modapts times on actual times for nine activities (using Mod = 0.129 sec).

Mod Value. The regression equation at the top of Figure 20 indicates that the slope of the regression line is 1.131. The two-tail t-test, using the mean square residual variance of the regression (0.001), indicated that for $p = 0.05$, the slope was significantly different from unity. The equation for the analysis was:

$$t = \frac{(1.131-1)}{\sqrt{\frac{0.001}{9}}} = 12.428 \qquad \text{where } t_{(0.05/2,\,7)} = 2.365. \qquad (7)$$

A simple regression using the number of Mods for each activity was performed subsequently on the actual times for the activities. This regression produces the identical plot shown in Figure 20, except with a different scale on the abscissa. The value of the slope equals the value of the "best–fit" Mod. The slope for the regression equation (i.e.,. best Mod value) was 0.146. The regression equation is given by:

$$y = -0.059 + 0.146\,x \qquad R^2 = 0.963, \tag{8}$$

where:

y = actual times for the movements, and
x = the number of Mods assigned to each movement.

The overt actions shown in Figure 16 were implemented in the Micro Saint model. Activities that were not directly observable from the tapes were assigned the Modapts equivalent times, or times were assumed from alternate sources (as explained in the detailed descriptions of the nodes). PUBS A, PUBS B, and PUBS C searches were assigned Seek elements based on approximated numbers of half pages and whole pages searched by the operator. Moreover, the number of Reads were added to the times based on the number and type of parameters needed for each portion of the search. Those assignments are included in Appendix E.

Model Distributions

For the movements listed in Figure 15, and for the three different pubs searches, K-S test results showed that lognormal distributions were inadequate ($p \leq 0.000$), but gamma distributions provided good fits ($p > 0.10$) to the distributions of the actual data. The two-tail probabilities for each of the tests are listed in Table 19. A comparison of gamma frequency distributions against actual data frequency distributions are illustrated in Appendix F.

TABLE 19
Summary of Kolmogorov-Smirnov Test Results for Movement and Search Distributions: Lognormal and Gamma

Movement	Number of cases	p	
		Lognormal	Gamma
Grasp Mouse	117	< 0.001	0.998
Press 2nd & 3rd EFX numbers Return key	290	< 0.001	0.632
Leave Shift key and press Designate ID FAB	110	< 0.001	0.994
Leave pubs and press Shift Key	114	< 0.001	0.773
Leave Sig Sel FAB and grab Pubs	114	< 0.001	0.449
Leave Return key and grab Mouse	114	< 0.001	0.367
Leave Mouse and press Sig Sel FAB	114	< 0.001	0.773
PUBSA search	87	< 0.001	0.747
PUBSB search	29	< 0.001	0.738
PUBSC search	29	< 0.001	0.994

Model Standard Deviations

Original data files used for creating the time values of Modapts were unavailable; hence, the data are not available for determining standard deviations (G. C. Heyde, personal communication, September, 1991).

Existing databases created for Aegis systems were sought. Osga (1987) collected human performance data for the purpose of describing the human-computer interaction within command communication centers aboard ship. Osga recognized the importance of creating a database for facilitating the human-computer interface in workstation design and he used Micro Saint models to compare different Aegis systems.

Osga (1987) established an approach for collecting data for the purpose of creating a design database, but his work has limited application. The data collected for Osga's work were collected by computer; that is, performance times were collected from some input by the operator into the system until the next input. Osga lists these start and stop points of each activity, but the location of buttons/keys in relation to each other was not readily apparent. Therefore, it was not possible to determine: (1) body motions involved in the reported activities, (2) sensory activities used in completing the reported activities, and (3) information regarding other non-button push activities internal to the reported activities. In short, movement time data presented in Osga (1987) are valuable for comparing Aegis systems already studied and built, but are limited in terms of application to unknown systems unless the simulation model-builder is familiar with the Aegis systems.

It was impossible to calculate standard deviations from the original Modapts data, and an adequate source of movement times (and standard deviations) was not found in literature pertaining to related displays and controls consoles. Therefore, the actual data collected from the participants were assigned. Standard deviations for each model node are included in Appendix E.

Workload Methodology

Scaling solution. A Kendall's Coefficient of Concordance of 0.77 was computed, indicating substantial agreement among participants regarding the ordering of the cards and, thus, the relative importance of the three dimensions. Reid, Potter, and Bressler (1989) specified that for an obtained coefficient of concordance greater than 0.75, the group can be treated as a homogeneous group. Therefore, a single group SWAT scale was used. The relative importance of the three dimensions was found to be 31.33% for the Time dimension, 27.50% for the Effort dimension, and 41.18% for the Stress dimension. The group scaling solution is shown in Appendix G.

Workload scores. Regression of the predictors Format, Color condition, Symbol set, Density, mean time to process an emitter, and mean number of hook errors on mean workload scores obtained by Group 1 accounted for 79 percent of the variation of mean workload scores (this is the R^2 adjusted to account for the number of factors used in the regression). The results from the multiple regression are listed in Table 20.

The obtained beta coefficients and intercept were built into the IT model for generating workload scores. All factors were used in estimating the workload score even though only density and format were significant ($p = 0.05$). No further attempt was made to uncover variables influencing the workload scores.

TABLE 20
Multiple Regression: Workload Scores

Count	r	R^2_{adj}	
24	0.89	0.79	$p < 0.0001$

Variable	Beta (obtained)	Beta (Std)	p
FORMAT	0.014	0.014	0.9352
SYMBOL	-0.237	-0.237	0.0990
COLOR	-0.036	-0.036	0.8124
DENSITY	0.803	0.803	0.0030
TIME	0.037	0.037	0.8555
ERRORS	0.047	0.047	0.8263
INTERCEPT	23.538		

Model Verification

The linear regression through the origin of mean performance scores generated by the model accounted for 78 percent of the variance in performance scores obtained from the participants (p<0.0001). Figure 21 presents the best–fit line for the regression. The following t-test formulation (Montgomery and Peck, 1982) for hypothesis testing in linear regression was used to determine if the regression slope (1.06) was significantly different from unity:

$$t = \frac{B_1 - 1}{\mathrm{se}(B_1)}. \qquad (9)$$

In this equation, B_1 is the slope of the regression equation, 1 is the value (unity) that the slope is tested against, and se(B_1) is the standard error of B_1. The absolute value of the calculated t-statistic (t = 1.50) did not exceed the tabled t-value for $t_{(0.05/2,\ 22)}$, indicating that the slope was not significantly different from unity.

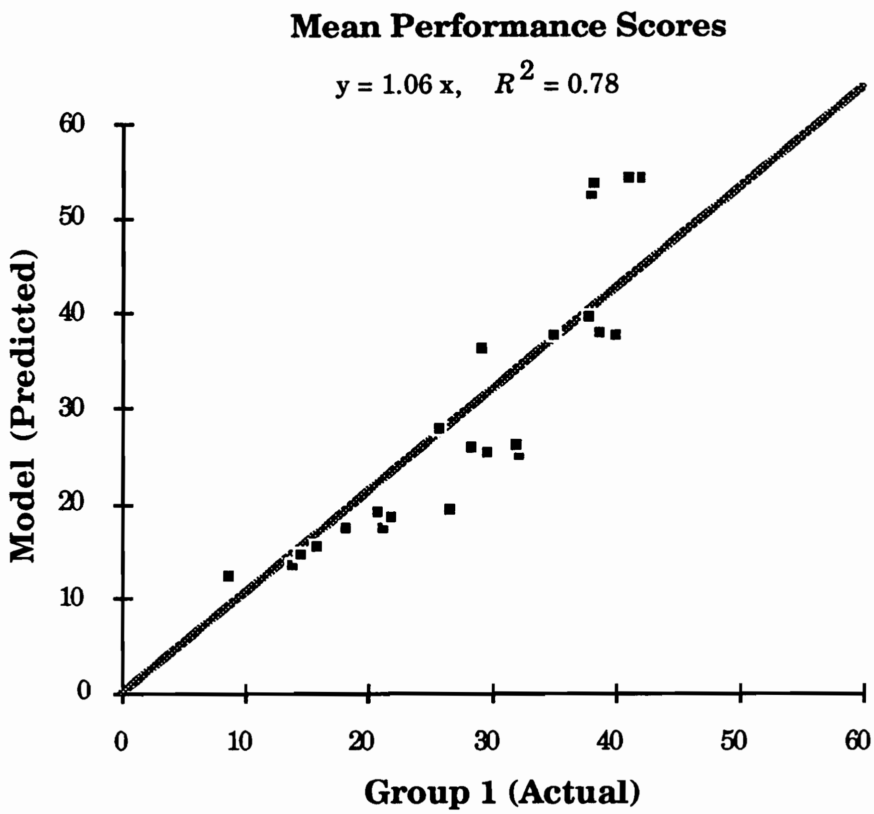


Figure 21. Regression of modeled mean performance scores on Group 1 mean actual performance scores across all 24 display configurations.

Regression through the origin of model mean workload scores on Group 1 mean workload scores across all 24 configurations is illustrated in Figure 22. Mean workload scores generated by the model accounted for 74 percent of the variance in performance scores obtained from Group 1 ($p <$ 0.0001).

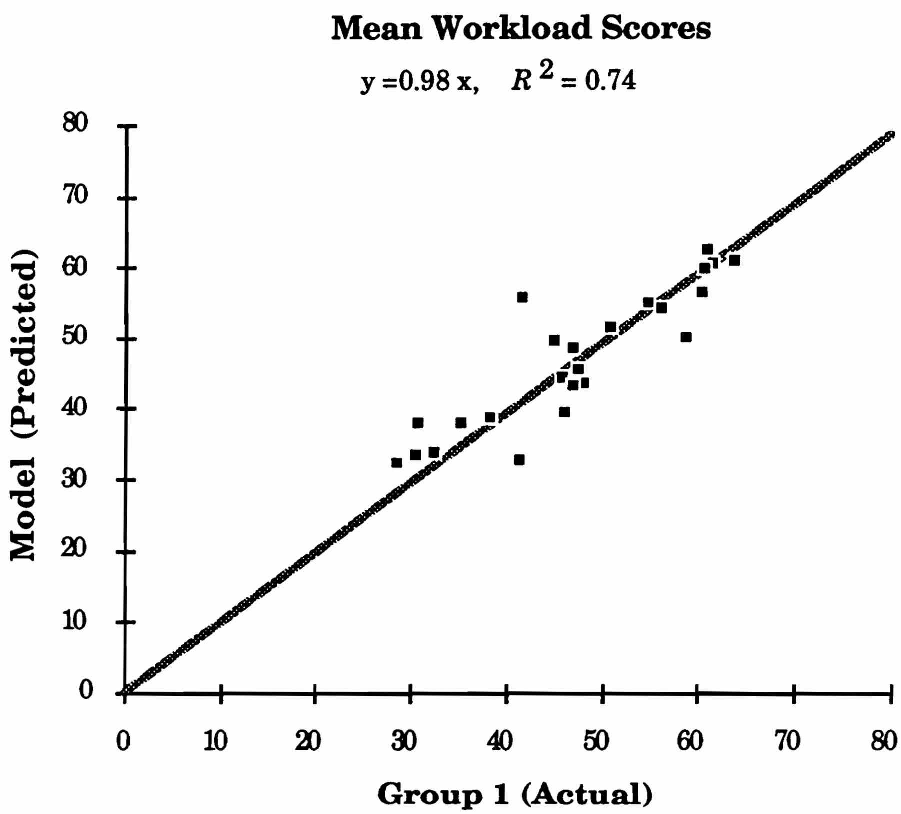


Figure 22. Regression of model mean workload scores on Group 1 mean workload scores across all 24 display configurations.

Using Equation 9 to test the workload regression slope (0.98) against unity, the absolute value of the calculated *t*-statistic (t = 1.0) did not exceed the tabled *t*-value for $t_{(0.05/2,\ 22)}$, indicating that the slope was not significantly different from unity.

Sensitivity analysis. R^2 values for regressing the model generated performance scores on the Group 1 performance scores were calculated for each of three different weighting schemes used in defining the performance variable (Equation 1). The results are shown in Table 21.

Comparisons between the R^2 values using z-tests (Equation 2) on the R^2 values did not uncover significant differences between the three weighting schemes (p = 0.05). The z-tests results are shown in Table 22.

TABLE 21

$R2$ values for Model Performance Scores regressed on Group 1 Performance scores using three different weights in the Performance Metric

Weighting Scheme	Time Weight (b_1)	Error Weight (b_2)	R^2 (Model Performance Scores on Group 1 Performance Scores	p
1	0.75	0.25	0.836	<0.0001
2	0.50	0.50	0.802	<0.0001
3	0.25	0.75	0.720	<0.0001

TABLE 22

Comparison of R^2 Values of Model Performance Scores on Group 1 Performance Scores for Three Weighting Schemes

Compared Weighting Schemes	z
1 - 2	0.0323
1 - 3	0.0573
2 - 3	0.0376

Model Validation

Linear regression through the origin of mean performance scores generated by the model on the mean performance scores obtained from Group 2 across all 24 configurations is shown in Figure 23. The results of the regression indicate that mean performance scores generated by the model accounted for 78 percent of the variance in performance scores obtained from the Group 2 ($p < 0.0001$).

Using the t-test in Equation 9, the t-statistic was calculated to compare the slope of the regression (1.07) against unity. The calculated t-statistic ($t = 1.75$) did not exceed the tabled t-values for $t_{(0.05/2,\ 22)}$, thus indicating that the slope was not significantly different from unity.

Linear regression through the origin of mean workload scores generated by the model on the mean performance scores obtained from Group 2 across all 24 configurations is shown in Figure 24. The results of the regression indicate that mean performance scores generated by the model accounted for 76 percent of the variance in performance scores obtained from the Group 2 ($p < 0.0001$).

Using the t-test in Equation 9, the t-statistic was calculated to compare the slope of the regression (0.86) against unity. The absolute value of the calculated t-statistic ($t = -7.0$) did exceed the tabled $t_{(0.05/2,\ 22)}$ value, thus indicating that the slope was significantly different from unity.

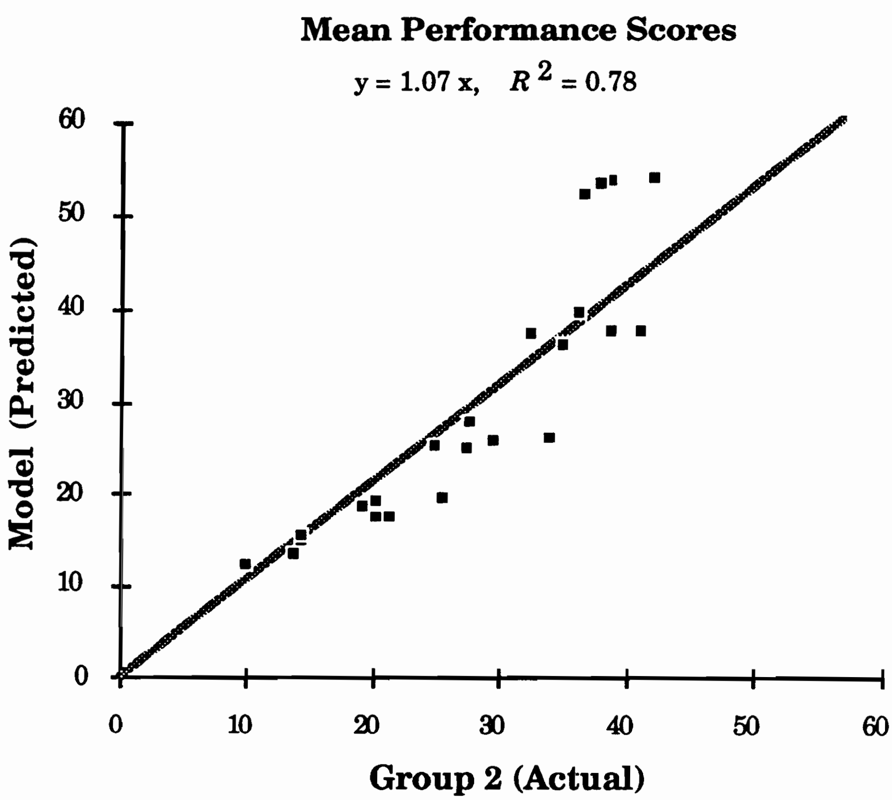


Figure 23. Regression of model mean performance scores on Group 1 mean performance scores across all 24 display configurations.

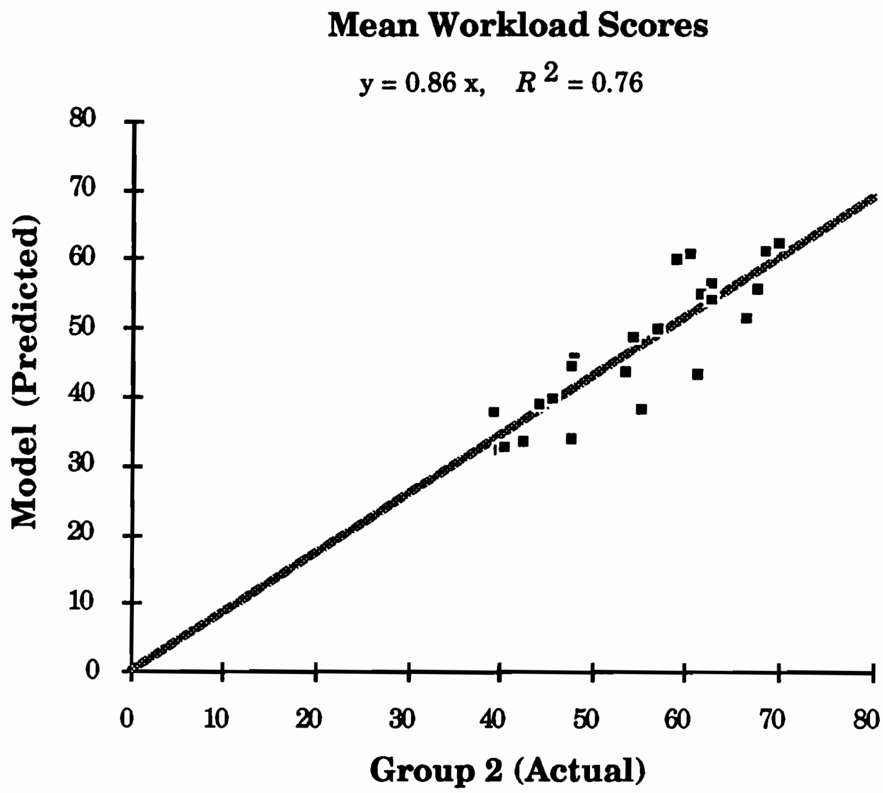


Figure 24. Regression of model mean workload scores on Group 2 mean workload scores across all 24 display configurations.

Cross validation. Correlation analyses between Groups 1 and 2 on mean performance scores and mean workload scores across all 24 configurations indicated that there was a linear relationship between Group 1 and Group 2 scores ($r = 0.82$, and $r = 0.72$, respectively).

A regression through the origin was calculated for regressing Group 2 mean performance scores on Group 1 mean performance scores. The equation for that regression best–fit, illustrated in Figure 25, shows that the Group 1 and Group 2 are very similar ($R^2 = .98$). Using Equation 9, a t-statistic comparing the slope of the regression against unity was calculated ($t = -1.0$). The absolute value of the calculated t-statistic did not exceed the tabled value for $t_{(0.05/2,\ 22)}$, indicating that the slope of the regression line was not significantly different from unity.

A regression through the origin was calculated for regressing Group 2 mean workload scores on Group 1 mean workload scores. The equation for that regression best–fit, illustrated in Figure 26, shows that the Group 1 and Group 2 are not similar ($R^2 = 0.19$) with respect to workload ratings. Using Equation 9, a t-statistic comparing the slope of the regression against unity was calculated ($t = 3.61$). The absolute value of the calculated t-statistic did exceed the tabled value for $t_{(0.05/2,\ 22)}$, indicating that the slope of the regression line was significantly different from unity. Thus it can be concluded that Groups 1 and 2 were dissimilar in workload ratings across the 24 configurations.

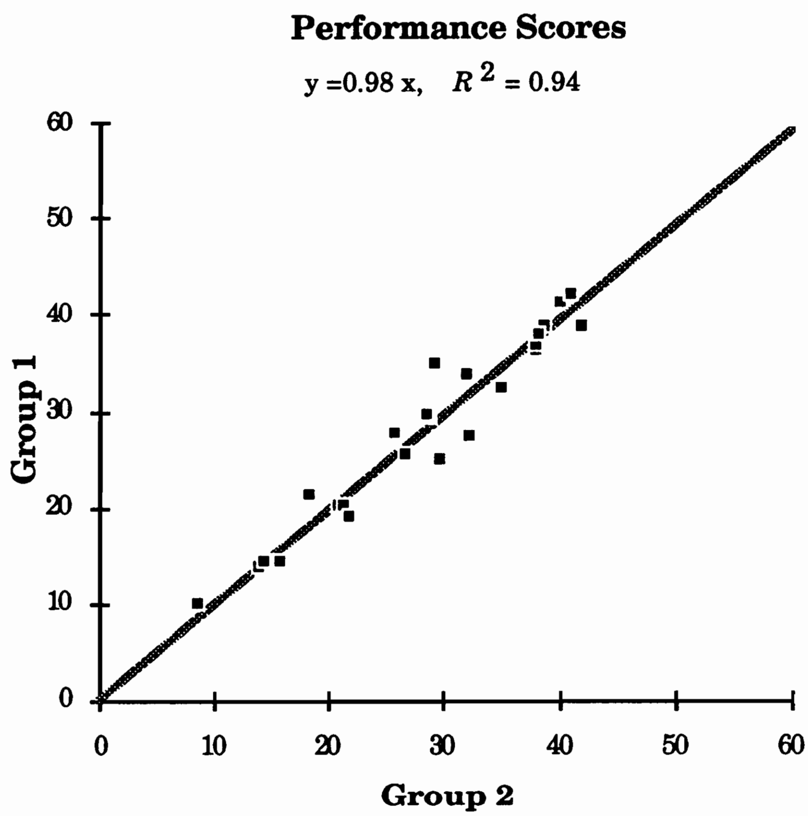


Figure 25. Regression of Group 2 mean performance scores on Group 1 mean performance scores across all 24 display configurations.

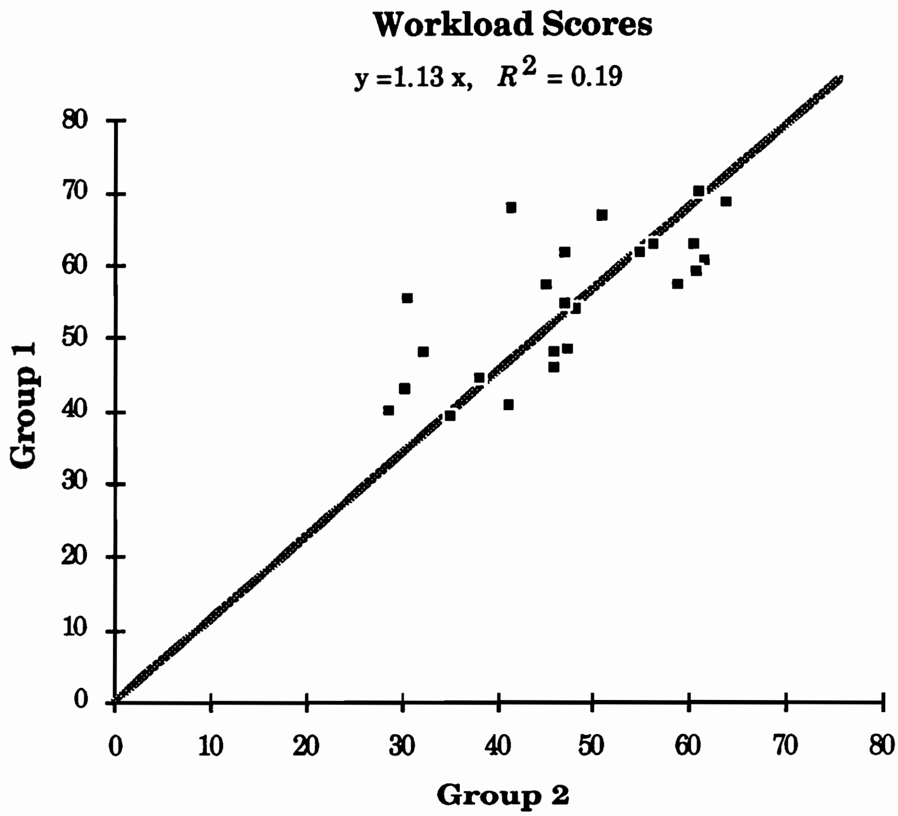


Figure 26. Regression of Group 2 mean workload scores on Group 1 mean workload scores across all 24 display configurations.

To take advantage of the data collected, the IT model was changed slightly to modify the regression equations for determining errors and error standard deviations using the data for all 12 participants. The regression equation showed that the variables listed in Table 21 accounted for 73 percent of the variance in mean number of errors ($p < 0.0001$). The error function used in the IT model was adjusted to reflect the obtained regression intercept and beta coefficients shown in Table 23.

TABLE 23
Multiple Regression: Errors (n = 12)

Count	r	R^2_{adj}	
24	0.88	0.73	$p < 0.0001$

Variable	Beta (Obtained)	Beta (Std)	p
FORMAT	-3.988	-0.664	< 0.0001
SYMBOL	-1.479	-0.242	0.0362
COLOR	-1.456	-0.246	0.0389
DENSITY	1.697	0.461	0.0005
INTERCEPT	3.456		

The regression equation using the variables of format, symbol, color, and density on the mean error standard deviations across all 24 configurations accounted for 65 percent of the variation in standard deviations obtained from all 12 participants ($p = 0.0003$). The error standard deviation regression results are shown in Table 24. The function for determining error standard deviations in the model was adjusted to reflect the obtained regression intercept and coefficients shown in Table 24.

TABLE 24
Multiple Regression: Error Standard Deviations (n = 12)

Count	r	R^2_{adj}	
24	0.81	0.59	$p < 0.0001$

Variable	Beta (Obtained)	Beta (Std)	*p*
FORMAT	-2.095	-0.593	0.0003
SYMBOL	-0.839	-0.046	0.0919
COLOR	-0.163	-0.237	0.7332
DENSITY	1.081	0.500	0.0014
INTERCEPT	1.944		

The models were supplemented with the equations for errors and standard deviations for all participants, then run 144 times in each of the 24 configurations. Random number seeds were generated using a random number generator. Regressions were performed on the actual and modeled mean performance scores indicating that the model accounted for 76 percent of the variability in mean performance scores obtained from the participants (p=0.0001). The regression plot is shown in Figure 27. The slope of the best-fit regression equation was not significantly different from unity (t = 1.75) at p = 0.05. Simple observation comparing the R^2 for this regression indicates that the model did as well representing all 12 participants as it did representing Group 1 or Group 2 in performance.

The model refined to reflect all 12 participants accounted for 84 percent of the variability in mean workload scores obtained from all 12 participants ($p < 0.0001$). The regression plot is shown in Figure 28. The slope of the best–fit regression equation was not significantly different from unity ($t = -1.0$) at $p = 0.05$. A summary of the actual data for the 12 participants and the model-generated data is listed in Table 25.

An illustration of the comparison of configurations using the scores from the model of all 12 participants is shown in Figure 28. For consistency, the mean workload scores were inverted (e.g., subtracted from 100). Figure 29 shows that in terms of color, symbol set, and format, the difference between configurations is more pronounced for performance than for workload.

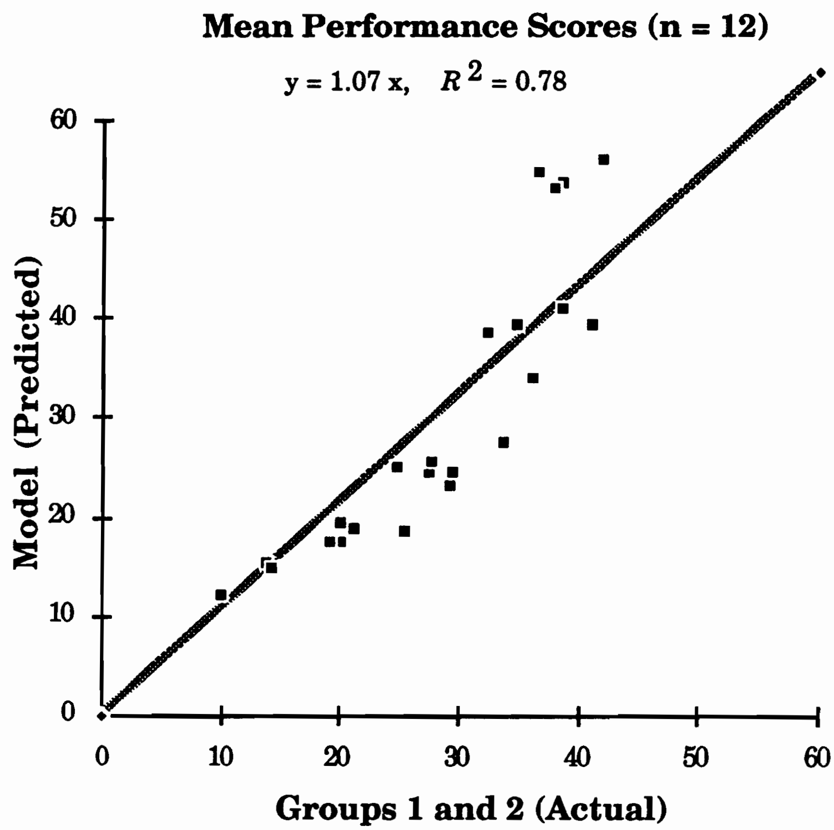


Figure 27. Regression of mean modeled performance scores on mean actual performance scores for Groups 1 and 2 across all 24 configurations (n = 12).

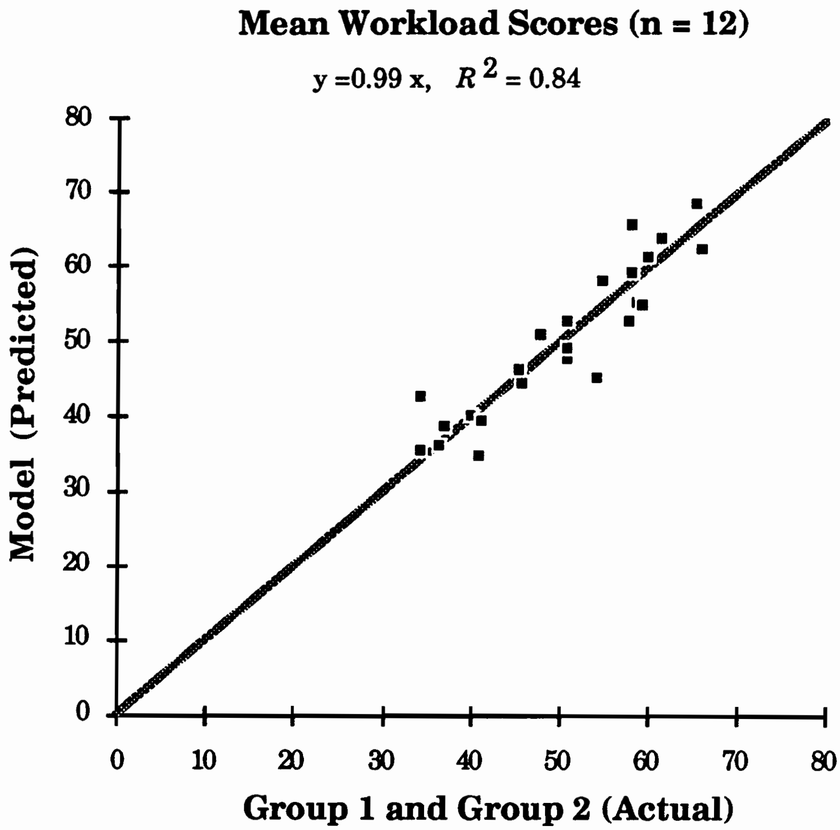


Figure 28. Regression of mean modeled workload scores on mean actual workload scores for Groups 1 and 2 across all 24 configurations (n = 12).

TABLE 25

Summary of Mean Performance and Workload Scores (n=12)

Format	Symbol	Color	Density	Performance		Workload	
				Actual	Model	Actual	Model
Polar	Old	B/W	Car	20.171	19.475	45.91	44.45
Polar	Old	B/W	Gulf	13.797	15.247	58.72	55.02
Polar	Old	B/W	Arm	10.029	12.152	65.43	68.20
Polar	Old	C	Car	27.481	24.351	34.32	42.55
Polar	Old	C	Gulf	21.289	18.882	50.89	52.58
Polar	Old	C	Arm	14.406	14.838	58.22	65.49
Polar	New	B/W	Car	29.245	23.249	40.12	40.16
Polar	New	B/W	Gulf	20.186	17.591	47.82	50.60
Polar	New	B/W	Arm	14.276	14.877	61.57	63.49
Polar	New	C	Car	36.205	33.994	34.33	35.46
Polar	New	C	Gulf	24.850	24.917	50.89	47.71
Polar	New	C	Arm	19.119	17.402	58.22	59.02
Range	Old	B/W	Car	38.771	41.007	41.33	39.46
Range	Old	B/W	Gulf	33.883	27.394	57.86	52.49
Range	Old	B/W	Arm	25.439	18.674	66.13	62.37
Range	Old	C	Car	36.721	54.693	37.21	38.54
Range	Old	C	Gulf	41.172	39.202	50.83	49.10
Range	Old	C	Arm	27.692	25.615	59.87	61.05
Range	New	B/W	Car	38.662	53.592	36.56	36.15
Range	New	B/W	Gulf	34.944	39.296	45.41	45.95
Range	New	B/W	Arm	29.563	24.335	54.76	57.78
Range	New	C	Car	42.038	56.005	40.98	34.73
Range	New	C	Gulf	37.943	52.947	54.28	45.19
Range	New	C	Arm	32.456	38.397	59.52	54.57

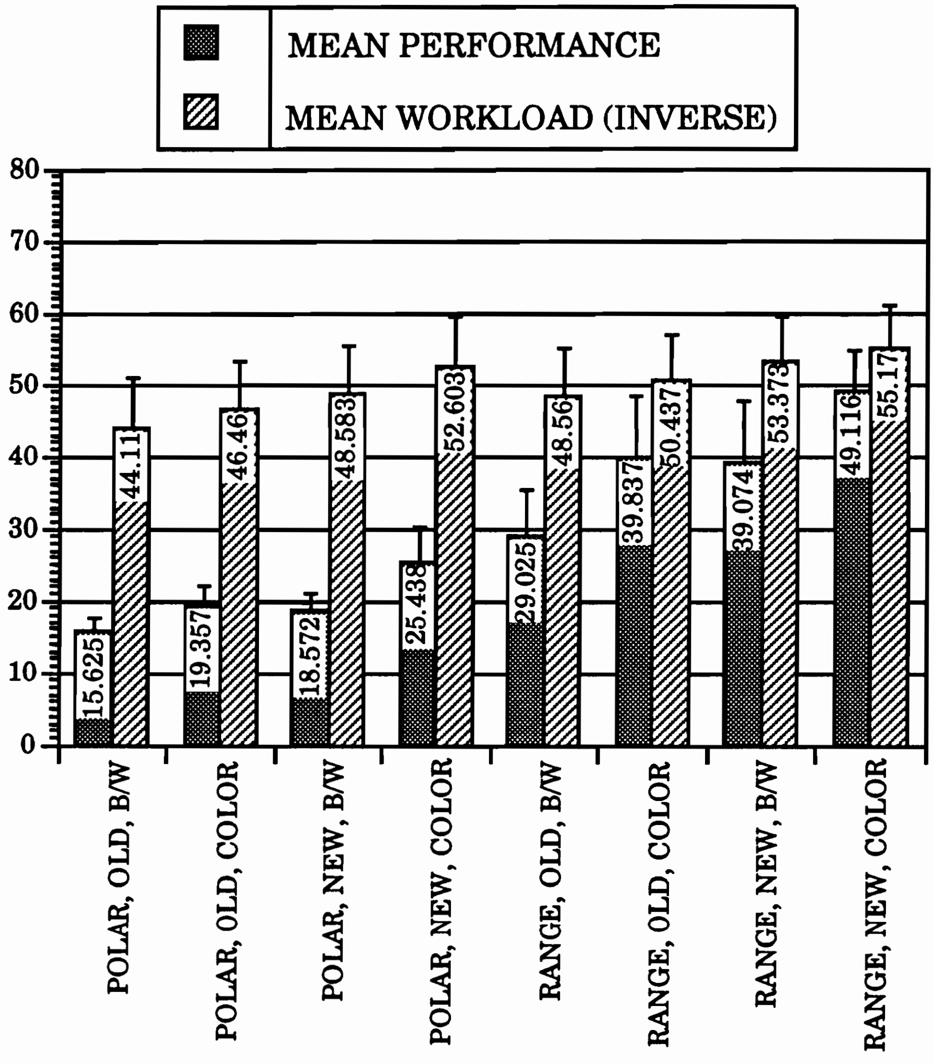


Figure 29. Summary of mean performance scores and mean workload scores generated by the IT Micro Saint model. Scores are collapsed over density levels, and standard error bars are shown.

DISCUSSION AND CONCLUSIONS

Overview

This research developed and tested the methodology for implementing task completion times into a task network model using a PTS. Moreover, this work examined and the feasibility of implementing a workload metric into the task network model. The methodologies were based on modeling prototypes of a naval display and control console. The generalizability of the findings are made with the knowledge that this methodology was accomplished by modeling a Macintosh *prototype* of the displays and controls console. This discussion specifically addresses issues of the *methodology*, and generally address design impacts for the AN/SLQ-32(V) system.

The discussion of this dissertation research will be separated into three sections, as follows:

(1) Time Methodology,

(2) Workload Methodology, and

(3) General Discussion of Impact.

Time Methodology

In the context of the IT, times for individual movements or activities defined by the Modapts system and those pulled from actual data indicated a high correlation (R^2 = 0.96). This suggests that for the individual movements included in the IT, the Modapts system was effective as a method for assigning completion times.

With respect to mean performance scores, results of the verification analysis of the Micro Saint Model indicated that the model of the IT task was able to account for a large portion of variation in scores ($R^2 = 0.78$). Because the slope of the regression equation best-fit line did not differ significantly from unity ($p = 0.05$), the high R^2 indicates that the model was effective in simulating Group 1 performance. Since the model was constructed using the basic principles and rules of the Modapts philosophy, the effectiveness of the model in representing the operator supports using the Modapts methodology.

The performance variable was designed to ascertain operator effectiveness. The metric is based on both normalized mean time to process an emitter and normalized mean number of hook errors. Equal weights were assigned to each. A sensitivity analysis revealed that varying the weights did not significantly affect the R^2 values between actual and predicted scores ($p = 0.05$), thereby eliminating the possibility that the R^2 values obtained in comparisons reflected an artifact created by the weighting scheme.

The true utility of the time methodology is the ability of the methodology to predict performance, thus eliminating the need for expensive prototypes and the associated time required for such a method. With respect to mean performance scores (as defined in Equation 1), the IT model accounted for 78 percent of the variation in mean performance of Group 2 indicated that the Micro Saint/Modapts methodology was able to account for variability in performance scores of a group of operators other than the ones used to construct the methodology. The slope of the regression best-fit line of modeled mean performance scores on Group mean performance scores did not differ significantly from unity (p = 0.05), revealing that the model was effective in simulating performance.

Regression of Group 2 mean performance scores on Group 1 mean performance scores showed that the two groups were homogenous with respect to performance across the 24 configurations (R^2 = .0.98). The slope of the best-fit regression equation was not significantly different from unity (p = 0.50), thus supporting the interpretation of the R^2 value. The similar R^2 values achieved by regressing the model mean performance scores on Group 1 and Group 2 (0.79 and 0.78, respectively) appears to corroborate the effectiveness of the model.

While the IT model was verified and validated in terms of performance, it is necessary to examine the advantages, disadvantages, and limitations of the Modapts system for use in modeling displays and controls consoles.

For the displays and controls arena, the efficiency of the Modapts methodology in specifying the physical motions of a task is the key advantage of its use. The limitations of the methodology extend to specific parameters for controls and, of course, to the assumptions of the Modapts system.

For the physical movements required in the IT, Modapts values were assigned readily. However, the methodology is not tailored to the detailed levels of the perceptual and cognitive processes relative to the issues of displays–issues not related to location of targets. Specifically, the times required to search the display in various configurations of format, color, symbol set, and density (nodes 2.1.1.1, 2.1.1.2, and 2.2.1) were obtained by interpolating the search times from the overt actions. Modapts asserts that complex tasks must occur in series. The interpolation of search time assumed that the operator did not overlap searching for a target with moving the mouse to the target. This is a reasonable assumption.

The models were built intentionally to avoid a level of abstraction that would involve theoretical underpinnings of how people search or how the visual system processes information. The intent was to use a more pragmatic level, where different displays and controls consoles could be evaluated empirically for design or redesign. At this level of modeling, appropriate data do not exist in a database, nor can they be accounted for by the Modapts methodology. This point brings two issues to light.

First, it is known that effects of color on performance are not consistent. Color research is vast and varied in terms of effects on performance. In practice, the use of color in a display can have beneficial effects or can deteriorate performance, depending on a number of factors. Color must be judiciously used in displays and controls consoles, with as few levels as necessary. In addition, color should only be reserved as a redundant code and implemented in a manner which logically facilitates the operators' tasks. However, the methodology used in the IT example is limited in its generalizability. For example, it is not known whether the use of a set of symbols coded with two levels of color provides better or worse performance than one that uses four. The methodology does not allow for direct examination of the impact of color implementation. Along these same lines, the methodology does not generalize to density levels that were not defined in the experimental conditions.

Shortcomings of the methodology related to limited generalizability raise a second issue. That is, the need for a more complete database for visual tasks to supplement the methodology. Sources such as the *Engineering Data Compendium (1988)* and various technical reports for different display sizes, symbols, densities, information content, present data in which the search times are internal to a task completion time. While they are useful in the present application, adequate information is not always given to subtract out the non-relevant portions of the time.

Alternately, data could be obtained in small studies which would not require the construction of complete prototypes. Illustrating this point within the AN/SLQ-32(V) redesign study, the Micro Saint models could have been built without the level of fidelity presented in the Macintosh prototypes (had the methodology been validated previously). Instead, a small study that required only a set of short, simple tests for hooking could have been completed and used to supplement the nodes of the model that could not have been derived from the Modapts methodology. The smaller study would also have provided the data for the anticipated hook errors that were associated with each display configuration. Thirty hours of participation by each participant could have been reduced to approximately eight hours.

In terms of manual controls, the methodology does not provide for times to use different input devices. The assigned time to use a mouse for selecting a chosen target was formulated on research done by Card et al. (1983). Based on Fitt's Law, Card et al. provided several equations for estimating the amount of time it takes to select text using various input devices. While various other input devices have associated equations for completion time, there is no specification for control/display ratios. This is offered more as a limitation of the level of modeling abstraction, rather than as a limitation of using the methodology.

In the present study, the chosen measurement of performance was a function of both mean time to process an emitter and mean number of hook errors. Although variance data were not used in this measure, there are often occasions when the variance data are important in making decisions. It could be the case that a more reliable (or consistent) configuration produces a slightly inferior mean performance. Modeling work with task network simulations has been considered by the Navy as a means for establishing Measures of Efficiency (MOEs) for operators (Osga, 1989). In this application, it is especially prudent to know variance information.

Unfortunately, one of the major disadvantages of the proposed methodology is the lack of standard deviation data. Osga's (1989) compilation of times and standard deviations for specific activities associated with various Aegis systems is a step toward building a database. However, due to the specificity to the systems used, and due to a lack of information to convert the activities into a common metric that could be used for other displays and controls consoles, the data have limited utility for other systems. The Modapts methodology offers one standard method for defining activities so that standard deviation data can be accumulated to build a generalizable database.

The Mod value appears to be very reliable. After a random sampling of movements in the context of the AN/SLQ-32(V) prototype showed that an allowance factor could be applied, the Micro Saint model using the original value produced a better fit (by observation). The Mod value suggested by a regression analyses (0.146) was suspect for two reasons. First, due to the nature of the task, it was reasonable to assume the motivation, and thus speed, of the operator would yield a value equal to, if not be less than, the established Mod value. Second, the Mod value was based on a vast database; hence, using something other than the established Mod value would need substantial support. It is interesting to note that the established Mod value falls within the 90 percent confidence interval around the regression-derived Mod value. The bounds of the 90 percent confidence interval are 0.126 – 0.166.

Workload Methodology

The workload metric in the IT Micro Saint metric was formulated on the beta coefficients produced by regressing the variables of format, symbol set, color condition, density, total emitter process time, and number of errors on Group 1 mean workload ratings. The R^2 value, adjusted to account for the six factors, was 0.78, indicating that using the regression to model operator workload scores accounts for a significant portion of variation in the Group 1 workload scores. In this study, the fit did not warrant further investigation for integrating workload factors into the model. In other words, the regression equation proved satisfactory for modeling workload for relative comparisons. For absolute analyses, accuracy of the regression needs to be assessed by investigating the slope and intercept of the regression model.

Verification of the workload metric produced by the model was accomplished using a regression through the origin of model mean workload scores on Group 1 mean workload scores. The results indicated that the model accounted for a significant portion of the variation in workload scores given by Group 1 ($R^2 = 0.74$). The slope of the best-fit regression equation was not significantly different from unity ($p = 0.05$) indicating that the model mean workload scores were good estimators of the workload scores obtained from Group 1.

The validation of the workload metric using a regression through the origin of mean model workload scores on Group 2 workload scores showed that the workload metric accounted for less of the variation in obtained workload scores for Group 2 than for Group 1 ($R^2 = 0.76$). However, testing the slope of the regression best-fit line uncovered that the slope was significantly different from unity ($p = 0.05$). This indicates that a linear relationship exists between the mean model workload scores, but that the magnitude of the scores were not accurate or that the scores at the end of the range covered were not consistent between the actual and model obtained scores. Therefore, the R^2 values obtained from regressions of the model on Group 1 and Group 2 cannot be compared directly.

The cross-validation of workload scores accomplished by regressing Group 2 mean workload scores onto Group 1 mean workload scores across all 24 configurations uncovered a small correlation between Group 1 and Group 2 with respect to mean workload scores ($R^2 = 0.19$). In addition, the slope of the best–fit regression line was significantly different from unity ($p = 0.50$). These findings indicate that either the methodology of determining workload was not adequate or that the groups were heterogeneous in terms of workload ratings. The latter explanation is supported by the small number of participants in each group.

A comprehensive workload investigation would include a series of experiments in which the same IT task is performed but with different factors varied. Examples include using a larger screen, automating the publications for on–line searches, and changing the way the operator manipulates the interface. These variables provide data that allow factors such as number of motions and number of times the operator had to hold items in short–term memory to be examined.

There is merit to using the regression equation technique in a predictive model by identifying beta weights for variables using projective workload techniques. There are several projective techniques in existence (Vidulich, Ward, and Schueren, 1991). A projective version of the SWAT (Pro-SWAT) has been used with some success by having raters predict ratings they would assign based on a description of the task (Reid, Shingledecker, Hockenberger, and Quinn, 1984). These projective ratings could be used to "predict" the workload.

Impact

The configurations of the AN/SLQ-32(V) used in the present study were compared in order to identify redesign options for the actual system. The objective of the experiment was to identify the best configuration among the alternatives. The nature of the IT requires the operator to quickly process an emitter in any tactical environment. In other words, the operator may have as few as 40 seconds to process an incoming missile whether he is in a dense emitter environment (e.g., Armageddon) or in a relatively void emitter environment (e.g., Caribbean). With respect to the IT, the configuration which produces the best performance and the least amount of workload is the optimal selection.

After the model had been refined to reflect all 12 participants, the model accounted for 84 percent of the variability in mean workload scores ($p < 0.0001$). The slope of the best–fit regression equation was not significantly different from unity ($p = 0.05$), indicating that the poor fit in the validation stage was due in part to the small number of participants and the differences that existed between Groups 1 and 2 in terms of workload rating scores.

The regressions between actual and modeled performance and workload scores produced high R^2 values with the slope of the best–fit regression lines not significantly different from unity indicate that either actual data or model generated data scores can be used to choose among the different design configurations ($R^2 = 0.78$ and 0.84, respectively).

Observation of mean performance scores and mean workload scores for each configuration, collapsed over density levels, indicates that the configuration consisting of the Range format, New Symbol set, and Color condition is the best alternative (Table 26 and Figure 29).

The intended purpose of creating the time–and to a lesser extent workload–methodologies is to allow a designer to quickly create simulation models to estimate performance and, thereby, make design decisions. Subsequently, the designer would use either inspection or statistical analyses to investigate the impact of the independent variables which comprised the different configurations. For the sake of completeness, such analyses (i.e., 2 x 2 x 2 x 3 within-subjects Analysis of Variance) are included in Appendix H for Performance and Appendix I for Workload. Because the results of the analyses are not the direct topic of this research, post hoc tests and subsequent discussions are left for other research efforts; however, means tables and interaction plots for significant effects ($p > 0.10$) are included in Appendices H and I.

Figure 29 emphasizes the fact that workload is not the same as performance, nor is it synonymous to performance (Linton, Plamondon, Bittner, and Christ, 1989). By definition, SWAT scores are relative ratings based on rank-ordering workload dimensions. The performance measure is a normalized metric which combines hook errors and time to complete a task. In practice, neither metric is suitable for use as an absolute criterion.

The goal of this research effort was twofold. The first goal was to develop and validate the Micro Saint/Modapts methodology. In terms of performance (as defined in this work), this goal was accomplished. The second goal was to explore a methodology for integrating workload parameters into Micro Saint task network models. The goal was met with limited success. Mean workload scores were attributed largely to density condition and the total time to process and emitter. The implementation of a workload regression equation into the model can be accomplished, but offers little information that could not be computed by the regression alone.

The direct application of this research impacts the ability to make design decisions for the AN/SLQ-32(V) DCC. The Micro Saint/Modapts methodology has now been validated. The next step in this line of work is to adapt the model to simulate the actual DCC and not the Macintosh prototype. Several proposed redesign options can be examined for comparative analyses, such as automating the library, and changing the manipulation style of the DCC, without all the time, money, and labor that was invested in the building prototypes and training participants. Approximately 50 percent of the time to process an emitter (about 12 out of 20-24 seconds) are used to search the publications, suggesting that automating the publications is a likely opportunity for improvement of the AN/SLQ-32(V).

To the extent that the modeling methodology explored in this dissertation is developed, potential future applications are not limited to design and redesign recommendations. The logic of the task network could be incorporated into training programs, specifically automated training.

REFERENCES

Biferno, M. A. (1985). *Mental workload measurement: Event-related potentials and ratings of workload and fatigue*. (NASA CR 177345) Washington D.C.: National Aeronautics and Space Administration.

Beaton, R. J., Dyess, B., Miller, R. H., and Moscovic, S. A. (1991, July). *Phase A report – Human engineering evaluation of AN/SLQ–32(V) display device console: Critique of existing system.* Technical Report. Displays and Controls Laboratory Blacksburg, VA: Virginia Polytechnic Institute and State University.

Beaton, R. J., Dyess, B., Miller, R. H., and Moscovic, S. A. (1992, April). *Phase B final report – Human engineering evaluation of AN/SLQ–32(V) display device console: Redesign of existing system.* Technical Report. Displays and Controls Laboratory Blacksburg, VA: Virginia Polytechnic Institute and State University.

Blake, B. (1991). *Jane's radar and electronic warfare systems (3rd Ed.) 1991-1992*. Surrey, UK: Jane's Information Group.

Boff, K. R., and Lincoln, J. E. (1988). Human performance reliability. In *Engineering data compendium: Human perception and performance* (pp. 1365-1400). OH: Harry G. Armstrong Aerospace Medical Research Laboratory Wright-Patterson Air Force Base.

Bortolussi, M. R., Kantowitz, B. H., and Hart, S. G. (1986). Measuring pilot workload in a motion base trainer: A comparison of four techniques. *Applied Ergonomics, 17*(4), 278-283.

Card, S. K., Moran, T. P., and Newell, A. (1983). *The psychology of human-computer interaction*. New Jersey: Erlbaum.

Chaffin, D. B. (1987). Biomechanical aspects of workplace design. In G. Salvendy (Ed.), *Handbook of human factors* (pp. 601-619). New York: Wiley.

Chubb, G. P., Laughery, K. R., Jr., and Pritsker, A. B. (1987). Simulating manned systems. In G. Salvendy, (Ed.), *Handbook of Human Factors* (pp. 1298-1327). New York: John Wiley and Sons.

Damos, D. L. (1984). Classification systems for individual differences in multiple-task performance and subjective estimates of workload. *Proceedings of the 20th Annual Conference on Manual Control.* (NASA CP 2341). Washington D.C. : National Aeronautics and Space Administration.

Eggemeier, F. T. and O'Donnell, R. D. (1982). A conceptual framework for development of a workload assessment methodology [Summary]. *Proceedings of the 1982 American Psychological Association Annual Meeting*. Washington D.C.: American Psychological Association.

Gartner, W. B. and Murphy, M. R. (1976). *Pilot workload and fatigue: A critical survey of concepts and assessment techniques*. (NASA-TN-D-8365) Washington D.C.: National Aeronautics and Space Administration.

Gerber, D. L. and Heyde, G. C. (1985). Modapts plus: A superior system for work analysis. In *Proceedings of the Annual International Industrial Engineering Conference and Show*.

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

Hart, S. G. (1978). Subjective time estimation as an index of workload. *Proceedings of the Symposium on Man-System Interface: Advances in Workload Study*. Washington D.C.: Air Line Association.

Hart, S. G. (1987). Theory and measurement of human workload. In J. Zeidner (Ed.), *Human productivity enhancement*. New York: Praeger.

Hart, S. G., Battiste, V., and Lester, P. T. (1984). POPCORN: A supervisory control simulation for workload and performance research. *Proceedings of the 20th Annual Conference on Manual Control* (pp. 432-454). Washington D.C.: National Aeronautics and Space Administration.

Hartman, B. O. (1980). *Evaluation of methods to assess workload.* (AGARD-AR-139) France: North Atlantic Treaty Organization Advisory Group for Aerospace Research and Development.

Johanssen, G., Moray, N., Pew, R., Rasmussen, J., Sander, A., and Wickens, C. (1979). Final report of the experimental psychology group. In N. Moray (Ed.), *Mental workload: Its theory and measurement*. New York: Plenum Press.

Karger, D. W., and Bayha, F. H. (1987). *Engineered work measurement (4th ed.)*. New York: Industrial Press.

Kirkpatrick, M., Malone, T. B., Heasly, C. C., and Baker, C. C. (1990). Manpower, personnel, training, safety (MPTS) simulation tools: Network and simulation for workload assessment and modeling (SIMWAM). In *Proceedings of the Human Factors Society 34th Annual Meeting*. Santa Monica, CA: The Human Factors Society.

Krantz, D. H. and Tversky, A. (1971). Conjoint measurement analysis of composition rules in psychology. *Psychology Review, 78*, 151-169.

Laughery, K. R., Jr. (1989). Task network modeling as a basis for analyzing operator workload. In *Proceedings of the Human Factors Society 33rd Annual Meeting* (pp. 110-114). Santa Monica, CA: Human Factors Society.

Laughery, K. R., Jr., and Drews, C. (1984). Micro Saint: A computer simulation designed for human factors engineers. In *Proceedings of the Human Factors Society 28th Annual Meeting*. Santa Monica, CA: Human Factors Society.

Laughery, K. R., Jr., and Drews, C. (1990). An enhanced environment for task network modeling: The development of paradigm - a Macintosh version of Micro Saint. In *Proceedings of the Human Factors Society 34th Annual Meeting* . Santa Monica, CA: Human Factors Society.

Laughery, K. R., Jr., and Hegge, F. (1984). A methodology for mapping laboratory performance research to the effects on military performance. (Available from the Walter Reed Army Institute of Research, Washington D.C.).

Law, A. M., and Kelton, W. D. (1991). *Simulation modeling and analysis* (2nd Ed.). New York: McGraw-Hill and Company.

Linton, P. M., Plamondon, B. D., Dick, A. O., Bittner, A. C. Jr, and Christ, R. E. (1989). Operator workload for military system acquisition. In G.R. McMillan, D. Beevis, E. Salas, M.H. Strub, R.Sutton, and S. Van Breda (Eds.), *Applications of human performance models to system design.* New York: Plenum.

Long, J. L. (1974). A stereo tape recorder technique for observational data. *Human Factors, 16*(2), 154-160.

McCloy, T. M. (1987). Performance group meeting [Summary]. *Proceedings of the Workshop on the Assessment of Crew Workload Measurement Methods, Techniques, and Procedures*, 317-320.

Meister, D. (1985). *Behavioral analysis and measurement methods.* New York: Wiley.

Montgomery, D. C., and Peck, E. A. (1982). *Introduction to linear regression analysis*. New York: Wiley.

Mundel, M. E. (1985). *Motion and time study improving productivity (6th ed.)*. New Jersey: Prentice-Hall.

Nygren, T. E. (1981). *Development of a manual of use for conjoint scaling techniques* [Final Report]. Columbus Ohio: Ohio State University USAF - SCEEE Summer Faculty Research Program.

Nygren, T. E. (1982). *SWAT: User's guide* [Computer program]. Columbus, Ohio: Ohio State University, Department of Psychology.

Nygren, T. E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors, 33*(1), 17-33.

Osga, G. A. (1989). *Measurement, modeling, and analysis of human performance with combat information center consoles*. (NOSC TD 1465). San Diego, CA: Naval Ocean Systems Center.

Rees, R. L. D. (Ed.). (1977). *Performance modelling and prediction 1: Analysis and bibliography*. England: Infotech International.

Reid, G. B., Eggemeier, F. T., and Nygren, T. E. (1982). An individual approach to SWAT scale development. *Proceedings of the Human Factors Society 26th Annual Meeting*, 639-642.

Reid, G. B., Potter, S. S., and Bressler, J. R. (1989). *Subjective workload assessment technique (SWAT): A user's guide (U)* (Technical Report AAMRL-TR-89-023). Wright-Patterson Air Force Base, OH: Armstrong Aerospace Medical Research Laboratory.

Reid, G. B, Shingledecker, C. A., and Eggemeier, F. T. (1981). Application of conjoint measurement to workload scale development. *Proceedings of the Human Factors Society 25th Annual Meeting,* 522-526.

Reid, G. B., Shingledecker, C. A., Eggemeier, F. T., & Nygren, T. E. (1981). Development of multi-dimensional subjective measures of workload. *Proceedings of the 1981 International Conference on Cybernetics and Society*, 403-406.

Reid, G. B., Shingledecker, C. A., Hockenberger, R. L., and Quinn, T. J. (1984). A projective application of the subjective workload assessment technique. *Proceedings of the IEEE 1984 National Aerospace and Electronics Conference* (pp. 824-826). New York: Institute of Electrical and Electronics Engineers.

Reinhart, W. F., Glynn, C. D., Dye, C., Takahama, M., and Snyder, H. L. (1988). *The role of short-term memory in operator workload*. (AAMRL-TR-88-024). Dayton, OH: Armstrong Aerospace Medical Research Laboratory.

Sheridan, T. B., and Simpson, R. W. (1979). *Toward the definition and measurements of the mental workload of transport pilots*. (FTL R-79) Massachusetts: Massachusetts Institute of Technology Flight Transportation Laboratory.

Shinnick, M. D. (1985). Improving the effectiveness of rehabilitation services through a common language for practitioners. *Journal of Rehabilitation*, July-August-September, 33-38.

Shinnick M. D. (1987). *Modapts Plus resource manual*. Blacksburg, VA: Dynamics Research Group.

Shinnick, M. D., and Erwin, W. W. (1989). Work measurement system creates shared responsibility among workers at Ford. *Industrial Engineering*, August 89, 28-30.

Shinnick M. D., and Gerber, D. L. (1985). A common language for analyzing work. *Journal of Systems Management*, *36*(4), 8-13.

Tan, G. P. (1990). *Modeling a cim system with Micro Saint*. Unpublished masters thesis, Virginia Polytechnic Institute and State University, VA.

Tijerina, L., and Treaster, D. (1990). Model validation, sensitivity analysis, and utilization with Micro Saint: A case study. In *Proceedings of the Human Factors Society 34th Annual Meeting*. Santa Monica, CA: Human Factors Society.

Tripartite Working Group, (1976). *An introduction to predetermined motion time systems*. London: Her Majesty's Stationary Office.

Turner, W. C., Mize, J. H., and Case, K. E. (Eds.). (1987). *Introduction to industrial and systems engineering* (2nd Ed.). New Jersey: Prentice-Hall.

Vidulich, M. A., and Tsang, P. S. (1985). Techniques of subjective workload assessment: A comparison SWAT and NASA-Bipolar methods. *Proceedings of the Human Factors Society 29th Annual Meeting* (pp. 71-75). Santa Monica, CA: Human Factors Society.

Vidulich, M. A., Ward, G. F., and Schueren,J. (1991). Using the subjective workload dominance (SWORD) technique for projective workload assessment. *Human Factors, 33*(6), 677-691.

Vidulich, M. A., and Wickens, C. D. (1986). Causes of dissociations between subjective workload measures and performance. *Applied Ergonomics, 17*(4), 291-296.

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

Wickens, C. D., Larish, I., and Contorer, A. (1989). Predictive performance models and multiple task performance. In *Proceedings of the Human Factors Society 33th Annual Meeting* (pp. 96-100). Santa Monica, CA: Human Factors Society.

Williges, R. C. and Wierwille, W. W. (1979). Behavioral measures of aircrew mental workload. *Human Factors, 21*, 549-574.

Zandin, K. B. (1980). *MOST work measurement systems*. New York: Marcel Dekker.

APPENDIX A

MODAPTS CHARTS

	SMALL/LIGHT OBJECTS	LARGE/HEAVY OBJECTS
Move	**M1** Finger, 1" **M2** Hand, 2" **M3** Arm, 6" **M4** Wholearm, 8" **M5** Extended Arm, 18" **M7** Trunk, 30"	**M2** Hand 1 Hand: 2", Wt = 2 lb **M3** Arm 1 Hand: 6", Wt = 9 lb **M4** Wholearm 1 Hand: 12", Wt = 18 lb 2 Hand: 24", Wt = 35 lb **M5** Ext. Arm 1 Hand: 18", Wt = 18 lb 2 Hand: 35", Wt = 35 lb **M7** Trunk 2 Hand: 39", Wt = 136
Get	**G0** Contact or Touch **G1** Simple finger closing **G3** Complex finger closing	**G2** One Hand engagement **G4** Two Hand engagements **G8** Three Hand engagements **G12** Four Hand engagements **G16** Five Hand engagements
Put	**P0** To general location **P2** To specific location **P5** To exact location	**P0** To general location **P2** To specific location **P5** To exact location, one additional hand engagement **P10** To exact location, two additional hand egagements
Load Factor	**L0** < 4.4 lbs **L1** > 4.4 lbs < 13.3 lbs **L2** >13.3 lbs < 17.6 lbs	**L0** < 35 lbs **L1** > 35 lbs < 44 lbs **L2** > 44 lbs < 53 lbs
	1. > 17 use other table. 2. Divide by 2 for 2 hands. 3. Divide by 3 if slid.	1. > 53 consult manual. 2. Data assumes 2 hands used. 3. Divide by 3 if slid.

Walk

W5	Per pace, obstructed
W4.5	Per pace, unobstructed

Foot Action

F3	Heel remains on floor

Bend and Rise

B17	Hand goes below knees

Sit & Stand

S30	Production work
S48	Office work

Crank

C3	Wrist–Up to 3.5 " diameter
C4	Arm–Above 3.5" diameter

Juggle

J2	To gain better control

Extra Force

X4	A hesitation (not visible)

Use Tool

U.5	Finger motions
U1	Hand motions
U2	Arm motions
U3	Wholearm motions
U4	Not applicable
U5	Not applicable

Eye Control

E2	Eye fixation
E2	Eye travel
E4	Eye focus

Read

R2	One word–general reading
R3	One word–careful reading
R3	Reading up to three digits

Vocalize

V3	For each word spoken

Decide

D3	For the unusual case

Number/count

N3	Per item, arranged
N6	Per item, disarranged

Handwrite

H4	One character, continuous
H5	One character, discontinuous
H6	One upper case, continuous
H7	One upper case, discontinuous
H21	One word, continuous
H26	One word, discontinous
H35	One word, upper case

Arithmetic

A5	Add, subtract, simple multiplication
A18	Difficult multiplication
A24	Difficult division

APPENDIX B

PARTICIPANT INFORMED CONSENT FORM

You are being asked to participate in a research project. The purpose of the study is to determine performance differences given various computer prototypes. You will be asked to complete 30 separate trials. It is anticipated that the entire experiment will take approximately 30 hours total of your time, spread over *10 days*.

This research is being conducted by the Human Factors Laboratory of the Department of Industrial and Systems Engineering. ISE research team members on this project include:

Bo Dyess, Masters Student (231-3323)
Sandy Moscovic, Doctoral Student (231-3323)
Richard Miller, Doctoral Student (231-3323)
Dr. Robert J. Beaton, Faculty Member (231-5936)

The study consists of two parts: introduction/screening, and training/ experimental sessions.

<u>Screening tests:</u> You will be required to complete a screening procedure. This screening procedure includes testing with Dvorine Color Plates to ensure that you do not have color-deficient sight. This procedure takes approximately one minute.

After the screening, you will be asked to perform a card sort for the Subjective Workload Assessment Technique. This card sort entails rank ordering 27 different cards based on your judgment of cognitive workload.

Introduction/ consent form/ SWAT	60 minutes

<u>Experimental sessions:</u> The experimental sessions require training you on a task using the computer. The task will be taught first by components then in whole. You *must* reach a specified level of proficiency at the task before testing.

There will be eight different experimental testing conditions. Since you must learn the task before the first of these, the first training session will take the longest. The approximate length of the first training session is expected to be about three to four hours.

Training on scan types	15 minutes
Training on symbol set	15 minutes
Training on threat level	15 minutes
Training on hooking targets	60 minutes
Training on the task	60 minutes
Testing	30 minutes
Total Time	195 minutes (about 3 hours 15 min.)

Each of the seven subsequent test trials will last approximately 3 hours:

Review of scan types	5 minutes
Training on symbol set	15 minutes
Training on threat level	15 minutes
Training on hooking targets	60 minutes
Training on the task	45 minutes
Testing	30 minutes
Total Time	170 minutes (about 2 hours 50 min.)

These 8 test trials will occur in consecutive days according to your schedule. YOU MUST BE ON-TIME FOR ALL YOUR SCHEDULED SESSIONS due to the necessity to schedule participants one after the other.

If you decide to participate, you will be paid approximately $3.50 per hour ($10.00 per day) for the actual time you participate. If you complete all trials, you will be paid an additional bonus of $50.00.

Your participation will require you to be available for the following days:

Day 1	SWAT Training	60 minutes

(At some point, within a few days of your first session you will need to return for 8 **consecutive** days):

Day 2	1st configuration	3.5 hours	$10.00
Day 3	2nd configuration	3 hours	$10.00
Day 4	3rd configuration	3 hours	$10.00
Day 5	4th configuration	3 hours	$10.00
Day 6	5th configuration	3 hours	$10.00
Day 7	6th configuration	3 hours	$10.00
Day 8	7th configuration	3 hours	$10.00
Day 9	8th configuration	3 hours	$10.00

In a few weeks you will be asked to participate in the final session;

Day 10	9th & 10th configuration	6 hours	$70.00

You will be paid $10.00 at the completion of Days 2-9. After completing the session on Day 10 you will be paid $20.00 for that day, and a bonus of $50.00.

It is important that once you commit to the schedule accepted by the research team and you, that you keep to that schedule. **If you miss more**

than one day, or are late to any sessions, you might NOT be eligible for the $50.00 bonus.
The first 9 days could start on any day of the week. Depending on what day you start you will have to have sessions on at least one Saturday and Sunday. There is some flexibility within a few-day window to help you with scheduling, but you need to attempt to participate on the second through ninth days consecutively. There will be several days (and as much as a few weeks) between Day 9 and Day 10.

Risks and Rights: There are no known risks associated with this research. The only known discomfort to which you will be exposed is possible fatigue resulting from the length of the study. However, you will be permitted to take rest breaks.

As a research participant, you have certain rights:

1. It is your right to withdraw from the study at any time for any reason.
2. Members of the research team will answer any questions you may have concerning this research. You should not sign this consent form until you are satisfied that you understand all the terms involved.
3. You have a right to see your data and withdraw it from the study if you so desire. If you desire to withdraw your data, please inform the site monitor immediately. Otherwise, identification of your data will not be possible since it is separated from the participant in order to ensure anonymity.
4. If you wish to receive a summary of the results of this research, please include your address (where you expect to be living three months from now) with your signature below. Please do so only if you are truly interested in seeing the results. If you desire more detailed information after receiving the synopsis, please contact the Human Factors Laboratory and a full report will be made available to you.
5. Should any further questions or problems arise, you may contact any of the research team members. If you have any concerns about the way the research is being conducted or the way you are being treated, you may contact Dr. E.R. Stout, Chairman of the Institutional Review Board (231-5281).

Your participation is greatly appreciated and we hope that will find the study to be an interesting experience. Your signature below indicates that you have read this document in its entirety, that your questions have been answered, and that you will not discuss participation in this study with anyone until April, 1992 when the study is to be completed.

____________________ ____________________________

Print Name Signature Date

APPENDIX C

SWAT INSTRUCTIONS

Verbal Card Sort Instructions

When we speak of "mental workload," we are clearly referring in some sense to mental effort. Our ideas about mental workload are affected by our experiences with physical workload. We can easily think of the effort one must expend to lift a heavy object, dig a ditch, or participate in our favorite sport. Physical work has been quantified in many ways including carbon dioxide production, heart rate, or amount of work performed in a unit of time. Mental workload, on the other hand, has proven to be more difficulty to measure because it is something that occurs within the person and isn't directly observable.

We might think of mental workload as the amount of concentration required to write a paper, work simple addition problems, or solve complex algebra problems. There is probably unanimous agreement that the amount of work required to solve complex algebra problems would be greater than the amount of work required to solve simple addition problems. While some mental tasks are clearly "harder" than others, measures which quantify this phenomenon have been difficult to develop and validate.

The experiment you are participating in is concerned with mental, not physical workload, and we will deal with methods of measuring the amount of workload experienced while performing a computer-controlled laboratory task.

There are several ways in which your mental workload could be measured, the first of which is to assess your performance on the task. For example, if the task is driving a car, the precision of following a desired track or reaction time to something suddenly entering your visual field, etc., could be an indicator of your workload. Frequently, however, there isn't any change in this type of observable performance, although two people (or the same person under two different conditions) may experience differing degrees of effort expenditure to achieve this performance.

Another way to attempt to measure your mental workload would be to simply ask you to rate your workload on a scale, say from one to ten, for whatever task you are performing. If daydreaming was labeled as "1" and intense concentration as "10", you could probably give a rating corresponding to the workload you were experiencing in performing a task. However, this approach does not give us much information about WHY you gave a particular rating, and we would not be sure that each person intended to describe the same level of workload even though the numbers used were the same.

Another approach is to break up mental workload into several dimensions, or factors, which are generally considered to comprise workload. In this approach, you would not actually be giving ratings on workload, but you would rate the amount or degree of each factor that exists in a given task situation. An application of this approach has been developed and is called the Subjective Workload Assessment Technique

(SWAT). This approach has an added feature of obtaining information from the people using the scale about how the identified factors go together to create their perception of mental workload.

At this time, read the written instructions that have been provided to you, and which you will be able to keep as you go through this card sort.

(Give the participants time to read the written instructions.)

We will now go over the technique you have just read about in more detail. The SWAT technique describes subjective workload as being composed primarily of three dimensions: Time Load, Mental Effort Load, and Psychological Stress Load.

It is important that you understand the meaning associated with the three dimensions and how they relate to the definition of workload. Let's go into a little more detail about the dimensions.

Time Load: Time Load refers to the amount of time pressure experienced in performing your task. This includes the fraction of total available time that you are busy and the degree to which different aspects of the task overlap or interfere with one another. Under high amounts of time load, you are unable to complete the task due to a shortage of time or interference created by the overlap activities.

In a classroom test situation , there could be a high degree of Time Load caused by having a large number of problems (e.g., 100 versus 10) to solve and in a very limited amount of time (e.g., 30 minutes). Notice that we are not considering anything about how much effort is involved in solving the problems or the stress level involved in this situation.

Mental Effort Load: Mental Effort Load refers to the amount of attention and/or concentration required to perform a task. Tasks that require Mental Effort Load include storing and recalling things from memory, decision making, calculations, and problem solving. High levels of Mental Effort Load are required in situations that demand total concentration, whereas lower levels of Mental Effort Load are required when your mind wanders or your attention is distributed over more than one "easy" task component.

Mental Effort Load could involve memorizing items, performing calculations on numbers, concentrating on listening to a speaker for important points, or making very difficult decisions. In the test situation, the problems to solve could be very difficult, requiring you to remember a formula, conversions, and complicated solution procedures. Or, they could be very easy with the solution to the problem being immediately obvious. The difficulty of the problems is not necessarily related to the amount of time provided to complete the test.

Psychological Stress Load: Psychological Stress Load refers to the presence of confusion, frustration, and/or anxiety which hinders completion of your task.

Psychological Stress Load includes such things as pressure to excel, anxiety over physical dangers, tension, fatigue, general state of health or feelings, and comfort factors such as temperature or noise. In the test situation, if your course grade was to be determined by your performance on a certain test, there would probably be quite a high level of stress. However, in a situation where your grade was fairly well determined, the stress level would undoubtedly be less, regardless of the time pressure or the amount of concentration required. Also, if construction was going on near the test room, noise and distractions could affect your ability to concentrate and therefore impose psychological stress. In a driving situation, stress could be produced by obscure road signs, heavy traffic, or inclement weather, which could cause you to become lost, frustrated, or concerned for your safety.

There are three levels of each dimension which can be used to give a rating. One is associated with the lowest degree of each of the dimensions, three is associated with the highest degree and two is a middle degree. Verbal descriptors are provided to define how you should evaluate the levels of each of the dimensions.

While it can be seen that these three dimensions contribute to mental workload, you may be able to think of other dimensions that may have an effect on the workload involved in performing a task. While this may be true, we believe that these three dimensions can be used to cover most of what most people are referring to when they speak of workload.

Another interesting aspect of SWAT is that the procedure provides a mathematical technique for combining your ratings on the three dimensions into a single workload scale.

SWAT

TIME LOAD

1. Often have spare time. Interruptions or overlap among activities occur infrequently or not al all.
2. Occasionally have spare time. Interruptions or overlap among activities occur frequently.
3. Almost never have spare time. Interruptions or overlap among activities are very frequent or occur all the time.

MENTAL EFFORT LOAD

1. Very little conscious mental effort or concentration required. Activities almost automatic requiring little or no attention.
2. Moderate conscious mental effort or concentration required. Complexity of activity is moderately high due to uncertainty, unpredictability, or unfamiliarity. Considerable attention is required.
3. Extensive mental effort and concentration are necessary. Very complex activity requiring total attention.

PSYCHOLOGICAL STRESS LOAD

1. Little confusion, risk, frustration, or anxiety and can easily be accommodated.
2. Moderate stress due to confusion, frustration or anxiety noticeably adds to workload. Significant compensation is required to maintain adequate performance.
3. High to very intense stress due to confusion, frustration, or anxiety. High to extreme determination and self-control required.

Later you will perform a certain task, and give a rating for each of the three dimensions at the end of each task trial. These three numbers will then be combined as in this illustration into a value from 0 to 100. This will then serve as your workload score for that particular trial.

For now, we need to find out information from you so that we can develop the mathematical model which will be used to combine the ratings you will give later. The information we are looking for concerns the importance you place on the three dimensions. You will tell us this by taking this deck of cards, each of which has a combination of the three dimensions on it, and order them from that combination which represents the lowest workload to the combination which represents the highest workload. Do this by imagining a situation that you have experienced which could be described by the combination of the dimensions on a particular card and making a relative judgment about the workload associated with accomplishing this task and rank the cards accordingly. As you can see, with three levels of each dimension there is a total of three times three times three, or 27 possible combinations which could be given. In this way, 27 cards are created which you must rank-order.

Notice that in doing this ranking procedure, the difficult part involves the trade-off decisions which have to be made. Suppose, for example, that you are comparing card 1-2-1 for time, effort, and stress, respectively, with card 1-1-2. Now in one situation the effort is higher, while in the other situation stress is higher. You must decide which situation you would choose as lower in workload. To do this, you need to decide which dimension, effort or stress, has a greater impact on the overall workload to you. Similar situations will arise with different combinations of all three dimensions. There is no right or wrong answer to these decisions, since each person feels differently about the importance of the dimensions. Some people feel that time has the greatest impact, others feel it is effort, and still others feel that stress is the most important.

Because we are trying to determine what constitutes mental workload for you, we would rather not supply you with examples of workload that are represented by each combination. We ask that you supply your own examples. This is necessary because a situation that is very demanding for one person might be very easy for another. Likewise, since we all have different backgrounds, we may or may not be able to relate to a specific example. For example, if I give examples that are related to flying an airplane, that could be very meaningful to a skilled and experienced pilot while most of you in this room might not have any such experience. Therefore, your impressions of the workload involved in such a task would be dependent upon impressions provided by other people.

To avoid this, you are asked to read the descriptors from a card and try to think of something you have experienced that this set of descriptors would have accurately described. Then take another card and repeat the procedure. By comparing the events which you have recalled determine which of the events had the highest workload for you. Repeat this process until you arrive at an order for the 27 cards that begins with the description of the lowest workload event and ends with the description of the highest workload event.

As you try to imagine situations which could be described by the combinations on the cards, there may be combinations for which you have a hard time imagining a situation that fits. In these cases, we could provide you with an appropriate situation, but it is more beneficial if you assume that such a situation or event does exist and try to determine where it would rank in relation to the other situations.

Pay attention to the verbal descriptors on the cards to make your judgments, as it is important that you become comfortable with the levels of the dimensions. This will help you later when you make your ratings.

Written Card Sort Instructions

SWAT CARD SORT INSTRUCTIONS FOR PARTICIPANTS

During the course of this experiment, you will be asked to quantify the mental workload required to complete the tasks you will be performing. Mental workload refers to how hard you work to accomplish some task, group of tasks, or an entire job. The workload imposed on you at any one time consists of a combination of various dimensions which contribute to the subjective feeling of workload. The Subjective Workload Assessment Technique (SWAT) defines these dimensions as (1) Time Load, (2) Mental Effort Load, and (3) Psychological Stress Load.

For the purposes of SWAT, the three dimensions have been assigned three levels. The dimensions and their levels are described in the following paragraphs.

<u>Time Load</u>

Time Load refers to the amount of spare time that you have available (fraction of the total time that you are busy). When Time Load is low, sufficient time is available to complete all of your mental work with some time to spare. As Time Load increases, spare time drops out and some aspects of performance overlap and tasks interrupt one another. This overlap and interruption can come from performing more than one task or from different aspects of performing the same task. At higher levels of Time Load, several aspects of performance often occur simultaneously, you are busy, and interruptions are very frequent.

Time Load may be rated on the three point scale below:

1. Often have spare time. Interruptions or overlap among activities occur infrequently or not al all.
2. Occasionally have spare time. Interruptions or overlap among activities occur frequently.
3. Almost never have spare time. Interruptions or overlap among activities are very frequent or occur all the time.

Mental Effort Load

As described above, Time Load refers to the amount of time one has available to perform a task or tasks. In contrast, Mental Effort Load is an index of the amount of attention or mental effort required by a task regardless of the number of tasks to be performed or any time limitations. When Mental Effort Load is low, the concentration and attention required by a task is minimal and performance is nearly automatic. As the demand for mental effort increases due to task complexity or the amount of information which must be dealt with in order to perform adequately, the degree of concentration and attention required increases. High Mental Effort Load demands total attention or concentration due to task complexity or the amount of information that must be processed.

Mental Effort Load may be rated using the three point scale below:

1. Very little conscious mental effort or concentration required. Activities almost automatic requiring little or no attention.
2. Moderate conscious mental effort or concentration required. Complexity of activity is moderately high due to uncertainty, unpredictability, or unfamiliarity. Considerable attention is required.
3. Extensive mental effort and concentration are necessary. Very complex activity requiring total attention.

Psychological Stress Load

Psychological Stress Load refers to the contribution to total workload of any conditions that produce anxiety, frustration, or confusion while performing a task or tasks. At low levels of stress, one feels relatively relaxed. As stress increases, confusion, anxiety, or frustration increases and greater concentration and determination are required to maintain control of the situation.

Psychological Stress Load may be rated on the three point scale below:

1. Little confusion, risk, frustration, or anxiety and can easily be accommodated.
2. Moderate stress due to confusion, frustration or anxiety noticeably adds to workload. Significant compensation is required to maintain adequate performance.
3. High to very intense stress due to confusion, frustration, or anxiety. High to extreme determination and self-control required.

Each of the three dimensions just described contribute to workload during performance of a task or group of tasks. Note that, although all three factors may be correlated, they need not be. For example, one can have many tasks to perform in the time available (high Time Load) but the tasks may require little concentration (low Mental Effort Load). Likewise, one can be anxious and frustrated (high Psychological Stress Load) and have plenty of spare time between relatively simple tasks. Since the three dimensions contributing to workload are not necessarily correlated, please treat each dimension individually and give independent assessments of the Time Load, Mental Effort Load, and Psychological Stress Load that you experience in performing the following tasks.

One of the most important features of SWAT is its unique scoring system. SWAT uses a procedure to find separate scoring weights for each level of a dimension. Then, it determines a distinctive workload scale for each person or group. This scaling system greatly improves the precision of the workload ratings you will give later.

In order to develop your individual scale, we need information from you regarding the amount of workload you feel is imposed by various combinations of the dimensions described above. We get this information by having you rank order to workload associated with each of the combinations.

In order for you to rank order the workload for each of the combinations, you have been given a set of 27 cards with the combinations from each of the three dimensions. Each card contains a different combination of levels of Time Load, Mental Effort Load, and Psychological

Stress Load. Your job is to sort the cards so that they are ranked according to the level of workload represented by each card.

In completing your card sorts, please consider the workload imposed on a person by the combination represented in each card. Arrange the cards from the lowest workload condition through the highest condition. You may use any strategy you choose to order the cards. One strategy that is useful is to arrange the cards into three preliminary stacks representing "high,", "moderate," and "low" workload. Individual cards may be exchanged between stacks, if necessary, and then rank ordered within stacks. Stacks can then be recombined and checked to be sure that they represent your ranking of lowest to highest workload. However, the choice of strategy is up to you and you should choose the one that works best for you.

There is no "school solution" to this problem. There is no correct order. The correct order is what, in your judgment, best describes the progression of workload from lowest to highest for a general case rather than any specific event. That judgment differs for each of us. The letters you see on the back of the cards are to allow us to arrange the cards in a previously randomized sequence so that everyone gets the same order. If you examine your deck you will see the order on the back runs from A through Z and then ZZ.

Please remember:

(1) The card sort is being done so that a workload scale may be developed for you. This scale will have a distinct workload value for each possible combination of Time Load, Mental Effort Load, and Psychological Stress Load. The following example demonstrates the relationship between the card sort and the resulting workload scale:

Time	Effort	Stress	Workload Scale
1	1	1	0.0
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
3	3	3	100.0

(2) When performing the card sorts, use the descriptors printed on the cards. Please remember not to sort the cards based on one particular task (such as flying an airplane) or what you anticipate that you will be doing in this study. Sort the cards according to your general view of workload and how important you consider the dimensions of Time Load, Mental Effort Load, and Psychological Stress Load to be. Base these decisions on all types of experiences and task situations.

(3) During the actual experiment, you will accomplish the desired task. Then, you will provide a SWAT score based on your opinion of the mental workload required to perform the task. This SWAT score will consist of one number from each of the three dimensions. For example, a possible SWAT score is 1-2-2. This represents a 1 for time Load, a 2 for Mental Effort Load, and a 2 for Psychological Stress Load.

(4) We are not asking for your preference concerning Time Load, Mental Effort Load, and Psychological Stress Load. Some people may prefer to be "busy" rather than "idle" in either Time Load, Mental Effort Load, or Psychological Stress Load dimension. We are not concerned with this preference. We need information on how the three dimensions and the three levels of each one will affect the level of workload as you see it. You may prefer a 2-2-2 situation instead of a 1-1-1 situation. However, you should still realize that the 1-1-1 situation imposes less workload on you and leaves a greater reserve capacity.

The card sort procedure will probably take 30 minutes to an hour. Please feel free to ask questions at any time. Thank you for your cooperation.

Swat Event Scoring Instructions (Written)

REFRESHER SUMMARY

This summary outlines the purpose of the Subjective Workload Assessment Technique (SWAT) and the procedure for giving SWAT ratings. SWAT is a quantitative method for measuring mental workload using subjective ratings. Remember that this method uses the ranking information which you provided with your card sort to create a workload scale. This scale has a distinct workload value for each possible combination of Time Load, Mental Effort Load, and Psychological Stress Load, the three dimensions which comprise SWAT. The definitions of the dimensions are as follows: Time Load refers to the amount of task interruption or overlap; Mental Effort Load is the amount of attention or concentration required to perform a task; and Psychological Stress Load refers to the degree of confusion, anxiety, or frustration involved in performing a task. As you may recall, each dimension has three levels of verbal descriptors: low, moderate, and high. The levels of each dimension can be thought of as a three-point scale, with level 1 being the lowest point or least amount of a dimension and level 3 being the highest point or greatest amount of a dimension.

DO YOU HAVE ANY QUESTIONS ABOUT SWAT?

SWAT DIMENSIONS

TIME LOAD

1. Often have spare time. Interruptions or overlap among activities occur infrequently or not at all.
2. Occasionally have spare time. Interruptions or overlap among activities occur frequently.
3. Almost never have spare time. Interruptions or overlap among activities are very frequent or occur all the time.

MENTAL EFFORT LOAD

1. Very little conscious mental effort or concentration required. Activities almost automatic requiring little or no attention.
2. Moderate conscious mental effort or concentration required. Complexity of activity is moderately high due to uncertainty, unpredictability, or unfamiliarity. Considerable attention is required.
3. Extensive mental effort and concentration are necessary. Very complex activity requiring total attention.

PSYCHOLOGICAL STRESS LOAD

1. Little confusion, risk, frustration, or anxiety and can easily be accommodated.
2. Moderate stress due to confusion, frustration or anxiety noticeably adds to workload. Significant compensation is required to maintain adequate performance.
3. High to very intense stress due to confusion, frustration, or anxiety. High to extreme determination and self-control required.

APPENDIX D

SCREENING/TRAINING INSTRUCTIONS

SCREENING

1. Show AN/SLQ-32(V) OTD, and give a synopsis of the EW job, console on a ship, etc. (10 Min)

2. Dvorine Color Plate Test (1 Min).

2. SWAT Card Sort (60 Min).

TRAINING

1. General Scenario Description.

2. Training on scan types (sounds). (Show and explain; 5 trails, 25 emitters each; Criteria: 23/25)
   Pass: go to next training
   Fail: repeat training on symbols

3. Training on emitter symbols. (Show and explain; 5 trails, 25 emitters each; Criteria: 23/25)
   Pass: go to next training
   Fail: repeat training on symbols

4. Training on emitter threat levels. (Show and explain; 5 trails, 25 emitters each; Criteria: .23/25)
   Pass: go to next training
   Fail: repeat training on threat levels

5. Training on hooking. (Explain. 40 trails–5 sets of 5–, 15 emitters each, to task stabilization—no more than .5 second mean change in five consecutive trials, with the exceptions of outliers caused by error; error-induced outliers will be defined as twice the mean error level for that participant).

6. Training on first configuration to task stabilization (Explain, 20 five minute trails, mean time doesn't improve more than 2.00 seconds for five consecutive trials.)

GENERAL SCENARIO DESCRIPTION

You are being trained to do one of the primary duties of a naval Electronic Warfare officer, or EW. In this scenario, you are on a ship and your responsibility is to observe the environment around the ship through the use of a radar-type scope to detect, identify and classify the emitters (such as missiles, planes, ships and subs) that you encounter. The reason that you do this task is that although the computer is very good at detecting signals from other vessels, it is poor at making correct identifications. However, you are able to take several pieces of information and make the correct identification.

In general, you will have a primary display that shows where the emitters are in relation to your ship. You must process each and every one quickly and accurately using the functions available to you to evaluate certain parameters associated with each emitter. Those parameters include:

-the emitter's frequency (a number ranging from 100 to 20,000)
-the threat level (a number ranging from 0 to 7)
-the PRF, or pulse repetition frequency (another number from 100 to 20,000)
-the scan type of the emitter
-and the bias factor (the type of emitter; Hostile or Friendly)

The role you play in performing this task is critical to the protection of you, the ship, and everyone on board. For instance, if a missile has been launched at your ship, you may have as few as thirty seconds to identify the emitter as hostile. Clearly, the consequence of not making a correct identification is great. Therefore, in the performance of this task, SPEED AND ACCURACY are of the essence.

In the upcoming sessions, we will train you on the components of the task and subsequently, the whole task. Once you have learned the task and can perform it as quickly and accurately as possible, we will test you on your performance.

Do you have any questions at this time?

SOUNDS
(Scan Types)

[Use the computer trial sessions.]
The emitters each have a sound that identify the scan type of that emitter. Put differently, the way an emitter emits "radar" waves is called a scan type, and helps to identify what the purpose of that emitter is. For our purposes, there are four scan types that you will need to identify.

Those types are standard (STD)*, Circular (CIR)*, Conical (CON)*, and Sector (SEC)*.
[Run the trials, and explain each sound as you get to it.)

See how standard (STD) signal sounds like a constant tone, with a regularly appearing signal. Notice how the picture represents these signal spikes appearing out of the constant tone.

This is a circular (CIR) scan type. It sounds like an object is putting out a signal and rotating it around it's entire body, so that the sound seems to get louder as it comes near you, fades as it moves away, then gets louder as it comes back around. The picture looks like this.*

This scan type is a Conical (CON) type. The sound sounds like an emitter is continuously pointing a beam at you, and moving that beam around in small circles (but still pointing at you). This would be illustrated by a cone coming out of the front of an emitter. Listen to what a conical sound sounds like. Not the illustration of the scan type.

This scan type is a Sector or SEC type. The sound represents a beam pointing out at you scanning back and forth is a small sector over you. The signal sounds like it is going back and forth over you...it has a "tin-ny" quality to it. Notice how that scan type looks when graphed.

We'll just run through the first session of these sounds together to get you familiar with the scan types, then I will let you repeat sessions on your own to practice. In order to participate in the study you must maintain the highest level of accuracy. Sounds will be presented in sets of 25. You will have to go though (5 - pretest or 1-session) set(s) of these sounds. Try not to make any mistakes. Do you have any questions?

EMITTER SYMBOLS
(Configuration specific)

Configurations 1 & 5
Old NATO/B/W

(*-Denotes pointing to the object.)

Here (show card) are the eight different types of symbols that you will need to work with. Note that there are hostile types, which are designated with "v" shapes* around a dot; and there are friendly types, which are designated with half circles* around a dot. In the SLQ, hostile is abbreviated "hos" and friendlies are "fnd".

A half-square* over a dot represents an unknown ("unk") bias factor. A bias factor is the classification, such as a hostile aircraft, a friendly sub, etc.

The two overlapped "x's"* represents a missile---which is always considered hostile. The abbreviation for missile is "msl"

For aircraft (or "air"), the dot is always represented with a "hat" over it* , the shape of which depends on whether the emitter is hostile or friendly.

For surface ships ("sur") or land-based emitters (simply, "land"), the dot is surrounded by the corresponding shape: one on top and one underneath*.

For submarines ("subs"), the friendly half circle or hostile v-shape is underneath the dot*.

You may have as much time as you would like to study these emitters. Let me know when you would like to practice identifying these emitters. Take as much time as you like.

Do you have any questions?

EMITTER SYMBOLS
(Configuration Specific)

Configurations 2 and 6
Old NATO/Color

(*-Denotes pointing to the object.)

Here (show card) are the eight different types of symbols that you will need to work with. Note that there are hostile types, which are designated with "v" shapes* around a dot; and there are friendly types, which are designated with half circles* around a dot.

A half-square* over a dot represents an unknown ("unk") bias factor. A bias factor is the classification, such as a hostile aircraft, a friendly sub, etc.

The two overlapping "x's" * represents a missile---which is always considered hostile (Missile is abbreviated as "MSL" by the SLQ.).

For aircraft ("air"), the dot is always represented with a "hat" over it*, the shape of which depends on whether the emitter is hostile or friendly.

For surface ships ("sur") or land-based emitters (simply, "land"), the dot is surrounded by the corresponding shape: one on top and one underneath*.

For submarines ("sub"), the friendly half circle or hostile v-shape is underneath the dot*.

In addition to the shape of the emitter, the emitters are coded with color. All missiles are red.* All other hostiles (air, subs, sur/land) are an orange or amber color.* All friendlies are coded green *.
Unknown emitters are yellow*.

You may have as much time as you would like to study these emitters. Let me know when you would like to practice identifying these emitters. Take as much time as you like.

Do you have any questions?

EMITTER SYMBOLS
(Configuration Specific)

Configurations 3 and 7
New NATO/B/W

(*-Denotes pointing to the object.)

Here (show card) are the six different types of symbols that you will need to work with. The shape of the emitter depicts what kind of an emitter is being depicted:*

-missiles look like missiles ("msl")
-aircraft ("air")
-surface ships ("sur")
-submarines ("subs)
-land-based emitters, or "land" (look like satellite dishes)
-and unknowns ("unk") are boxes with an "x" inside

Note that missiles are always "hostile" ("hos"). The remainder of the symbols may be either hostile or friendly (fnd)--which you will have to determine by reading the "bias factor". We will train you on those threat levels in the next session.

For now, just memorize the symbols. You may have as much time as you would like to study these emitters. Let me know when you would like to practice identifying these emitters. Take as much time as you like.

Do you have any questions?

EMITTER SYMBOLS
(Configuration specific)

Configurations 4 and 8
New NATO/Color

(*-Denotes pointing to the object.)

Here (show card) are the ten different types of symbols that you will need to work with. The shape of the emitter depicts what kind of an emitter is being depicted:*

- -missiles look like missiles ("msl")
- -aircraft ("air")
- -surface ships ("sur")
- -submarines ("subs")
- -land-based emitters, or "land" (look like satellite dishes)
- -and unknowns ("unk") (boxes with an x inside)

Note that missiles are always "hostile" and red. The remainder of the symbols may be either hostile or friendly ("fnd")---which you will have to determine by the color. All hostile emitters that are not missile will be coded orange or amber in color*.

All friendlies are green*. Unknown emitters are yellow.

For now, just memorize the symbols. You may have as much time as you would like to study these emitters. Let me know when you would like to practice identifying these emitters. Take as much time as you like.

Do you have any questions?

THREAT LEVELS

(Session immediately following training of emitter types except for New NATO/Color)

(Show appropriate configuration-specific card.)

Now that you know and can recognize the emitter symbols, you need to be learn the threat level, often abbreviated "TL" for each type.

Threat levels range from zero to seven, with zero being no threat, and seven the highest level of threat.

All friendlies, regardless of type, are considered "zero"*.

There is no "1" or "2" level.

All unknowns are considered intermediate levels of threat---a "3"*.

All hostile land based threats are "4"*.

All hostile surface ships or hostile submarines are a threat level of "5"*.

All hostile aircraft are a threat level "6"*.

Finally, all missiles are hostile, and are a threat level of "7".

You may now study the TLs for as long as you like, and when you are ready we will practice identifying them.

Do you have any questions?

HOOKING
(Configuration Specific)

POLAR CONFIGURATIONS

Now we are going to practice the first part of the task, which means that you are sitting at your console and encounter new emitters .

An alert will sound when new emitters appear. The SLQ will try to identify the new emitters, and will present what it believes to be the correct identification on the Polar display* AND will list them here* in a "Threat Summary List". You must try to locate and identify each of the emitters that appears in the threat summary list. Each new emitter is registered on this summary, and you will try to process them in the order they appear – or more simply, top to bottom.

[POLAR ONLY]
In this type of display, you have what is called a Polar display. Each of the bearing lines represents a location or bearing in degrees around you (your ship is in the center*). Zero or 360 is ahead of you, or the top of the display. 90 degrees is to the right of you, 180 behind you (or at the bottom of the screen), and 270 is to the left of you. To help you, the three regions of the display are used by the computer to portray what kind of emitter is on that bearing.

These rings on the Polar display* do NOT portray distance from your ship. Instead, they are meant to assist you by displaying the emitters according to threat. Because you are mostly concerned with threats, the largest section is this * intermediate section, which is where all missiles (TL of 7) will appear. In the center of the display, all friendlies (TL of 0) will appear. Do not be confused because the innermost circle does not have bearing lines--- you simply must imagine the bearing lines to extend into the center. In the very outermost ring, all unknowns and non-missile hostiles will appear (TLs 3-6).

The way to find the emitter is to read the information in the threat summary list, correlate it with what you've been taught about symbols and threat levels, and then go out and "hook" the emitter. Hooking means you place the cursor over the emitter you think is the right one, then you "click" the mouse button.

When any emitter has been selected by you in this manner, the computer does what is called "placing the emitter in close control", which means that it displays here * all the parameters of the emitter that you have selected. You may determine by reading those parameters whether you have or have not "hooked" the emitter you intended to identify. There are two things that happen when you hook the correct emitter:

(1) the emitter is "greyed out" on the threat summary list
(2) the close control parameters match* the threat summary description of that emitter.

Only 15 new emitters can be listed in the threat summary* list at a time. If there are more than 15 new emitters, the emitters are located on subsequent "pages". The number of pages of emitters is given here*. If you want to go to other pages, you use the "up" arrow* to go to the next page and the "down" arrow to go to a previous page.

Remember, you must process the emitters from top to bottom of the list. Also remember that SPEED and ACCURACY are essential! Be sure to use all the relevant information to help you---such as which ring on the primary display contains what emitters, etc. Your mean time will appear on the screen after each trial.

Do you have any questions?

I will now ask you to practice hooking emitters to the best of your ability. Remember, you must work as quickly and accurately as possible. As you go through these practices, try to think of strategies that will enable you to go as quickly as possible.

HOOKING
(Configuration Specific)

RANGE CONFIGURATIONS

Now we are going to practice the first part of the task called hooking. You are sitting at your console and encounter new emitters .

An alert will sound, and new emitters will appear on the Range display and will be listed here in the Threat Summary List*. You must try to identify each of the emitters that appears in the threat summary list*. Each new emitter is registered on this summary, and you will try to process them in the order they appear---or more simply, top to bottom.

In this type of display, you have bearing lines* which represents a location of an emitter with respect to degrees around you (your ship is in the center). Zero or 360 is ahead of you, or the top of the display. 90 degrees is to the right of you, 180 behind you (or at the bottom of the screen), and 270 is to the left of you. To help you, there are concentric rings which mark distance or range away from you*.

The way to find the emitter is to read the information in the threat summary list, and then go out and "hook" the emitter like you did in the previous training session.

When any emitter has been selected by you in this manner, the computer does what is called "placing the emitter in close control", which means that it displays here * all the parameters of the emitter that you have selected. You may determine by reading those parameters whether you have or have not "hooked" the emitter you intended to identify. There are two things that happen when you hook the correct emitter:

(1) the emitter is "greyed out" on the threat summary list
(2) the close control parameters match the threat summary description of that emitter.

Only 15 new emitters can be listed in the threat summary list at a time. If there are more than 15 new emitters, the emitters are located on subsequent "pages". The number of pages of emitters is given here*. If you want to go to other pages, you use the "up" arrow* to go to the next page and the "down" arrow to go to a previous page.

Remember, you must process the emitters from top to bottom of the list. Also remember that SPEED and ACCURACY are essential! Your mean times will be displayed on the screen after each of the trials.

Do you have any questions?

I will now ask you to practice hooking emitters to the best of your ability. Remember, you must work as quickly and accurately as possible. As you go through these practices, try to think of strategies that will enable you to go as quickly as possible.

CONFIGURATION TASK STABILIZATION

Now we are going to work on the entire task.

One of your "tools" you have to help you in this identification and classification task is a set of Publications or pubs*. [Show pubs.] The pubs are a reference that list all the existing emitters. Emitters are listed by order of their Cross-reference number in the front of the pubs and by frequency in the back.

Each emitter has a unique 3-digit number associated with it which is used to identify that emitter. The number is referred to as an EFX (emitter file index) and is located on the first few pages in the pubs*.

I will come back to how to use the pubs in a bit.

You know that on the Polar/Range display, when the SLQ detects new emitters, an alert sounds, the new emitters appear, and the emitters are listed on the threat summary list. These emitters on the Polar/Range display represent the SLQ (computer's) guess at the correct identification.

When your SLQ detects new emitters you will hear an auditory alert . Starting at the top of the list, you will need to locate the corresponding emitter on polar/range display. Once you have hooked the correct emitter, you will need to evaluate the emitter on each of the following parameters to see whether or not the computer has correctly identified and classified the emitter:

- Threat level
- Bias Factor
- Frequency
- Pulse Repetition Frequency (PRF)
- Scan Type
- Scan

The computer can always be relied on to give you the correct location and frequency of an emitter, but it is not always accurate with any of the other parameters.

Sometimes you will notice that there are inconsistencies between the symbol threat level or type in the threat summary list and symbol displayed on the Polar/range display.

Sometimes you will see that after you have hooked an emitter, the parameters that show up in the close control box* are inconsistent with the parameters listed in the threat summary list.

Or perhaps there are inconsistencies between the threat summary list, polar/range display, and the close control information.

Any inconsistency is a clue that the computer (SLQ) has made an incorrect designation. In fact, everything could be correct in the list but the Scan number. Unfortunately there is no way to check to see if the Scan number is correct, so you must go to the pubs to figure out the correct EFX.

On a real ship, the EW can check the scan type, the pulse repetition frequency (PRF) and scan of an emitter by using another piece of equipment located on the ship. This other piece of equipment is MORE ACCURATE than the SLQ at identifying scan type and PRF. It presents an audio display of the scan type so the EW can listen to and determine scan type; it provides a readout of the PRF.

In this scenario, we give you this "other piece of equipment" on this same MACINTOSH screen with the press of a button located here*. The button is labeled "Sig Sel" for Signal Select. When you wish to check the accuracy of the SLQ classification of PRF and scan type, you "use" this other piece of equipment. To turn off this other piece of equipment, you hit the Shift Key. Before you do any further function, you must hit the Shift Key which stops the SIGNAL SELECT actions. (* Demonstrate the ULQ-Sig Sel.)

During the session you might find the information on the ULQ and the close control area to match, even so you need to figure out the correct sound in order to designate the correct EFX number. Since the first few pages of the publications contain the correct EFX numbers you need to figure out the correct cross-reference number. The correct cross-reference number is found by looking up the Frequency and PRF listed in the main part of the publications finding the correct entry in the back of the publications* (demo). Then use the X-REF number to search the listing in the front of the publications . Each X-REF number will have two entries, you must use your knowledge of the emitters sound type to determine which of the two entries is correct* (demo). Once you have picked what you think is the correct entry find the EFX listed. Enter the EFX number in the close control box by hitting the DesigID button, and entering the 3-digit EFX and then hit the Return Key. Note how the Design ID will bring up a box for the EFX entry here* (demo). The DesigID button will not function with the ULQ window open, as shown here* (demo).

So, after you have hooked the correct emitter, there are several things that you will want to check (these items are not in any order---you will need to adopt your own strategies):

- TL, emitter symbol, name, and bias factor are consistent across all displays (threat summary list, close control, and main display)
- Scan Type and PRF

You must consult the pubs for the correct EFX #.

If you need to use the pubs then simply pick up the book, and search for the correct emitter.

After you have found the correct EFX (by searching for the correct designation), you must hit the DESIG key, enter the three-digit EFX, then the Return Key.

Remember after processing each emitter, you must Designate the EFX, even if it is already correct.

Also note that after you have hooked the emitter at the top of the list, it is "greyed" out of the emitter summary list (just like in the hooking practice trials).

As you process emitters from top to bottom of the threat summary list, you may occasionally hook an emitter out of order and see it grey out further down the list. In this case, it is up to you to either continue to identify the emitter you've accidentally hooked, and then resume with the top of the list– OR– you may simply continue to try and hook the emitter at the top of the list.

The order and strategy you use to go through this process is entirely up to you. You must remember that the critical factors are SPEED AND ACCURACY. The white box above the display is to provide you with feedback by displaying a mean time for each correctly identified emitter that you have processed.

The feedback is for you to use to decrease your times and to increase your accuracy. If you realized that you have not correctly identified an emitter because a time did not appear in the window, do not try to re-identify the emitter. Simply continue to the next emitter and do the best you can.

Do you have any questions?

(Answer all questions---without telling them a specific strategy.
Run trials.)

INTEGRATED TASK TEST SESSION

This session will be very similar to the training session you just completed. You will now be given three, five-minute sessions. Each one will have a different number of emitters on the screen, and a different number of emitters in the threat summary list. Remember that if you complete the emitters on the first page use the arrow keys to go to the second page of emitters. The session will time out, so go as FAST as possible and designate as many emitters that you can–and as ACCURATELY as you can. At the end of the session you will need to answer three questions relating to the level of workload that you experienced during the session.

Take a minute or two to read this event scoring information for the SWAT workload measure.

Again, at the end of the session, you will be prompted by the computer to give a time load (1, 2, or 3), a mental effort rating (1, 2, or 3), and a psychological stress load rating (1, 2, or 3). Simply click the mouse over the appropriate number and then click "Okay".

After the session you will have a short questionnaire to fill out concerning your experiences with the interface.

APPENDIX E

NODE PARAMETERS FOR MICRO SAINT IT MODEL

```
Task Number:  1.1            Name:  BEGIN TASK
Upper Network:  0
Upper Name:  IT
Time Distribution:            Normal
Expressions:
    {Release Conditions}     1;
    {Beginning Effect}
                             FORMAT:=POLAR,     {POLAR,RANGE}
                             SYMBOL:=OLD,         {OLD, NEW}
                             COLOR:=BW,            {BW, C}
                             DENSITY:=1,           {1-CARIBBEAN,
                                                      2-GULF
                                                      3-
                             ARMAGEDDON}
                             MOD:=0.129              {MOD  value];
   {Mean Time}               0 ;{Dummy Variable}
   {Standard Deviation}      0;
   {Launch Effect}           n/a
Decision Type:               Single
What Happens Next:           Probability:
            1.2 Get Mouse    Always
```

```
Task Number:  1.2            Name:  Get Mouse
Upper Network:  0
Upper Name:  IT
Time Distribution:            Gamma
Expressions:
    {Release Conditions}     1;
    {Beginning Effect}
   {Mean Time}               ({M}5+{G}0)*MOD;
   {Standard Deviation}      0.268;
   {Launch Effect}           n/a
Decision Type:               Probabilistic
What Happens Next:           Probability:
         2.1.0 Dum_Var       1;
                2 Hook       1;
```

Task Number 2.1.1	Name: (P) Threat Summary
Upper Network: 2	
Upper Name: IT	
Time Distribution:	Gamma
Expressions:	
{Release Conditions}	1;
{Beginning Effect}	
{Mean Time}	(2*{R}3+2*{R}2)*MOD;
{Standard Deviation}	
{Launch Effect}	n/a
Decision Type:	Multiple
What Happens Next:	Probability:
2.1.1.1 Search B/W	Color == 0;
2.1.1.2 Search C	Color == 1;

Task Number 2.1.0	Name: Dum_Var
Upper Network: 2	
Upper Name: IT	
Time Distribution:	Gamma
Expressions:	
{Release Conditions}	1;
{Beginning Effect}	ERROR:= ERRORS, ERRORSTDEV:=ERRSTDEV, NO:= gamma(ERRORS, ERRSTDEV);
{Mean Time}	(2*{R}3+2*{R}2)*MOD;
{Standard Deviation}	
{Launch Effect}	n/a
Decision Type:	Multiple
What Happens Next:	Probability:
2.1.1 (P) Threat	FORMAT==POLAR;
2.1.2 (R) Threat	FORMAT==RANGE;

Task Number 2.1.1.1	Search(B/W)
Upper Network: 2	
Upper Name: IT	
Time Distribution:	Gamma
Expressions:	
{Release Conditions}	1;
{Beginning Effect}	1; {Eyes to polar display, look up bearing}
{Mean Time}	1.964-.492*SYMBOL+ 0.717*DENSITY; {REGRESSION}
{Standard Deviation}	2.56; {AVERAGE OF ALL POLAR ST DEVS FOR NODE}
{Launch Effect}	n/a
Decision Type:	Single
What Happens Next:	Probability:
2.2	1;

Task Number 2.1.1.2	Search(B/W)
Upper Network: 2	
Upper Name: IT	
Time Distribution:	Gamma
Expressions:	
{Release Conditions}	1;
{Beginning Effect}	
{Mean Time}	1.964-.0871-0.492*SYMBOL+0.717*DENSITY;
{Standard Deviation}	2.56; {AVERAGE OF ALL POLAR ST DEVS FOR NODE}
{Launch Effect}	n/a
Decision Type:	Single
What Happens Next:	Probability:
2.2	1;

Task Number 2.1.2	(R)Threat Summary
Upper Network: 2	
Upper Name: IT	
Time Distribution:	Gamma
Expressions:	
{Release Conditions}	1;
{Beginning Effect}	
{Mean Time}	(2*{R}3 +{R}2)*MOD;
{Standard Deviation}	2.56; {AVERAGE OF ALL POLAR ST DEVS FOR NODE}
{Launch Effect}	n/a
Decision Type:	Single
What Happens Next:	Probability:
2.2	1;

Task Number 2.1.2.1	Search(Range)
Upper Network: 2	
Upper Name: IT	
Time Distribution:	Gamma
Expressions:	{TIME IS FROM BRG/RANGE READS AND BETA WTS}
{Release Conditions}	1;
{Beginning Effect}	1; {Eyes to polar display, look up bearing, Iinfo theory for 8 unique symbols}
{Mean Time}	({R}3+2*{N}3)*MOD+(0.713-0.104*COLOR-0.441*SYMBOL+ 0.682* DENSITY);
{Standard Deviation}	1.922; {AVERAGE ST DEV FOR RANGE HOOKS}
{Launch Effect}	n/a
Decision Type:	Single
What Happens Next:	Probability:
2.2	1;

Task Number 2.2	Hook
Upper Network: 2	
Upper Name: IT	
Time Distribution:	Gamma
Expressions:	{Time to position mouse and click using tpos=1.03 + logbase2(D/S +.5)sec}
{Release Conditions}	1;
{Beginning Effect}	1; {Eyes to polar display, look up bearing, Iinfo theory for 8 unique symbols}
{Mean Time}	1.4;
{Standard Deviation}	0.07;
{Launch Effect}	n/a
Decision Type:	Single
What Happens Next:	Probability:
2.3 Hooked?	1;

Task Number 2.3	Hooked?
Upper Network: 2	
Upper Name: IT	
Time Distribution:	Gamma
Expressions:	{operator decides if the correct emitter is hooked}
{Release Conditions}	1;
{Beginning Effect}	
{Mean Time}	{Modapts only award decision time for the rejections}
{Standard Deviation}	
{Launch Effect}	n/a
Decision Type:	Multiple
What Happens Next:	Probability:
2.4 Wrong Emitter	NO > 0.5;
3.1 Sig Sel	NO <= 0; {yes, correct one}

Task Number 2.4	Wrong Emitter
Upper Network: 2	
Upper Name: IT	
Time Distribution:	Gamma
Expressions:	{The operator has hooked a wrong emitter}
{Release Conditions}	1;
{Beginning Effect}	NO-=1, if FORMAT==0 then RETRYA:=27, RETRYB:=3, RETRYC:= 90 else RETRYA:=77, RETRYB:= 3, RETRYC:= 40;
{Mean Time}	if FORMAT==0 then {D}3*MOD else 0;
{Standard Deviation}	none
{Ending Effect}	DUM_ERR+=1;
Decision Type:	Probabalistic
What Happens Next:	Probability:
2.4 Close Emitter	RETRYA;
2.1.0 Dum_Var	RETRYB;
2.6 Overlap Retry	RETRYC;

Task Number 2.5	Close emitter retry
Upper Network: 2	
Upper Name: IT	
Time Distribution:	Gamma
Expressions:	{Time based on the Tpos=1.03+0.096logbase2(D/S+.5) equation for same targets size, d=1.27cm}
{Release Conditions}	1;
{Beginning Effect}	
{Mean Time}	1.09;
{Standard Deviation}	0.07;
{Ending Effect}	DUM_ERR+=1;
Decision Type:	Single
What Happens Next:	Probability:
2.3 Hooked	1;

Task Number 2.6	Overlap retry
Upper Network: 2	
Upper Name: IT	
Time Distribution:	Gamma
Expressions:	{for overlapping, close emitters} {Tpos= 1.03 + 0.096logbase2(d/s+0.5) d=0.15cm}
{Release Conditions}	1;
{Beginning Effect}	
{Mean Time}	MOD; ;
{Standard Deviation}	0.07;
{Ending Effect}	DUM_ERR+=1;
Decision Type:	Single
What Happens Next:	Probability:
2.3 Hooked	1;

Task Number 3.1	Sig Sel
Upper Network: 3	
Upper Name: IT	
Time Distribution:	Gamma
Expressions:	Operator turns on o-scope by hitting the Sig Sel FAB}
{Release Conditions}	1;
{Beginning Effect}	
{Mean Time}	({G}1+{M}5+{G}0)*MOD;
{Standard Deviation}	0.253;
{Ending Effect}	
Decision Type:	Single
What Happens Next:	Probability:
3.2 Read CC	1:
3.3 Grab Pubs	1;

Task Number 3.2	Read CC Parameters
Upper Network: 3	
Upper Name: IT	
Time Distribution:	Gamma
Expressions:	{Read close control parameters: Freq, PRF}
{Release Conditions}	1;
{Beginning Effect}	
{Mean Time}	4*{R}3*MOD;
{Standard Deviation}	
{Ending Effect}	READ:=1;
Decision Type:	Single
What Happens Next:	Probability:
3.4 Dum_Pubs	1:

Task Number 3.3	Grab PUBS
Upper Network: 3	
Upper Name: IT	
Time Distribution:	Gamma
Expressions:	
{Release Conditions}	1;
{Beginning Effect}	
{Mean Time}	({M}4+{G}0)*MOD;
{Standard Deviation}	0.269;
{Ending Effect}	GRAB:=1;
Decision Type:	Single
What Happens Next:	Probability:
3.4 Dum_Pubs	1:

Task Number 3.4	DUM_PUBS
Upper Network: 3	
Upper Name: IT	
Time Distribution:	Gamma
Expressions:	
{Release Conditions}	(GRAB==1) & (READ==1);
{Beginning Effect}	
{Mean Time}	
{Standard Deviation}	
{Ending Effect}	
Decision Type:	Probablistic
What Happens Next:	Probability:
3.5 PUBSA	10;
3.6PUBSB	2;

Task Number 3.5	PUBSA
Upper Network: 3	
Upper Name: IT	
Time Distribution:	Gamma
Expressions:	{SEARCH STRATEGY 1}
{Release Conditions}	
{Beginning Effect}	
{Mean Time}	(4*{SS}6 + {SS}10+5 +{SS}10+8 +2*{SS}6 +{SS}10+4) *MOD;
{Standard Deviation}	4.25;
{Ending Effect}	
Decision Type:	Single
What Happens Next:	Probability:
4.1 Shift Key	1;

Task Number 3.6	PUBSB
Upper Network: 3	
Upper Name: IT	
Time Distribution:	Gamma
Expressions:	{{1ST PART OF SEARCH STRATEGY 2}
{Release Conditions}	
{Beginning Effect}	
{Mean Time}	(4*{SS}6 + {SS}10+5 +{SS}10+8) *MOD;
{Standard Deviation}	3.962;
{Ending Effect}	
Decision Type:	Single
What Happens Next:	Probability:
4.1 Shift Key	1;

Task Number 3.7	Shift
Upper Network: 3	
Upper Name: IT	
Time Distribution:	Gamma
Expressions:	
{Release Conditions}	
{Beginning Effect}	
{Mean Time}	({M}4+{G}0)*MOD;
{Standard Deviation}	0.206;
{Ending Effect}	
Decision Type:	Single
What Happens Next:	Probability:
3.8 Desig ID	1;

Task Number 3.8	Desig ID
Upper Network: 3	
Upper Name: IT	
Time Distribution:	Gamma
Expressions:	{Hits Desig ID FAB}
{Release Conditions}	
{Beginning Effect}	
{Mean Time}	({K}3.3)*MOD;
{Standard Deviation}	0.143;
{Ending Effect}	
Decision Type:	Single
What Happens Next:	Probability:
3.9 PUBSC	1;

Task Number 3.9	PUBSC
Upper Network: 3	
Upper Name: IT	
Time Distribution:	Gamma
Expressions:	{Strategy 2- Look up EFX in x-ref}
{Release Conditions}	
{Beginning Effect}	
{Mean Time}	({E}4+2*{SS}6 +{SS}10+4) *MOD;
{Standard Deviation}	2.874;
{Ending Effect}	
Decision Type:	Single
What Happens Next:	Probability:
3.10 1ST EFX	1;

Task Number 4.1	Shift Key
Upper Network: 4	
Upper Name: IT	
Time Distribution:	Gamma
Expressions:	{Operator found EFX, turns off o-scope}
{Release Conditions}	
{Beginning Effect}	
{Mean Time}	({M}4+{G}0)*MOD;
{Standard Deviation}	0.226;
{Ending Effect}	
Decision Type:	Single
What Happens Next:	Probability:
4.2DESIG ID	1;

Task Number 4.2	DESIG ID
Upper Network: 4	
Upper Name: IT	
Time Distribution:	Gamma
Expressions:	
{Release Conditions}	
{Beginning Effect}	
{Mean Time}	({K}3.3)*MOD;
{Standard Deviation}	0.226; ;
{Ending Effect}	
Decision Type:	Single
What Happens Next:	Probability:
4.3 EFX-1	1;

Task Number 4.3	EFX-1
Upper Network: 4	
Upper Name: IT	
Time Distribution:	Gamma
Expressions:	{Enters first digit of EFX}
{Release Conditions}	
{Beginning Effect}	
{Mean Time}	({K}3.3)*MOD;
{Standard Deviation}	0.226; ;
{Ending Effect}	
Decision Type:	Single
What Happens Next:	Probability:
4.4 EFX-2	1;

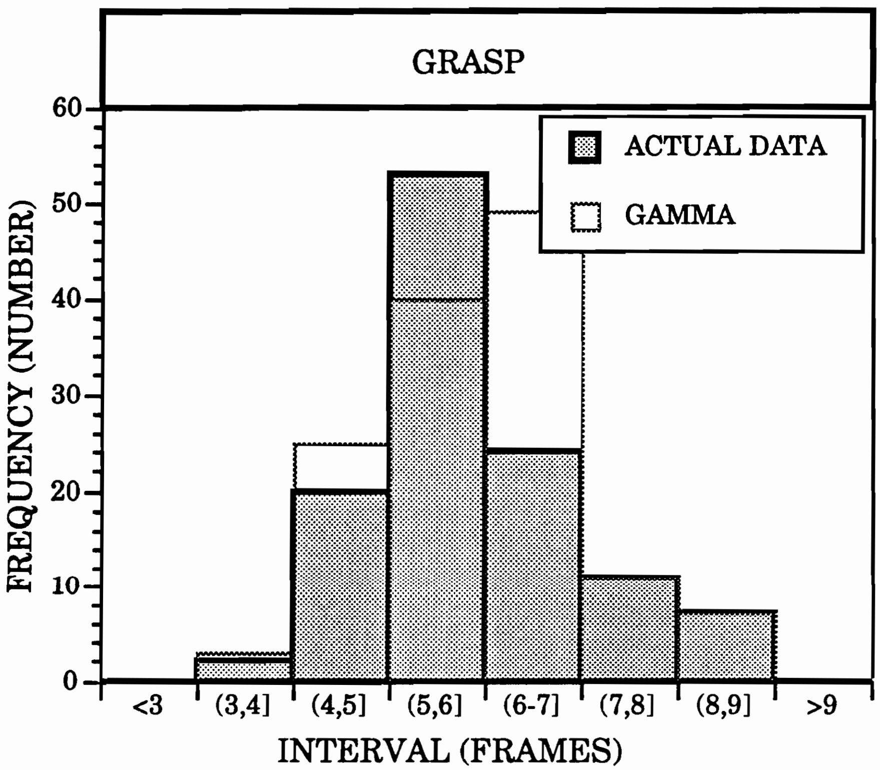
Task Number 4.4	EFX-2
Upper Network: 4	
Upper Name: IT	
Time Distribution:	Gamma
Expressions:	{Enters second digit of EFX
{Release Conditions}	
{Beginning Effect}	
{Mean Time}	({K}3.3)*MOD;
{Standard Deviation}	0.226; ;
{Ending Effect}	
Decision Type:	Single
What Happens Next:	Probability:
4.5 EFX-3	1;

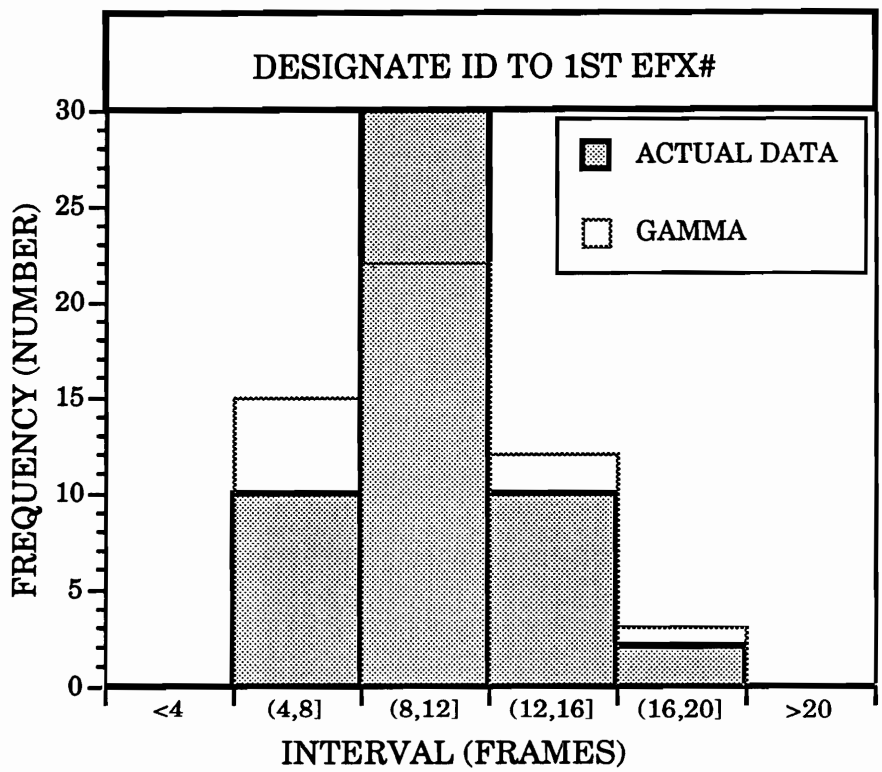
Task Number 4.5	EFX-3
Upper Network: 4	
Upper Name: IT	
Time Distribution:	Gamma
Expressions:	{Enters third digit of EFX
{Release Conditions}	
{Beginning Effect}	
{Mean Time}	({K}3.3)*MOD;
{Standard Deviation}	0.226;
{Ending Effect}	
Decision Type:	Single
What Happens Next:	Probability:
4.6. Return	1;

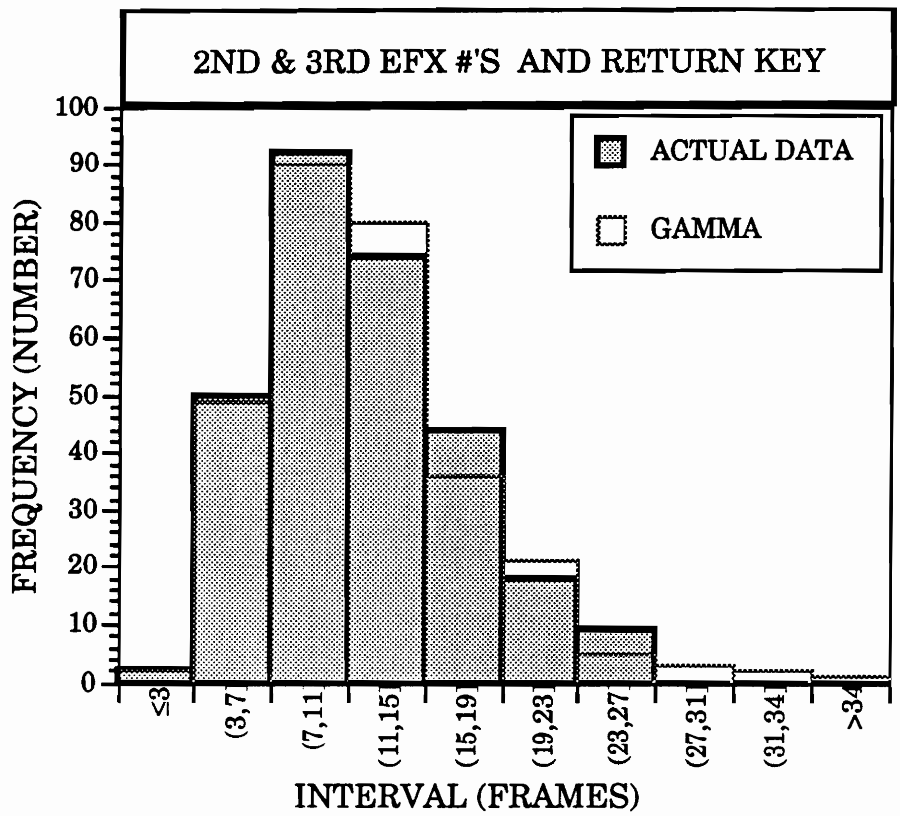
Task Number 4.6	Return
Upper Network: 4	
Upper Name: IT	
Time Distribution:	Gamma
Expressions:	{Enters second digit of EFX
{Release Conditions}	
{Beginning Effect}	
{Mean Time}	({K}3.3)*MOD;
{Standard Deviation}	0.226;
{Ending Effect}	WKLD:=-0.891+ (1.519)*(clock)-(0.164)*(DUM_ERR)-(0.984)*FORMAT-(3.063)*SYMBOL-(0.017)*(COLOR)+(8.66)*(DENSITY);
Decision Type:	None Following
What Happens Next:	Probability: 1;

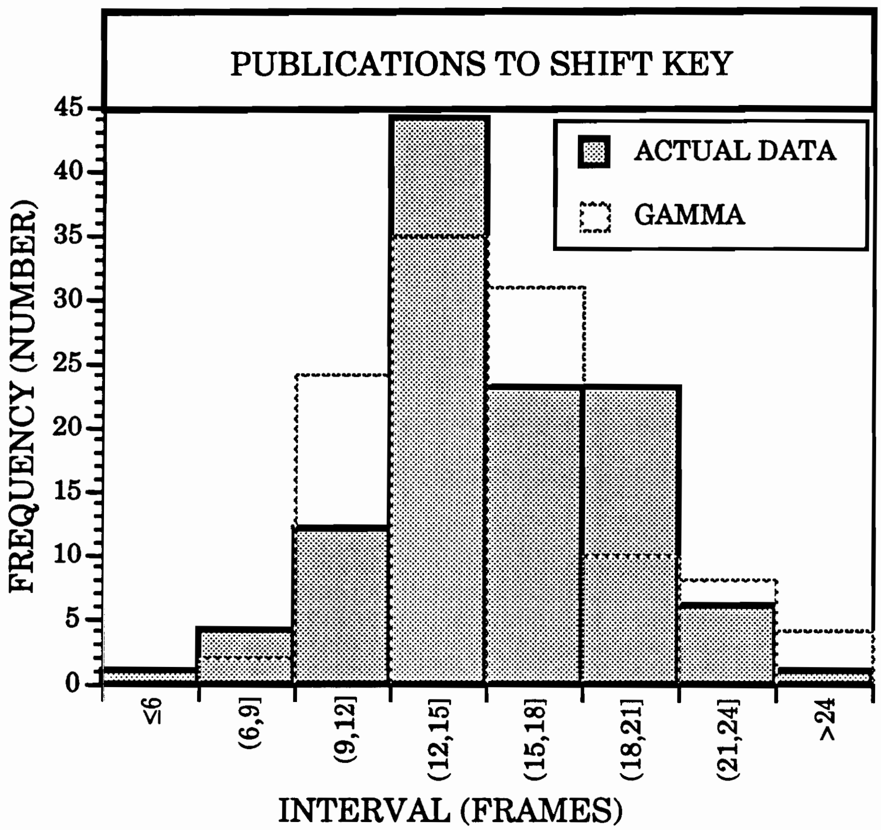
APPENDIX F

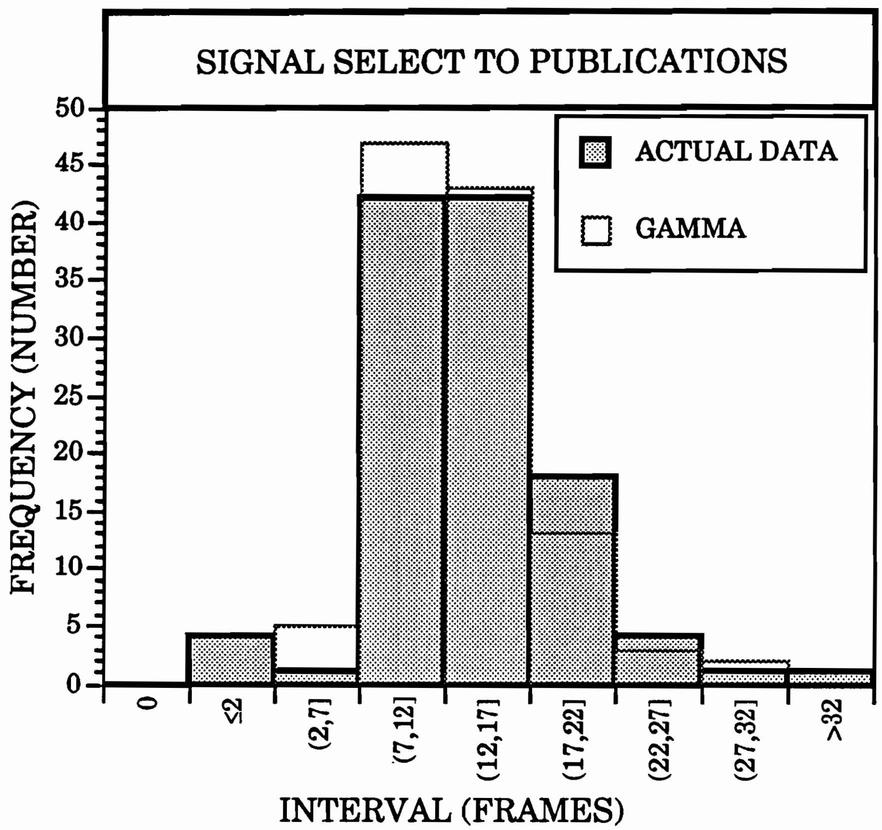
FREQUENCY DISTRIBUTIONS

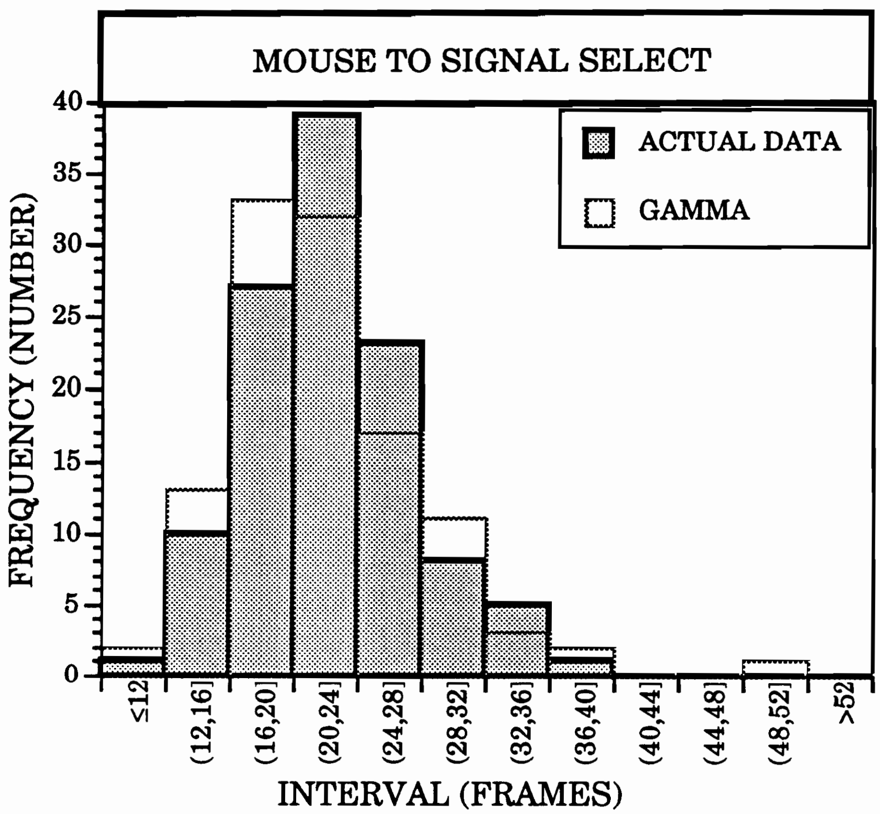


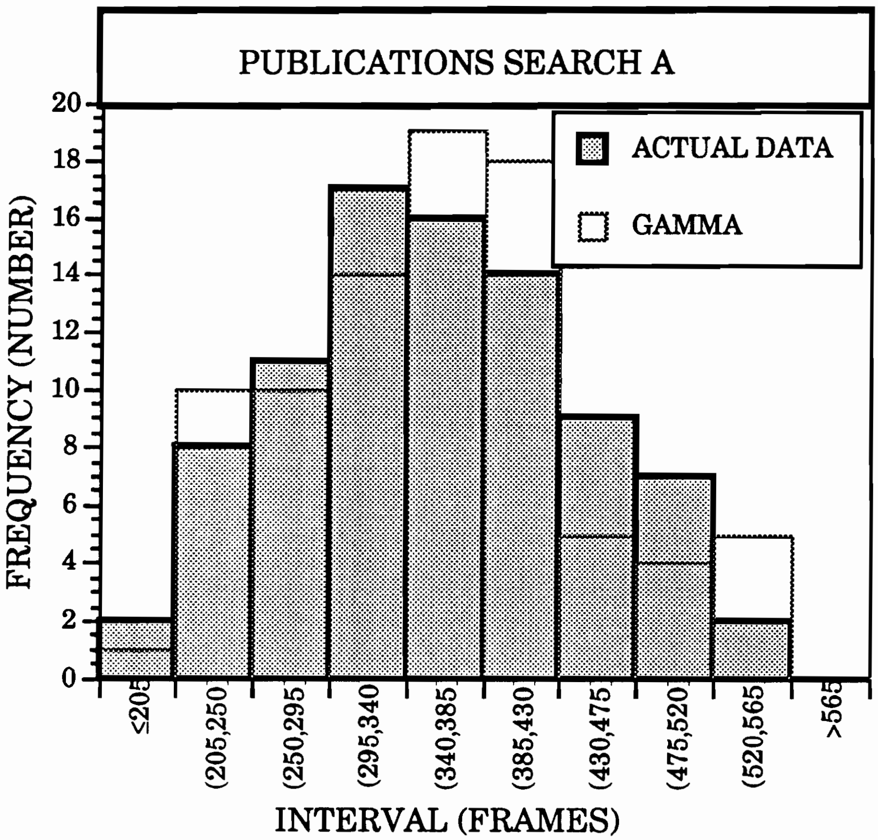

DESIGNATE ID TO 1ST EFX#
ACTUAL DATA
GAMMA
FREQUENCY (NUMBER)
30
25
20
15
10
5
0
<4
(4,8]
(8,12]
(12,16]
(16,20]
>20
INTERVAL (FRAMES)

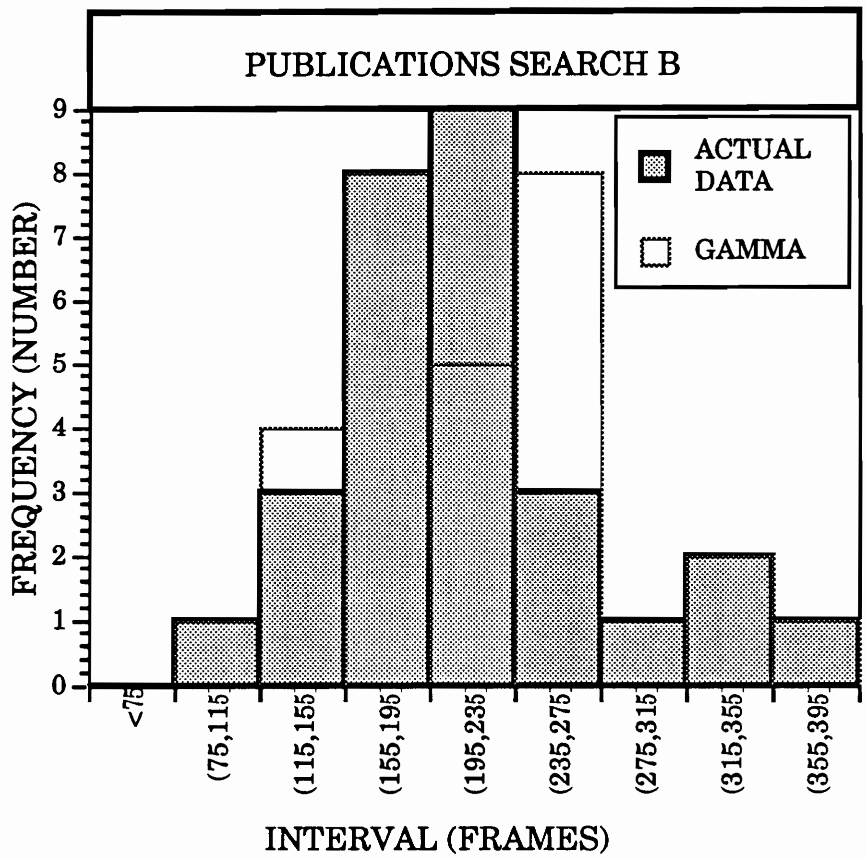

2ND & 3RD EFX #'S AND RETURN KEY
ACTUAL DATA
GAMMA
FREQUENCY (NUMBER)
100
90
80
70
60
50
40
30
20
10
0
≤3
(3,7
(7,11
(11,15
(15,19
(19,23
(23,27
(27,31
(31,34
>34
INTERVAL (FRAMES)

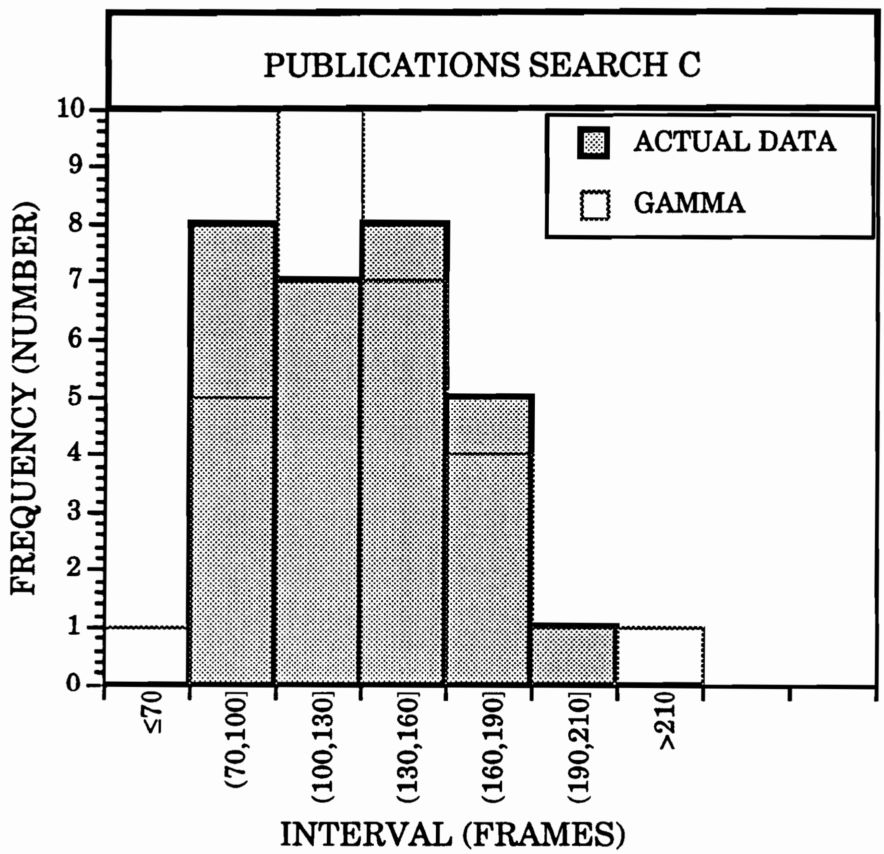

PUBLICATIONS TO SHIFT KEY
ACTUAL DATA
GAMMA
FREQUENCY (NUMBER)
45
40
35
30
25
20
15
10
5
0
≤6
(6,9]
(9,12]
(12,15]
(15,18]
(18,21]
(21,24]
>24
INTERVAL (FRAMES)

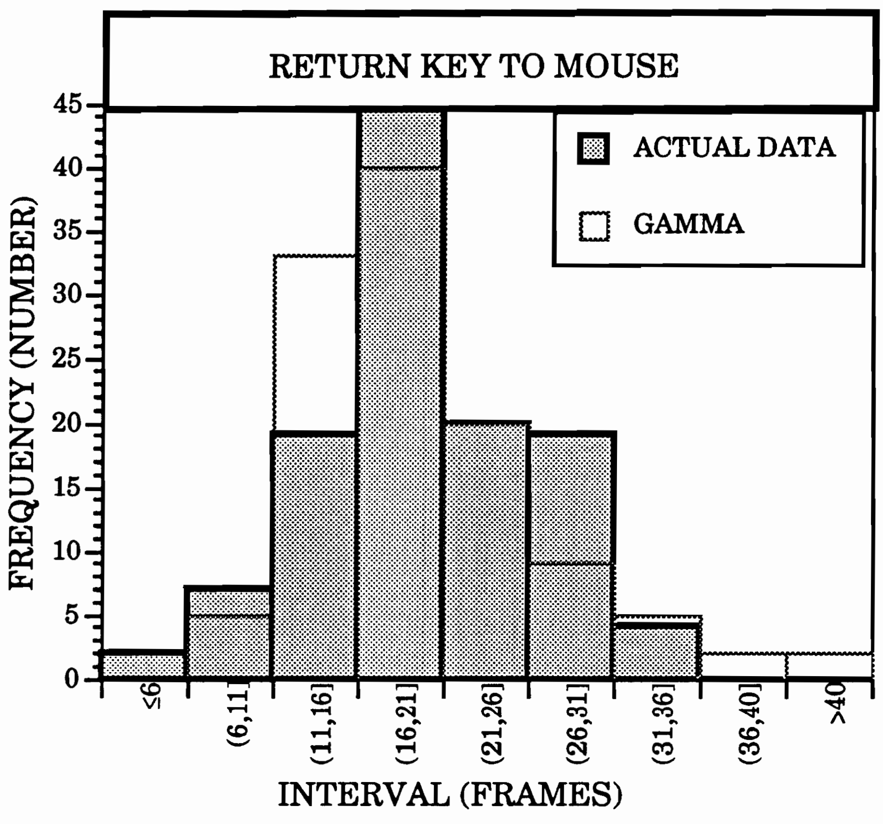

SIGNAL SELECT TO PUBLICATIONS
ACTUAL DATA
GAMMA
FREQUENCY (NUMBER)
50
45
40
35
30
25
20
15
10
5
0
0
≤2
(2,7]
(7,12]
(12,17]
(17,22]
(22,27]
(27,32]
>32
INTERVAL (FRAMES)

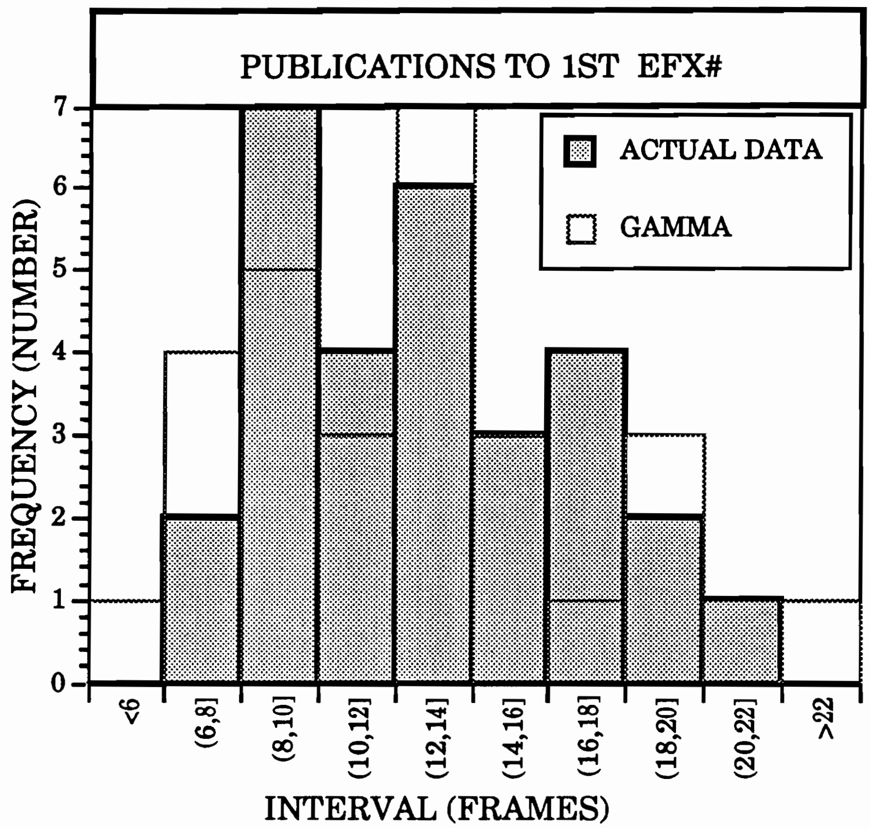

MOUSE TO SIGNAL SELECT
ACTUAL DATA
GAMMA
FREQUENCY (NUMBER)
40
35
30
25
20
15
10
5
0
≤12
(12,16]
(16,20]
(20,24]
(24,28]
(28,32]
(32,36]
(36,40]
(40,44]
(44,48]
(48,52]
>52
INTERVAL (FRAMES)

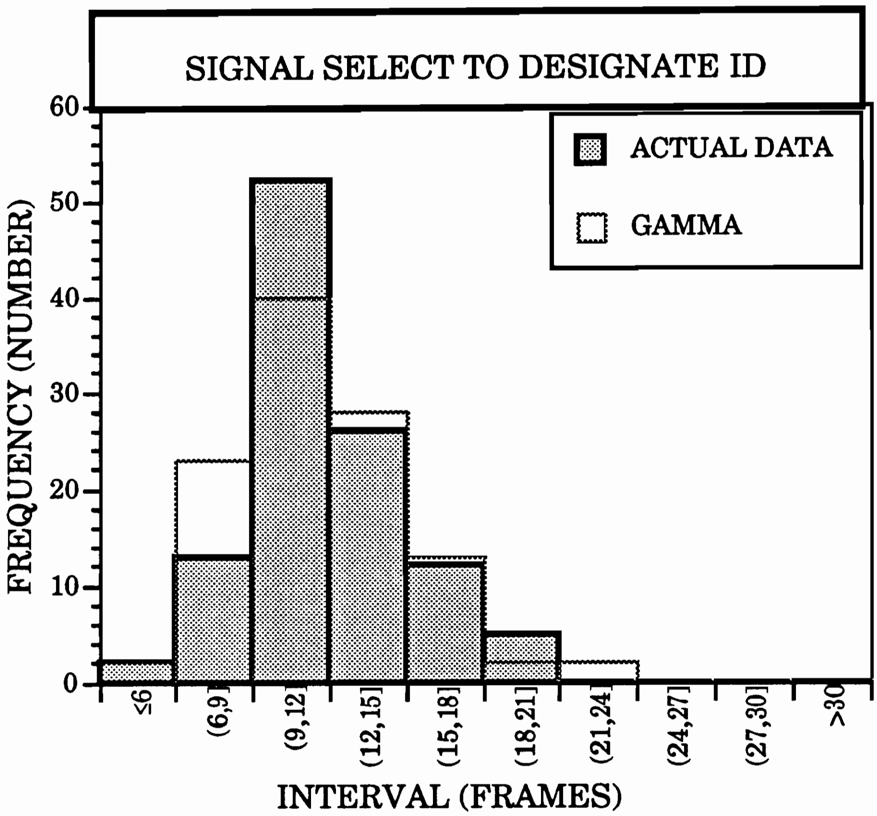

PUBLICATIONS SEARCH A
ACTUAL DATA
GAMMA
FREQUENCY (NUMBER)
0
2
4
6
8
10
12
14
16
18
20
≤205
(205,250
(250,295
(295,340
(340,385
(385,430
(430,475
(475,520
(520,565
>565
INTERVAL (FRAMES)


PUBLICATIONS SEARCH B
ACTUAL DATA
GAMMA
FREQUENCY (NUMBER)
0
1
2
3
4
5
6
7
8
9
<75
(75,115
(115,155
(155,195
(195,235
(235,275
(275,315
(315,355
(355,395
INTERVAL (FRAMES)


PUBLICATIONS SEARCH C
ACTUAL DATA
GAMMA
FREQUENCY (NUMBER)
0
1
2
3
4
5
6
7
8
9
10
≤70
(70,100]
(100,130]
(130,160]
(160,190]
(190,210]
>210
INTERVAL (FRAMES)


RETURN KEY TO MOUSE
ACTUAL DATA
GAMMA
FREQUENCY (NUMBER)
0
5
10
15
20
25
30
35
40
45
≤6
(6,11]
(11,16]
(16,21]
(21,26]
(26,31]
(31,36]
(36,40]
>40
INTERVAL (FRAMES)


PUBLICATIONS TO 1ST EFX#
ACTUAL DATA
GAMMA
FREQUENCY (NUMBER)
0
1
2
3
4
5
6
7
<6
(6,8]
(8,10]
(10,12]
(12,14]
(14,16]
(16,18]
(18,20]
(20,22]
>22
INTERVAL (FRAMES)


SIGNAL SELECT TO DESIGNATE ID
ACTUAL DATA
GAMMA
FREQUENCY (NUMBER)
0
10
20
30
40
50
60
≤6
(6,9]
(9,12]
(12,15]
(15,18]
(18,21]
(21,24]
(24,27]
(27,30]
>30
INTERVAL (FRAMES)

APPENDIX G

SWAT GROUP SCALING SOLUTION

TIME LOAD	EFFORT LOAD	STRESS LOAD	WORKLOAD SCORE
1	1	1	00.0
1	1	2	20.4
1	1	3	40.8
1	2	1	12.4
1	2	2	32.8
1	2	3	53.2
1	3	1	26.4
1	3	2	46.8
1	3	3	67.2
2	1	1	10.3
2	1	2	30.7
2	1	3	51.1
2	2	1	22.7
2	2	2	43.1
2	2	3	63.5
2	3	1	36.7
2	3	2	57.1
2	3	3	77.5
3	1	1	32.8
3	1	2	53.2
3	1	3	73.6
3	2	1	45.2
3	2	2	65.6
3	2	3	86.0
3	3	1	59.2
3	3	2	79.6
3	3	3	100.0

APPENDIX H

PERFORMANCE SCORES: ANOVA TABLE

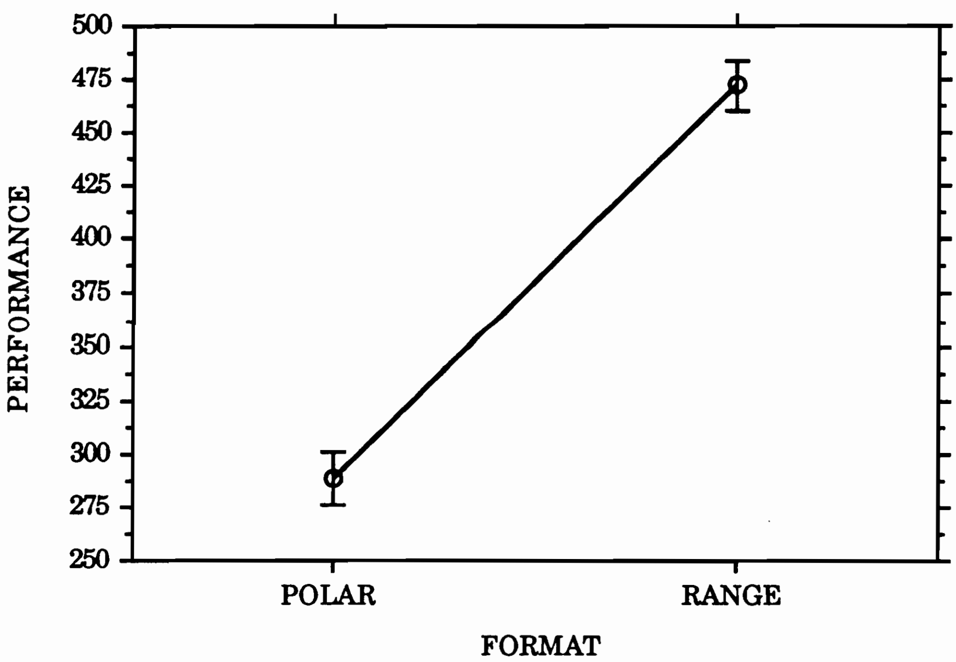
Source	*df*	*e**	*MS*	*F*	*p*
Subject	11		43038.325		
Format	1	1.000	2428792.710	104.683	.0001
Format *Subject	11		23201.419		
Symbols	1	1.000	145950.461	10.400	.0081
Symbols*Subject	11		14033.566		
Color	1	1.000	220972.092	10.685	0075
Color *Subject	11		20679.661		
Density	2	.906	556850.983	45.862	.0001
Density *Subject	22		12141.837		
Format *Symbols	1	1.000	83491.189	6.589	.0262
Format*Symbols*Subject	11		12671.991		
Format*Color	1	1.000	38137.192	1.924	.1929
Format*Color*Subject	11		19825.220		
Symbols*Color	1	1.000	12332.499	.831	.3815
Symbols*Color*Subject	11		14840.986		
Format*Density	2	.784	18875.293	1.709	.2084
Format*Density*Subject	22		11047.550		
Symbols*Density	2	.944	5061.735	.418	.6635
Symbols*Density*Subject	22		12110.417		
Color*Density	2	.824	2771.447	.183	.8235
Color*Density*Subject	22		15127.014		
Form*Sym*Col	1	1.000	2728.245	.104	.7531
Form*Sym*Col*Sub	11		26237.432		
Form*Sym*Dens	2	.847	7905.740	.393	.6770
Form*Sym*Dens*Sub	22		20138.623		
Form*Col*Dens	2	.899	2996.919	.225	.8003
Form*Col*Dens*Sub	22		13319.346		
Sym*Col*Den	2	.939	9142.256	.873	.4317
Sym*Col*Den*Sub	22		10472.879		
Form*Sym*Col*Den	2	.970	6120.700	.386	.6844
Form*Sym*Col*Den*Sub	22		15861.614		

*All F-test results were corrected for sphericity using the Greenhouse-Geisser (1959) adjustment procedure. Epsilon denotes the correction factor and tabulated probabilities reflect the adjusted F-tests.

Means Table
Effect: Format
Dependent: Normalized Performance

	Count	Mean	Std. Dev.	Std. Error
POLAR	144	288.380	148.604	12.384
RANGE	144	472.046	142.648	11.887

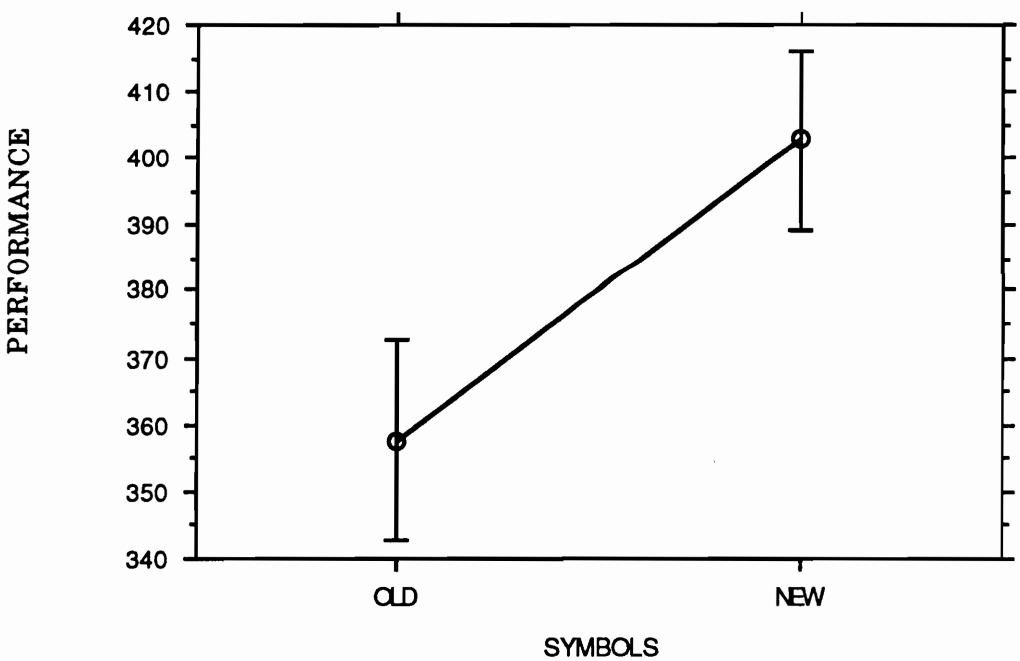
Means Plot
Effect: Format
Dependent: Normalized Performance
With Standard Error Bars



Means Table
Effect: Symbol
Dependent: Normalized Performance

	Count	Mean	Std. Dev.	Std. Error
OLD	144	357.701	180.805	15.067
NEW	144	402.725	160.327	13.361

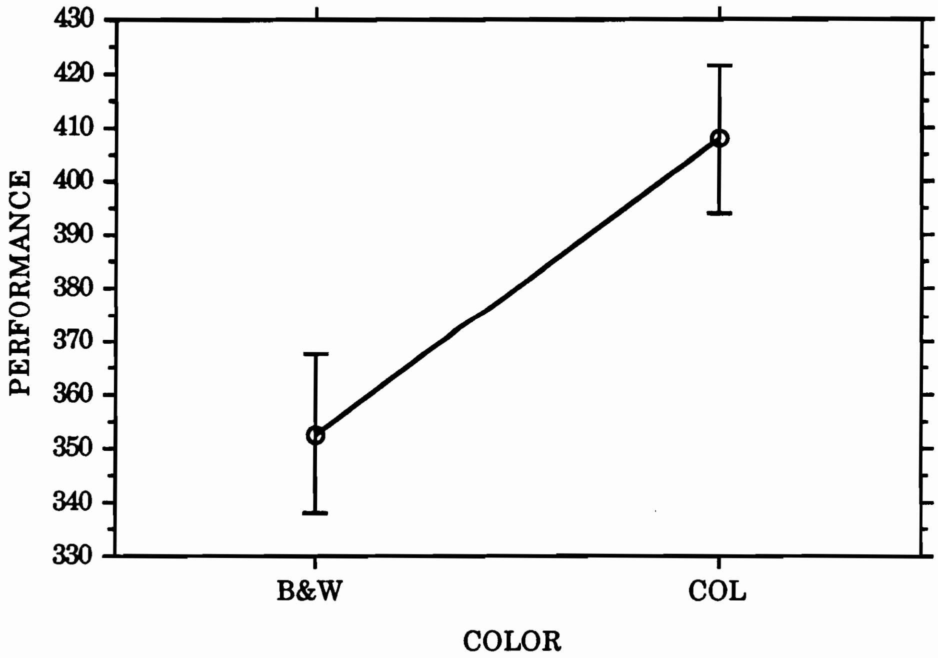
Means Plot
Effect: Symbol
Dependent: Normalized Performance
With Standard Error Bars



Means Table
Effect: Color
Dependent: Normalized Performance

	Count	Mean	Std. Dev.	Std. Error
B&W	144	352.513	174.660	14.555
COL	144	407.912	165.423	13.785

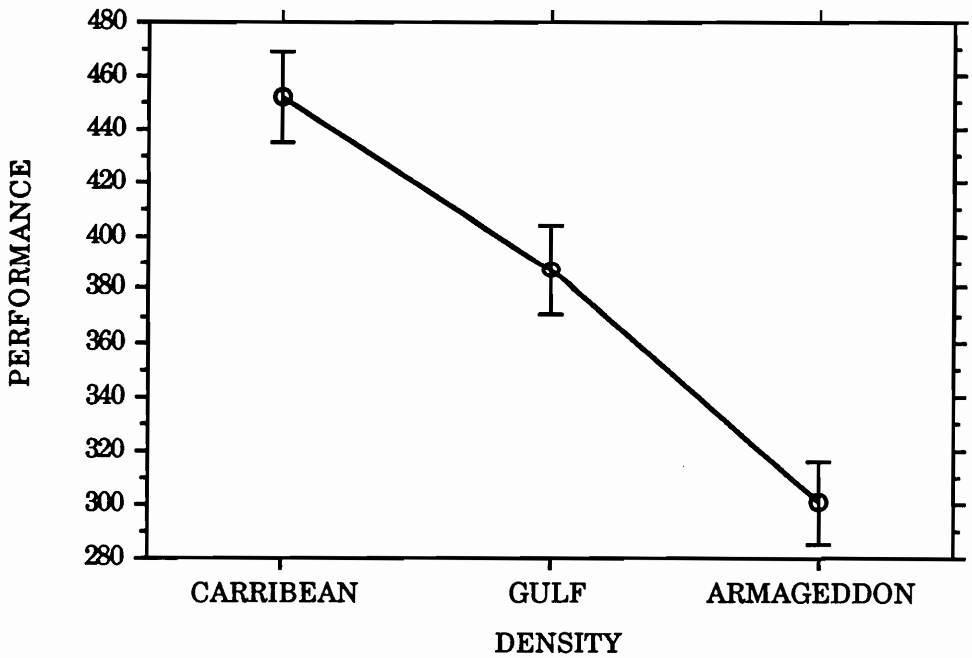
Means Plot
Effect: Color
Dependent: Normalized Performance
With Standard Error Bars



Means Table
Effect: Density
Dependent: Normalized Performance

	Count	Mean	Std. Dev.	Std. Error
CARRIBEAN	96	452.281	162.703	16.606
GULF	96	387.827	166.470	16.990
ARMAGEDDON	96	300.531	153.393	15.656

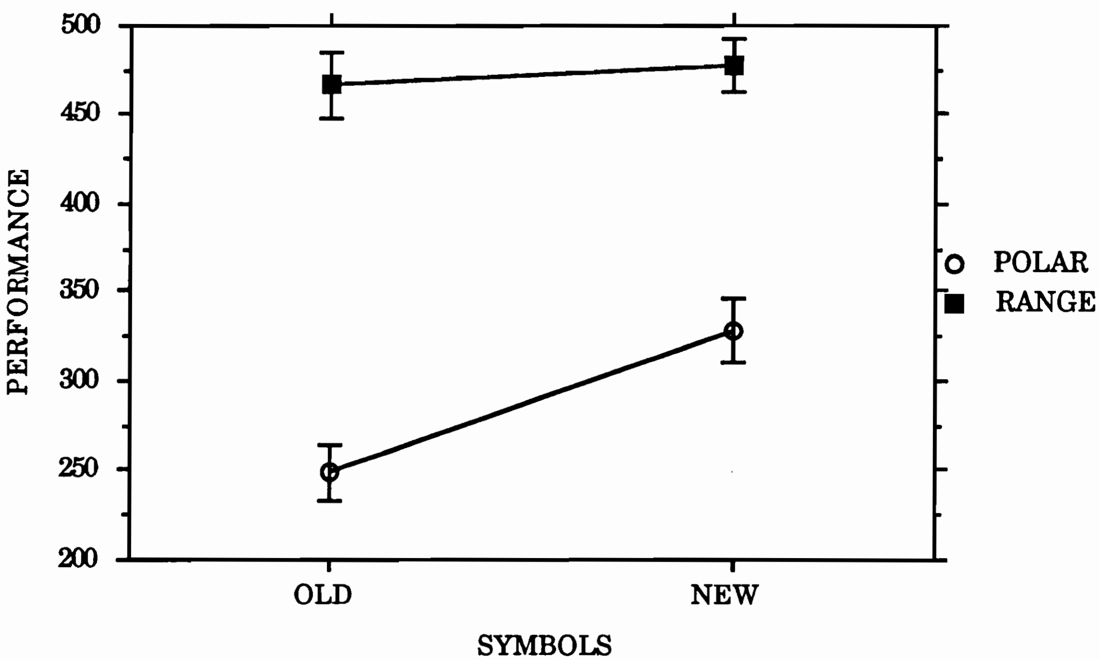
Means Plot
Effect: Density
Dependent: Normalized Performance
With Standard Error Bars



Means Table
Effect: Format *Symbol
Dependent: Normalized Performance

	Count	Mean	Std. Dev.	Std. Error
POLAR, OLD	72	248.842	132.771	15.647
POLAR, NEW	72	327.918	153.878	18.135
RANGE, OLD	72	466.561	155.495	18.325
RANGE, NEW	72	477.531	129.397	15.250

Means Plot
Effect: Format *Symbol
Dependent: Normalized Performance
With Standard Error Bars.



APPENDIX I

WORKLOAD SCORES: ANOVA TABLE

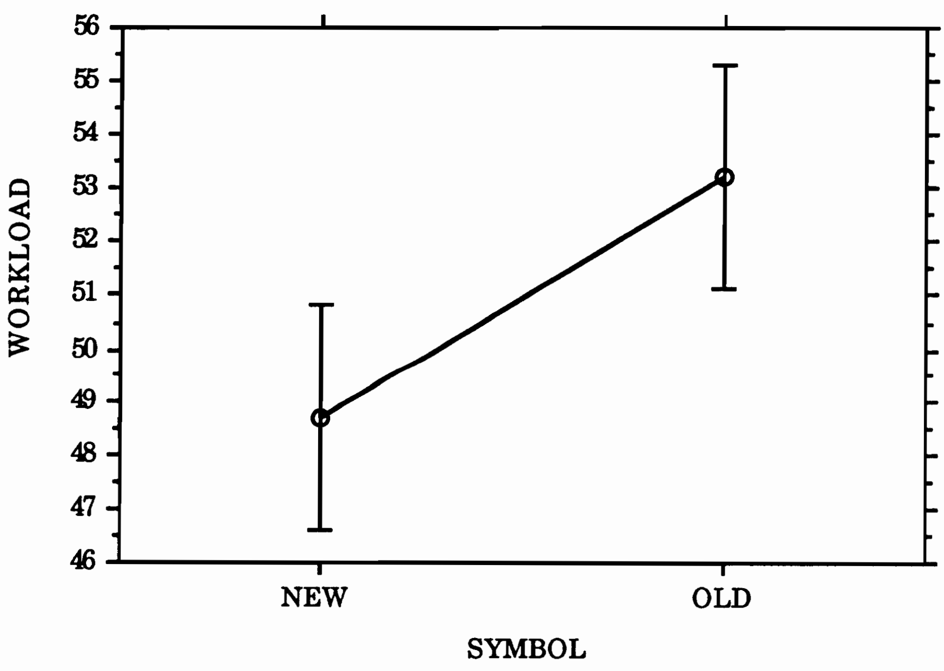
Source	*df*	e^*	*MS*	*F*	*p*
Subject	11		8430.935		
Format	1	1.000	90.878	.631	.4438
Format *Subject	11		144.046		
Symbols	1	1.000	1460.791	5.787	.0349
Symbols*Subject	11		252.447		
Color	1	1.000	207.774	2.673	.1303
Color *Subject	11		77.737		
Density	2	.555	10572.828	7.706	.0147
Density *Subject	22		1372.100		
Format *Symbols	1	1.000	56.198	.162	.6949
Format*Symbols*Subject	11		346.460		
Format*Color	1	1.000	234.758	1.730	.2151
Format*Color*Subject	11		135.667		
Symbols*Color	1	1.000	982.648	14.927	.0026
Symbols*Color*Subject	11		65.828		
Format*Density	2	.894	22.338	.168	.8235
Format*Density*Subject	22		132.998		
Symbols*Density	2	.795	7.994	.044	.9274
Symbols*Density*Subject	22		182.358		
Color*Density	2	.781	18.707	.158	.8308
Color*Density*Subject	22		118.579		
Form*Sym*Col	1	1.000	353.381	.909	.3607
Form*Sym*Col*Sub	11		3.403		
Form*Sym*Dens	2	.831	130.353	.991	.3758
Form*Sym*Dens*Sub	22		131.516		
Form*Col*Dens	2	.965	6.805	.020	.9778
Form*Col*Dens*Sub	22				
Sym*Col*Den	2	.964	173.856	1.508	.2441
Sym*Col*Den*Sub	22		115.310		
Form*Sym*Col*Den	2	.714	16.190	.192	.7529
Form*Sym*Col*Den*Sub	22		84.120		

*All F-test results were corrected for sphericity using the Greenhouse-Geisser (1959) adjustment procedure. Epsilon denotes the correction factor and tabulated probabilities reflect the adjusted F-tests.

Means Table
Effect: Symbol
Dependent: Workload

	Count	Mean	Std. Dev.	Std. Error
NEW	144	48.703	25.278	2.107
OLD	144	53.208	25.291	2.108

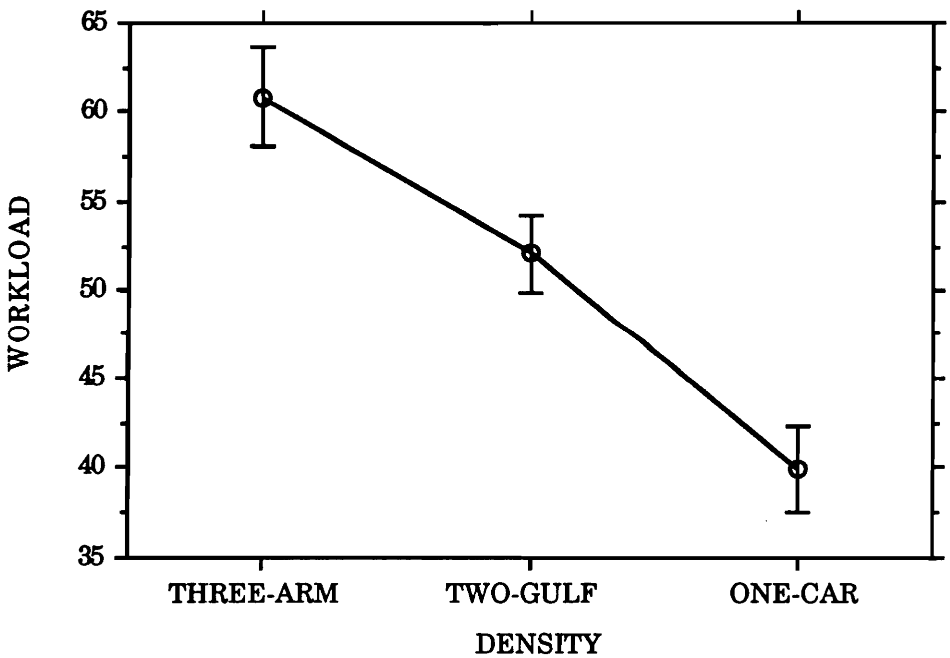
Means Plot
Effect: Symbol
Dependent: Workload
With Standard Error Bars



Means Table
Effect: Density
Dependent: Workload

	Count	Mean	Std. Dev.	Std. Error
ARMAGEDDON	96	60.826	27.088	2.765
GULF	96	52.108	21.073	2.151
CARIBBEAN	96	39.932	23.229	2.371

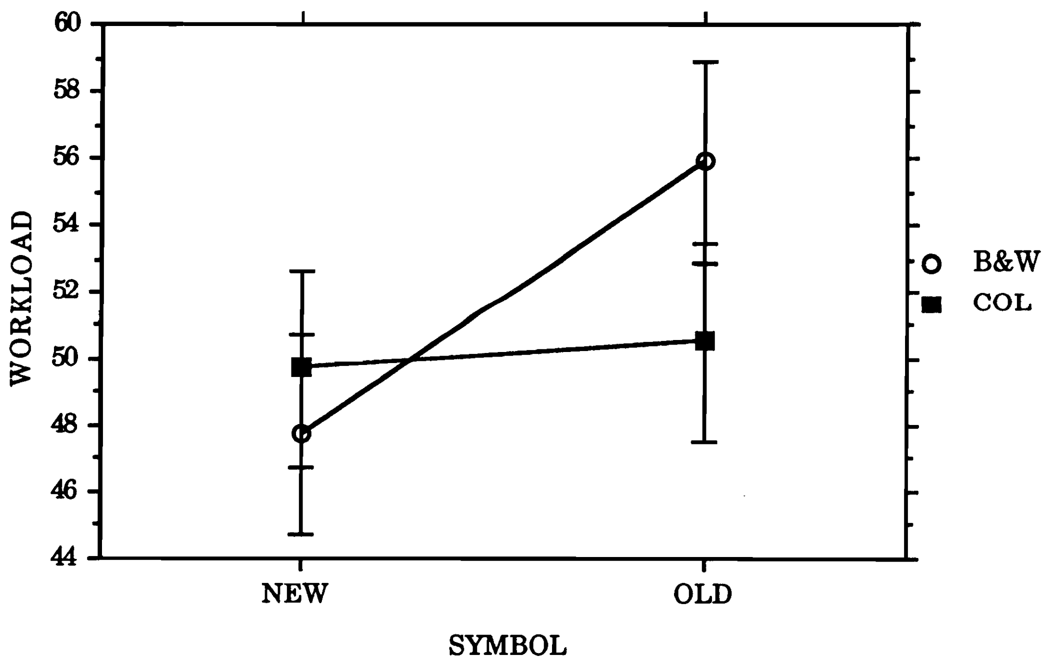
Means Plot
Effect: Density
Dependent: Workload
With Standard Error Bars



Means Table
Effect: Color * Symbol
Dependent: Workload

	Count	Mean	Std. Dev.	Std. Error
B&W, NE...	72	47.706	25.382	2.991
B&W, OLD	72	55.904	25.196	2.969
COL, NEW	72	49.701	25.312	2.983
COL, OLD	72	50.511	25.272	2.978

Means Plot
Effect: Color * Symbol
Dependent: Workload
With Standard Error Bars



VITA

Sandra A. Moscovic was born on December 6, 1963. She received a B.S. in behavioral sciences from the United States Air Force Academy (1986). She received her M.S. in industrial psychology with an emphasis in human factors from St. Mary's University in San Antonio, Texas (1989).

Capt. Moscovic is an active duty Air Force officer. She has served in the Air Force as a behavioral scientist researcher from 1986 to 1987 and as a human factors engineer until the present time. Her current interests include displays and controls interface design, primarily on military systems. She is a student member of the Human Factors Society (HFS), VPI & SU HFS student chapter, and Society for Information Display.

Capt. Moscovic will continue her career as a human factors engineer with the Air Force as an instructor at the United States Air Force Academy in Colorado Springs, Colorado.

Sandra A. Moscovic

Signature