Operator Workload Measurement Validation for the Mark IV-A DSCC Monitor and Control Subsystem

M. Le May
Montclair State College, New Jersey

E. E. Hird
Operations Sustaining Engineering Section

B. Y. Rodriquez
Goldstone Operations Section

The validation of some proposed measures of operator workload in a large interactive computer system is described. The tests were conducted on Mark III Deep Space Network (DSN) equipment using experienced operators as subjects. In addition to identifying the significant operator workload measures, some useful guidelines have also been obtained on operator task scheduling.

I. Introduction

The updating and increasing automation of large, interactive computer systems is a common process where such systems are in use. The DSN is currently undertaking such an update by introducing a new Mark IV-A monitor and control subsystem at its three station complexes located approximately 120 degrees longitude apart around the earth. The purpose of the Mark IV-A system is to reduce operating costs by centralizing operations at each station, and increasing the amount of automation (Ref. 1). This will result in new combinations of operators, and new ways of operating the system, with as yet unknown workload demands.

Therefore, it was considered necessary to develop measures of operator workload that would underscore those subtasks on which the load is likely to exceed scheduled times and increase operator stress, especially if anomalies occur during the task performance. The extensive literature on human workload measurement (Refs. 2, 3, and 4) yielded a number of possible techniques. An evaluation of their effectiveness in various circumstances, also based on the literature, led to a choice of three measures that might prove useful. These are:

(1) Sequential operator ratings of workload on subtasks.

(2) The ratio of the time required to the time available to do a particular task.

(3) Initialization time, which is defined as the time taken to initiate the physical (keystroke) portion of a subtask.

These measures are described in detail in Ref. 1. The salient feature of the particular technique used here is that the operator ratings are taken at the end of each small subtask of a whole task (e.g., at the end of the load of a particular program such as the Metric Data Assembly (MDA), in a total "load and go" task) so that ratings and times are obtained for each subtask. In this way, subtasks that are likely to exercise an operator's capacity near its limit are rated and timed in the context of the whole task. This is especially useful during the design and operational testing phase of a new large interactive computer system such as the Mark IV-A, since it may be possible to make design changes or devise workarounds to alleviate difficulties with particular subtasks during this phase.

Although the three proposed measures were culled from an extensive literature on the subject, and should be usable on this basis, their use in large interactive computer systems has not been tested. Moreover, the technique of inserting ratings at intervals in the task and using ratings to time subtasks is also unique in this context. Therefore, the study described here was undertaken to assess the validity and reliability of the proposed workload measures in the operational environment of an existing Mark III Deep Space Station (DSS). The experiment used a "5 × 3 within subjects" design, in which each subject completed a whole computer "load and go" task consisting of five subtasks under each of three workload conditions: low, medium, and high workload. (The low workload condition involved only four subtasks.) In this way, the sensitivity of the measures to an objective increase in workload, and to differences among subtasks, could be determined since, if the measures were valid, they could be expected to rise with actual workload increases and with more difficult tasks.

In addition to the rating of each subtask, operators were asked to rate a number of aspects of the working environment on a separate questionnaire.

II. Method

A. Subjects

Nine operators of the ground tracking and data acquisition equipment at DSS 12 at Goldstone, California, served as subjects. All had some experience and were thoroughly familiar with the job.

B. Apparatus

The operating elements of the DSN include the Network Operations Control Center (NOCC) at JPL and three Deep Space Communications Complexes (DSCCs) located at Canberra, Australia; Goldstone, California; and Madrid, Spain. The NOCC controls, monitors, and coordinates the DSCCs' operational activities in response to the requirements placed upon the DSN by the various Flight Projects. The DSCCs provide ground-based communications links with Flight Projects' spacecrafts in interplanetary flight or in highly eccentric earth orbit. In addition, they support some radio astronomy tasks and very long baseline interferometry measurements. Each of the DSCCs currently operates a 64-m and a 34-m antenna. The 64-m and 34-m antennas at Goldstone are operated independently of each other at separate locations, while similar antennas at Canberra and Madrid are colocated and conjointly operated. A number of computerized subsystems, some of which were used in collecting data for the study reported here, are used in a partially automated manner in conjunction with some other manually operated subsystems and a selected antenna to acquire data from the various spacecraft in accordance with a predetermined schedule.

Time was scheduled in four-hour blocks at DSS 12 at Goldstone to conduct the prepared operator workload trials. One operator could always be tested in a time block, and it was sometimes possible to test two. A simulated "load and go" task that required one operator was designed for the study. The task involved initializing seven different computers so that they were ready to track a designated spacecraft. A "load and go" may be characterized as the reinitialization of each subsystem computer for the next spacecraft to be tracked during the period of time needed to repoint the antenna. It was assumed that subsystem elements that need calibration will not change significantly during the upcoming support activity.

C. Procedure

There were three versions of the task, designed to impose a low, medium, or high workload on the operator. The low workload condition was a simple "load and go" task with no complications; its scheduled time for completion was 12 minutes. In the medium workload condition, a ranging operation requiring additional equipment was added. The scheduled time for this was 8 minutes in addition to the 12 minutes required for the original task. In the high workload condition the same task was used (with ranging included), and two anomalies on two different computers were added by the investigators. The scheduled times were set by the experimenters, and provided just enough time to complete each task under normal working conditions.

Before any data was collected, the purpose and methods of the study were explained to all the potential subjects in a written form, and in a brief discussion session. Then, at the start of each session, subjects read a standard set of instructions which directed them to rate, on a scale from one to seven, the amount of mental effort (how hard they had to think) and the amount of time stress (how rushed they felt) involved in the subtask they had just completed. Each of the ratings was given a verbal label. These labels were printed on a card that was mounted on the console for easy visibility during the procedure. A reproduction of the card is illustrated in Fig. 1. They were also informed of the scheduled time of 12 minutes, and of the 8-minute time "pad" allowed for the medium and high workload conditions.

Following the instructions, subjects were given a short practice session in which to familiarize themselves with the procedure, so that the workload measures themselves would not intrude on the timing of the subtasks.

Then, when they were sure they understood the procedure thoroughly, the high, medium, and low workload conditions were arranged in a random order and the trials were begun. Each operator input and computer response was obtained on a printout, along with its time, so that operator ratings and times could be read for each subtask. In a continuous task such as the one studied here, the end of one subtask signals the start of the next; thus, the operator's ratings could be used to obtain the total time for that subtask and the time for the operator to start the next one. This latter was the time between the keystroke that gave the rating and the next operator input. Total subtask time was the time between one rating and the next.

The low workload condition was chosen so that operators could accomplish the task easily, with no time pressure; it was possible to accomplish the medium workload task in the allotted time, given the 8-minute "pad", and it was possible to complete the task in the time allowed under the high workload condition if the operator understood the system thoroughly and worked fast.

After the completion of each condition, the operator was asked to complete a 14-item workload questionnaire on which a number of aspects of the working environment were also rated on a scale from one to seven. This is reproduced in Fig. 2.

Following the completion of all of the conditions, a debriefing session was held in which the operators were asked to comment on the study and explain the ratings they had just given.

III. Results

Three measures of operator workload were obtained from the printouts. The first was, of course, the operator's rating of subtask difficulty and time stress. This was obtained directly from the printout. The notation on the printout of the times when these were given made it possible to calculate the amount of time the operator actually required to perform the subtask. This was then used to obtain the second measure, which was the ratio of the time required to the time allowed.

$$W = \frac{T_r}{T_a}$$

where

W = workload ratio

T_r = actual time required by the operator

T_a = time allowed, or scheduled time

The scheduled time is set by NOCC for each whole task, e.g., one-half hour is allowed for a complete "load and go" operation. Actual times to perform the task were obtained from printouts for four recent operational runs on which no anomalies were encountered. The task was found to occupy an average of 38 percent of the scheduled time under these conditions. Average times were then obtained for each subtask under operational conditions, and these were divided by 0.38 to obtain the scheduled time, T_a for that subtask. Thus, the scheduled time for each subtask was assigned on the basis of its average proportion of the scheduled time for the whole task. It was necessary to do this because scheduled or allowed times are given only for whole tasks, and detailed data on subtasks was desired.

The third measure, initialization time, was the time between the rating of one task and the first input on the next.

Each of the measures was subjected to a "3 (workload conditions) by 5 (subtasks)" analysis of variance to determine whether differences observed among the conditions were significant. The two ratings (mental effort and time stress) were analyzed separately. The time stress ratings and the workload ratio varied significantly with both subtask and condition; the mental effort rating varied only with subtask, and the initialization time was not related to workload variations. No significant interactions were observed, i.e., the effect of increasing workload was the same over all subtasks.

The means of each of the measures for all subtask and workload conditions are shown in Tables 1 through 4, where each of the processors used in program loading are indicated. The mean ratings for "mental effort" rise with increasing workload, and vary with the subtask, but these differences are not significant. The means of the ratings for each load condition over all subtasks for the "time stress" measurement, however, do rise significantly with the objective increase in workload, and also change with the subtask. This indicates that different subtasks impose differing amounts of time stress, and that the measure is sensitive to objective workload increases. Figure 3 illustrates the overall rise in the ratings for low, medium, and high workload conditions. The data has been normalized for easier comparison with the workload ratio data presented in Fig. 4.

The means of the workload ratio (W) for each condition are presented in Table 3, and the significant rise in workload ratio with increasing objective workload is illustrated in Fig. 3. The workload ratio is somewhat more sensitive to workload conditions than is the "time stress" rating, although both measures rise and fall similarly with the same subtasks. This is consistent with the lack of significant interaction terms. The Spearman rho correlation between "time stress" ratings and workload ratio is 0.78. It appears that both measures are measuring the same subtask characterisics, and both are sensitive to workload increases.

Variations in initialization time did not prove significant, either for subtask or for imposed workload. These are presented in Table 4.

The number of errors (see Table 5) made by operators during task performance was relatively small (43 in almost 8 hours of operation) so no statistical analysis could be done, but they increased from 9 under the low workload condition to 17 under the medium and high workload conditions. The number of operator inputs relative to the minimum number of required inputs also increased from 1.09 at low workload to 1.15 at medium workload and 1.34 at high workload.

The mean ratings for each of the statements on the Operator Workload Questionnaire are presented in Fig. 1. None of the ratings seem to indicate that operators have particular problems with the working environment of the system as it now stands. During the exit interview, the operators agreed that the task chosen was representative of typical tasks at the DSS. They also felt that the taking of the workload measures was somewhat intrusive on task performance. No other problems in operating the system were mentioned.

IV. Discussion

The significant findings for the time stress ratings and the workload ratio mean that these two measures do, in fact, vary with objective workload and can therefore be used together to measure operator workload at the DSN monitor and control consoles. Moreover, the fact that the subsystem has a significant effect on both measures means that subtasks vary in the amount of workload. Tasks with higher measured workloads are those on which errors leading to data loss might be expected to occur. This knowledge itself should help operators in becoming oriented to a new system (operators probably are aware of "trouble spots" in an older system without any measures), and may provide information useful for hardware and software improvements in the system during operational testing.

The "mental effort" ratings did rise with increasing workload, but differences were too small to be significant. This is not too surprising, since it is not usually necessary to exert much mental effort in the performance of a very familiar operation. This measure probably would be sensitive to workload during the operation of a new system. At any rate, since the sensitivities of time stress ratings and workload ratio were demonstrated with the mental effort measure in use, it should be left in place in future use, in order not to change the cognitive aspects of the situation, i.e., its removal might have an effect on the other two measures, which were significant.

The initialization time measure was an attempt to measure the time required for mental processing of the information needed to perform a control function. The reasons for its failure to respond to objective workload increases are beyond the scope of this paper. However, in view of the recent finding of Wierwille and Connor (1983) that only 3 out of 20 workload measures were actually sensitive to workload increases on a particular task, the lack of sensitivity of 1 such measure is not too surprising.

The high correlation between the rating and the workload ratio measures indicates that they are essentially measuring the same thing, and they therefore act as a check on each other's reliability. They are not redundant, however. The rating is extremely easy to use and interpret, and can be obtained immediately for use in possible "on the spot" diagnosis of troublesome tasks. Moreover, it is necessary in order to obtain the times used in calculation of the workload ratio.

The workload ratio is a particularly useful measure, since it provides information on the time to perform each subtask as well as the whole task, and on the proportion of task time needed for each subtask. It is the measure most sensitive to workload variations. Moreover, although the mean ratios appear rather low for all the conditions, there is considerable variability. In particular, the presence of anomalies, even simple ones such as those introduced by the experimenters, can raise the ratio near or over 1.00. This means that the scheduled time has been exceeded on that task. This happened seven times, or five percent of the total number of subtasks during the medium and high workload conditions, and the ratio was over 0.80 on another seven occasions. An examination of the raw data showed that these were the occasions on which errors were likely to be made. Moreover, nine "volunteer" equipment failures (not induced by the experimenters) occurred during the testing, and these also seemed related to increased workload, since seven occurred during the high workload condition. Thus, both equipment and operator failures seem to be related to operator workload. In addition, the simple fact of exceeding the scheduled time can mean lost data. Thus, it is important for system operation that the operator workload be kept at a moderate level and the first step towards achieving this goal is the use of the techniques developed here for operator workload measurement.

V. Conclusions

1. Two workload measures that are sensitive to increasing objective workload were developed for use in operating a large, interactive computer system (the DSN monitor and control subsystem). They are:

   a. Operator rating, on a scale from one to seven, of the "time stress" encountered in a particular subtask.

   b. The ratio of the time required to do a particular subtask to the time allowed for it.

2. The average workload ratio in the (typical) task studied was approximately 0.40; i.e., the tasks occupied 40% of the scheduled time. This ratio led to exceeding the scheduled time on five percent of the subtasks.

3. Time stress ratings vary from one individual to another. However, subtasks that have a mean rating of three or above from at least three operators are considered problematic.

4. To ensure that the 0.40 workload ratio is not exceeded, workload ratings and ratios can be taken whenever new equipment or a new program or procedure is introduced, so that reasonable scheduled times can be established. This work would be done after operators are somewhat familiar with the system. Also it may need to be done more than once, as time (T_r) will become shorter as operators become more familiar with the system.

References

1. Le May, M., Chafin, R., and Hird, E. E., "The Measurement of Operator Workload in the Mark IV-A DSCC Monitor and Control Subsystem." *The Telecommuncations and Data Acquisition Progress Report, 42-72,* October-December 1982, pp. 177-185, Jet Propulsion Laboratory, Pasadena, California, February 15, 1983.

2. Wierwille, W. W., Williges, R. C., and Schiflett, S. G., "Aircrew Workload Assessment Techniques," in Hartman, B. O. and McKenzie, R. E., *Survey of Methods to Assess Workload.* AGARDograph No. 246, August, 1979.

3. Moray, N., *Mental Workload: Its Theory and Measurement.* Plenum Press, New York, 1979.

4. Wierwille, W. W. and Connor, S. A., Evaluation of 20 Workload Measures Using a Psychomotor Task in a Moving-Base Aircraft Simulator. *Human Factors,* Vol. 25, No.1, pp. 1-16, 1983.

Table 1. Mean operator ratings of "mental effort" required for each subtask under low, medium, and high workload conditions

Load Condition	Subtask				
	CPA[a]	MDA[b]	TPA[c]	DCO[d]	PRA[e]
Low	1.67	1.55	1.78	1.67	
Medium	2.11	1.77	1.77	1.77	2.77
High	2.22	1.77	1.77	1.66	3.00

[a]Command Processor Assembly (two processors were initialized).
[b]Metric Data Assembly.
[c]Telemetry Processor Assembly (two processors were initialized).
[d]Digitally controlled oscillator.
[e]Planetary Ranging Assembly.

Table 2. Mean operator ratings of "time stress" imposed by each subtask under low, medium, and high workload conditions

Load Condition	Subtask				
	CPA	MDA	TPA	DCO	PRA
Low	2.00	1.67	1.78	1.55	
Medium	2.22	1.55	1.66	1.77	2.88
High	2.22	1.77	1.66	1.88	3.77

Table 3. Mean workload ratio measure ($W = T_r/T_a$) for each subtask under low, medium, and high workload conditions

Load Condition	Subtask				
	CPA	MDA	TPA	DCO	PRA
Low	0.56	0.29	0.32	0.39	
Medium	0.59	0.19	0.24	0.56	0.41
High	0.65	0.27	0.31	0.59	0.60

Table 4. Mean initialization time in seconds for each subtask under low, medium, and high workload conditions

Load Condition	Subtask				
	CPA	MDA	TPA	DCO	PRA
Low	14.4	6.5	14.9	11.3	
Medium	17.4	6.5	16.4	7.5	14.3
High	17.4	11.8	14.9	13.3	15.7

Table 5. Number of errors and ratio of actual inputs (I_a) by operators to minimum number of required inputs (T_r) under each workload condition

Parameters	Load condition		
	Low	Medium	High
Errors	9	17	17
I_a/T_r	1.09	1.15	1.34

MENTAL EFFORT	TIME STRESS
The task was:	There was:
(1) Extremely easy to do.	(1) Plenty of time to spare.
(2) Very easy to do.	(2) Some time to spare.
(3) Easy to do.	(3) A little time to spare.
(4) Neither hard nor easy to do.	(4) Time allowed is just enough to get the job done.
(5) Hard to do.	(5) Hard to finish in the time allowed.
(6) Very hard to do.	(6) Very hard to finish in time allowed.
(7) Extremely hard to do.	(7) Impossible to finish in the time allowed.

MENTAL EFFORT WILL ALWAYS BE YOUR FIRST RATING AND TIME STRESS YOUR SECOND.

Fig. 1. Verbal labels for rating scales given to operators during the testing procedure

OPERATOR WORKLOAD QUESTIONNAIRE

Please indicate, in the space provided, your agreement or disagreement with the following statements on a scale from 1 to 7 where:

1 = very strongly disagree
2 = strongly disagree
3 = disagree
4 = neither agree nor disagree
5 = agree
6 = strongly agree
7 = very strongly agree

	Statement		Mean Operator Rating
1.	The workload required of the operators of this system is too high.	[ ]	2.78
2.	The tasks in this run were excessively complex.	[ ]	3.00
3.	I was familiar with all the commands I needed to use on this task.	[ ]	5.88
4.	It was easy to find the information I needed in the system documentation.	[ ]	5.00
5.	I had confidence in the actual equipment in the string that was made available for this task.	[ ]	4.77
6.	The lighting arrangements were appropriate for the tasks.	[ ]	4.77
7.	The heating and/or air conditioning were adequate.	[ ]	4.77
8.	The physical arrangement of the equipment was appropriate.	[ ]	4.0
9.	The system provides all the displays that are needed for the task.	[ ]	4.77
10.	Each display provides adequate information for its task.	[ ]	4.89
11.	Prompts make clear what action is to be taken.	[ ]	4.44
12.	Subsystem error messages describe each fault in sufficient explanatory detail.	[ ]	3.67
13.	Subsystem error messages make clear what action is to be taken.	[ ]	3.89
14.	There were other problems in operating the system (specify).	[ ]	

Fig. 2. Operator workload questionnaire sample with mean ratings of nine operators for each statement

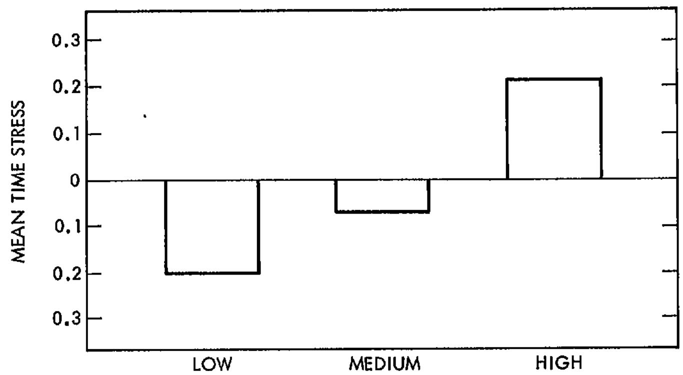

Fig. 3. Mean normalized operator "time stress" rating under low, medium, and high workload conditions

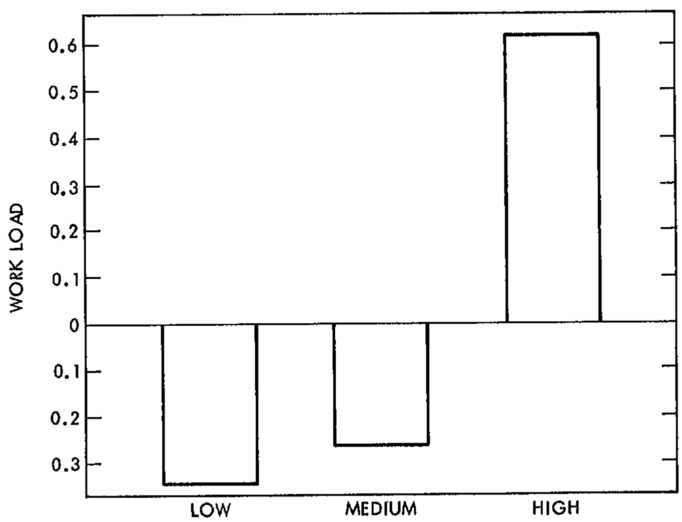

Fig. 4. Mean normalized workload ratio ($W = T_r/T_a$) under low, medium, and high workload conditions