A System Performance Throughput Model

Applicable to Advanced Manned Telescience Systems

Richard F. Haines
Research Scientist

/N-14-CR
43060
P23

January 1, 1990

Research Institute for Advanced Computer Science
Ames Research Center - NASA

RIACS Technical Report 90.1

RIACS

Research Institute for Advanced Computer Science
An Institute of the Universities Space Research Association

A System Performance Throughput Model

Applicable to Advanced Manned Telescience Systems

Richard F. Haines

Research Institute for Advanced Computer Science
Ames Research Center - NASA
Moffett Field, California 94035

ABSTRACT

As automated space systems become more and more complex, autonomous, and opaque to the flight crew it becomes increasingly difficult to determine whether the total system is performing as it should. This paper addresses some of the complex and interrelated human performance measurement issues that are related to total system validation. It presents an evaluative "throughput model" which can be used to generate a human operator-related benchmark or figure of merit for a given system which involves humans at the input and output ends as well as other automated "intelligent agents." The concept of sustained and accurate command/control data-information transfer is introduced. The first two input parameters of the model involve nominal and off-nominal "predicted"events. The first of these calls for a detailed task analysis while the second a contingency event assessment. The last two required input parameters involve actual (i.e., measured) events, namely human performance and continuous semi-automated system performance. An expression combining these four parameters was found using digital simulations and identical, representative, random data to yield the smallest variance. Manned simulations are underway to further evaluate this throughput model.

This work was supported in part by Cooperative Agreement NCC 2-387
from the National Aeronautics and Space Administration (NASA)
to the Universities Space Research Association

LIST OF ABBREVIATIONS

n	number of events
r	probability of event occurrence
CCDI	command control data/information
CEIP	contingency event impact parameter
CSPM	continuous system performance monitoring
Ec	contingency event
I	input (command)
Ir	impact rating
MTBF	mean time between failure
MTTR	mean time to repair
MTM	mean time to monitor
O	output (system response)
Pm	performance metric
Pr	event (probability)
Tp	throughput

This paper was presented at the HCI International '89 meeting, September 18-22, 1988 Boston, Massachusetts and a shorter version was published under the title "An Information Throughput Model for Complex, Transparent, Telescience Systems" in the official proceedings "*Designing and Using Human-Computer Interfaces and Knowledge Based Systems*, G. Salvendy and M.J. Smith (Eds.), vol. 12B, Pp. 354-360, Elsevier, New York, 1989.

INTRODUCTION

Telescience is the effective conduct of science through the use of remote resources including other people. In order to fully capitalize upon the many benefits which telescience offers (cf. Leiner, 1989) it will be necessary to prove that the theoretical advantages claimed are actually achieved. Indeed, it is one thing to design and build advanced computing and communication technologies and another to be able to show that the completed systems' throughput not only meets all specifications but actually contributes to productivity, flexibility, morale, lower costs, and safety. The present paper addresses this need for an approach to validate complex manned telescience systems.

As operational systems become larger, more complex, opaque and autonomous, it is likely that the operator(s) will be less and less able to play an effective role in monitoring and even controlling them, particularly when they malfunction. It will become increasingly important, then, to understand very early in the design process of a new telescience system what kinds of impacts the proposed system may have on user productivity, safety, and quality of total system performance. Advanced rapid prototyping approaches can be used to study these impacts. This paper presents an evaluative model with which to compare information throughput (Tp) of one candidate telescience system with another using both digital and manned simulation data. The model generates a human operator-related benchmark or figure of merit for a given system.

THROUGHPUT MODEL

The main objective here is to formulate a single number indicating a *performance throughput* benchmark or "figure of merit" for a given system. The present paper presents a preliminary equation that can be used to evaluate one system configuration with another. Work is progressing elsewhere on some of the present input parameters and associated modeling [eg. Anderson (1983), Card et al., (1983), Dreyfus and Dreyfus (1986), Newell and Card (1985), Rouse and Morris (1987)], however, no one is attempting to model the complete telescience system with users-in-the-loop.

Chin et al. (1987) and Gallagher (1974) have reviewed the literature dealing with the use of subjective evaluation measurement tools and have found weaknesses in many of them ranging from low reliabilities (Larcker and Lessig, 1980) to no validation (Gallagher, 1974). Clearly, use of objective human performance measurement methods is preferred. Indeed, a complete model must also provide useful insights into how and why operator errors are made (cf. Goldbeck and Ferrante, 1979; Rouse and Rouse, 1983).

THE CONCEPT OF THROUGHPUT

Throughput is defined as *the mean sustained rate of accurate "command/control-data/information" (CCDI) transfered from one place to another during a normal work period using telecommunications hardware.*

This definition begins with the recognition that all humans involved in a semi-automated and/or computerized system must be considered as integral components of the CCDI transfer process. Both the "sender" who types on a keyboard as well as the person who "receives", uses and/or relays the information represent important, potentially predictive parameters in the final Tp equation. In the case of a stand-alone workstation, the sender and receiver is the same person.

This definition also recognizes that successful task accomplishment requires accurate transfer of data of many different kinds. Command data may refer to digital autopilot output signals to aerodynamic surface actuators on an airplane, to an astronaut's manual input to control the remote manipulator system in the Space Shuttle vehicle's cargo bay, etc. Whatever its form, such data is considered to be equivalent to system-related information.

In highly complex, linked, compartmentalized systems, interface units translate one kind of information into another kind which the linked sub-system(s) needs to accomplish its task(s). In such situations throughput implies more than mere connectivity. Throughput incorporates all of the interrelated characteristics of complex systems that contribute to total system operability.

The idea of *sustained* CCDI transfer refers to the quantity or bit rate of information transfered per unit time within a normal work period. For Space Station Freedom operations, for example, if a duty period of a crew member is eight hours, and a given procedure or experiment is scheduled to last two hours, then the normal work period (here) will be two hours. This distinction is made because of the likelihood that different experiments will call for different work periods as well as different kinds of information. Indeed, it is not appropriate to compare a given performance measure obtained on an orbital astronomy experiment, for example, against use of the same measure obtained during an active on orbit robotic control maneuver which has fuel, physical impact, and other constraints.

The concept of *accurate* CCDI transfer refers to the quality of the information transfered. Algorithms have been developed with which one can sample various features of information input and output to compare them statistically. However, for more complex kinds of systems, e.g., audio/visual presentations given by a group of people, quantitative measures of quality are fewer and involve many interactions; that is, many more assumptions and qualifications are required. In addition, some semi-automated systems provide alternate paths to the same solution. Accurate CCDI transfer takes into account the adequacy of task accomplishment not whether a specific (or even a preferred) solution was followed.

AN OVERVIEW OF THROUGHPUT ANALYSIS

The approach to Tp analysis presented here involves four steps. The first two, deal with nominal and off-nominal *predicted* events while the second two deal with *actual*, measured events. The first two are: (A) Traditional human factors task analysis, and (B) Contingency events impact analysis.

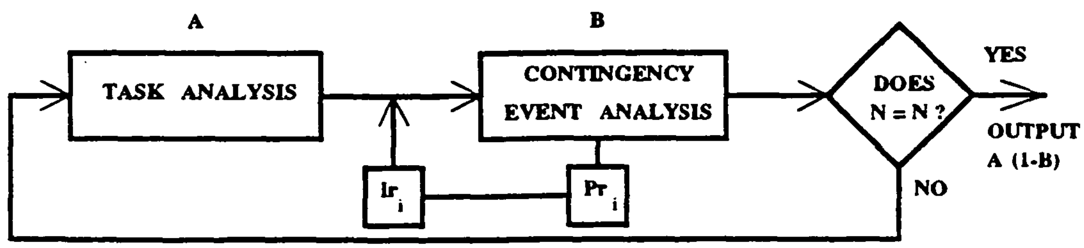
The task analysis is a well known method for systematically considering each and every task required in carrying out a process, operation or task. The results of such an analysis provide insights about how one or more sub-tasks contribute to sucessful system operation (or to system failure), how long the process or experiment will take, and many other useful insights. Contingency events impact assessment considers off-nominal, low probability of occurrence events that might not scrub the experiment or activity but could delay its completion, introduce data input errors, produce user frustration, increase workload, or other untoward effect(s).

In the second part of the Tp equation *actual* data obtained from simulations or actual flight are used. Two kinds of input data are required: (C) *a human performance* datum using one pre-specified performance metric (Pm) (i.e., human response) at a time, and (D) *continuous system(s) performance monitoring* (CSPM) data where the operator is not necessarily involved. An expression for combining all four steps is presented later.

Figure 1 presents a simple diagram to illustrate the (A) - (B) relationship which is in the form of an iterative loop used to calculate the system's predicted response.

Figure 1

Flow Diagram for Determining a System's Predicted Response



THROUGHPUT ANALYSIS: THE APPROACH

Now let us consider each of the four steps involved in a typical Tp analysis in greater detail.

Step 1. Task Analysis

A time-proven approach not only for assessing human performance associated with operation of a complex system but also for characterizing manned system operations is that of task analysis (cf., Morgan et al., 1963). What is usually done in a task analysis is that a trained system user, principal investigator, and/or system designer lists all of the behaviors that are involved in operating the system under all expected operational situations. This requires that a formal decomposition of the main process, operation or task into sub-units be performed; this generates a clearly defined, ordered sequence of user actions.

Task analysis is usually carried out at one of two different levels. A *macro task analysis* (assuming nominal environmental conditions) might be something like: "Turn on main power supply switch," "Turn on sub-systems power supplies for processes A, B, and C," "Visually verify that Closed Circuit TV cameras 1 and 2 are on, " "Boot computer up using internal operating system," "Run alignment calibration program AP-7," "Pre-set event timer to zero," "Manually adjust the TV image from each camera," "Initiate data collection," etc.

In contrast, a *micro task analysis* of the same operation might consist of the following sub-tasks:

A1. "Call Engineering Officer to verify that module power bus L44 has been pre-designated for this operation,"
A2. "Turn on main power supply switch,"
A3. "Visually verify that main power circuit breakers are in 'SET' position,"
A4. "Turn on sub-system power supply for process A, B, and C,"
B1. "Verify that rack internal temperature is within green zone,"
B2. "Verify that CCTV Camera 1 is on,"
B3. "Verify that CCTV Camera 2 is on,"
C1. "Turn on computer and verify that a valid command prompt is present on screen,"
C2. "Load program AP-7 - "Run Alignment Calibration Program,"
D1. "Manually pre-set rack AP event timer to zero,"
E1. "Inspect/adjust CCTV Camera 1 image for best contrast,"
E2. "Use CCTV Camera 1 joy stick control to center field of view upon black cross-hairs inside animal-holding facility,"
E3. "Repeat steps E1 - E2 for CCTV Camera 2,"
F1. "Start specimen data collection as per experimental protocol."
Etc.

Clearly, a micro task analysis can involve more user reading time but less memorized information concerning each step.

Each sub-task is listed in temporal order so that various estimated or measured performance metrics, determined through pre-flight simulation, may be associated with each one. A Pm is a specific behavioral or performance measure that can be quantified and used in one of a series of iterative Tp calculations, as will be discussed. The above listing of sub-tasks has been expanded further to illustrate the approach. The resulting micro task analysis table might look like that given in Table 1.

Table 1

Hypothetical Example of Space Station Freedom
Micro Task Analysis

Task	*Performance*		*Metrics*		
	Accom. plished? (yes, no)	**Time Reqd? (sec)**	**Elapsed time (min/sec)**	**Out of order (note 1.)**	**Refer to Manual (note 2.)**
	A	B	C	D	E
A1. Call Engineering Officer to verify power module bus L44 is pre-designated	Y	20	0' 20"	0	0
A2. Turn main power supply switch ON	Y	5	0'25"	0	0
A3. Visually verify main power circuit breakers are in 'SET' position	Y	15	0'40"	0	0
A4. Turn on sub-system power supply for process A, B, and C	Y	20	1' 00'	0	0
B1. Verify rack internal temp. is within green zone	Y	10	1'10"	0	0
B2. Verify CCTV camera 1 is ON	Y	5	1'15"	0	0
B3. Verify CCTV camera 2 is ON	Y	5	1'20"	0	0
C1. Turn on computer and verify valid command prompt is visible on screen	Y	8	1'28"	0	0
C2. Load Program AP-7 "Run Alignment Calibration Program"	Y	20	1' 48"	0	0
D1. Manually Pre-set rack AP event timer to zero	Y	4	1'52"	0	0
E1. Inspect/adjust CCTV Camera 1 image for best contrast (Note 3)	N	-	1'52"	0	0
E2. Use CCTV Camera 1 joy stick, center FOV on black cross-hairs (animal-holding facility)	Y	12	2'04"	0	0

E3. Inspect/adjust CCTV Camera 2 image for best contrast	Y	10	2'14"	1	0
E4. Use CCTV Camera 2 joy stick, center FOV on black cross-hairs (animal-holding facility)	Y	12	2'26"	0	0
F1. Start Specimen data collection (per exp. protocol)	Y	1	2'27"	0	0
G1. Specimen 1. Place on sled and obtain mass by depressing labelled keys on keypad	Y	25	2'52"	0	0
G2. Spec. 1. Visually verify normal leg reflexes using procedures in experiment protocol	Y	30	3 '22"	0	0
G3. Spec. 1. Apply stimulating electrodes to right rear gastrocnemius muscle and ground electrode to shaved spinal area on midline as per manual	Y	2'45"	6'07"	0	0
G4. Spec. 1. Activate stimulation Program AP-8	Y	22	6'29"	0	0
G5. Spec. 1. (At end of pre-programmed stimulation sequence depress computer "tag" switch	Y	14	6'43"	0	0
G6. Spec. 1. Remove electrodes and stow in pocket	Y	1'30"	8'13"	0	0
G7. Spec. 1. Wipe animal's leg with cleansing solution, dry with dry cotton cloth	Y	30	8'43"	0	0
G8. Spec. 1. Visually verify leg reflexes using procedures in experimental protocol	Y	30	9'13"	0	0
G9. Spec. 1. Replace animal in specified holding compartment	Y	16	9'29"	0	0
H1. Specimen 2. (Repeat Steps G1-G9)	Y	6' 37"	23'06"	0	0
I1. Activate communications data link with on-board computer using link program AP-44	Y	25	23'31"	1	0
I2. Cross-check mission elapsed time on master clock with manual entry value just prior to data dump	Y	12	23'43"	0	0

I3. Verify that data printer is ON	Y	6	23'49"	1	0
I4. Verify that CCTV Camera 1 digitizer is ON	Y	6	23'55"	0	0
I5. Verify that CCTV Camera 2 digitizer is ON	Y	10	24'05"	0	0
I6. Depress "DATA DUMP" button (Note that green light comes ON. If not, repeat steps I1 - I2)	Y	10	24'15"	1	0
J7. Deactivate data communications data link with on-board computer with BIT setting O11OO1	Y	10	24'25"	0	75
K1. Call Communications Officer to verify that Earth POIC link has been established and is ready for down-link data dump	Y	50	25'15"	1	0
L1. Initiate down-link data dump (bus 21M-7)	Y	12	25'27"	0	0
L2. (At confirmed conclusion of successful data dump) power down CCTV camera 1	Y	8	25'35"	0	0
L3. Power down CCTV Camera 2	Y	8	25'43"	0	0
L4. Power down Rack AP computer	Y	10	25'53'	0	0
L5. Turn Main Power Switch OFF	Y	6	25'59"	0	0
END OF PROCEDURE Summary	37Y; 1N		25 min 59 sec	5	75

Notes: 1. Column D contains hypothetical data concerning the number of times an event (sub-task) was performed out of the correct order.
2. Column E contains hypothetical data concerning the total time (sec) spent referring to a written protocol manual.
3. Step E1 was skipped on purpose. It was assumed that the user knew that the contrast setting had not changed since last use.

The above sub-task listing plays a number of important roles. First, at the global level, it shows quickly and unmistakably whether all critical tasks have been completed. Second, the Time Required column (B) provides quantitative values useful for integrating this particular experiment into the master, crew-event mission time line. Third, the total elapsed time (column C) is useful in the same way. Columns A through E represent sample performance metrics (see discussion under step 3 below).

Step 2. Contingency Events Impact Assessment

The second major contributor to the Tp equation is called a "*Contingency Event Impact Parameter*" (CEIP). It is represented by box B in Figure 1. The present Tp determination calculations provides quantitative estimates of most (or all) low probability events that could adversely influence Tp. As mentioned previously, this step yields useful insights about subtle and unplanned factors that can delay or otherwise influence the successful accomplishment of the procedure, introduce errors into the data stream, lead to performance omissions or additions, distract the operator from concentrating hard enough on the task(s) at hand, etc. Four steps are involved in determining CEIP:

(*1*) List all probable (contingency) events (Ec). These are events that could occur during the activity which would affect system Tp adversely.

(2) Assign a probability value (Pr) to each contingency event that it will occur within the normal work period.

(*3*) Assign an impact rating value (Ir) to each event such that, if it did occur, sub-system or full system performance would be reduced by a predetermined, arbitrary amount (r).

Ir can range from 1 to 5 where: 1 = no impact of any kind, 2 = low negative impact with no lasting or adverse consequences on completion of task, 3 = moderate negative impact in terms of delays in accomplishing the task, 4 = high negative impact bordering on aborting the planned operation, and 5 = very high negative impact causing the task to be aborted.

(*4*) Set the parameter "r" at a constant value, e.g., 0.01. Under most circumstances it is best to use a constant value of r for a given application of this Tp equation.

CEIP is expressed in terms of the mean Pr x Ir over a typical event series "i-n", where i is the "i-th" input value and n is the total number of events.

$$\text{CEIP} = \text{Sum}_{i\text{-}n}(\text{Pr} \times \text{Ir}) / n \qquad (1)$$

When all contingency events are included in equation one, CEIP represents an estimate of the integrated impact they will have on Tp. A simple illustration is in order. Consider the micro task analysis given in Table 1. There are many different possible contingency events that could occur which would impact the ultimate success of the operation. To illustrate the procedure, Table 2 presents these hypothetical events and estimates for Pr and Ir for each event. The value r is assumed to be 0.01 for this illustration.

Table 2

Illustration of Hypothetical
Contingency Event Impact Assessment

Hypothetical Contingency Events	Pr	Ir	(Pr x Ir)
1. Intercom will malfunction during step A1	.001	4	.004
2. Power Module Power bus L44 has erroneously been designated to another experiment.	.0005	4	.002
3. Engineering Officer is off-duty (unavailable).	.008	2	.016
4. Main Power switch fails	.0002	5	.001
5. Main Power circuit breakers malfunction	.0002	4	.0008
6. Rack thermometer malfunctions (needs replacing)	.0013	2	.0026
7. Visual obstruction to seeing thermometer	.05	1	.05
8. Ambient illumination is so low green color on gauge cannot be discriminated readily	.007	2	.014
9. CCTV Camera 1 'LED' "ON" light burned out	.0002	3	.0006
10. CCTV Camera 1 monitor malfunctions (requires replacement (Interacts with item 22)	.0004	5	.002
11. CCTV Camera 1 has been removed for another use	.0055	3	.0165
12. CCTV Camera 2 'LED' "ON" light burned out.	.0002	2	.0004
13. CCTV Camera 2 monitor malfunctions (requires replacement)	.0004	5	.002
14. CCTV Camera 2 has been removed for another use.	.005	3	.015
15. Computer program AP-7 will not load properly. due to disc-read (input) error	.008	4	.032
16. Computer program AP-7 will not load properly. due to tape drive malfunction	.004	4	.016
17. Rack AP computer will not boot properly due	.005	5	.025

to insufficient power			
18. Rack AP computer will not boot properly due to improper operating sequence	.008	4	.032
19. Rack AP computer will not boot properly due to electronic component failure	.0002	4	.0008
20. Rack AP event timer malfunctions (requires replacement)	.001	2	.002
21. Operator forgets to reset event timer	.01	2	.02
22. CCTV Camera 1 electronics malfunction/degrade so that contrast function doesn't work (interacts with item 10)	.001	3	.003
23. CCTV Camera 1 monitor electronics malfunction	.001	4	.004
24. Operator forgets to check CCTV Camera 1 contrast	.08	1	.08
25. CCTV Camera 1 joy stick malfunctions	.005	3	.015
26. CCTV Camera 2 electronics malfunction/degrade so that contrast function doesn't work	.001	3	.003
27. CCTV Camera 2 monitor electronics malfunction so that contrast cannot be seen or varied	.001	3	.003
28. Operator forgets to check CCTV Camera 2 contrast	.08	5	.4
29. CCTV Camera 1 joy stick malfunctions	.005	5	.025
30. Specimen inside animal-holding facility blocks camera view of cross-hair target	.007	6	.042
31. CCTV Camera 2 joy stick malfunctions	.005	5	.025
32. Specimen inside animal-holding facility block camera view of cross-hair target	.007	6	.042
33. Specimen processing procedure is delayed for "n" minutes for any reason	.005	2	.01
34. POIC personnel require unplanned variation in experimental protocol	.05	2	.1

35. POIC to Space Station real-time communications link is broken during data collection	.007	5	.035
36. Space Station Total Power level varies due to unanticipated solar panel occlusion longer than on-board storage cells will permit	.001	5	.005
37. Ambient air pressure change will occur and force delay of experiment until corrected	.0002	4	.0008
38. Solar flare will occur and cause malfunction of microcomputer or associated hardware	.0004	5	.002
39. Specimen required for experiment will be damaged during pre-experiment set-up	.002	5	.01
40. POIC to Space Station communications link will be rerouted causing a time delay of "x" sec. in each direction	.003	4	.012
41. POIC to Space Station communications link will introduce data drop outs on random basis with mean frequency of X bits/video frame	.008	4	.032
Summary: Pr (maximum = 0.01; minimum = 0.0002)	CEIP Sum = 2.6205 CEIP = 0.0639		

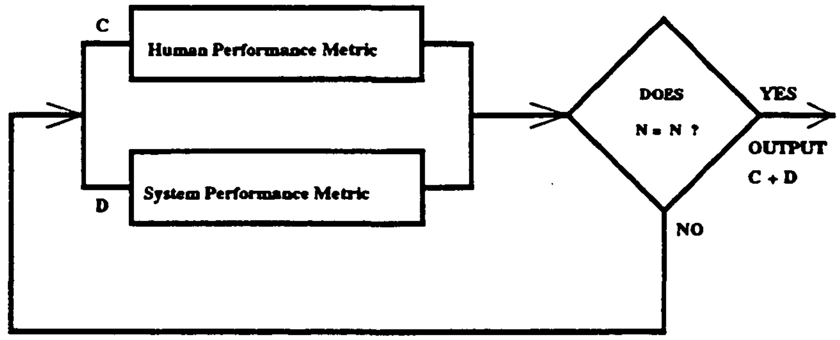
Note: Hypothetical probability estimates for the events listed above have been included only for purposes of illustrating the basic technique.

Step 3. Continuous Human Performance Monitoring

The third parameter contributing to the Tp figure of merit involves *continuous monitoring of actual crew performance* at the output end of the system (or subsystem) in question. It is represented by box C in Figure 2 which is a flow diagram for calculating the two actual performance parameters of this model. These data are used in several ways. One is to weight selected parameters in the Tp figure of merit equation. Use of unobtrusive (covert) performance monitoring equipment (e.g., hidden TV camera, recorder and voice tape) provides such quantitative data (Haines et al., 1989). In this regard, data based on interviews and questionnaires have been used to obtain new insights about how actual computer system users actively learn to solve problems that they have defined for themselves. For instance, Carroll and Mack (1985) reported that the design characteristics of the M-Mach interface is influenced by user motivation. A second use for such data is to identify subtle behavioral response patterns possibly indicative of procedural misunderstandings. For inanimate systems (e.g., telerobot) that are under manual control, TV based monitoring is extremely useful if not essential. A companion paper to the present one discusses a variety of human performance measurement and validation procedures applicable to manned telescience system.

Figure 2

Flow Diagram for Determining a System's
Actual Performance



Covert crew performance monitoring tends to yield more reliable data than when the crew know they are being monitored. Subtle glance exchanges of participants in a meeting as well as their non-verbal emphases implied by heavy handed mouse usage, for example, can be captured and analyzed by videotape techniques (Mantel, 1988). One experimental approach is to aim a concealed television camera, connected to a video recorder (with elapsed time capability), at the person(s) carrying out the task. There are many candidate human performance metrics that can thus be quantified. Table 3 lists some:

Table 3

Candidate Human Performance Metrics
Related to Covert TV Performance Monitoring

A. Task Accomplishment Metrics:
   1. Was each micro- or macro-task initiated at all?
   2. Was each micro- or macro-task completed?
   3. Number of times person had to abort entire procedure and start over
   4. Specific event/task at which person aborted the task
   5. Number of times person repeated an individual action (cf. Greenberg & Witten, 1988)

B. Temporal Metrics:
   1. Time required to carry out a given step or sub-task
   2. Time spent dealing with (i.e., correcting) each contingency event
   3. Time spent dealing with all contingency events combined
   4. Time between a given step or sub-task
   5. Total time spent referring to a written protocol (rather than performing the task from memory)

C. Response Accuracy Metrics:

1. Number of times a task was performed out of the correct order in a series
2. Number of events that were executed incorrectly for any reason
3. Number of errors made per task (cf., Haines et al., 1989)
4. Number of errors made per unit time (e.g., specified work period)
5. Number of times a tool or part was dropped by accident
6. Number of times person misidentified a contingency event

D. Response Movement Metrics:

1. Number of discrete hand/finger motions required to complete a task
2. Number of times the head or body moved up/down during a task in order to see all support hardware and related information
3. Total range of hand/finger motion required in preselected planes during task accomplishment
4. Number of times person activated remote-moveable manual controls located in one location versus another
5. Number of times person used two hands together to accomplish a task

E. Miscellaneous Metrics:

1. Number of verbalizations made to carry out a task
2. Number of questions asked to carry out a task
3. Number of user-induced contingency events which occurred during the task
4. Specific behavior that preceded an incorrect task operation
5. Number of times user discovered and used a new task/procedure (i.e., previously unused) that improved his productivity, workload, etc.
6. Number of eye contacts to another person nearby
7. Facial expressions of surprise, anger, determination, etc.

Once the choice of a Pm is made, the selected covert or overt, continuous monitoring procedure is applied to each event in the Micro Task list of Table 1. An example is in order.

Let us assume that the following Pm are selected from Table 3: (A2) *was task completed?* (B1) *time required to carry out the step or sub-task),* (C1) *number of times a task was performed out of the correct order in a series,* and (B5) *total time spent referring to a written experimental protocol rather than being performed from memory.* Hypothetical Pm values for these specific items are given in Table 1 to illustrate this approach. The basic approach is to perform separate iterative Tp calculations for each Pm selected.

Step 4. Continuous System Performance Monitoring (CSPM)

The fourth and final parameter contributing to the Tp figure of merit involves *continuous system performance monitoring*. This step is represented by box D in Figure 2. This type of analysis has been conducted in many forms over the years. As employed traditionally, this step often emphasizes an end-to-end, binary (pass-fail) system evaluation. A measure of the time over which a user may expect the system to operate before an incapacitating failure occurs is referred to as the mean time between failure (MTBF), a gross measure of continuous system performance indeed. In normal practice, hardware units from a production line are selected at random and operated until they fail (for any reason). A statistical description of the probability of unit failure is then cited as the MTBF, typically per "N" hundred thousand hours of operation. Analogous concepts include mean time to repair (MTTR) and mean time to monitor key functions (MTM). Clearly, such approaches must be modified for application to complex, user(s)-in-the-loop systems.

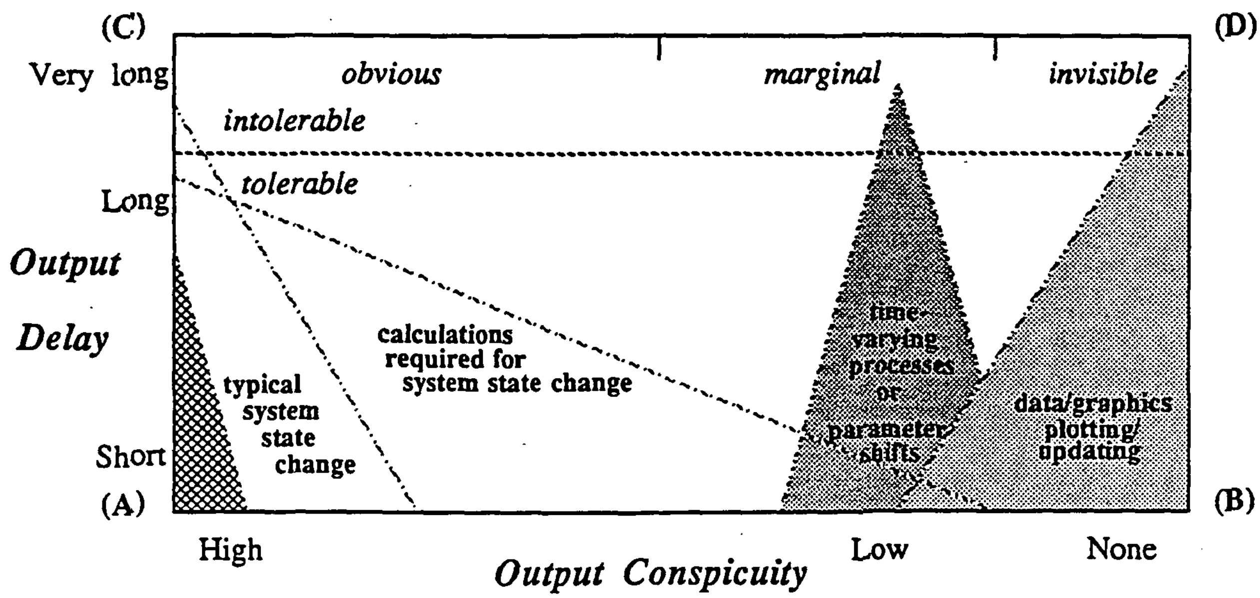
The general concept of CSPM involves recording specific nominal input (I) (command) information and also the output (O) (responses) which result from it. What is calculated is the input/output ratio [per unit time (ti)]. Thus:

$$\text{CSPM} = I/O_{ti} \qquad (2)$$

Complex system output can take many forms which can be diagrammed along various axes. Figure 3 illustrates a variety of kinds of output arrays presented along the conspicuity and delay axes. Conspicuity is defined as the degree to which the user can directly perceive the current status of the information being output in processed or unprocessed form. A computer screen's clock icon, indicating current percentage of available computational power *assigned directly to the operation in question*, represents a situation possessing high output conspicuity. Likewise, a screen message stating "data being calculated" would be an equally

Figure 3

Diagram of Hypothetical Output Conspicuity and Delay Relationships



high conspicuity output. In contrast, a blank or static screen represents no output conspicuity.

It can be seen that for, most manned systems, the preferred system output usually clusters near corner (A), being both highly conspicuous to the operator and possessing short delays. System performance characterized by the diagonally opposite corner (D) makes it very difficult to diagnose a system failure. Likewise, the (A) - (B) axis becomes more important when humans must monitor system output. Generally, very long delays can be tolerated better than can low or no output conspicuity. For every system there is a delay that becomes intolerable. It is usually defined as when the operator can no longer perform the required task. Less frequently, it is when the semi-automated system inhibits its own normal operations due to intolerably long delays. An important issue is whether or not the CSPM calculation should include processes that involve intolerably long output delays. It is recommended that each Tp determination include a clear definition of what constitutes an unacceptably long output delay. All processes taking longer than this value either should be corrected or excluded from the calculation.

THROUGHPUT CALCULATION

Most previous task analytic performance models consider only one parameter, e.g., the macro or micro tasks involved. Some success has been claimed for them in predicting training time, productivity of novice and skilled users, and transfer for text editing applications (Polson, 1987). In distinction to these earlier models, the present Tp figure of merit includes three other parameters which supports wider extrapolation to multinodal telecommunication situations in which low probability contingencies also may be anticipated.

It remains to combine the values derived from steps 1 through 4 above in the most useful manner. The following (preliminary) expression is presented:

$$\mathbf{Tp} = \mathbf{A\,(1\text{-}B)\,/\,(C\text{+}D)} \qquad (3)$$

Equation 3 yielded the smallest distribution standard deviation of various equations that were evaluated (cf. appendix). Using this expression, Tp is a dimensionless number that ranges from -1 to +1.

As discussed above, terms A and B in equation 3 represent optimal, predicted performance of the system. Term B may be considered as a factor which, every time it occurs, negatively impacts system performance; ideally B should be a small number. Using 1 - B reduces the magnitude of the impact B has on A since, for example, if B = zero then A x 1 = A. And if B = .02 then A x (1 - .02) = 0.98A, etc. Terms C and D represent actual, measured performance of the same system under the same operational conditions and performance metric(s). They are simply summed since either C and/or D could occur.

Digital simulations were run using randomly selected values for parameters A through D. These results are presented in Table 4. The following qualifications were placed upon the generation of the random numbers for use in the candiate equations: r = 0.01; 15% of the Pr values were between 0.0002 and 0.01, 85% of the Pr values were between 0.0002 and 0.001; 25% of the Ir values were between 2 and 5 and 75% were between 2 and 4. As noted, equation 4 produced the smallest standard deviation and smoothest overall distribution of Tp values.

Table 4

Statistical Parameters for Six
Candidate Throughput Equations

Calculational Approach	N-NN*	Mean	Median	S.D.
Individual (predicted/actual) cycles grouped and then summed	N	-2.024×10^{6}	-1.515×10^{6}	1.309
Individual (predicted/actual) cycles grouped and then summed	NN	0.1623	0.1613	0.0169
Predicted and actual cycles summed and then inserted in equation	N	-1.829×10^{2}	-1.800×10^{2}	0.177
Predicted and actual cycles summed and then inserted in equation	NN	5.477×10^{3}	5.521×10^{3}	0.514
Each parameter summed and then inserted in equation	N	1.942×10^{-3}	1.9075×10^{-3}	0.188
Each parameter summed and then inserted in equation	NN	5.168×10^{-3}	5.214×10^{-3}	0.481

* N = normalized data; NN = non-normalized data;

SUMMARY

A computationally simple expression has been offered with which investigators can compare the theoretical performance of one manned telescience system with another given different input parameters. Early Tp figures of merit based upon digital simulation input data can be refined by using subsequent manned simulation data. This iterative approach is presently being validated further using simulation data similar to that reported elsewhere (Haines et al., 1989).

REFERENCES

Anderson, J.R., *The Architecture of Cognition*. Cambridge, Harvard Univ Press, 1983.

Card, S.K., T.P. Moran, and A. Newell, *The Psychology of Human-Computer Interaction*. Hillsdale, New Jersey, Erlbaum, 1983.

Carroll, J.M., and R.L. Mack, Metaphor, computing systems, and active learning. *International J. Man-Machine Studies*, vol. 22, Pp. 39-57, 1985.

Chin, J.P., K.L. Norman, and B. Shneiderman, Subjective user evaluation of CF PASCAL programming tools. Technical Report (CAR-TR-304), Human-Computer Interaction Laboratory, University of Maryland, College Park, MD, 1987.

Dreyfus, H.L., and S.E. Dreyfus, *Mind over Machine*. New York, The Free Press, 1986.

Gallagher, C.A., Perceptions of the value of a management information system. *Academy of Management Journal*, vol. 17, no. 1, Pp. 46-55, 1974.

Goldbeck, R.A., and F.M. Ferrante, An information processing model for predicting error detection and correction: An analysis of operator performance. Ford Aerospace & Communications Corp., Final Report HC A04/MF A01, June 1979 - May 1981.

Greenberg, S., and I.H. Witten, How users repeat their actions on computers: Principles for design of history mechanisms. In Soloway, E., D. Frye, and S.B. Sheppard (Eds.), CHI'88 Conference Proceedings, Human Factors in Computing Systems, Omni Shoreham Hotel, Washington, D.C., Pp. 171-178, May 15-19, 1988.

Haines, R.F., V. Johnson, K.H. Vogelsong, and W. Froloff, Ames Life Science Telescience Testbed Evaluation. Research Institute for Advanced Computer Science, Technical Report 89.31, July 1989.

Larcker, D.F., and V.P. Lessig, Perceived usefulness of information: A psychometric examination. *Decision Science*, vol. 11, no. 1, Pp. 121-134, 1980.

Leiner, B.M., Telescience Testbed Pilot Program Final Report: Volume I, Executive Summary. Research Institute for Advanced Computer Science, Technical Report 89.7, February 1989.

Mantel, M.M., Video: Data for studying human-computer interaction. In Soloway, E., D. Frye, and S.B. Sheppard (Eds.), CHI'88 Conference Proceedings, Human Factors in Computing Systems, Omni Shoreham Hotel, Washington, D.C., Pg. 135, May 15-19, 1988.

Morgan, C.T., J.S. Cook, A. Chapanis, and M.W. Lund, *Human Engineering Guide to Equipment Design*. New York, McGraw-Hill Book Co., 1963.

Newell, A., and S.K. Card, The prospects for psychological science in human-computer interaction. *Human-Computer Interaction*, vol. 1., Pp. 209-242, 1985.
Polson, P.G., A quantitative theory of human-computer interaction. In J.M. Carroll (Ed.), *Interfacing Thought: Cognitive Aspects of Human-Computer Interaction*. Cambridge, Mass., MIT Press, 1987.

Rouse, W.B., and N.M. Morris, Conceptual design of a human error tolerant interface for complex engineering systems. *Automatica*, vol. 23, Pp. 231-235, 1987.

Rouse, W.B., and S.H. Rouse, Analysis and classification of human error. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-13, Pp. 539-549, July-Aug. 1983.

APPENDIX

CALCULATIONAL FORMULAE EVALUATED

A number of different calculational approaches employing equation 3 were examined using digital simulation. They are presented here. Randomly selected values were used for parameters A, B, C, and D (with the qualifications given above). The identical set of random numbers were inserted into each of the following equations to generate the data of Table 4. It may be noted that equations (4 and 5) are non normalized.

$$Tp = \frac{[A1(1-B1)] + [A2\ (1-B2)] + \ldots [An\ (1-Bn)]}{(C1 + D1) + (C2 + D2) + \ldots (Cn + Dn)} \qquad (4)$$

$$Tp = \frac{(A1 + A2 + \ldots An)\ \ 1-(B1 + B2 + \ldots Bn)}{(C1 + C2 + \ldots Cn) + (D1 + D2 + \ldots Dn)} \qquad (5)$$

In the following three equations a term has been added to normalize the resulting frequency distribution. Equation (6) is the normalized version of equation (3).

$$Tp = \frac{[A1\ (1-B1) / (C1 + D1)] + \ldots [An\ (1-Bn) / (Cn+ Dn)]}{[A1\ (1-B1)] + [A2\ (1-B2)] + \ldots [An\ (1-Bn)]} \qquad (6)$$

Equation (7) presents the normalized version of equation (4).

$$Tp = \frac{\{[A1(1-B1)] + [A2(1-B2)] + \ldots [An(1-Bn)]\} / (C1+D1) + (C2+D2) + \ldots (Cn+Dn)}{[A1(1-B1)] + [A2\,(1-B2)] + \ldots [An\,(1-Bn)]} \qquad (7)$$

Equation (8) is the normalized version of equation (5).

$$Tp = \frac{[(A1 + A2 + \ldots An)\ \ 1 - (B1 + B2 + \ldots Bn)] / [C1 + C2 + \ldots Cn) + (D1 + D2 + \ldots Dn)]}{[(A1 + A2 + \ldots An)\ \ 1 - (B1 + B2 + \ldots Bn)]} \qquad (8)$$

RIACS
Mail Stop 230-5
NASA Ames Research Center
Moffett Field, CA 94035
(415) 694-6363

The Research Institute for Advanced Computer Science
is operated by
Universities Space Research Association
The American City Building
Suite 311
Columbia, MD 21044
(301) 730-2656