RIACS

Research Institute for Advanced Computer Science
NASA Ames Research Center

IN-53
43053
P.35

Human Performance Measurement-Validation Procedures Applicable to Advanced Manned Telescience Systems

Richard F. Haines
Research Scientist

February 12, 1990

Research Institute for Advanced Computer Science
NASA - Ames Research Center

RIACS Technical Report 90.10

(NASA-CR-185447) HUMAN PERFORMANCE MEASUREMENT: VALIDATION PROCEDURES APPLICABLE TO ADVANCED MANNED TELESCIENCE SYSTEMS (Research Inst. for Advanced Computer Science) 35 p CSCL 05I G3/53

N92-10282

Unclas
0043053

Human Performance Measurement-Validation Procedures 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 telescience systems become more and more complex, autonomous, and opaque to their operators 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. The assumption is made that human interaction with the automated system will be required well into the Space Station Freedom era. This paper discusses candidate human performance measurement-validation techniques for selected ground-to-space-to-ground and space-to-space situations. Most of these measures may be used in conjunction with an information throughput model presented elsewhere (Haines, 1990). Teleoperations, teleanalysis, teleplanning, teledesign, and teledocumentation are considered as are selected illustrative examples of space-related telescience activities.

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 (USRA)

RIACS TR 90.10

LIST OF ABBREVIATIONS

di	device (input)
do	device (output)
pi	people (input)
po	people (output)
pu	processing unit
CAI	computer assisted instruction
CCTV	closed circuit television
MRMS	mobile remote manipulator system
MTBF	mean time between failure
MTM	mean time to monitor
NTSC	National Television Systems Committee
OMV	orbital maneuvering vehicle
ORU	orbital replacement unit
Pm	performance metric
PI	principal investigator
POIC	payload operations integration center
SOC	science operations center
SIRTF	space infrared telescope facility
Td	time to task accomplishment using degraded video
Tt	time to task accomplishment using normal video
Tp	throughput
UKIRT	United Kingdom infrared telescope

INTRODUCTION

Telescience is the effective conduct of science through the use of remote resources including other people. There are at least three generally recognized aspects of telescience: teleoperations, teledesign, and teleanalysis. *Teleoperations* can take many forms; a space robot that performs useful functions while being controlled from the ground or another spacecraft is an example. *Teledesign* refers to the effective combination of remotely located design tools and designers to develop something useful. An example might be a graphics plotter connected to a remotely located computer (containing an appropriate database) which is programmed and/or controlled by three remotely located architects to plan the layout for a new building. *Teleanalysis* refers to the capability to perform data integration and analysis remotely. An example might be that of a multidisciplinary group of "environmental" planners who need to develop a master plan for a huge wilderness area. A geologist may work on Landsat imagery data while integrating past flood coverage and minerological data. An urban, land-use planner may, at the same time, work on large scale, surface vehicle traffic-flow data while integrating projected water supply data. A transportation specialist may work on past and present air transport density plots for the area under study as well as for adjacent regions. Having a well designed teleanalysis capability means that all of these persons (and others) can share their data, edit and graphically modify them, and jointly produce useful designs and plans.

In order to capitalize fully 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 communications 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 one important aspect of this need for an approach to validate complex systems, namely *human performance* measurement and validation procedures.

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. I have developed an evaluative model which can be used to compare information throughput (Tp) of one candidate telescience system with another using both digital and manned simulation data (1990). The model generates a benchmark or figure of merit for a given manned system. One of the required input parameters for this model is a human performance metric (Pm). This paper presents various operator performance criteria, evaluative procedures, and related discussion that can be used to measure and validate human performance involved in rapid prototyping of telescience systems.

HUMAN PERFORMANCE VALIDATION PROCEDURES

How can complex telescience systems be evaluated from a human factors standpoint? What methods are available to study the effectiveness of a specific human-system interface? Despite a voluminous literature on human-computer interaction in general (Helander, 1988), relatively little has been written to date on the subject of how humans interact with their databases and with other humans *remotely* using telescience systems. There are many challenging procedural, training, hardware, and software design issues related to telescience.

Of primary interest here are methods and hardware which can provide practical understandings about how humans interact remotely with intelligent systems which have varying degrees of autonomy. Also discussed are different types of telecommunication links (audio, video, audio-visual, electronic data) and their relationship to human performance measurement. This discussion is presented in terms of five operational situations. These situations encompass the majority of future manned and unmanned space operations where telescience will find an immediate application. Table 1 lists them. Each should be considered as *two-way* tele-communications.

Table 1

Basic Operational Situations Relevant to
Human Factors Validation

Situation	Participants
A.	Person(s) (earth) to/from Person(s) (space)
B.	Person(s) (earth) to/from Machine(s) (space)
C.	Person(s) (earth) to/from Person(s) and Machine(s) (space)

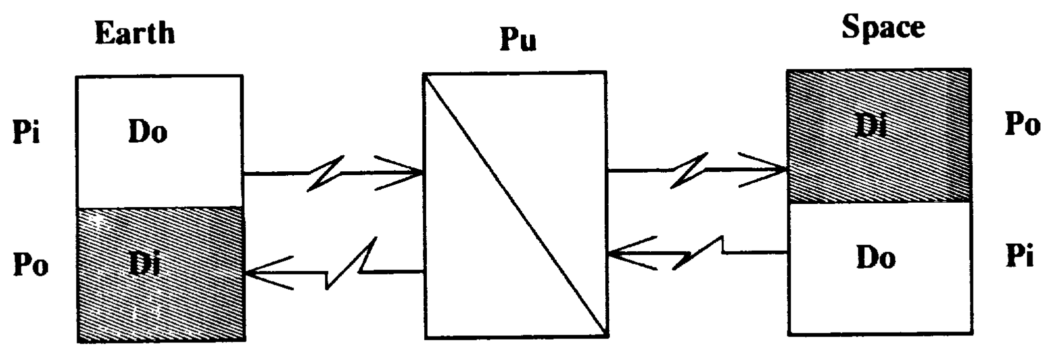
D Machine(s) (earth) to/from Person(s) and Machine(s) (space)

E. Machine(s) (space) to/from Machine(s) (space)

The simplest telecommunication information system may be characterized by one or more people (pi) and devices (di) at the input end, processing units (pu), and people (po) and devices (do) at the output end. Each of these components possesses its own inherent delays, bandwidths, and other operating characteristics. Figure 1 presents an element diagram of such a system.

Figure 1

Element Diagram of a Simple Earth-Space Telecommunication System



A traditional way of establishing the overall performance of the above three hardware elements (di, pu, do) is to measure how long it takes them to carry out "n" iterative calculations under rigidly specified conditions. Indeed, such benchmarks for computers (cf., Beeler, 1984; Brice, 1983; Emrick, 1983; Levy & Clark, 1982) (e.g., Baskett, Dhrystones, LINPACK, Livermore Loops, Whetstones, MIPS, FLOPS) and other hardware (Mello-Grand, 1984) support valuable inter-system comparisons. The ultimate usefulness of any benchmark rests upon the assumption that a high correlation exists between system performance on the benchmark(s) and performance on the everyday mixture of codes. Nevertheless, there are no such benchmarks available to evaluate total system performance for situations with the *human in the loop*.

One candidate approach for validating total system performance would be to calibrate pi and po and add these values to the hardware's benchmark value. While this approach would help control for the influence of individual differences among the users (cf. CHI'88 Panel, 1988), it would not (necessarily) cope effectively with hardware that is becoming more "intelligent" in its capability to compensate for human errors of omission and comission. As pu(s) become increasingly able to perform "smart" functions, the total system output metric would be biased, making the users appear to be performing better than they really are. But there is another general approach.

In this second approach the capability of the pu to compensate for pi and po information processing errors must be pre-determined. Then the ability of both pi and po to compensate for the limitations of the processing unit is determined under controlled conditions. Finally, the results of these two steps is integrated into a formula that results in one figure of merit for the total system. After the approach has been implemented and verified it would be possible to compare the total Tp performance of one system with another, with human operators present in the loop.

Another approach to validating total system performance is to measure common aspects of input and output and report the differences. This is the easiest and most common approach taken today. A current, if somewhat complex variation on this general theme is that of Barnard et al. (1987). They suggest an architecture for human information processing where there is no need for a central executive capability or working memory since the entire system is self-controlling by means of representations passed from one subsystem to another. This is accomplished by providing means for tagging unified activities at each stage of operation, from input to output. For such a system to generate and control overt actions accurately, each individual activity must act together in a coordinated way. Thus, the dynamic control of Tp requires characterizing the passage of these representations (tags) among the various subsystems. This approach is based on the assumption not only that the input was correct but also was what was intended.

Theoretical models of Rouse and Morris (1987) dealing with a proposed architecture for intelligent interfaces and of Barnard (1987) Barnard et al. (1987) should be consulted since they provide useful basic frameworks for developing validation schemes for complex systems.

The system performance throughput model which I developed (1990) involves four initial steps. The first two deal with defining and quantifying nominal (A) and off-nominal (B) *predicted* events. The second two deal with defining and quantifying *actual*, measured human performance (C) and system performance (D) events. The resultant Tp value is calculated using the equation $A(1-B)/(C+D)$. This model can be used to quantify system performance throughput of advanced manned telescience systems.

Illustrations of Space Related Telescience Activities

This section discusses five basic telescience operating modes (Table 1) with a brief description of related human performance measurement-validation procedures and activities for each one.

Situation A. Person(s) (earth) to/from Person(s) (space).

Supported by advanced telecommunications, principal investigators located at many different locations on Earth will be involved in many new remote activities. Some of them are listed in Table 2.

Table 2

Basic Earth to Space Telescience Modes
and Related Experiment/ Science Activities

Telescience Mode	Science Procedure or Activity
Teleoperations:	
	Monitoring of procedures and hardware setup
	Monitoring of experimental data collection
	Observation of related events
	Management of all resources, future event planning, etc. via teleconferencing
Teleanalysis/Telearchiving:	
	Manipulation of raw and processed data
	Manipulation of specimen(s)
	Optimal display of data
	Development/updating of software
Teleplanning Support:	
	Development of in-space activity timelines
	Decision to retest
	Decision to abort experiment or process
	Decision to extent experiment or process longer
	Decision to replace one experiment or process with another
Teledesign:	
	Send and receive strategic planning/design data
	Draft/edit drawings
Teledocumentation:	
	Computer-assisted instruction (CAI)
	Report preparation, editing, routing, distributing

As already discussed, teleoperations refers to activities that are controlled remotely. Teleoperations will be integrally involved in space telescience. For example, Young (1987) points out with regard to the Space Infrared Telescope Facility (SIRTF) operation that, "...for the most part, a large central institute for the support of the operation of SIRTF is not envisioned. In general the planning, monitoring of the observations, decisions with regard to continuing and/or modifying the observations and processing of the data could all be accomplished from a person's home institution provided the observer is adequately equipped...". This approach is typical of advanced space and life sciences experiments planned for the Space Station Freedom era. Such an approach will call for an adequately validated telescience support capability.

As used here the term validation refers to the process whereby an assigned system function or capability is compared to what is actually achieved under operational conditions.

A number of practical operational constraints to teleoperations are listed in Table 3 since they impact the design and conduct of validation testing.

Table 3

Selected Parameters Which Can Impact
or Constrain Teleoperations

Space environment (pressure, temperature, radiation, etc.)
Time constraints (orbital dynamics, inertial energy limits, human crew interaction constraints, etc.)
Energy constraints (on-board power, etc.)
Volumetric constraints (e.g., reach envelope of the MRMS on Space Station)(Anon, 1989)
Hardware reliability characteristics (e.g., MTBF, MTM))
Resupply/maintenance schedule constraints
Crew availability
Orbital Replacement Unit availability

Four different types of telecommunication links are discussed here with various human performance metrics that can be used to measure their Tp.

Type 1. (Audio Link Only)

The simplest example of a space telescience activity employing a two-way ("open microphone") audio link would involve an experimenter located at a payload operations integration center (POIC) communicating with one or more flightcrew about such topics as an ongoing experiment, routine station-keeping matters, and personal matters. Person to person communications have received the largest amount of basic and applied research of any of the categories listed here. Among the most important parameters to consider are transmission delay, frequency distortions over time, auditory quality of system output (headphones, speakers), signal/noise ratio of the transmitted audio signal, special squelch circuitry effects, and peculiar auditory characteristics of each speaker's voice.

Most techniques for analysis of verbal communication are ad hoc and can be used only with well constrained tasks e.g., protocol analysis (Ericsson and Simon, 1984). Bailey and Kay (1986; 1987) presented another approach known as 'verbal data structural analysis' for quantifying real world tasks involving human-computer system interactions. I have presented other techniques involving contextual analyses which may find use in this situation [1979(b)]. The interested reader should consult these references.

As is well known, people communicate with one another in many different ways, each of which calls for a somewhat different way of quantifying their behavior. For example, during what I will call the direct social conversation mode, two people within voice range of one another will tend to sit a certain distance apart facing in certain relatively fixed directions relative to one another. These geometric factors can be readily recorded using closed circuit television (CCTV) cameras and video recorders. Image analysis of the video tape is not accomplished readily, however, and requires a great expenditure of the analyst's time. As Mackay (1988) stated, video is a powerful medium for capturing and conveying information about how people interact with computers. The same can be said for how people interact with each other. It provides a record of sequential streams of often subtle behavior that is difficult or impossible to capture in any other form. Video also preserves the context in which the behavior takes place. Such data can be of inestimable value as new experimental hypotheses are generated and need to be tested (later) using the available video record.

The Visual Courseware Group at MIT has established Project Athena (Ibid.). Among its activities is one in which full frame rate National Television Systems Committee (NTSC) video signals are digitized and presented within an X window display on a high-resolution color graphics monitor. An objective of Project Athena is to support faster means of capturing, analyzing, and presenting video data. The user can create software "buttons" to tag various events for later analysis. The TV camera's output is fed to the workstation where it is visible within a dedicated window. The experimenter can use a mouse or light pen to quickly tag those persons or events seen which deserve later consideration. Textual annotations also can be made in real time. Upon replay, the experimenter can view the entire tape at any speed or see just those tagged events of interest. Such tags can also be programmed as video editing cues to produce a second generation copy in a new "sorted" order, e.g., list all "emotional aggression" tags first followed by all "physical aggression" tags second, etc. The MIT researchers have also provided for modifying old tags or creating new ones based on symbolic labels such as those listed in Table 4.

Table 4

Symbolic Labels Useful for Modifying
Existing or New Tags

Clock time or frame number
Symbolic labels that describe events
Recorded keystroke patterns
Frame by frame snapshots from the video record.
Textual patterns from a transcription from the audio track.

Recently completed NASA Space Station Freedom crew interaction research involved the use of four closed circuit television cameras operated simultaneously at full frame rate (50 Hz) (Haines et al., unpublished technical report, 1987). One overhead camera had a wide field of view (approx. 60 deg arc diam.) lens and was aimed vertically downward. It was used to observe and record crew movements within the simulated flight deck. Three horizontally aimed cameras oriented approximately 90 deg arc to each other were found to be effective in monitoring facial expressions, hand motions, and other fine motions over the five hour-long duty period. These geometric factors of human behavior are determined largely by such considerations as sight line convenience and comfort, personal volume envelop limits, ambient noise impact, lighting and shadow characteristics, visual acuity requirements to perceive facial expressions and "body language," eye-to-eye contact needs, available furniture and their location and moveability within the room, and other such factors.

In a remote social conversation mode, where individuals are separated by some distance, these same two people (as above) will tend to adopt significantly different postures, verbal and facial expressions, etc., all of which can be monitored and analyzed.

The fundamental differences that occur between the "direct" and the "remote" social conversation modes immediately suggest validation procedures. They include: (1) Analysis of the verbal content of persons in each situation using standard syntactical and related techniques [Haines, 1979(b)], (2) Accuracy of information communicated in each direction per unit time, (3) Volume of information communicated in each direction per unit time, (4) Changes in the user's understanding or cognition using a series of semantic differential scales (Osgood et al., 1957), (5) Resistance to distraction, (6) Judged workload before, during, and/or after the communication period, (7) Judged level of personal vigilance to unanticipated, secondary tasks, (8) Individual techniques used to cope with deliberate and unplanned communication ambiguities, (9) Individual techniques used to cope with expected and unexpected transmission problems (delays, dropouts, distortions, etc.), (10) Voice frequency and volume characteristics over time, (11) Voice frequency and volume characteristics during periods of perceived and real stress, and other such techniques. Yet two people also can communicate in other more complex ways that involve higher cognitive processes

From a social interaction standpoint, two people find themselves playing both fixed and changing social roles during verbal communications. During the direct social conversation mode each individual may attempt to control the direction of the conversation in order to achieve some desired end goal or agenda. Each may do this through body language, facial expressions, sitting taller than the other person, or otherwise trying to dominate the discussion in direct, physical ways. However, during the remote social conversation mode, these same two individuals may adopt very different communication patterns because they are not physically in each other's presence. The voice may be raised in pitch and/or volume, speech may become more rapid, an authoritative tone of voice may be used, etc. The participants' inability to see one another will tend to cause them to rely solely on the auditory cues available. Most of these cues can be recorded and analyzed off-line. The commercially available 'Psychological Stress Evaluator' is one such apparatus that has limited capability to detect the presence of voice stress [Haines, 1979(a)].

From the standpoint of two people trying to relay scientific data and related information back and forth verbally, here referred to either as *direct* data conversation mode or *remote* data conversation mode, all communication is verbal, carried out in real-time, and tends not to involve very many social conversation mode factors. That is, their discussions tend to be more emotionally neutral and often center on impersonal subjects, numbers, symbols, mathematical, engineering, scientific, or mechanistic issues. A prominant exception to this rule occurs when a new person enters the conversation who does not know the current social "rules of the game" that is being played. During a crew change, for example, conversations involving neutral experimental data may include subtle or obvious humor, irony, or other out-of-context statements which may function as "ice-breakers," or "social tension reducers."

There are a number of techniques with which to quantify informational throughput achieved per unit time applicable both to direct and remote verbal communications. They include: (1) Measuring the number of words transmitted, (2) Measuring the number of words or concepts that are repeated (for any reason), (3) Measuring voice quality (average pitch or loudness), and (4) Requiring the speaker to "think out loud" to try to identify procedural errors and the mental models one may be using to operate a system.

Type 2. (Video Link Only)

From a human factors point of view, another example of person to person communications in support of telescience involves a raster video link where there is one person at each end, each seeing, but *not hearing*, the other in realtime. A variation is where the human is watching the remote operations of a robot or other unmanned operation in order to assess how it is operating. Such a system can also support a PI on the ground who has "hands-on" control of an operation in space, thus freeing up the flight crew to carry out other duties. In addition, this type of telecommunications link can support a wide variety of on-orbit tasks such as personnel briefings where psychosocial personal interaction is involved, workload assessment under operational conditions, etc.

Use of a one- or two-way video link without voice is almost unheard of today because of the relative ease and low cost of incorporating a voice channel on the transmitted video signal. Nevertheless, this type of situation may occur and calls for some comment. As used here, the term "video" includes typical alpha-numeric information, graphic displays, and dynamic imagery. A typical application would be a TV camera located within an animal cage to permit continuous remote monitoring (Haines and Jackson, 1990). The relatively low bandwidth requirements for human voice make this type of situation infrequent today since voice can be added to a video signal with relative ease.

"Composite" video displays involving simultaneous vector (also referred to as "stroke" or "calligraphic") and raster graphics are also available today. Understanding the nature of remote, complex, three dimensional objects can be enhanced using such systems when computer-generated imagery can be used to provide target object perspective, rotation, zoom, artificial shading (etc.). It is likely that such understandings will enhance human productivity during future space operations, e.g., during complex proximity operations where the out-the-window scene will be supplemented with superimposed real time, computer generated virtual imagery. Also, exploratory viewing is supported by the use of composite displays via direct object manipulation and progressive refinement. Visual continuity of target movement can also be preserved when a target vehicle passes out of sight behind another opaque object(s) during proximity operations; computer graphics can be used to portray the exact position of the occluded object. It is likely that useful insights can be gained about a variety of remotely imaged phenomenon using a video only link if there is sufficient computer power available.

Another important operating parameter is that of delay in the telecommunication system. For reasonably slowly moving targets, manual tracking is known to shift from smooth, continuous tracking with zero delay to a strategy of "move and wait" for visual image transmission delays greater than about 0.25 seconds. Clearly, for remote operations that require highly precise manual control (e.g., optical alignment, focusing, microscopic stage location, tracking a moving object) little or no temporal delay can be tolerated. Of course there are a multitude of intermediate situations. For example, what temporal delay can be tolerated if viewers of a TV image must manually aim a camera at a moving animal/specimen and maintain a focused image of it? This is an experimental question that has received relatively little study to date. If the POIC to orbit transmission delay is 0.7 sec (or more) what impact will this parameter have on the ability to stabilize the camera image on targets moving at X deg/sec when the monitor's field of view (FOV) is X/2 deg wide? In such cases achieving sufficient viewing time becomes the prime consideration. Different tasks require different FOV sizes. Research is needed to define optimal FOV for different types of tasks. Table 5 provides proposed initial FOV standards based upon the author's past experiences.

Table 5

Minimal Proposed Television Camera
Field of View Size Standards

Static rigid object subtending 5 deg x 5 deg	10 deg x 10 deg
Static animate object (that is likely to move at some unexpected time) subtending 5 deg x 5 deg	15 deg x 15 deg
Object subtending 5 deg x 5 deg and moving (horizontally) linearly at X deg/sec	4X deg FOV (horiz.) x 2 X deg FOV (vert.)
Object subtending 5 deg x 5 deg and moving in random directions at X deg/sec.	4X deg FOV (horiz.) x 4X deg (vert.)

Note: At least four seconds worth of object movement time are available using these values which allows for nominal human recognition and motor reaction time.

Video bandwidth has been found to be directly proportional to the product of *resolution* (ht. x width pixels/frame), *frame rate* (frames/sec), and *gray scale* (bits/pixel). A study by Ranadive (1987) found that when the user varied one of these parameters at a time and tried to manually operate a remotely controlled manipulator device while watching his own movements via a TV display, he could carry out the assigned tasks relatively well even though the TV image was degraded significantly. All subjects were trained to asymptote levels of proficiency before data was collected. Performance was defined as the quotient Tt / Td where Tt is the time to accomplish the task using full video (no degradation) and Td is the time required to accomplish the task using degraded video. Thus, as long as only one of the three parameters was degraded performance was still acceptable down to a point where the task could no longer be accomplished at all. In addition, he found that frame rate and gray scale could be degraded by larger amounts than resolution before the critical performance limit was reached. For tasks employed in this study the limiting parameters were:

Resolution .... 64 x 64 pixels @ 28 frames/sec @ 4 bits/pixel
Frame Rate .... 3 frames/sec @ 128 x 128 pixels @ 4 bits/pixel
Gray Scale .... 1 bit/pixel @ (28 frames/sec @ 128 x 128 pixels)
(values in parenthesis are assumed)

This study provides a useful candidate experimental design for use in future video investigations involving remote manual control of robotic systems. The fact that resolution, frame rate, and gray scale trade off in an approximate 1:1:1 fashion, respectively, raises the question whether varying two parameters at the same time would show the same trade off ratio. Such studies need to be conducted.

In another study conducted by Deghuee (1987) the operator was permitted to adjust resolution, frame rate, and gray scale during manual (robot) control operations under total bit rate constraints. The study showed that the type of manipulation task undertaken yielded the most significant differences in performance. In addition, dynamically changing these three parameters in real time also influenced performance although lower bit rates did *not* result in reduced performance. Since only two bit rates were studied (10K and 20K), it is possible that they were not different enough to produce significant performance decrements. It also was noted that the operators did not adjust the three parameters to achieve an image with some "optimal" quality but, rather, set each parameter to achieve some predetermined combination of settings.

The above two studies seem to suggest that if the operator can obtain a high quality image of some remotely televised operation from time to time, overall manual control performance does not suffer from degradations in resolution, frame rate, or gray scale as long as some minimum threshold value is maintained. It remains to determine how often the "best" image should be updated under operational tasks and how good is "best"? McGrath of MIT (personal communication) has suggested that an automated system should be employed which permits the operator to choose the available bit rate that would optimally integrate these three parameters. If some average (default) bit rate is imposed on the system, for example, the operator could increase frame rate in order to better visualize rapid motion of a target vehicle while gray scale and resolution would decrease accordingly by predetermined amounts and in the proper sequence. In the study by Deghuee the software prioritized these trade offs as follows: (A) frame rate increases:gray scale decrease:then resolution decreases. (B) resolution increases:gray scale decreases:then frame rate decreases. and (C) gray scale increases:resolution decreases:then frame rate decreases. Another question is whether other combinations would lead to faster operator accommodation to such viewing conditions or other strategies for accomplishing the task(s).

If the imagery being transmitted to the ground is a realtime (i.e., not delayed or frame frozen) scan of the flight crew then a number of performance metrics are available. Several are presented in Table 6

Table 6

Candidate Performance Metrics for Realtime Video Only Link Transmissions

(a) Measurements of the ability of the sender(s) to send meaningful but randomly selected information to the receiver per unit time under realistic workload conditions

(b) Measurements of the general strategy or approach taken by a sender to organize information to be transmitted

(c) Manual control of a remote manipulator with time and accuracy as primary dependent measures

(d) Measurements of the accuracy of received information compared with what was sent; time-averaged error rate is useful along with error type classification

(e) Measurements of perceived workload during the transmission period both by the sender and receiver

(f) Measurements of subtle behavioral cues of the sender such as facial expressions, lip motions, eye fixation patterns, etc.

(g) Monitor all flight crew incursions into the personal volume of others and note changes in social behavior over time

Of course, for such data to be used in a scientifically precise sense, an accurate record must be kept of the sender's actual behavior. Two CCTV cameras connected to separate video recorders are often sufficient for this purpose. One should be aimed vertically downward with a black tape X-Y grid pattern on 12 or 24 inch centers on the floor filling its field of view. The second should be aimed horizontally and located at the operator's eye level.

If the imagery being transmitted to the ground is of a cage containing one or more animals inside then other records and measurements may be taken. In all cases it is essential that objective (e.g., video) records be kept of the animals' actual movements of interest for later comparison with movements and responses of the control group. If the imagery being transmitted to the ground is of an electronic rack of equipment or other nonmoving object which only (may) change in brightness and/or color, e.g., warning lights which flash on and off, then use of digital frame buffers and image difference comparators programmed to indicate only imagery that changes may be effective.

Use of gaming techniques such as charades can be useful during manned tests not only to identify those flight crew who are talented at communicating entirely through non-verbal means but also to discover what facial expressions, hand-body motions and other non-verbal cues are most effective in transmitting information. Experiments can be conducted where each flight crew person must attempt to communicate pre-defined information over the video media alone.

Type 3. (Video and Audio Link)

The best known telecommunication mode used to date in U.S. manned space missions is audio-visual. The flight crew and ground crew can see and hear each other during which time a wide variety of information can be shared. Standard quality NTSC video (typically >300 horizontal lines) with color can aid in assessing non-verbal cues e.g., facial expressions, body language, interpersonal reactions. This telecommunications capability supports numerous validation approaches some of which are discussed below.

Investigators and others interested in understanding the relation of task-related practice to support technology have turned increasingly to the use of audiovisual data recording. When time is a relevant dimension of the behavior of interest the audiovisual record provides an effective recording medium. This section considers full scan rate video and undistorted audio communications separately from slow scan rate video and distorted audio transmissions since the two situations differ significantly in terms of their potential impact on human performance. In the following section I will consider *active* human monitoring of full scan rate video and undistorted audio communications.

Full Scan Rate-Undistorted Video. There are many candidate human performance validation methods available to quantify the Tp of audiovisual systems. One general class of methods involves measuring the time required for a user to reach asymptote on a learning curve in order to become proficient in a new skill. This was done both with and without the video link *and* with and without the audio link in a recent remote coaching study (Haines et al., 1989). We found that when the PI could monitor the real time performance of remotely located, relatively inexperienced (surrogate) Mission Specialists, quality of science is significantly improved. Conversely, loss of video resulted in many errors that were not caught by the PI or the ground controller.

Another general class of research methods has to do with administering subjective attitude surveys to all parties before, during, and/or after an undistorted audiovisual transmission is made. Subjective attitudes regarding the judged adequacy of the transmission to support a required task are determined. In this kind of study it is imperative to try to hold as many of the extraneous variables constant as possible, e.g., distractions in the test environment and motivational factors.

A third experimental paradigm that is particularly suitable to a laboratory situation is to permit the user to vary each of a number of stimulus parameters independently until an acceptable level of display quality is achieved. This is done under operational conditions where, for instance, video bandwidth, grey scale, resolution, etc. may be less than optimal. This approach can provide useful insights about what level of information display quality the user feels is adequate as well as the amount of time that is needed to make such tradeoff judgments.

Slow Scan Rate (Freeze Frame) Video. Slow scan video refers to non-real time imagery. A typical situation would involve a raster line by raster line build up of the video image over the period of many seconds (e.g., 10-20 seconds per full image, black and white single-field, NTSC-like format for a 56 kbps circuit). The final image is static and may be color or black and white. This type of image display can impact how meetings are planned and carried out as well as how effectively they are judged to be later. Southworth (1986; 1987), for example, has described how such systems can be used effectively in science and engineering. McIntosh (1988) documented their effectiveness as a supporting sub-system in a number of rapid visual problem solving applications in business. Swift (1984; 1988) has also documented their use to link a Senator in Washington, D.C. with a University of Hawaii class in Honolulu and in other teaching situations. Keen (1986) presents an excellent historical overview of the use of freeze-frame video by America's mass media (TV and print). Finally, an unpublished paper entitled "Telemedicine and Slow-Scan Video" by Robert H. Jaros of the Department of Nuclear Medicine, Catholic Medical Center, Manchester, New Hampshire and Cynthia E. Keen of Colorado Video, Inc., Boulder, Colorado cites numerous examples of the effective use of slow-scan video in "telemedicine" (e.g., X-ray; electrocardiographic; body wounds; rashes; and eye injury imagery). The above cited practical applications of this technology provide a useful foundation from which further manned system evaluations may be carried out. A number of possible general and specific evaluation techniques are possible.

An experimental question of interest related to slow scan video media is how slow can the imagery be presented on the screen without leading to a complete breakdown in the effective flow of information because of user frustrations, misunderstandings, interpretive errors, premature responses or other potential problems. For example, one possible experimental protocol would require that a precise series of tasks must be accomplished that are imaged on the slow scan video. Various image scan rates would be presented (in random order), with each video frame containing information with matched difficulty and relevance to the task at hand. Each task that the user must carry out would be measured in terms of time to accomplishment and error rate and then related to scan rate.

Another protocol that is useful is to determine whether the user will act prematurely or will wait for the entire video image to be displayed before taking some action on the basis of the display. The degree to which scan rate is directly related to the need to display the *entire* screen full of information can be measured.

In a laboratory setting, *covert* monitoring of user responses to slow scan video may uncover overt behavior regarding how the user copes with the absence of a constantly updated visual image. For instance, he may become impatient and distracted or he may use the "dead" image period to plan for the next video image transmission.

Compressed Video. There are a growing number of techniques for suppressing or eliminating redundant video information, i.e., picture elements (pixels) which don't change.

The essence of acceptable video compression supporting remote scientific monitoring is to be able to provide adequate image resolution *and* motion fidelity. Of course, specifying what is adequate is not easy; usually it is best to permit a representative group of actual users to make this assessment under operational conditions as was done elsewhere (Haines and Jackson, 1990).

The assessment technique used by Haines and Jackson included the following steps: (1) a high quality, sixty-second-long video tape master was made of several scenes of interest, in this case three SpaceLab 3 rat cages containing seven small white rats, (2) each original scene was compressed to six different levels ranging from 384 kbps to 1,536 kbps (using commercially available hardware and proprietary compression algorithm) and then reassembled in a presentation tape with the compressed scenes presented in random order, (3) observers watched each scene and immediately rated it on overall acceptability as well as on the quality of image motion and resolution. It was found that for these levels of compression and the particular algorithm that was used, higher compression levels were acceptable if the motions of the rats being remotely monitored were slow (typ. <2" per sec) or of small amplitude (typ. <0.5"). As expected, acceptable image detail was inversely related to the magnitude of image compression.

When both the audio and video signals are compressed it may be possible to allocate different percentages of the available bandwidth to each in order to achieve an acceptable audio-visual transmission. One commercial system, which has a total bandwidth of 384 kbps, provides four different compression level combinations as shown in Table 7 (all values in kilo-bits per second).

Table 7

Audio-Visual Bandwidth Allocation
and General Quality of Transmitted Information
(Compression Labs Incorporated, San Jose, Ca.)

Video = 320	Image is sharp and clear with little motion blur visible
Audio = 64	Voice quality is higher than standard telephone service
Video = 352	Image contains very minor image distortion and blue
Audio = 32	Voice quality is good, some distortion of higher pitched voices
Video = 368	Image contains edge blurring during rapid motions, high illlumination level is needed
Audio = 16	Voice is similar to long distance telephone communication with frequency cut-off effects
Video = 376	Image appears fuzzy with poorer temporal and spatial resolution
Audio = 8	Voice is below long distance phone service, diction is difficult to perceive, speaker's personal identity is difficult to determine

Audio-visual communications are also used to present warning and system status information that must be monitored passively. Human performance assessment during *passive* system monitoring can take various forms, some of which are listed in Table 8.

Table 8

Selected Performance Metrics Useful During
Passive Audio-Visual System Monitoring

Assessment of attention capture
Measurement of attention span
Measurement of error detection capability
Identification of error correction strategies
Measurement of psychological/mental fatigue
Measurement of visual eye scan behavior

Attention during complex tasks usually changes so rapidly, is so subtle in its effects, and is so transparent in the processes it uses that it is very difficult to measure [cf. Kahneman (1973); Wickens (1980)]. It has been suggested that attention itself cannot be measured at all but only some correlated artifact of it. Typically, one's performance on a task can only be related indirectly to attention before and during the task. Nevertheless, some meaningful data can be obtained which is related to passive system monitoring through *attentional capture assessment*. One general approach is to present the viewer with a dynamic, real-life situation which must be attended to over prolonged periods of time in order to answer questions correctly. An observer on the ground might monitor an in-flight experiment via a televised transmission, for example. At some unexpected point a "target stimulus" is introduced and the observer is monitored to find out: (a) whether he identified its presence, (b) how long it took to perceive it, and (c) what response did it evoke. The literature on attentional capture and conspicuity in general is relatively large (cf. Fischer et al., 1980). Responses specifically related to errors, introduced at random intervals within an ongoing experiment or procedure, can be monitored and analyzed in realtime or after the experiment is over.

Table 9 presents other performance metrics which are useful in quantifying human performance during active control of remote operations.

Table 9

Selected Performance Metrics Useful During
Active Control of Remote Operations

Manual controllability of dynamic systems (e.g., robots)
Measurement of input control error type and rate
Assessments of subjective workload of selected components of the task
Measurement of selected psychophysiological responses (heart rate, galvanic

skin response, blink rate, etc.)
Accomplishment of primary/secondary task
Measurement of adequacy of task performance within a given period of time
Determination of task accuracy while coping with communication transmission delays
Measurement of performance ratings by operator(s)
Measurement of rerformance ratings by non-participatory observer(s)
Measurement of gross body movement (head, eyes, limbs, torso)
Determination of operator's monitoring capability (errors per unit time)

Decision making may be thought of as an unconscious sorting of available plans combined with a more formal, conscious, and overt comparison of available resources. While this second aspect of decision making usually can be monitored, the former unconscious aspect cannot. The decision making process takes place within the mind where neither introspection nor scientific method can discern it. All one can measure is its results.

Clearly, the process of human decision making is extremely hard to measure in operational settings. Advanced simulations are useful in helping to determine what behavioral correlates of decision making should be measured. A detailed task analysis is extremely helpful in such research since it can provide insights concerning the most likely decision-transition points in an ongoing sequence of actions.

Type 4. (Electronic Data Communication Link)

This situation refers to computer network-based systems where many people read and respond to alpha-numeric video displays that are linked with other systems. Other names for this general area include computer conferencing, electronic mail/bulletin boards, computer message system, simultaneous conferencing, and electronic information exchange. To support efficient and reliable experimental *data* transmission, different grades of communication services will be required, each carefully matched to the kind of application that is planned. This will be true between different ground personnel as well as between space and ground personnel. Further research is needed on the effects of transmission latency, bandwidth, and bit error rates on human productivity. It also must be mentioned that computers and communication are merging more and more; the human's cognitive use of each technology is becoming increasingly difficult to separate and measure. Carasik and Grantham (1988) point out that, "...the extended OS/2 on the new IBM PCs will support communications primitives to support transmission of voice, bitmapped graphics, and text within one framework." It is only a matter of time before conversational speech, virtual three-dimensional screen imagery, hyper-media informational formats, etc. will be added which will provide new solutions to old problems as well as new challenges to the human factors engineer.

Future telescience activities conducted in space will involve principal investigators located on the ground as well as in space. It is likely that the ratio of personnel who will need to communicate with each other between the Earth and Space Station Freedom will be anywhere from 1:1 to 10:1 or more, respectively, or more during a typical crew shift. Use of a person to person "pipeline" communication concept will help control information overload in space. In general, the following planning and execution factors (Table 10) will play a significant role in formulating the best ways of supporting person to person telescience activities and in determining how best to validate them.

Table 10

Selected Planning and Execution Factors Related to Person to Person Telescience Activities

Number of group/team meetings of the space crew scheduled per shift
Number of scheduled meetings that are rescheduled unexpectedly
Size of the space crew per meeting
Authoritarian status of each person on the ground and in space
Personal communications skills of each person on the ground and in space
Need for personal communication privacy and data security

Effective individual and group decisions are heavily dependent on accepted communication protocols, social conventions, judged uncertainties and adoption of an acceptable risk to reward ratio (Tversky and Kahneman, 1974). Nevertheless, system designers usually do not have first hand knowledge of these conventions and protocols for the wide range of environments in which telescience will increasingly find itself in the future. Several examples are in order: (a) When users' initial expectations are not met by a newly introduced automated system they tend not to use it as fully as they might, (b) a social hierarchy-structured work situation typically governs what type of information is transmitted and when, and (c) new office hardware that is forced upon the workforce without proper training or acceptance can govern what type of information is transmitted and when. Indeed, these research findings derived from traditional office environments can be used (with caution) in planning for support of science in laboratory environments.

Two separate NASA Ames' projects incorporated computer conferencing capability which deserve further comment (Vallee, 1984). They were: (1) a conference on "future transportation systems" involving NASA, industry, and university participants who needed to mutually assess current technology (as of late 1975), and (2) a "Communications Technology Satellite" (CTS) project involving six NASA centers and about 20 contractors over a four year-long period beginning in 1975. User statistics were collected in a number of categories. Both groups had access to entries typed into a keyboard made by other project participants on an ad lib basis, i.e., whenever they logged into their networked system. They could also send public and private messages. While the two groups differed significantly in their overall objectives, the percentage of system usage time in five categories was relatively similar as shown in Table 11. These five categories are useful in comparing one telecommunication system with another.

Table 11

Comparison of Computer Conferencing Usage
Percentages in Five Categories
(After Vallee, 1984)

Category	Future Transportation systems	Communications Tech. Satellite
Administrative	32	23
Procedural	24	19
Substantive	23	43
Learning	9	8
Social	12	7

Vallee points out that computer conferencing played several important roles. First, it replaced or supplemented other media, i.e., users confirmed information that they had received through other channels. Second, it helped deal with emergency situations in so-called crisis management situations. Third, it promoted an effective style of management, e.g., use of the public communication mode (during this conference) confirmed prior private group participant agreements. Fourth, it extended communications beyond normal working hours. The normal "telephone window" between the east and west coast was expanded to 12 - 13 hours, according to a conference participant.

In summary, computer conferencing will play an increasingly important role in advanced planning for Space Station Freedom as well as during its lifetime of complex operations. Further research is called for to identify how computer-assisted conferencing should be managed to the benefit of the flight and the ground crews.

Borrowing from Vallee (Ibid.), a matrix made up of three *modes of communication*: (1. talking to onesself, 2. talking to another person, and 3. talking to a group) and six *routes for human communication* (1. No delay-send, 2. No delay-receive, 3. No delay-send and receive, 4. Delay-send, 5. Delay-receive, and 6. Delay-send and receive) form an array of all possible communication patterns that is useful for discussing electronic data communication links. He also presents interesting data regarding how two NASA clients used a text-based computer conferencing system.

There are a number of validation techniques suitable for assessing electronic data communication links from a human information Tp point of view. One bottleneck to date has been the design of the user's input. Gould and colleagues, (1984, 1986, 1987), for instance, have shown that people read the same words/text more slowly from CRT display than from paper. They did show, however, that when the quality of the screen's images were improved over what is now considered the "standard" font (i.e., improving contrast, aliasing, and pixel size), reading speed between the two media became equivalent. In a similar vein, the present standard QWERTY keyboard layout has been shown to be slower and more prone to input errors than are other alphabetic keyboard layouts. The point is that this type of research on computer input devices provides many useful experimental techniques for validating existing and future hardware.

A human factors question of concern has to do with situations involving the need for synchronous reception of audio, visual, and data information. The situation is illustrated by an astronaut who may be performing an experiment on orbit under the verbal and video guidance of an expert on the ground (cf., Haines et al., 1989). What consequences will occur if there is asynchronous transmission of the audio and video data? Also, data rates and latency need to be realistically defined to support the large number of experiments under the constraint of limited ground-space-ground bandwidth. The communications system has to be robust enough to accommodate a range of grades of services with guaranteed minimum latency. Perhaps a communications "load levellor" scheduling algorithm is needed for all experiments using a given channel that has a fixed, maximum bandwidth. Thus, a group of high bandwidth experiments might share one channel having its own algorithm while another group of low bandwidth experiments could share another channel having a different algorithm, etc.

Another human factor area that deserves much more research in order that telescience assume its proper role in space operations is that of information display. There are many unanswered questions concerning how dynamically interacting information should be presented to users. Some of these questions are listed in Table 12.

Table 12

Some Unanswered Questions Related to the Optimal Presentation of Dynamically Interacting Information

1. What presentation format(s) elicits the highest comprehension rate? For instance, should all available information be presented visually or can some be presented in other sensory modalities?

2. What features of presentation format(s) support optimal perceptual detection and recognition of critical data? Can new ways be found to present massive data arrays in space and time that maximize one's ability to quickly and accurately identify critical features? (cf. Tufte, 1983)

3. Is the investigator able to view *and* interact with ultra-large data bases which involve experimental data and models so as to permit parameter changes to be made in real time and otherwise to interact with the experiment as it occurs?

4. Another human factors issue has to do with optimizing the networking design of complex distributed information systems. Some of the many unanswered questions here include:

a. How is experimental data to be accessed simultaneously by spatially dispersed experimenter teams?

b. What is the best way to support real-time decision-making to meet new, unplanned opportunities for novel data collection and data analysis?

c. How can the flight crew be assisted in trouble shooting onboard hardware?

While it is likely that the physical scientist will interact with generally well-known phenomena and will collect data that is largely numeric, the life scientist often will study unstable, dynamic phenomena and behavioral responses which cannot be preplanned. He will need a wide choice of imagery as well as data. He will probably also need a flexible communications downlink/uplink capability to permit timely and creative decision making support.

Situation B. Person(s) (earth) to/from Machine (space)

The Space Station Program will eventually incorporate a wide variety of systems with varying degrees of autonomy. Some of them will have to be monitored, diagnosed, actively controlled, commands cancelled, or othewise interacted with from different locations on the ground and also in space. Telescience will undergird most of these activities.

In this section I will briefly discuss man-machine interactions where the machines represent highly "intelligent", semi-autonomous systems. The term "PI in a box" has been applied in this context. Examples of this basic category of telescience activity are found in a number of autonomous operations where humans will periodically monitor system "health." In addition, other examples are found in remote systems operations from the ground, e.g., production and assembly of raw materials on and construction of Space Station Freedom, satellite servicing, active exploration of space and platform repair/maintenance. Indeed, autonomous systems including telerobots of all kinds will play a central role in such future operations (Brackman et al., 1986; Bronez, 1987; Bronez et al., 1986;).

The term *automation* is defined here as any pre-programmed, mechanized task that is initiated by some precondition (user resesponse or otherwise) and which is self-sufficient thereafter. New automation technologies are most likely to be used on Space Station Freedom when it can be shown that they lead to significant improvements in one or more of the following areas: increased payload accommodation, increased human productivity, increased safety and reliability, increased flexibility and growth capability, increased crew morale, decreased operator training and operating costs, decreased ground operating costs, and decreased on-orbit weight.

The introduction of automation to operational systems has not been without its problems; human error continues to play a predominant role in the safe operation of all large and complex systems. For example, according to an article in *Aviation Week and Space Technology* magazine (pg. 31, September 12, 1988), a Soviet ground controller made a manual input control error which was sent to their Phobos-1 Mars spacecraft. This error caused the vehicle to lose its antenna lock on Earth so that it no longer could react to any control signals. The mission may have failed because of this rather innocuous human error. Indeed, increasing levels of hardware reliability has been accompanied by a growing incidence of human error(s) in accident causality (Billings, 1989). *In some past situations the automated system has placed the human in a role for which he has not been adequately trained, for which he is poorly adapted personally, or which exceed his ability to adapt to and cope with taxing situations.* In other past situations the errors can be traced to poor man-machine interface design. A key question here is "exactly what role should the human play when interacting with automated systems?" Studies have suggested that humans who do not fully understand the internal components of a highly complex, automated system may do more harm than good in interacting with it, particularly if the interface has not been designed to do such things as: (a) continuously and consistently monitoring all faults, (b) annunciate all errors unambiguously to the operator(s), (c) provide an unambiguous and consistent logic diagram to follow in the event of system malfunction, (d) limit the consequences of wrong human input actions, (e) limit the consequences of hardware malfunctions, etc..

Clearly, more research is needed to properly match the cognitive (intellectual) and perceptual capabilities and limitations of the user with the automated systems interfaces. It will continue to be a non trivial challenge to find the optimal control interface (boundary) between the human's input and the automated system.

Fully autonomous systems in space must carry out a wide variety of tasks. It is instructive to list some of them here (Table 13) since they provide a foundation on which later examples of man-in-the-loop and man-out-of-the-loop, i.e., "autonomous" situations may be compared from a validation method standpoint. Many of these tasks are now performed by people in space and on earth using time consuming procedures.

Table 13

Some Space Tasks Involving Autonomous Systems

1. Use of heuristic rules in detecting failures. Using knowledge based on prior experiences (machine or human) to detect and diagnose system problems

2. Capability to use model-based or causal failure detection and diagnosis. Using second-order/model-based knowledge to diagnose system problems

3. Decision-making in uncertain situations. Making sensible decisions when knowledge of the status of other supporting system components or of the larger "world" knowledge base is imprecise or incomplete

4. Real time monitoring and correction of failures. Putting a plan of action into place

that continuously and accurately keeps track of system status with respect to nominal and off- nominal operating conditions

5. Planning for failure corrections. Developing a plan of action to repair some failure event that meets certain criteria

6. Resources usage scheduling. Planning how available resources should be allocated to potential users in real time

7. Operations scheduling execution. Capability to reformat task/user requests into executable system commands

8. Performing future trend analysis. Capability to visualize slowly developing trends that may include a high degree of noise and/or changing input parameters

9. Capability to Learn. Capability to change, add, or delete information from an operational knowledge base automatically as conditions change and new knowledge is added

---

An adequate telecommunication capability will be required on Space Station Freedom to allow humans to monitor the operations of intelligent systems and to decide when different functional abilities (e.g., heuristic rather than model-based reasoning) should be employed. In addition, humans will need to be able to quickly override decisions that have been made by autonomous systems that are likely to result in near-term and far-term malfunctions. In short, humans will require communications links that have negligible delays, adequate bandwidth, and data-stream integrity.

Situation C. Person(s) (earth) to/from Person(s) and Machine(s) (space)

The Space Station era will generate many new requirements related to hardware and software design as well as to the user interface. It has been pointed out that experimental complexity, diversity, and flexibility will increase as mission duration increases on Space Station Freedom. To more adequately support and exploit these new capabilities, ground-based investigators will require: (1) near real-time access to flight data, (2) high-speed computing power in support of data modeling, analysis and resource management, and (3) the ability to permit in-flight experimental modifications when unanticipated events occur. Telesciences will undergird all of these requirements.

Telecommunications will support the interaction of people on the ground with people and machines in space in a wide variety of ways. For example, the Space Infrared Telescope Facility (SIRTF) will be operated in a telescience mode by principle investigators located in many different home institutions on earth. A high Tp communications network will be required that is extremely efficient, reliable, and interactive. Data will be transferred between the laboratory in space, Science Operations Center (SOC), and the various research institutions. A recently completed telescience testbed pilot activity conducted jointly by the

University of Arizona, Smithsonian Astrophysical Observatory, and Cornell University (cf. Leiner, 1989) provided a valuable demonstration of the capability to carry out routine communications via networks to facilitate data transfer, software development and transfer, and general intercommunication among all participants.

Careful planning must be given to how the *total system* will be validated. This situation will assume increasing relevance in the space station era as so-called "intelligent systems" mature and are integrated into space based systems. An example is found where both manned and unmanned space hardware is involved and real time decision-making is required among all elements. The ground science and support crew may be involved in performing off-line calculations to support trade-studies that impact the crew in space. At the same time, the pre-programmed space hardware may be carrying out assigned operations that, if continued without interference, will eventually lead to a disastrous system failure. It is here that systems that are highly "fault-tolerant" can generate problems for both the ground and space crews. *Such fault-tolerant systems should never be completely opaque, i.e., hidden from the user.*

An example of a remotely located, automated system that requires manned assistance is found in the United Kingdom Infrared Telescope (UKIRT) that can be operated remotely from Edinburgh, Scotland. The telescope is located in Hawaii on Mauna Kea. Two technicians are required to be present at the telescope because the computers and power supplies must be turned on manually. In addition, telescope slewing (aiming) is also done by the site technicians for other reasons. While it is theoretically possible to connect the telescope in real time with the Scotland control center (with a five sec round trip transfer time) over an X.25 telenet connection, it is sent to a local disk (at Mauna Kea) and inspected off-line in Scotland. Thus, what is described as an automated system still is not entirely autonomous.

A useful human performance procedure in situations involving degraded video imagery from remote sites is to set up the visual monitoring situation so that a periodic decision must be made by the viewer based upon the degraded image. Image quality is systematically varied and the adequacy, completeness, delay, etc. of the decisions made are noted and related to image quality. Of particular relevance in such situations are image threshold conditions where there is a 50-50 probability of the decision going either way.

Ground personnel may need to communicate with the flight crew and flight systems (automated and non automated) simultaneously. Determining total system Tp in such highly complex and interactive circumstances is not easy. The possibility of unplanned hardware failures on the ground and in space, audio visual communication link degradations, and human errors on the ground and in space make for complex interactions indeed. For example, if the POIC electronically interrogates the Space Station Freedom's flight crew concerning the status of a sub-system and finds an obvious discrepancy between their assessment and what the automated sensor system is transmitting to the POIC's computer, how should the discrepancy best be resolved? Such situations are likely to occur as on board systems becoming increasingly "intelligent" and transparent to all users.

One general class of validation studies that can be done during pre-flight simulation of a mission has to do with insertion of deliberate system and/or sub system failures. This approach is in common use in training commercial airline pilots to cope rapidly and correctly with malfunctions in flight. For example, a space station mockup's thermal control safety cutoff switch could be programmed to stick in one position at time "X" well into a high workload period of the simulation. Crew behavior and performance is monitored using audio visual recordings before, during, and after this event to: (a) establish the ongoing baseline of workload, task accomplishment, interpersonal relations, communications patterns, etc. prior to the malfunction, (b) determine whether the crew noticed the malfunction and, if they did, how long did it take them, (c) determine what specific actions were taken to cope effectively with the malfunction, and (d) what "down stream" consequences occurred as a result of the malfunction and the crew's responses to it? Use of MIT's Athena Project video tagging methodology would be very useful in such validation studies.

Space Station Freedom's ability to support a broad range of scientific activities coupled with its projected lifetime of at least 30 years will call for creative solutions to providing flexible on-board training systems. It is here that telescience can play yet another significant role. Most of the flight crew will not be computer experts. Indeed, they will most likely be trained before the mission only on those specific skills that will be needed to carry out planned events. Herein lies a non-trivial challenge. How does one identify the best performance metrics to use in a Tp analysis in such instances? As Goransson et al. (1987) suggest, "adaptation to local circumstances and needs is usually a necessity." The flight crew are probably going to remain inherently more flexible than the computer. In addition, most of the automated systems will be hidden and many will not even be interacted with except during planned maintenance periods or malfunctions.

Advanced planning for off-nominal situations involving machines in space can take various forms, each of which rests upon a thorough understanding of the components of the system in question. Developing contingency plans for system failure, for example, often involves little more than restating how the system operates and how to insert a new element into an existing system architecture. Human memory and data base access play key roles. Table 14 presents a list of possible procedural steps for quantifying one's ability to cope with a system malfunction in a remote space system.

Table 14

General Procedural Steps for Evaluating Operators' Capabilities to Cope With a Remote System Malfunction

1. List all possible malfunctions
2. List all feasible solutions in real time
3. Time how long it takes to do step 1 and 2
4. Record which solution(s) was chosen
5. Interview decision makers concerning why they selected the solution(s) they did
6. Determine how successful that decision was through realistic simulations
7. Determine whether other decisions were tried first and found unsuccessful

8. Administer subjective workload rating immediately following decision making
9. Assess how members of the decision-making group were chosen
10. Determine how many different people actually contributed to the decision chosen
11. Determine the time required to derive a list of "n" alternative solutions to a particular off-nominal situation.

Situation D. Machine(s) (earth) to/from Person(s) and/or Machine(s) (space)

The autopilot of a modern airplane is an example of a fairly rudimentary automated system which provides navigational (i.e., long-term "guidance") and dynamic short-term "control") information in real time to the control surfaces of the airplane in such a way that the airplane flies itself; the pilot functions outside of this closed control loop, and is (thereby) able to perform other functions of a more global planning, strategic nature. Carefully designed human sensory alerting lights and auditory tones are used in the cockpit to signal the pilot when the airplane's autopilot is not operating within predefined limits. However, the introduction of new and complex hardware into the airplane's cockpit is not without its traps which must be carefully considered before final implementation [Curry, 1985; Rouse and Morris, 1987]. The example of an autopilot is useful for illustrating several points regarding telescience applications in future space operations.

In the space station program there are going to be a number of types of unmanned, semi-autonomous free-flyers, e.g., Orbital Maneuvering Vehicle (OMV), that are being designed to act as ground- or space-controlled, robot "tug boats". They will possess diverse, multifunctional sensory capabilities (e.g., radar and laser ranging systems, stereo television depth capability). In addition, they will be able to do rudimentary operations without direct human control through the use of various "smart" sub-systems. A design goal is to pre-program the on-board control system of such vehicles with an end objective and then simply push its "go" button. From then on the vehicle would (ideally) complete each assigned task accurately and in the correct order. Now let us turn to the automated "machine" on earth which would communicate with and control a free flyer of this kind. Let me call this a ground control station. (cf. Sary, 1989)

The ground control station for an OMV would need to include at least the following command, communication, and control (C^3) functions: (a) controls and displays for on-board sensory system operations to permit human override of automatic systems during unanticipated conditions in space, (b) effective means for displaying information that is related to deciding whether the human on the earth or the automated system in space should be given control authority, (c) a *general knowledge* data base which is sufficiently large to encompass all reasonable future (nominal and off-nominal) situations and flexible enough to be updated as needed, (d) an *experience* data base in which resides a constantly updated virtual memory to provide an input/output data trail of all input commands, their consequences, and all relevant operating conditions at the time the command was executed, (e) fully adequate communication links between the earth and space to support all C^3 functions, and (f) dynamic, error-tolerant strategies with which to cope with off-nominal situations.

Effective error *tolerant* design requires that there be measureable (i.e., quantifiable) effects of a causal chain that eventually will lead to a human error(s). In distinction, error *reduction* considers those factors which take place well before the hardware and software are built and deployed. They may include such factors as system architecture, maintainability, crew selection and training, and other issues. Clearly, many of these subtle considerations have to do with the human factor, a design element that should *not* be left to the end of the process. The point is that while the human may be designed out of the ground control station in a misguided attempt to save money, the long-term consequences are likely to be very costly if not catastrophic. A far more acceptable approach is to design the earth based segment of the C^3 system to permit the computer to do what it does best and the human to do likewise. Herein lies the continuing challenge, i.e., finding the best mutual allocation of these resources.

When the design characteristics of the ground control station are *completely integrated* into those of the remote autopilot a number of beneficial things will happen. First, the human operator will be able to play the role purely of system manager where his superior decision-making capabilities can be used in a more optimal way, i.e., he is able to carry out longer-term, global, strategic planning. He is not burdened with near-term, generally high workload, high distracting and fatiguing tasks. Second, the human's rate of cognitive and perceptual error generation will tend to decrease because of reduced attentional workload and divided attention, all things equal. Third, his overall productivity will tend to increase because of more efficient use of available resources. Indeed, effective resource management has been shown to contribute significantly to overall flight safety in commercial flight in America (Billings, 1989). Finally, the time between successive failures of the autopilot will lengthen to the point where the pilot's skill level in coping with the failure begins to degrade. When this point is reached telescience will be useful in supporting periodic, remote skill maintenance training for the pilot.

Situation E. Machine (space) to/from Machine (space)

It is likely that future developments in robotic systems will include so-called "smart front ends". This refers, among other things, to television and other sensory systems coupled to powerful, on-board decision-support hardware. Working together they will be able to carry out semi-autonomous missions. When this time comes the sensor output data and the decision-support hardware will not need to communicate with computational hardware on earth to support real-time system safety checks and current mission verification. Only then will such systems become autonomous. One type of validation technique that could be conducted would be to program a deliberate event to take place in the smart front end and measure the effect it has on the remote hardware's ability to deal with it. Time/accuracy of system response tradeoffs could be conducted.

A primary difference between situations D and E is that in the first there is a human present to make critical inspections, diagnoses, and actions that would tend to be very inefficient for a machine to carry out without extensive pre-programming and dedicated sensing hardware. In addition, the costs involved in providing the capability to perform maintenance and repair in space are also high. What the ultimate limit is of using telecommunications to connect one machine in space with another machine in space remains to be seen. A few such roles are offered here: (a) video imaging of flight hardware components on vehicle A to look for obvious evidence of damage on vehicle B, (b) remote video imaging during rendezvous operations by a repair craft that is transfering changeout hardware to the mal- or non-functioning vehicle, and (c) data transmission to/from the mal- or non-functioning vehicle's self-diagnosing sub systems.

SUMMARY AND CONCLUSIONS

Five different situations relevant to human performance validation are discussed in terms of typical telescience activities (e.g., teleoperations, teleanalysis, teledesign). These five situations involve humans and machines on earth communicating with humans and/or machines in space. Specific examples of candidate human performance measurement-validation techniques for audio, video, audio-visual, electronic data communications are provided. It is pointed out that rapid prototyping of candidate systems has already shown itself to be a cost- and time-effective means for verifying the adequacy of new, untried approaches, developing and evaluating new user interfaces, performing trade-off studies on selected variables, taking quick looks at real time data, evaluating advanced system architectures, and other activities.

REFERENCES

Anon, *Man-Systems Integration Standards.* NASA-STD-3000, Rev. A, National Aeronautics and Space Administration, Washington, D.C., October 1989.

Bailey, W.A., & E.J. Kay, Towards the standard analysis of verbal data. *IEEE Systems, Man, and Cybernetics Proceedings*, Pp. 582-587, 1986.

Bailey, W.A., and E.J. Kay, Structural analysis of verbal data. Pp. 297-301 In Carroll, J.M., and P.P. Tanner (Eds.), Proceedings,of the Human Factors in Computing Systems and Graphics Interface CHI+GI, Toronto, Canada, April 5-9, 1987.

Barnard, P., Cognitive resources and the learning of human-computer dialogues. In Carroll, J.M. (Ed.), *Interfacing Thought: Cognitive Aspects of Human Computer Interaction.* Chpt. 6, Cambridge, MA, MIT Press, 1987.

Barnard, P., M. Wilson, and A. MacLean, Approximate modelling of cognitive activity: towards an expert system design aid. Proceedings of Human Factors in Computing Systems and Graphics Interface, Pp. 21-26, Meeting, Toronto, Canada, April 5-9, 1987.

Beeler, M., Beyond the Baskett Benchmark. *J. Computer Architecture News*, vol. 12, No. 1, Pp. 20-31, March 1984.

Billings, C.E., A historical perspective: Toward human centered automation. Proceedings of the Fifth International Symposium on Aviation Psychology, Hyatt Regency Hotel, Columbus, Ohio, April 17-20, 1989.

Brackman, E.J., W.T. Park, and L.S. Peters, A conceptual approach to designing robotic sensor systems for space. SRI International, Project No. 8620 (September 26, 1986).

Brice, R., Benchmarking your benchmarks: a user's perspective. *J. Computer Performance*, vol. 4, No. 2, Pp. 73-79, June 1983.

Bronez, M.A., Vision concepts for space-based telerobotics. Rockwell International, Space Station Systems Division, Rept. SSS 86-0165, March 29, 1987.

Bronez, M.A., M.M. Clarke, and A. Quinn, Requirements development for a free-flying telerobot - ROBIN. Proceedings of the 1986 International Conference on Robotics and Automation, San Francisco, CA, April 7-10, 1986.

Carasik, R.P., and C.E. Grantham, A case study of CSCW in a dispersed organization. Proceedings of the Conference on Human Factors in Computing Systems, Pp. 61-65, Omni Shoreham Hotel, Washington, D.C., May 15-19, 1988.

CHI'88 Panel, Dealing with diversity: Approaches to individual differences in human-computer interaction. Procedings of the Conference on Human Factors in Computing Systems, Pp. 79-81, Omni Shoreham Hotel, Washington, D.C., May 15-19, 1988.

Curry, R.E., The introduction of new cockpit technology: A human factors study. NASA Technical Memorandum 86659, 1985.

Deghuee, B.J., Operator-adjustable frame rate, resolution, and gray scale trade-off in fixed bandwidth remote manipulator control. Mass. Institute of Technology, thesis, 1987.

Emrick, R.D., Evaluating CPU performance with synthetic benchmarks. *J. EDP Performance Review*. Vol. 11, No. 9, Pp. 1-6, September 1983.

Ericsson, K.A., and H.A. Simon, *Protocol Analysis Verbal Reports as Data*. Cambridge, MIT Press, 1984.

Fischer, E., R.F. Haines, and T. A. Price, Cognitive issues in head-up displays. NASA Technical Report 1711, Washington, D.C., 1980.

Goransson, B., M. Lind, E. Pettersson, B. Sandblad, and P. Schwalbe, The interface is often not the problem. Pp. 133-136 In Carroll, J.M., and P.P. Tanner (Eds.), Proceedings of the Human Factors in Computing Systems and Graphics Interface - CHI+GI, Toronto, Canada, April 5-9, 1987.

Gould, J.D., and N. Grischkowsky, Doing the same work with hardcopy and with cathode ray tube (CRT) computer terminals. *Human Factors*, 26, 323-337, 1984.

Gould, J.D., and N. Grischkowsky, Does visual angle affect reading speed? *Human Factors*, 28, 165-173, 1986.

Gould, J.D., L. Alfaro, R. Finn, B. Haupt, A. Minuto, and J. Salaun, Why reading was slower from CRT displays than from paper. Proceedings of 1987 Human Factors in Computing Systems and Graphics Interface. Toronto, Canada, April 5-9, 1987.

Haines, R.F., Voice stress analysis in UFO witnesses. *Journal of UFO Studies*, vol. 1, no. 1, pp 1-7, 1979(a).

Haines, R.F., *Observing UFOs: An investigative handbook.* Nelson-Hall, Chicago, 1979(b).

Haines, R.F., An information throughput model applicable to advanced manned telescience activities. Research Institute for Advanced Computer Science, Ames Research Center-NASA, Technical Report 90.1, January 1990.

Haines, R.F., and R.W. Jackson, Video image compression and remote life science experiment monitoring. Proceedings of 1990 annual Aerospace Medical Association meeting, New Orleans, La., May 13-17, 1990.

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

Helander, M. (Ed.), *Handbook of Human-Computer Interaction.* North-Holland, New York, 1988.

Kahneman, D., *Attention and Effort.* Prentice-Hall, Englewood Cliffs, 1973.

Keen, C.E., Freeze-frame TC can be cost-effecrtive choice when motion doesn't have to be understood. *Communications News*, April 1986

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

Levy, H.M., and D.W. Clark, On the use of benchmarks for measuring system performance. *J. Computer Architecture News*, 10(6):5-8, 1982.

Mackay, W.E., Video: Data for studying human-computer interaction. Proceedings of the Conference on Human Factors in Computing Systems, Pg. 136, Omni Shoreham Hotel, Washington, D.C., May 15-17, 1988.

McIntosh, D.K., The rapid visual problem solver. *Procomm Enterprises Magazine*, (P.O. Box 886, San Anselmo, CA 94960), January 1988.

Mello-Grand, S., The art of benchmarking printers. *J. Byte*, Pp. 193-215, February 1984.

Osgood, C.E., G.J. Suci, and P.H. Tannenbaum, *The Measurement of Meaning*. University of Illinois Press, Urbana, 1957.

Ranadive, V., Video resolution, frame rate and gray scale tradeoffs under limited bandwidth for undersea teleoperation. Mass. Institute of Technology, thesis, 1987.

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

Sary, C., Space station instrument control system study. Computer Sciences Corp., Beltsville, MD 20705, Rept. DSTL-89-013, CSC/SD-89/6187, November 1989.

Southworth, G.R., Slow-scan television in science and engineering. *Telecommunications*, July 1986.

Southworth, G.R., Still picture communication systems. *Photomethods*, Pp. 15-18, March 1987.

Swift, D.W., Do-it-yourself videoconferencing. *Telephony*, August 20, 1984.

Swift, D.W., Senator lauds slow-scan TV. *Telephone Engineer & Management*, February 15, 1988.

Tufte, E.R., *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, Conn, 1983.

Tversky, A., & D. Kahneman, Judgment under uncertainty: heuristics and biases. *Science*, vol. 185, Pp. 1124-1131, 1974.

Vallee, J., *Computer Message Systems*. McGraw Hill Publications, New York, 1984.

Wickens, C.D., The structure of attentional resources. In R.S. Nickerson (Ed.), *Attention and Performance VIII*, (pp. 239-255), Erlbaum, Hillsdale, NJ, 1980.

Young, E.T., A proposal for the Telescience Testbed Pilot Program applied to science operations for the Space Infrared Telescope Facility (SIRTF). University of Arizona, submitted to USRA, 1987.