The Effects of Task Interruption on User Performance in a Multitasking Environment: Implications for Computerized Reminders

by

Darlene Merced

Thesis submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Master of Science

in

Industrial and Systems Engineering

APPROVED:

Robert C. Williges, Chairman

Paul T. Kemmerling

Marianne Rudisill

February 1, 1991

Blacksburg, Virginia

c.2

LD
5655
V855
1991
M472

The Effects of Task Interruption on User Performance in a Multitasking Environment: Implications for Computerized Reminders

by

Darlene Merced

Committee Chairman: Robert C. Williges
Industrial and Systems Engineering

(ABSTRACT)

Large and complex operations environments such as the one being designed for the Space Station Freedom will undoubtedly place many requirements on crewmembers to simultaneously perform multiple tasks; i.e., "multitasking." Moreover, it is unlikely crewmembers will finish a task without being interrupted several times. Depending on the nature of these interruptions, crewmembers may be forced to suspend their activities. One danger in suspending activities is forgetting (e.g., forgetting what the activities were, or one's position within the activities, or certain procedures, etc.). Forgetting to complete even a single activity in space may have serious consequences. As a countermeasure to forgetting, the present study introduced the use of computer- based "reminders." The purpose of this study was to: (1) establish a baseline measure of multitasking, (2) examine the effects of interruptions on task performance, and (3) explore the use of computer-based "reminders."

The results revealed subjects could perform three and five simultaneous tasks with a response accuracy of 98 percent. A large portion of that performance was attributed to practice and response pacing. In addition, the interruption task chosen for this study did not affect reaction time or response error performance on the main tasks. The use of a computer-based checkmark reminder proved to be beneficial in reducing menu search time regardless of the number of items to be located. Search time decreased considerably when the reminder was available to subjects as opposed to when it was absent. This study concludes that with practice and pacing, people can perform at least five simple tasks simultaneously. Furthermore, the benefits of incorporating computer-based reminders should not be limited only to multitasking environments.

ACKNOWLEDGEMENTS

The author wishes to extend a special thanks to Professor Paul T. Kemmerling for arranging an unforgettable ten month work tour with NASA's Man-Systems Division in Houston, Texas. I also wish to thank Dr. Marianne Rudisill for allowing me to conduct this study in the Human-Computer Interaction Laboratory and for the advice she provided. This research was supported by the Office of Aeronautics, Exploration and Technology. Also, a special thank you is extended to Dr. Tina Holden of Lockheed. Most of all, I would like to thank my committee chairman Dr. Robert C. Williges for his support and advice throughout the data analysis process.

On a more personal level, I wish to thank several people for providing ten months of continuous moral support. Special thanks to: Tandi Bagian; Nicholas Pausback; Fran Wuamett; Ileana Natali-Lugo; Laurie & Sammie Weaver, for providing a cheerful side to the madness; and my parents for their continuous support. Finally, I would like to express a lifetime of thankfulness to Nathan Moore.

TABLE OF CONTENTS

INTRODUCTION 1
Supporting Technology 1
Multitasking Examples 2
Managers and Office Personnel 2
Pilots and Astronauts 4
Multitasking Limitations 5
Background Information 6
Space Station Freedom 6
Research Domain 8
Research Purpose 10
LITERATURE REVIEW 11
Information Processing 11
Working Memory 13
Components 13
Limitations 15
Attention 18
Selective Attention 19
Divided Attention 21
Spatial Proximity 21
Object Integrality 22
Task Difficulty 22
Multiple Resources 22
Mental Workload 25
Multitasking 26
Systems Approach 26
Psychological Perspective 27
Conscious and Subconscious Control 29
Interruptions 31

Types of Multiple Activities....34
Reminders....36
Characteristics....36
Effective Reminders....38
Design Options....39
Windows....40
Symbolic Cues....40
Spatial Cues....41
System Design Support....42
METHOD....43
Experimental Design....44
Subjects....46
Hardware....50
Software....50
Task....50
Procedural Tasks....50
Interruption Task....58
Display Menu....62
EXPERIMENTAL PROCEDURE....64
Training....64
Part I....64
Part II....65
Test Protocol....71
DEPENDENT MEASURES....73
Objective....73
Subjective....75
DATA ANALYSIS and RESULTS....77
Pre-Interruption....77
GN&C Interruption Task....81
Post-Interruption....81
Menu Search Time....81
Viewing Task-Display....88
Workload Ratings....88
DISCUSSION....95

Multitasking Abilities....................95
Decision Complexity....................99
Pacing....................90
Response Complexity....................100
Interruption Effects....................100
Practice....................101
Reminder Effects....................104
Perceived Workload....................107
CONCLUSION....................**109**
REFERENCES....................**112**
APPENDIX I. Space Station Freedom Conceptual Workstation Mock-up Design....................**118**
APPENDIX II. User Interface Display Unit Design Requirements Specification....................**120**
APPENDIX III. Subject Questionnaire....................**122**
APPENDIX IV. Participant Consent Form....................**125**
APPENDIX V. Modified Cooper-Harper Rating Scale....................**127**
APPENDIX VI. Experimental Instructions....................**129**

LIST OF FIGURES

Figure 1. Human Information Processing Model....12
Figure 2. Working Memory Processing Codes and Associated Display Format....14
Figure 3. Recall and Retention Performance as a Function of Time....17
Figure 4. A Representation of Selective Memory Single-Channel Processing....20
Figure 5. Multiple Resource Processing Model....23
Figure 6. Multiple Activities Classification 35
Figure 7. Experimental Design: 3x2x2 Within Subject Design....45
Figure 8. Functional Use of Macintosh Computers Cited by Subjects....48
Figure 9. Query into Computer Multitasking User Habits....49
Figure 10. Operations Data File -- In-Flight Maintenance Task....51
Figure 11. Environmental Health Facilities Database -- Dynamic Environmental Measurement System Task....52
Figure 12. Fluids Management System -- Fluid Transfer System Evaluation Task....53
Figure 13. Environmental Control and Life Support System -- Water Management Task....54
Figure 14. Propulsion Module Assembly -- Resistojet Power Level Task....55
Figure 15. Communications and Tracking -- TDRSS Communications Link Task....56
Figure 16. GN&C Emergency Interruption Task(not at actual size)....60
Figure 17. Resource, Systems, and Documents Menu Bar....63
Figure 18. Six-Task Display Screen Layout (not at actual size)....68
Figure 19. Miniature Task-Display Screen....69
Figure 20. Average Pre-Interruption Reaction Time Performance by Number of Tasks....97
Figure 21. Average Post-Interrruption Reaction Time Performance by Number of Tasks....98

Figure 22 Pre-Interruption Average Reaction Time Performance by Replication........102
Figure 23. GN&C Interruption Tasks Average Reaction Time by Replication........103
Figure 24. Average Menu Search Time by Reminder........105
Figure 25 Average Menu Search Time by Number of Tasks 106

LIST OF TABLES

Table 1. A Description of the Space Station Freedom Information Systems.....7
Table 2. Guidance, Navigation, and Control System.....59
Table 3. ANOVA for Pre-Interruption Reaction Time.....78
Table 4. Student-Newman-Keuls Test for Pre-Interruption Reaction Time.....79
Table 5. ANOVA for Response Error Pre-Interruption.....82
Table 6. ANOVA for GN&C Reaction Time (emergency interruption).....81
Table 7. ANOVA for GN&C Response Error (emergency interruption)....83
Table 8. ANOVA for Post-Interruption Reaction Time.....84
Table 9 ANOVA for Response Error Post-Interruption.....85
Table 10. Student-Newman-Keuls Test for Post-Interruption Reaction Time.....86
Table 11. ANOVA for Total Menu Search Time.....87
Table 12. Student-Newman-Keuls Test for Total Menu Search Time.....90
Table 13. ANOVA for Time Spent Viewing the Minature Task-Display Screen.....91
Table 14. Student-Newman-Keuls Test for Time Spent Viewing the Miniature Task-Display Screen.....92
Table 15. ANOVA for the Subjective Mental Workload Ratings.....93
Table 16. Student-Newman-Keuls Test for Subjective Mental Workload Ratings.....94

LIST OF ACRONYMS

CRH	Common Resource Hypothesis
CSA	Canadian Space Agency
C&W	Caution and Warning
ECLSS	Environmental Control and Life Support System
ESA	European Space Agency
GN&C	Guidance, Navigation and Control
HCIL	Human-Computer Interaction Laboratory
JSC	Lyndon B. Johnson Space Center
NASA	National Aeronautics and Space Administration
NASDA	Japanese Space Agency
OSSA	Office of Space Science and Applications
RAM	Random Access Memory
SAPPHIRE	Screen Allocation Package Providing Helpful Icons and Rectangular Environment
SSF	Space Station Freedom
SSFP	Space Station Freedom Program
SSIS	Space Station Information System
WMT	Water Management Task

INTRODUCTION

The ability for an operator to perform several computerized tasks simultaneously is becoming more of a reality with advances in hardware and software technology. Unlike the microprocessors of several years ago, computers today have greater capacity for memory storage, real-time processing, and user-oriented interface features (e.g., windows, selectable menu options, icons, etc.). With these new system capabilities, users can simultaneously perform multiple computerized tasks (Billingsley, 1988). User simultaneous interaction with more than two related or unrelated tasks is referred to as multitasking. Other synonymous terms found in the literature are interleaving activities (Cypher, 1986), time-sharing (Wickens, 1987; Wickens, 1984), and concurrent tasks (Boff and Lincoln, 1988). The research effort presented here addressed multitasking, task interruptions, and the utility of computerized "reminders." The main purpose was to document multitasking performance and the effects of task interruptions.

Supporting Technology

Software applications such as Multifinder ™ and Microsoft ™ Windows permit users to open and switch among several applications, depending on the RAM (random access memory) availability. However, having several applications or windows open does not mean the system is multiprocessing. In the computer industry, multiprocessing refers to the computer system's capacity to parallel process multiple command inputs or multiple applications simultaneously. Most personal computers do not support this type of internal computing processing. For the most part, the use of display windows simulates system multiprocessing. However, the lack of true computer-supported multiprocessing does not imply humans will not engage in multiple activities.

Newer operating systems, such as UNIX, support multiprocessing and, therefore, human multitasking in the true sense (Cypher, 1986; Newman, Stephens, and Sweetman, 1985). For example, UNIX will allow users to simultaneously work on a text editing task while reading and answering messages as well as run programs. Usually, this multiple interaction is facilitated with the help of a window-supported workstation, such as the SUN or STAR workstations. Each application or task can be visible to the user as a separate window. In turn, each window may have its own window-management capabilities (e.g., scrolling, resizing, moving, etc.) defined by the system or application parameters.

Multitasking Examples

Managers and Office Personnel

The technological support for human multitasking may be changing the way we work. We will more than likely engage in several simultaneous activities due to an increase of multiprocessing computers in the workplace. Consider, for example, the following hypothetical office scenario: Mr. Brown begins his work day by logging onto his computer to read his daily messages. While Mr. Brown reads, he remembers he has not completed a budget report. Mr. Brown decides to finish the report while the ideas are fresh in his mind and to read his messages later. For the time being, Mr. Brown has suspended his reading task for a more important one. He proceeds to open and work on the report. While working on the report, he is momentarily side-tracked with thoughts about the budget.

In addition to his originally planned activity, Mr. Brown has now engaged in working on several minor subtasks indirectly related to the original report. In the midst of his work, he is interrupted by a message flashing across his monitor. The message is an urgent request from Mr. Brown's supervisor. At his supervisor's request, Mr. Brown suspends all his tasks. After several hours, Mr. Brown returns to finish the report. However, he realizes he cannot remember all the other subtasks. In addition, Mr. Brown has forgotten about his morning computer messages.

This example helps to illustrate the type of multitasking activities found in managerial job settings today. It also exemplifies the need for memory aids or "reminders." In most industries, computers are essential for accomplishing every day work activities. The ease with which information can be manipulated and accessed with computers almost seems to promote or encourage multitasking. Multitasking, however, is not exclusive to computer-supported activities. Certain occupational settings require people to perform multiple tasks simultaneously throughout the day without the necessity for computer assistance (e.g., nurses, switch-board operators, school teachers).

On the other hand, there are occupations which rely heavily on computerized systems as well as the human ability to successfully multitask.

For example, nuclear control room operators, pilots, and astronauts are often required to multitask. In these settings, the margin for error is very small and operators cannot risk forgetting to complete certain tasks or procedures as Mr. Brown did.

Pilots and Astronauts

The multiple cognitive, perceptual, and motor skill task requirements demanded of pilots have been well documented in Wiener and Nagel (1988). A pilot, for example, is required to monitor flight instrumentation (a visual task), control the aircraft (a motor task), listen to incoming air traffic reports (an auditory task), and troubleshoot (a decision-making task). In addition, the complexity of the pilot-aircraft system itself contributes significantly to the number of simultaneous tasks placed on pilots. It is a delicate system, and a breakdown of any components can adversely affect the pilot-aircraft performance. An instrument misreading, for example, may lead to an incorrect pilot decision and, perhaps, a catastrophic accident. This delicate system balance is also observed in the space industry.

The multiple tasks required of astronauts are similar to those of pilots in the sense that they are both required to systematically monitor numerous displays, process information from different sources, and are susceptible to various forms of interruptions (e.g. ground control communications, instrumentation warnings, etc.). Much of the work performed by astronauts and pilots can be classified as procedural-based tasks or checklist-types. Norman (1990) discusses the role of checklists and why they are used. Items on a checklist serve as: (1) reminders, (2) triggers or signals. and (3) status checks. According to Norman (1990), checklist items communicate information to others and also help to minimize future workload by distributing and organizing procedures. Checklists are also used as safety measures. Items are included to ensure system components are functioning properly and within their specified boundaries.

Interruptions are somewhat problematic in that they may divert human attention for long periods of time. Given the criticality and time constraints, operators can be forced to suspend activities at the risk of not properly completing them (similar to Mr. Brown's illustration). Unfortunately, pilots and astronauts do not have the luxury of forgetting to complete even simple tasks. As stated earlier, the margin for error is very small. In both instances, human lives and millions of dollars are at stake.

Multitasking Limitations

Independent of any particular occupation or operating environment, multitasking places great demands on the human: (1) information processing system, (2) available attentional resources, and (3) motor skills. Simply stated, we cannot do everything at once. Task interruptions, in addition, contribute to the demands placed on the human attentional resources and memory. In a typical office environment we find many forms of "reminders" such as post-it™ notes, to do lists, calendars, memos, and office organizers, etc. Some people choose to remind themselves by placing urgent work on top of their desk or in specially marked folders. Others may ask friends to call and remind them. The point is, we cannot *remember* everything there is to do and we often structure objects in our environment to aid us in our tasks. Therefore, we tend to rely on external reminders (Intons-Peterson and Fournier, 1896).

Background Information

A multi-year research effort has been undertaken by the Human-Computer Interaction Laboratory (HCIL) at NASA's Johnson Space Center (JSC) to examine user/system function allocation in a space multitasking environment. The objectives are to analyze the "system" composed of one human and multiple computer applications for the Space Station Freedom Program (SSFP) and related spacecraft applications; propose and evaluate methods for improving human-computer interaction in multitasking; and develop HCI design guidelines for the space multitasking environment.

Space Station Freedom

Space Station Freedom (SSF) is an international partnership among the United States, the European Space Agency (ESA), the Japanese Space Agency (NASDA), and the Canadian Space Agency (CSA). Space Station Freedom represents the first step for long duration manned space missions. When completed, the Space Station will be a permanent, multipurpose facility serving as a laboratory, an observatory, a servicing and assembly facility, a manufacturing plant, and a storage warehouse. The SSF's computer systems will operate under the guidance of three information systems, described in Table 1. Each information system has a primary purpose and several secondary support functions.

Table 1. A Description of the Space Station Freedom Information Systems (edited version, taken from NASA USE-1000, V. 2.1)

---

1. **Space Station Information System (SSIS)** - Provide information management across the Space Station Freedom.

   Secondary Functions

   * Provide information transfer among flight and ground elements.
   * Provide communications link between crewmembers and payload customers.
   * Provide information transfer in a standard, reliable, transparent fashion.
   * Facilitate operations and processing between customers and their payloads independent of geographical location.

2. **Software Support Environment (SSE)** - Provide automated rules and tools to minimize the cost and risk associated with program software development.

   Secondary Functions

   * Provide common environment for software development and maintenance.
   * Provide open access to SSF software development information.
   * Provide program-wide approved standards and methodologies.
   * Minimize cost of software ownership.

3. **Technical and Management Information System (TMIS)** - Provide automated rules and tools to facilitate management of program development.

   Secondary Functions

   * Maximize the effectiveness of technical and management processes.
   * Maximize the effective use of system engineering practices.
   * Facilitate the management of information resources.
   * Provide technical and management interfaces with SSF users.

The current SSF design specifies there will be three computer workstations available to the crewmembers. Each will contain a keyboard, hand controllers for controlling several robotic devices, a cursor control device, an audio and video disk player and recorder, and a printer among other devices (see Appendix I). Both the workstations and the user display unit interfaces will be designed to strict specifications. The user display unit interface design requirements are specified in two documents (NASA, USE 1000 and DR SY-45.1). Appendix II illustrates a few of the interface design requirements, such as the dedicated display bar for caution and warning signals.

Research Domain

Preliminary multitasking research is currently being conducted at JSC's Human-Computer Interaction Laboratory (HCIL) to investigate how window management capabilities affect user performance as the number and type of tasks vary. For example, "How will window management capabilities, such as resizing, overlapping, moving, and dragging affect performance of several simultaneous tasks?" Accordingly, "How will performance change given an increase in the number of display windows as a consequence of increasing the number of tasks?" This type of investigation is only one of many in a sequence of studies designed to investigate human factors related computer interface display issues in multitasking.

The need to study how humans perform multiple computer tasks is twofold. First, and perhaps of immediate value to SSFP, comprehensive design guidelines are needed to ensure the user-interface involving any multitasking activity is adequately designed. The second need for studying human multitasking performance is theoretically based. We need to understand (1) the types of cognitive demands multitasking activities produce, and (2) the effect on crewmember performance in terms of monitoring activities, decision-making, and keeping track of current, awaiting, and suspended activities.

Moreover, the human-to-computer relationship is becoming more of a cooperative interaction as evident by the numerous discussions related to computer supported cooperative work (Grief, 1988). Particularly with the SSF, a crewmember's ability to successfully perform mission tasks will be a team interaction with several computer systems. This team interaction may be a joint effort with other crewmembers or ground control personnel. However, team interaction should not lessen the need for memory or reminding aids. In fact, reminders would probably help to keep team members informed of each other's job progress as it may impact their own duties.

Empirical literature in the area of designing display interfaces for a multitasking environment is practically non-existent and speculative. At best, the literature contains unsupported design recommendations. With the exception of Norman, Weldon, and Shneiderman's (1986) discussion of cognitive processes, very little information is available on how windows should support multiple tasks. In terms of guidelines, Gilmore, Gertman and Blackman (1989) provide designers of complex process control systems with eleven windowing design guidelines. Unfortunately, the reference sources for these guidelines are not all empirically based. The authors state that few comprehensive guidelines are available and that research support for their suggested guidelines is low. There is clearly a need for more applied design research as well as theoretical inquiry involving all aspects of multitasking, not just space applications.

Research Purpose

Complex computer systems such as the SSIS will undoubtedly have many multitasking requirements. Moreover, it is unlikely crewmembers will finish a task without several interruptions. In fact, it is well known that the astronauts on Space Shuttle missions are constantly interrupted. It is also unlikely the number of interruptions on Space Station Freedom missions will be any less than those observed on Shuttle missions. The purpose of this study was to document user performance in a multitasking situation given interruptions and the availability of a computer-based "reminder." The goal of this research was to establish the foundation for future studies involving computerized reminding techniques.

LITERATURE REVIEW

Information Processing

A discussion on multitasking and reminders would not be complete without first reviewing the fundamentals of human information processing and related limitations. A generally accepted model of human information processing is represented in Figure 1 (Wickens, 1987; Wickens, 1984). According to the diagram, a stimulus is transmitted via our sensory modalities directly to our perceptual recognition process. Having identified the stimulus, a decision must then be made as to what type of action to take. A choice is made to respond immediately or to temporarily maintain the information in working memory. Once in working memory, the information may be transferred to long-term memory, reserved for a later response action, or forgotten altogether. Card, Moran, and Newell (1986) describe working memory functionally as the system where "mental operations receive their operands and leave their outputs" (p. 45-7). To complete the human information processing model, a feedback loop is depicted stemming from the chosen response action as new stimuli input to the system and a reservoir of attentional resources is shown extending to each of the processing structures.

The processing activities described in Figure 1 are constrained by the capacity of the various mental operations involved. Wickens (1987) describes the various mental operation capacities in two generic forms. First, each operation has limits on the amount and speed of information that can be

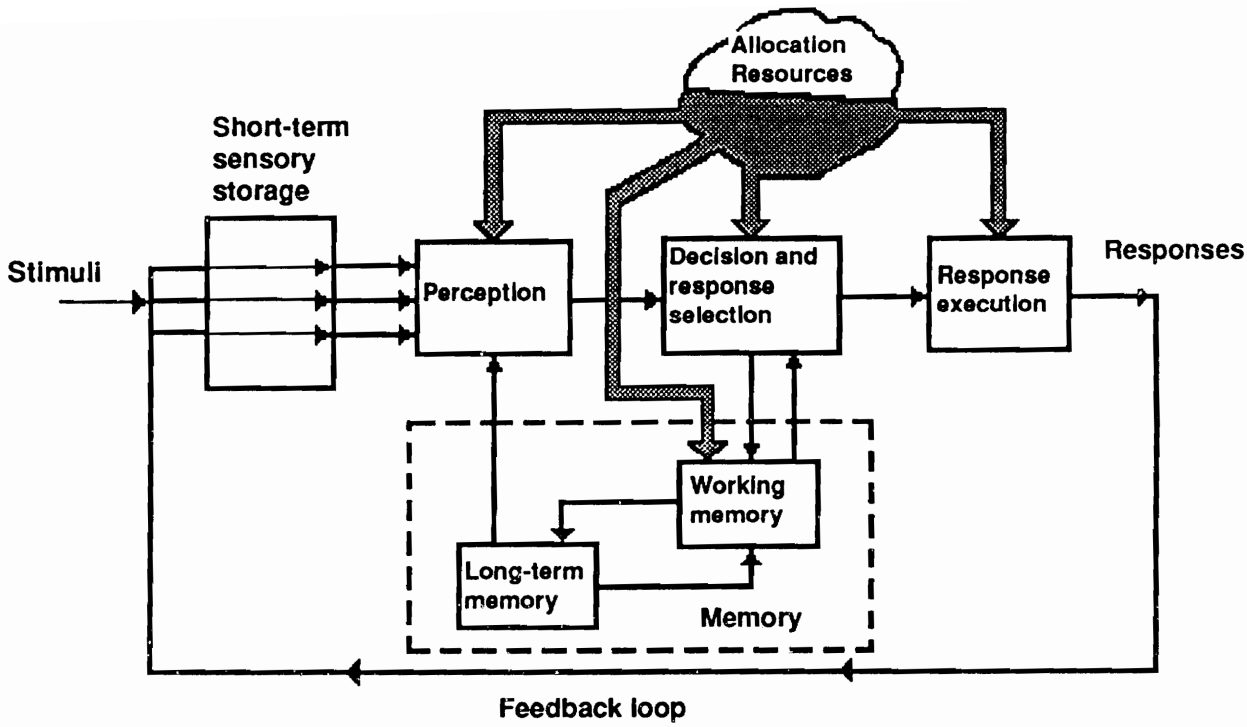


Figure 1. Human Information Processing Model (reproduced without permission from Wickens, 1987)

processed at one time. Second, there are limits to the total attention resource pool or "mental energy" available to the human information processing system. Much of the processing described in the model requires some level of attention to function efficiently, depicted as the cloud of attentional resources in Figure 1. The information processing literature review for this study's purpose focused primarily on the limitations of attentional resources and working memory.

Working Memory

Components

Working memory is believed to be composed of two complementary, yet independent, processing subsystem codes labeled "verbal" and "spatial" (see Figure 2). These two anchor-points represent a continuum along which mental transformations necessary to perform a task occur (Baddelely and Hitch, 1974). The verbal/linguistic-symbolic subsystem is responsible for processing alphanumeric codes, sets of instructions, directions, and logical operations. Conversely, the spatial-analog subsystem is responsible for processing information concerning location, analog transformation, and continuous motion analysis. An interesting point to note is that these two subsystems share general characteristics of rapid information loss and limited storage capacity (Wickens, 1987).

Inputs to working memory may come from several sources: sensory codes, long-term memory, and working memory itself. A common characteristic of working memory, however, is the continuous need for processing resources (Wickens, 1987, p.200). Without some form of attention, the information in

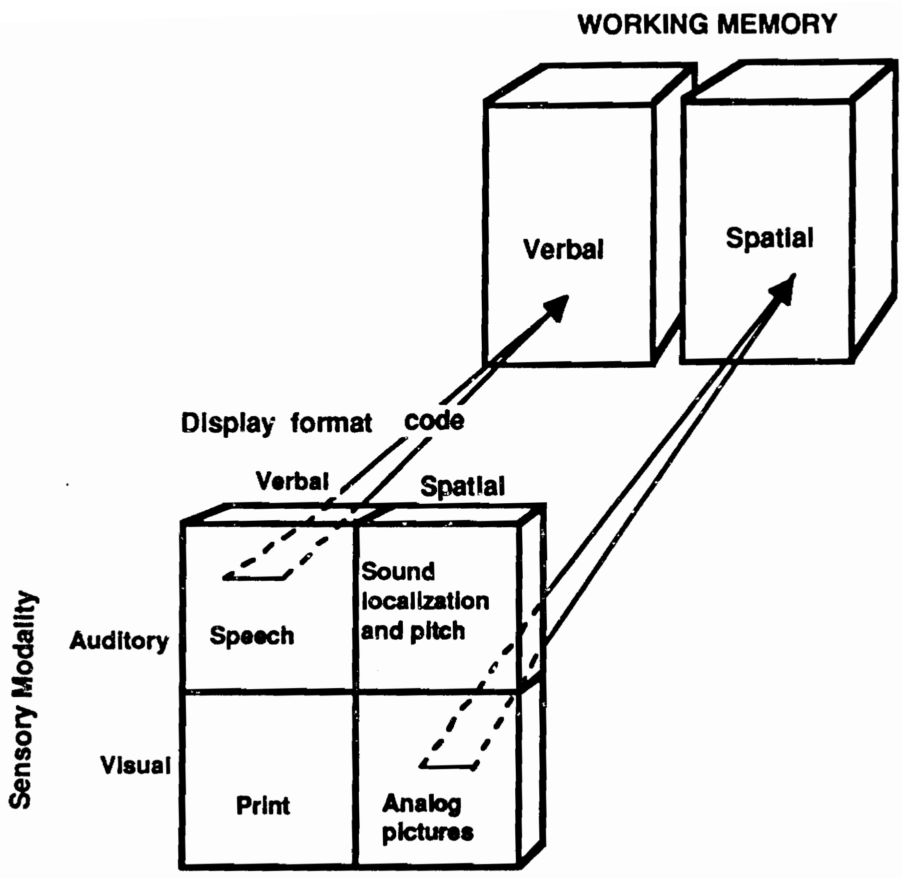


Figure 2. Working Memory Processing Codes and Associated Display Format Codes (reproduced and without permission from Wickens, 1987)

working memory will degrade. The importance here is not only in the implications for multitasking capabilities, but also for designing computer interface displays and reminders to compensate for some of these limitations. The display format codes, shown in Figure 2, associated with the two working memory codes, provide a basic guideline for designing information displays as defined by the input modality (auditory or visual). According to the figure, the optimum format for the respective memory code would be verbal-speech and spatial-analog.

Limitations

Because it is possible for working memory to receive input from sensory channels via the sensory codes, much of the limitation attributed to short-term memory has also been associated with the working memory's limitations. A classically cited example is rehearsing a new phone number for immediate use, which is quickly forgotten thereafter. Much of the work addressing working memory limitations has come from research on iconic and echoic code interference, insofar as these codes influence the inputs to working memory.

Unlike long-term memory, the two components or codes comprising working memory are subject to interference based on the similarities between concurrent activities. The codes themselves have been shown to have limited capacity for storing large amounts of information at any given time, thereby contributing to the working memory's limitation. Information held in working memory without rehearsal will most likely be forgotten within 10-30 seconds (referred to as the Brown-Peterson paradigm--findings summarized in Adams, 1980 and Wickens, 1987). These limitations are demonstrated by graphically plotting performance as a function of time. Figure 3 depicts the hypothetical relationship between recall performance and retention interval as the number of items held in working memory increases. The general trend observed in these curves is that of decay. Not surprisingly, increasing the number of items held in working memory will result in faster decay (represented by curves *b* and *c*). In terms of recall performance, a decaying function is indicative of some limitation within the working memory process.

Miller's (1956) magical seven plus or minus two recall span has been generally regarded as the working memory's limitation for storing unrelated items. However, in a real work environment, seven plus or minus two may in fact be an optimistic estimate. More recent investigations seem to indicate the number is closer to three or four items (Broadbent, 1975; Chase and Ericsson, 1981).

The associated limitations of our working memory are not strictly limited to short-term span memory retention intervals less than 30 seconds or seven plus or minus two unrelated items. Research has shown that with rehearsal and performance strategies--chunking and mnemonic techniques--items can be retained for longer periods of time (Bower, 1972; Chase and Ericsson, 1981). Chase and Ericsson (1981) demonstrated that with practice and time their subject had managed to increased his recall span from 7 digits to 80 digits. Although impressive, it should be noted that rehearsing and similar mnemonic techniques are still competing with other cognitive and perceptual activities for limited attentional resources.

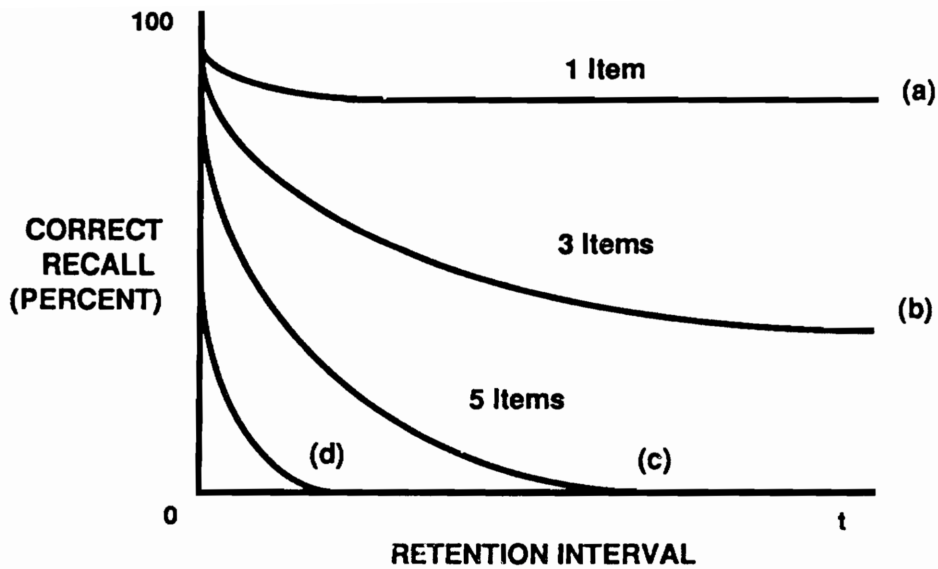


Figure 3. Recall and Retention Performance as a Function of Time (reproduced without permission from Wickens, 1987).

The assertion that short-term memory limitations are fundamentally related to the working memory's information processing has been disputed by some researchers. Klapp, Marshburn, and Lester (1983), coined the term "common resource hypothesis (CRH)" to reflect the assumption that the limit in short-term memory span (seven plus or minus two chunks) is equivalent to the limit in working memory space. In a series of experiments, the authors denied the CRH seven-chunk capacity measures as being the working memory processing space. The authors did not, however, deny the assumption of a limited-capacity working memory. The controversy of "how much" is available for working memory information processing may never really be resolved. Nevertheless, researchers agree there is a limit to working memory. Furthermore, this limit may be influenced by the types of attentional demands being placed on the working memory. The following section will discuss attentional demands as they apply to working memory and multitasking.

Attention

Thus far the discussion has centered around the limitations of working memory as defined by short-term span. The following section will address how the different roles of attention and attentional resources interact with working memory and an operator's ability to perform multiple tasks.

Conceptually, attention refers to the global resources available as well as the limitations of these resources to time-share activities and process information in parallel (Wickens, 1987 and Wickens, 1984). Wickens describes three general characteristics of attention: selective attention, divided attention, and related mental workload issues. Selective and divided attention reflect two differentiated hypotheses of how attentional resources are distributed among tasks. Selective attention emphasizes a single reservoir of attentional resources, while divided attention postulates separate reservoirs. Each hypothesis, however, is in agreement with respect to the finite capacity of mental resources available from the attentional reservoir.

Selective Attention

Selective attention refers to a person's ability to fixate and process one source of information at a time. In essence, a person is distributing his or her resources and acting as a single-channel processor (see Figure 4). For instance, an industrial control room operator may use selective attention to monitor the most critical panels of instruments under emergency situations. The operator is essentially "tuning out" what is perceived as irrelevant information.

Selective attention requires that operators switch or alternate their attention between various "events." In the process control room setting, an operator is sampling the environment for expected or unexpected event conditions (Sheridan, 1988). Unfortunately, the limitation of our working memory renders sampling an imperfect technique warranting external "reminding aids" as summarized in the following quote (taken from Wickens 1984, p. 252):

> ...people tend to sample information sources more often than they would need to if they had perfect memory about the status of an information source when it was last sampled. Also because of limits of memory people may forget to sample a particular display source if there are a

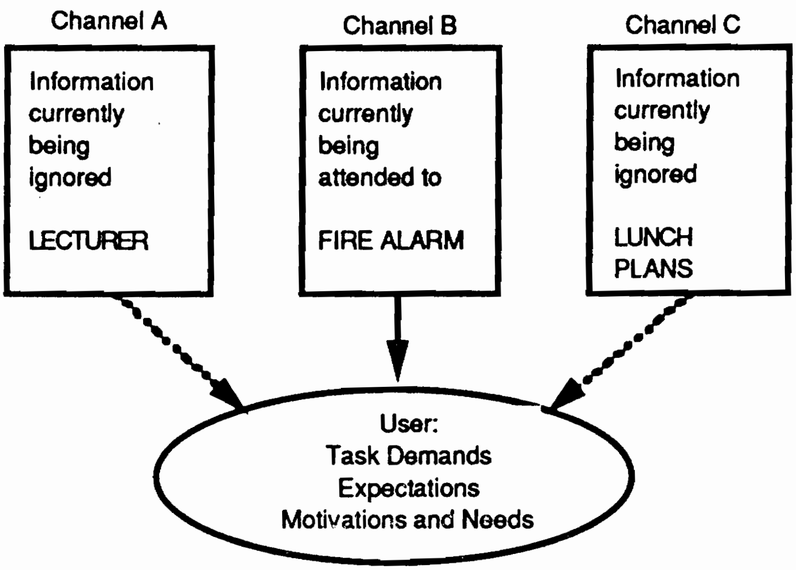


Figure 4. A Representation of Selective Memory Single-Channel Processing

> multitude of possible sources. This, for example, might well be the case for the monitor of a nuclear process control console. These limitations in memory suggest the usefulness of computer aiding to present "sampling reminders" (Moray, 1981).

The important issue here is that our working memory and selective attention abilities, remarkable as each may be, do have limitations requiring system "reminding" support. In a multitasking situation, an operator will more than likely be required to do more than selective event-sampling. For instance, an operator may be required to make critical decisions based on a number of informational sources. To that extent, an operator engaged in multitasking will probably divide his or her attention among the various activities.

Divided Attention

In contrast to selective attention single-channel processing, divided attention refers to a two channel parallel processing capability. The most obvious example of parallel processing is an automobile driver's ability to focus on the main road yet process peripheral cues that provide velocity information. Parallel processing, however, is not always perfect. The ability to divide ones attention may depend, to a large extent, on four factors: spatial proximity, object integrality, task difficulty, and multiple operations resources (Wickens, 1984). The underlying assumption is that divided attention is still dependent on the availability of limited resources.

Spatial Proximity

The ability to perceive visual stimuli is associated with their spatial location relative to other stimuli in the visual field. Thus, by moving two stimuli closer together within the visual field, the more likelihood there is for parallel processing. However, visual spatial proximity will not always guarantee parallel processing. In fact, extreme spatial proximity may cause confusion within a sensory modality as well as visual clutter on a display.

Object Integrality

Using various single-object attributes, such as color, size, and shape, is likely to facilitate parallel processing. The various attributes combine to form a single multidimensional integral object which is processed much more rapidly by our perceptual recognition system. This is particularly beneficial if multiple variables must be compared and processed together, but not if the variables are to be processed independently (e.g., compared or related to each other).

Task Difficulty

The level of task difficulty will in most instances affect the demand for attentional resources and operator performance. If the demand for additional attentional resources is met, then performance will be less likely to deteriorate. Conversely, if the amount of resources cannot be met with increases in task difficulty, performance is almost guaranteed to deteriorate.

Multiple Resources

Multiple resource theory argues that human's have resource capacities which can be defined along three dimensions (Wickens, 1984). The three dimensions illustrated in Figure 5 are: (1) processing stages, (2) sensory modalities, and (3) processing codes. The processing stage dimension is divided according to the underlying functionality difference between the responding mechanisms and the other variables. Time-sharing of tasks is better accomplished by cross-modal channels (e.g., eye and ear). Treisman

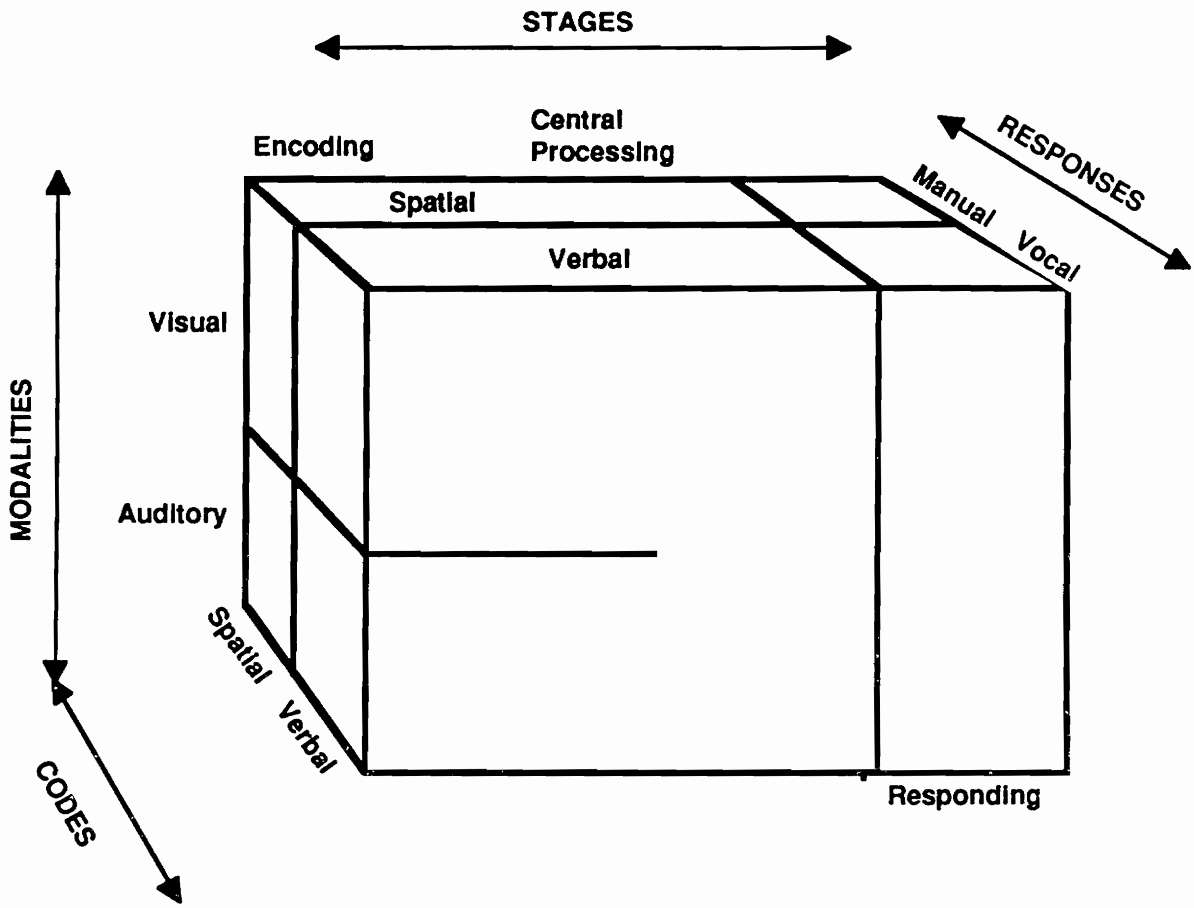


Figure 5. Multiple Resource Processing Model (reproduced without permission from Wickens, 1987)

and Davies (1973) found subjects increased their ability to divide their attention across modal channels. However, cross-modal divided attention was still less effective than focused attention. The last dimension, processing codes, affects the degree of interference between the working memory's spatial and verbal subsystems. Numerous studies supporting this multiple resources model are summarized in Wickens (1987 and 1984).

The underlying theorem is that, since not all tasks compete for the same "undifferentiated pool" of resources, an operator can time-share various activities. The multiple resource theory postulates different types of tasks may tap into different resource reservoirs. Tasks which rely on the same reservoirs for processing will cause interference. On the other hand, if tasks demand different resources on any of the three dimensions, time-sharing will be easier. In turn, changes in task difficulty will be less likely to disrupt performance in another task. Therefore, the extent to which tasks compete for, or demand, the same resources will determine an operator's ability to time-share effectively. Hence, this model would seem to support multitasking activity insofar as the tasks do not interfere with each other.

The general agreement with respect to attentional resources is that these resources are somehow limited in their capacity. In a recent study, Hirst and Kalmar (1987) supported the notion of several distinct resources, the total number of which was unknown. The authors proposed that, in addition to the three multiple resource theory dimensions, attentional resources should be further differentiated along a semantic dimension. This added dimension would characterized dual-task interference in terms of the semantic similarities between competing messages as opposed to the similarity of central processing. Another view of divided attention suggests the difficulty to divide one's attention may be a function of an individual's level of skill and not due to a fixed number of resources available. Hirst and Spleke (1980) found that, with practice, subjects could learn to divide their attention between a reading task while writing dictated messages. Subjects in their experiments were able to extract the meaning of the dictated message and integrate semantic information over a series of sentences. The authors acknowledge that, at first, it may be difficult to do two things at once. They note that performance is constrained only by "obvious peripheral factors," such as visual acuity, joint articulation, and the overall mechanics of effectors. With practice, they claim, "...people can learn to do indefinitely many things indefinitely well (p.114)." Their conclusion should, however, be qualified to situations that require simultaneous reading and writing activities.

Mental Workload

Mental workload, per se, is not a form of attention. Mental workload has been described as a multi-dimensional construct related to the proportion of resources expended in the performance of a particular task. Mental workload is assumed to be caused by increases in task demands which inversely affect attention and, consequently, multitasking. In addition, mental workload is assumed to create mental effort, thereby affecting the quality of operator performance. There are many techniques available which presumably measure workload (e.g., pupil dilation, event-related brain potentials, muscular activities, and subjective measures).

Increases in workload will cause a decrease in the amount of attention given to each task. Hence, there will be a perceived increase in mental workload and, ultimately, task performance will suffer. Tulga and Sheridan (1980), for example, using a queuing theory approach found that multitasking performance was a monotonically decreasing function of task workload. For this particular study, task workload was defined as time required to do an average task divided by the time affordable to do an average task.

Multitasking

Systems Approach

Since the 1970s, much of the multitasking work has been in mathematically modelling human decision-making, allocating functions between humans and computers, and graphically describing complex system monitoring performance (Greenstein and Rouse, 1982; Henneman and Rouse, 1986; Rouse, 1977; Tulga and Sheridan,1980). Rouse (1977) used methods of queuing theory and various decision-making variables to conduct a FORTRAN simulation that determined the affect of multitasking decision-making on system performance. His purpose was to determine the allocation of decision-making responsibilities between the human and computer. He concluded that the allocation of multitasking decision-making responsibility is situation dependent. An important statement made by Rouse (1977) related to the development of methods that would improve the communication between the human and computer. In his view, the problem was that the human-to-computer system did not know what the other was doing at any given moment.

An earlier study investigated people's ability to "keep track of several things at once" (Yntema, 1963). Yntema varied the number of attributes an object possessed and compared recall performance. In one experiment, subjects were given a message describing the attributes of one object at a time. Over a period of time the object's attributes would change. Subjects were required to keep track of the object's changing attributes. Then, subjects were interrupted and asked to recall the current state of a particular object. The results showed that recall performance was better when an object possessed many attributes versus having many objects with few attributes. Performance was observed to be much better when each attribute had its own unique set of states. Overall, Yntema found that: (1) people's capacity to store random information was low--that mistakes are made when keeping track of two or three at once, and (2) recall performance does not always improve when a variable follows a predictable sequence of states.

Psychological Perspective

Despite the proliferation of literature with respect to attention and multiple resource theory, little is still known about multitasking abilities and the underlying cognitive processes. The majority of the psychological experimental studies have investigated dual-channel task performance across modalities emphasizing recognition of patterns, words, and auditory signals. It has even been suggested that multiple task performance may be a function of individual differences. According to Damos and Smist (1983), dual-task performance can be attributed to three individual response strategies: (1) simultaneous response, in which the responses to two stimuli lie within some arbitrarily small interval (usually less than 100 ms); (2) alternating strategy, in which alternating responses are made between tasks that exceed the simultaneous response strategy interval; and (3) massed response strategy, in which two or more responses are made consistently to one particular task before switching to another.

Damos and Smist (1983) state that individuals do not simply select response strategies at random. In a series of experiments, a subject's natural response strategy was identified and classified according to one of the three strategies described. Subjects were then given a dual-task to perform, first using their natural response strategy and, secondly, using a different strategy. If the selection of response strategy were truly random, forcing subjects to adopt a different strategy would not have affected performance. However, if the response strategy represented some fundamental individual difference in information processing, large disruptions in performance would be observed. The results showed not all selection of response strategy was random. Subjects who naturally adopted the mass response strategy performed poorly when forced to adopted the simultaneous strategy. Interestingly enough, subjects who naturally adopted the alternating or simultaneous strategy performed equally as well with the other strategy, which suggests the selection between these two may be somewhat random (Damos and Smist, 1983). In fact, alternating response subjects performed better when forced to adopt the simultaneous strategy.

On a different note, Miyata and Norman (1986) provide us with what they call an "approximate theory" of the relevant psychological issues involving multiple activities. Much of the basis for their theory is related to the psychological issues discussed in the information processing section of this literature review (memory and attention). However, Miyata and Norman discuss other areas which have not been addressed thus far, such as conscious and subconscious control of activities, interruptions, and types of multiple activities.

Conscious and Subconscious Control

There are some activities which are so well learned that little conscious effort is required to accomplish them simultaneously (e.g., driving and talking). Well learned skills are typically said to be controlled by a subconscious mechanism. But, not all activities are second nature. Some activities require selective attention and conscious processing. Conscious control has limitations, particularly because it relies on the working memory's resources, whereas subconscious control mechanisms do not appear to use these resources (Miyata and Norman, 1986).

Conscious behavioral control can (but does not always) affect a person's ability to multitask. Conscious control may severely limit the amount of resources available to perform other activities. A task may be so novel or complex that it would demand greater attention, thereby preventing simultaneous interaction with other activities. A detailed theoretical discussion of conscious-driven behavior is provided by Shiffrin and Schneider (1977).

Shiffrin and Schneider (1977) described conscious controlled behavior as related to detection, search, and attention. According to their two-processing theory, attention and memory scanning (search and detection) reflect two processes: automatic detection and controlled searches. Automatic detection involves well learned routines in long-term storage. For the most part, these well learned automatic routines do not require conscious effort or attention.

More importantly, automatic processing facilitates a person's ability to divide his or her attention, whereas controlled searches hinder this ability.

Controlled searches are deliberate sequences of processing steps that place large demands on short-term memory and which require some level of conscious effort and attention. Controlled searches are serial in nature and are presumably initiated and terminated by the person. Controlled searches are also affected by cue similarities. Items from the same-category take much more time to detect and produce more errors than mixed-category cues. Detecting a house in an unfamiliar residential neighborhood is much more difficult, for example, because the surrounding cues are all homes (same-category). Therefore, in order to detect a specific house, a person would need to control his or her search by concentrating on specific details or instructions. On the other hand, locating a corner store in the same neighborhood versus a home would be much easier to detect because it is considered a mixed-category item.

In the above same-category and mixed-category example, the person is still utilizing a controlled search process due to the unfamiliar neighborhood. This novel situation will limit the number of activities the person will engage in simultaneously. Essentially, attention will be focused only on locating the house or corner store. However, as the neighborhood becomes more familiar, the person will find it easier to engage in conversations, listen and tune the radio, and enjoy the scenery with little to no conscious effort. Given time, navigating through the neighborhood will become routine--an automatic detection process. It should be noted that the transition between controlled processing and automatic processing is not instantaneous. Shiffrin and Schneider (1977) state, "automatic and controlled processes can proceed in parallel with one another " (p. 161). The authors proceed to state that controlled processing is commonly used to initiate automatic processing, particularly in complex processing situations.

With the exception of subconscious control, Miyata and Norman's (1986) analysis of conscious control share some similarities with Shiffrin and Schneider's (1977) theory. However, Miyata and Norman explicitly state that conscious control is primarily used in four situations: (1) while performing novel or ill-learned tasks, (2) when tasks are perceived to be difficult, dangerous, or critical, (3) while an operator overrides an automatic system, and (4) during interruptions and conflict resolution.

Interruptions

The fact that most activities are rarely completed without interruptions suggests that people may forget to finish some tasks. Miyata and Norman (1986) state, "...one suspends work on the current activity at the risk of losing track of the current activity by failing to resume the work where it was interrupted" (p. 268). There are basically two forms of interruptions, internal or external. Internal interruptions are provoked by our own thoughts. These may be new ideas or sudden revelations which divert attention away from the current activity. External interruptions, on the other hand, are produced by environmental or situational events (e.g., telephone ringing, activated alarms, etc.). In the SSF environment, for example, external interruptions may be computer-based warnings or ground control communication uplinks.

Miyata and Norman (1986) describe two types of processing modes that are affected by interruptions, task-driven processing and interrupt-driven processing. A person is said to be engaged in a task-driven processing mode when he or she is primarily attending to a single task and ignoring all other events. Sensitivity to external events is selectively decreased. In some instances, an external event may even go unnoticed. In other instances, an event may be noticed but its meaning or content is not fully processed. Task-driven processing is similar to Shiffrin and Schneider's (1977) controlled search processing in the sense that attention is being focused on one item or task at a time. In the task-driven mode, interruptions have little impact on the current activity. The authors write, "...if people become too engrossed in the task to which they are paying conscious attention, they will not process other events that occur on the computer screen" (p. 269). In contrast to the task-driven processing mode, a person engaged in an interrupt-driven processing mode is more receptive to interruptions and changing activities. Interrupt-driven mode processing would, therefore, most likely facilitate multitasking when compared to processing under the task-driven mode.

A recent study by Gillie and Broadbent (1989) helps to clarify why some interruptions are disruptive. In a series of experiments, the effects of an interruption's time length, similarity to the main task, and the interruption's degree of complexity were examined. Subjects were required to memorize a list of either five or seven items as well as the location of those items in the contents of a computer game. For the first experiment, the length of the interruption task--an arithmetic task--was 30 seconds. The results showed no significant main effect of interruption, nor did the results show a significant memory load for either the five or seven item list condition. There was, however, a significant effect for the amount of help requested after the interruption. The authors concluded that short, simple, and dissimilar interruptions do not have a disruptive effect on performance. In view of these findings, the authors conducted a second experiment which increased the interruption task length to 2.75 minutes. Once again, there was no significant effect of interruption or interaction with pre- and post-interruption performance.

A third experiment was conducted which eliminated the subject's opportunity to rehearse list items before the interruption. In addition, the interruption task was changed from a dissimilar arithmetic task to a processing-intensive free recall task. The length of the free recall interruption task remained 2.75 minutes. The results for the third experiment showed a significant main effect of interruption, memory load (five and seven list items), and pre- and post-interruption performance. However, the interaction between interruption and task performance before and after did not influence performance.

A final experiment was conducted that varied the complexity involved in processing the interruption task while allowing rehearsal of the list items to occur. An arithmetic task was again used but, instead of using numbers, letters were used to represent certain numbers. The average interruption time interval was 51 seconds (14.8 standard deviation). The results showed that having an opportunity to rehearse did not guard against an interruption's disruptive effect. The final conclusion reached by the authors was that the similarity of the interruption task to the main task would not always prove to be disruptive--that task complexity was another factor which influenced task performance.

Types of Multiple Activities

The discussion thus far has laid the psychological groundwork for describing Miyata and Norman's (1986) classification of various multiple activities and the controlling actions (see Figure 6). Miyata and Norman classify activities as being either under *current* control or *suspended* control. Current activities are the ongoing primary activities. Suspended activities are all those activities not under current control. They are relevant to the activity but do not control performance.

Current activities are further classified as "*foregrounded*" or "*backgrounded*" activities. Foreground activities are under our conscious-attentional control, typically reserved for the primary activity. On the other hand, background activities are active tasks that receive little conscious attention. They are somewhat automatic. Background activities can be external (tasks performed by another person or computer) or internal (within the person). Internal background tasks are so well learned that they require minimum conscious control and attentional resources. Thus, foreground and background activities allow other tasks to be performed simultaneously.

Of interest for this present study is Miyata and Norman's classification of suspended activities and their implications for designing a multitasking computer interface with "reminder" functions. Activities may be suspended for many reasons, such as interruptions (external or internal), lack of or need for more information, fatigue, and boredom. Hence, suspended activities are those activities which are "awaiting the appropriate time to resume execution" (p.273).

Bannon and his colleagues (1983) have observed many instances of users suspending activities for other tasks similar to Mr. Brown's scenario.

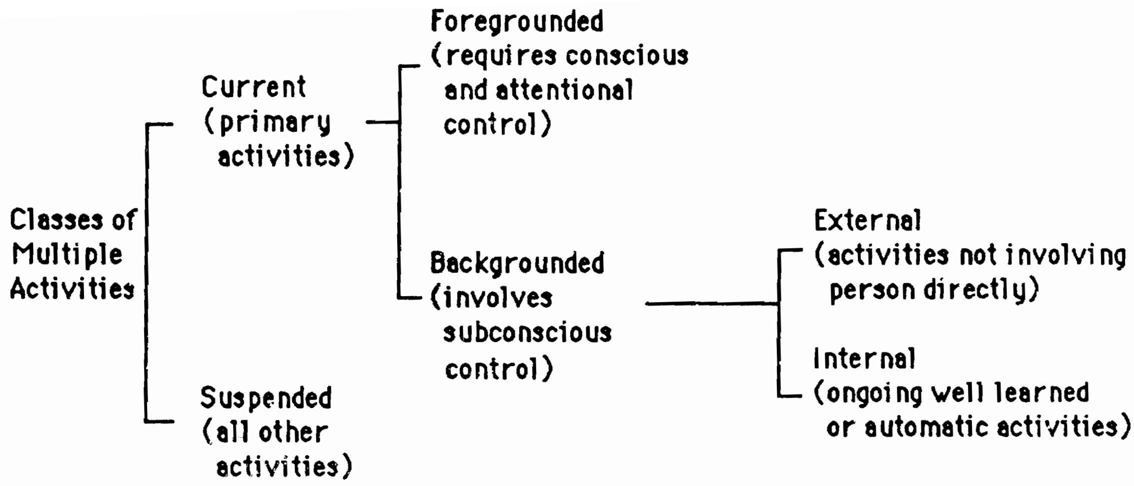


Figure 6. Multiple Activities Classification (Miyata and Norman, 1986)

There are many forms of external reminders available to us, such as notepads, calendars, and other people, as well as internal memory aids (Intons-Peterson and Fournier, 1986). For any given task or set of tasks, we may use several of these reminding techniques. In general, reminders help reduce the working memory workload. The use of reminders presents two main concerns: (1) how to most effectively remind users there are suspended activities, and (2) how to best facilitate resumption. Cypher (1986) discusses the cognitive necessity for helping users remember or keep track of their activities. Similar statements are found in Card, Pavel, and Farrell (1987), Card and Henderson (1987), and Bannon, et al (1983).

Reminders

Characteristics

As stated above, reminders are needed when activities have been suspended and must be resumed at a later time. According to Miyata and Norman (1986), there are two aspects of a reminder which may be independent of each other. First, reminders may be used as signals or indicators to represent an activity, for example, the auditory signal in most cars to remind passengers to fasten their seat belts. Second, reminders can also be descriptive in that they help persons retrieve the activity to be remembered by relaying descriptive cues or information. A photograph of a messy desk is a description that may remind a person to clean or organize the desk. Intons-Peterson and Fournier (1986) found that, in addition to being preferred over internal (within self) aids, subjects classified external memory aids as being more dependable, accurate, and easier to use.

A reminder may succeed in one aspect but fail in the other. Miyata and Norman provide an example of this using the UNIX computer system. UNIX allows users to suspend tasks by way of a "stop job" system command. When users attempt to logoff the system, they are confronted with a message stating, "There are stopped jobs." Although the message succeeds in signalling or indicating suspended tasks, two problems were noted. First, the "stop job" message was only displayed during th logoff process--not during task execution. Second, the message failed to provide descriptive information as to what the incomplete tasks were.

Findlay, Davies, Kentridge, Lambert, and Kelly (1988) conducted a pilot study to investigate the optimum computer display arrangement to present visual reminders and the tradeoffs involved. They used a text editing task and experimented with various ways of reminding the user which editing mode they were using (e.g., insert versus overtype mode). The reminder, in this case the cursor status information, was either placed within close proximity relative to the task's central viewing location or in the display periphery. Unfortunately, no significant difference in performance was found between the visual reminders. However, subjective reports indicated that subjects felt "lost" without the cursor status information.

Ross (1984) investigated a different form of reminding. He focused on the influences that reminding as a thought process had on learning a cognitive skill. He used previously learned "content" as reminders for new and similar test situations. The hypothesis was that subjects would use the same strategy in the test condition that was used in earlier learning episodes because the "context" would serve as a reminding agent. Three separate experiments were conducted and the author concluded that remindings have predictable effects on performance.

In general, the results for the first experiment showed that the earlier learned text content affected the choice of editing method used in the test condition. When remindings occurred, the test method chosen was consistent with the initially learned context. Remindings were also shown to be strongly related to observed task difficulty. In addition, practice appeared to reduce the effectiveness of remindings; however, this issue remained inconclusive. In the second experiment, the appropriateness of the reminding material (i.e., appropriate, unrelated, and inappropriate reminding) is varied to show that reminding can help as well as hinder performance. Being reminded of appropriate learning episodes helped performance, whereas, inappropriate reminding hindered performance. Finally, the third experiment demonstrated that the effects of remindings are subject to interference when irrelevant features are introduced at the problem-solving stage.

Effective Reminders

Given the literature on memory, attention, and multitasking, it seems that an effective visual reminder would be one that: (1) is visible and accessible at all times to the user, (2) is distinguishable from other forms of display signals and cues, (3) provides the necessary information to trigger the "reminding effect", and (4) allows some degree of user control. More importantly, reminders should not be disruptive in that their presence will prevent users from accomplishing other activities. In addition, reminders should be used with discretion in order to preserve their effectiveness. The use of too many reminders may decrease their value.

Design Options

There are various design and presentation techniques available for developing computer supported reminders. The following are some design possibilities: display windows; auditory or visual signals; symbolic or spatial cues; historical information or back-up views; color; and animation. The following subsections address several of these "reminding" design options, namely display windows, symbolic cues, and spatial cues. The discussion is not intended to be an exhaustive treatment involving the design of computerized reminders. More so, the purpose of this discussion is to describe existing human-computer interface design techniques and to provide some preliminary feasibility analysis. In the future, the goal of this research effort will be to systematically and experimentally evaluate these design options having established user performance given interruptions and no reminders. It should also be noted that there is no empirical evidence supporting the use of these techniques as reminders or when they should be used. The value of this study's results will provide basic knowledge (in terms of performance degradation) for deciding when reminders would be most beneficial.

<u>Windows</u>

In general, display windows allow multiple interactions with a variety of types of information. A display window may be used as a reminder in two ways. First, information contained in windows may be used as the reminding agent. Second, the display window itself may be a source of reminding due to the inherent characteristics of the window (i.e., shape, size, etc.). Perhaps the most effective reminder would be a combination of these two methods. This would be a dedicated window containing all the reminders or reminding information. This design would serve as a reminder indicator in addition to providing descriptive information regarding the nature of the suspended activities.

Although an attractive option, display windows are not without their problems. Issues such as window management and the use of overlapping versus tiled windows are still unresolved and very much task-specific (Bly and Rosenberg, 1986; Gaylin,1986). One potential disadvantage in using display windows it that they create visual display clutter (a subjective concept), thereby, reducing the effectiveness of the "reminder window".

Symbolic Cues

Symbolic cues such as icons may be an effective means of reminding, provided they are meaningful. Roger (1989) describes four forms of icons: resemblance, exemplar, symbolic, and arbitrary. Resemblance icons depict the underlying referent by the use of analogies (e.g., folder icons). Exemplar icons represent a general class of objects (e.g., a printer icon). Symbolic icons depict the underlying referent using higher levels of abstraction (e.g., the Macintosh trash can icon). Arbitrary icons have no prior associations to any referent.

A single icon may provide multidimensional information. For example, in SAPPHIRE (Screen Allocation Package Providing Helpful Icons and Rectangular Environment) an icon may contain two or six individual pieces of information related to a window or processing activity (Myers, 1984). One of the disadvantages with using icons is that they may take up considerable display space. In addition, choosing the appropriate "reminding" iconic representation may be difficult. In fact, designing icons is still more of an art than a science.

Spatial Cues

Spatial location cues may be useful reminders provided the location is predefined. If reminders are free to vary within the screen display, the user will have no frame of reference. Without predefined reference locations, the reminder's effectiveness is likely to decrease. Spatial cues have been explored and used for data management purposes. Dumais (1988) summarizes several studies pertaining to textual information retrieval using spatial cues for large databases. The results indicated marginal improvements in retrieval of textual information using location cues.

It seems spatial cues would have some beneficial effects as reminders (perhaps marginal) if the analogy between information retrieval and reminding is accepted. This is not to imply the analogy is perfect. There are underlying differences between information retrieval and reminders. One obvious difference is that the person, in one case, is actively seeking to retrieve information and in the other case he or she is being reminded to retrieve forgotten or suspended information. Spatial reminding cues, however, could serve these two distinct purposes. For example, a spatial location may be assigned to a specific window which would represent a certain type of information, such as all books pertaining to space flights. In addition, the same spatial cue may trigger a reminding effect to return those books to the library. This example is similar to the desktop or office metaphors (i.e., using special folders or reserved desktop space for "work-to-be-done").

System Design Support

According to Cypher (1986) and Bannon et al. (1983), the ideal system support for a multitasking environment would treat related user-activities as independent self-supporting workspaces. Each workspace would provide visual and contextual cues to reorient the user. Therefore, a person resuming an activity would know where he or she was prior to suspending that particular activity. Under these conditions, reminders would have to be program, application, and workspace independent. Recall that, in order for reminders to be effective, they must be visible to the user. Hence reminders should not be designed or placed in locations or applications that would be hidden from the user via the use of standard exiting, quitting, and closing procedures.

METHOD

Unfortunately, very little empirical data are available regarding user performance in computerized multitasking activities, or in the use of computer-based reminders. This study was a baseline investigation that examined task and environmental multitasking performance issues. This experiment was designed to investigate: (1) a person's ability to monitor and perform several simultaneous activities before and after interruption, and (2) a person's ability to reestablish those tasks after being interrupted both with and without computerized reminders. The objective was to compare and document performance between a reminder and a no-reminder multitasking condition.

Subjects were essentially required to perform several simultaneous tasks. They were then interrupted, at which point they were required to perform an emergency task. After the interruption task was completed, the subjects were required to reestablish and complete the tasks they were performing prior to the emergency. The hypothesis postulated in this study was that performance would degrade when the number of simultaneous tasks increased. In addition, the total length of time to reestablish the tasks after the interruption would be greater without the aid of a computer-based reminder. The experiment was designed in such a way as to always force subjects to suspend their activities in order to perform the interruption task. This technique allowed the effects of multitasking, task interruption, and computerized reminders to be assessed equally often.

Experimental Design

The experimental design was a 3x2x2 (Number of Tasks x Reminder x Replication) complete within subject factorial design shown in Figure 7.

Number of Tasks referred to the number of simultaneous activities subjects would be performing (1, 3, or 5) from a selection of six different tasks. These three levels provide: (1) a baseline measure--multitasking data of this nature was not found in the literature, (2) a general representation of the number of tasks most often performed, and (3) a midpoint value which aids in describing multitasking performance. The single task condition served as the controlled baseline comparison level. The total length of time subjects would spend per treatment condition was the main consideration for not including all the possible number of tasks subjects could perform up to five (i.e., 1, 2, 3, 4, and 5). The inclusion of all these factorial levels in the experimental design would have increased the number of treatment combinations and the experimental time sessions by a considerable amount. Hence, in order to reduce the problem associated with most within subject experimental designs (subject fatigue) the Number of Tasks factor was limited to the three levels shown in Figure 7.

Reminder referred to the availability of a computer-based checkmark reminder. On half the trials, a reminder was available for use by subjects and on half the trials no reminder was available. A simple checkmark embedded within a hierarchical menu structure was selected for initial investigation to determine the extent of its benefit in reducing menu search time performance. When available, the checkmark would appear next to the task's abbreviated

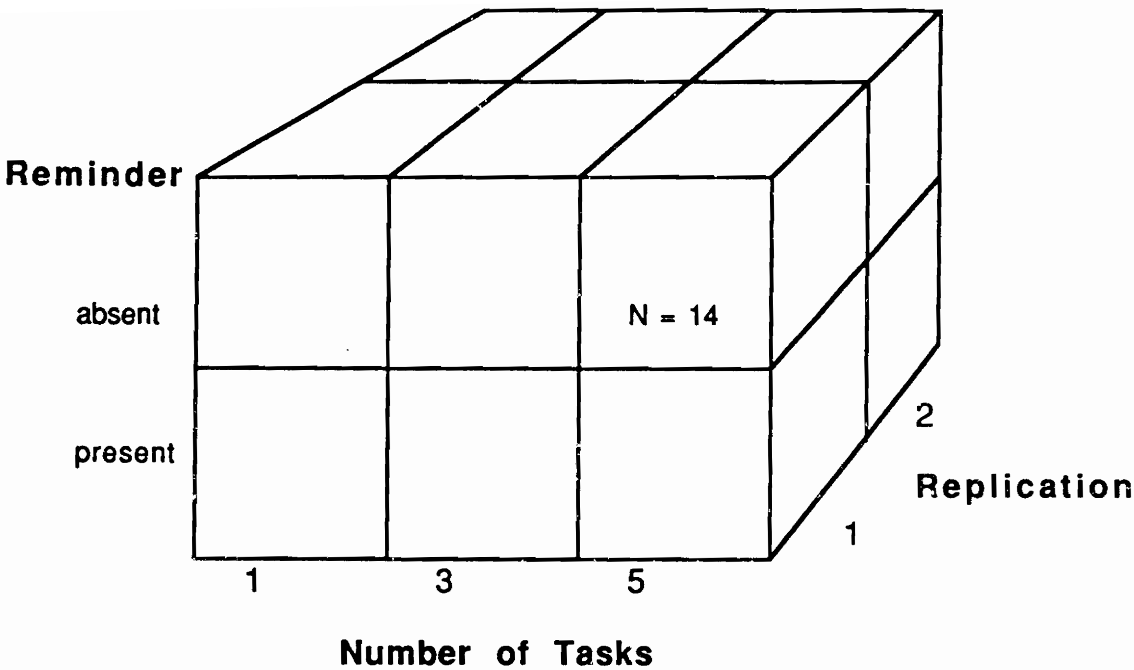


Figure 7. Experimental Design: 3x2x2 Within Subject Design

name to help subjects identify the task(s) they had been instructed to perform. A more detailed discussion will be provided in the Display Menu section.

By definition, Replication referred to the number of times each Number of Tasks and Reminder treatment combination was repeated. Randomized treatment combinations of 1, 3, and 5 tasks were repeated twice, once each for the reminder and no-reminder condition. Therefore, subjects participated in six experimental test conditions per Replication. Replication was used to detect practice effects that would otherwise not be shown with only one observation per treatment combination.

To counterbalance possible order effects, half of the subjects received the no-reminder treatment the first day of testing and half received the reminder treatment first. Consequently, those subjects in the no-reminder condition the first day received the reminder treatment the second day. Likewise, subjects in the reminder condition the first day, subsequently received the no-reminder treatment condition the second day of testing. Furthermore, the specific tasks to be performed were randomly selected from the six tasks. In all, subjects had six experimental test runs the first day and six the second day for a total of 12 experimental trials.

Subjects

Fourteen NASA Cooperative Education students volunteered to participate in this study. In all, there were nine male and five female participants. The age range of the subjects varied from 19 to 41. Excluding the outlier (age 41), the mean age was 22.8. Prior to participating in the study, prospective subjects were asked to complete a questionnaire shown in

Appendix III. Candidate subjects were allowed to participate in the experiment if they had normal or corrected vision and physical impairments that would prevent them from using a mouse input device.

In addition to the standard information requested in the questionnaire, participants were asked to describe their computer experience. This included the types of computers they most often used and the purposes for using them. Eleven subjects reported use of Macintosh computers (three daily, six occasionally, and two very little). The remaining three subjects stated they used Main Frame, Compaq, Apple 2E, IBM, or Sun Workstation computers occasionally. Figure 8 provides a percentage comparison of the functional use of computers by the subjects. Word processing (71%) and basic graphical uses (64%) were the two most commonly cited functions. Other uses included spreadsheets, 3-dimensional modelling, and animation.

The questionnaire also addressed specific multitasking issues such as the average number of computer applications used simultaneously, the average number of tasks, and the average number of display windows opened at any one time. Figure 9 shows that, for the most part, subjects limited their multitasking activities to one computer application, task, and window. To assess how well the subjects could judge their multitasking abilities, participants were asked to judge their perceived ability to "keep track of several changing things at once." Nine subjects felt confident they could do it, but would prefer to focus their attention on fewer than three changing things. Three subjects reported they would probably have some difficulty. Two subjects reported they would probably have no difficulty.

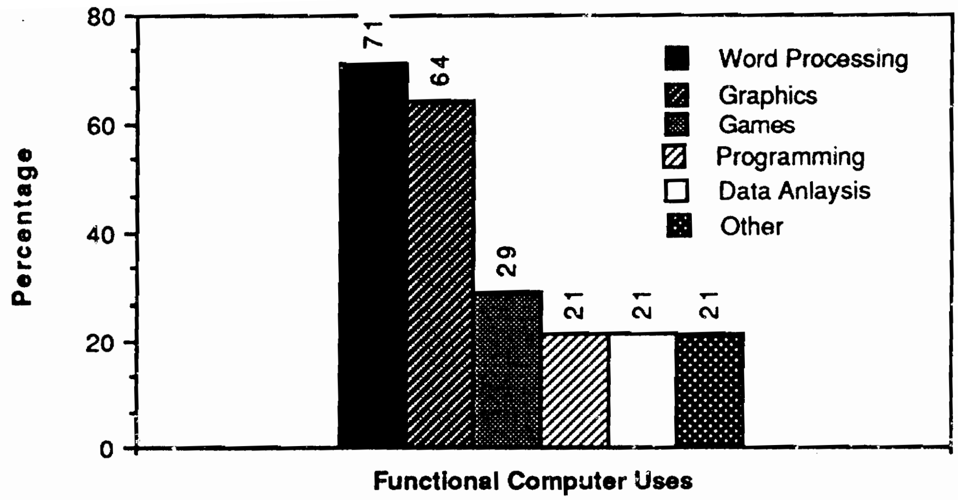


Figure 8. Functional Use of Macintosh Computers Cited by Subjects

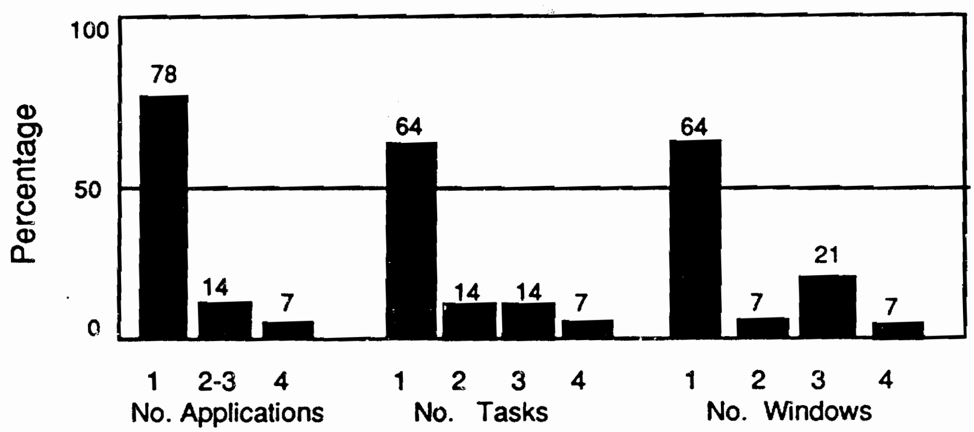


Figure 9. Query into computer multitasking user habits

Hardware

A Macintosh IIfx computer with eight megabytes of RAM was used to run the multitasking experimental simulation tasks. The computer was equipped with an Apple 13-inch diagonal high resolution (640 x 480) RGB monitor, an Apple extended keyboard, and a standard Apple Macintosh II mouse.

Software

The multitasking simulation was programmed in SuperCard™ Version 1.5 (developed by Silicon Beach). SuperCard provided a means of creating and customizing a Macintosh prototype application that would simulate a multitasking environment rapidly and at a low cost. SuperCard's unique programming language, SuperTalk, was used to record and time-stamp reaction time and error performance data.

Task

Procedural Tasks

Six hypothetical Space Station Freedom tasks were designed to represent several procedural tasks involving monitoring activities. Although the specific tasks to be performed on SSF have yet to be defined, each task was designed in such a way as to simulate a modularized subactivity or subsystem within a larger system. Each task is depicted and discussed in Figures 10 -15.

The design of each task had several common features. The most common feature was that each task looked and behaved like a small display window. However, unlike most Macintosh display windows, these could not be moved or resized by subjects. All task windows had a grey menu bar, a close

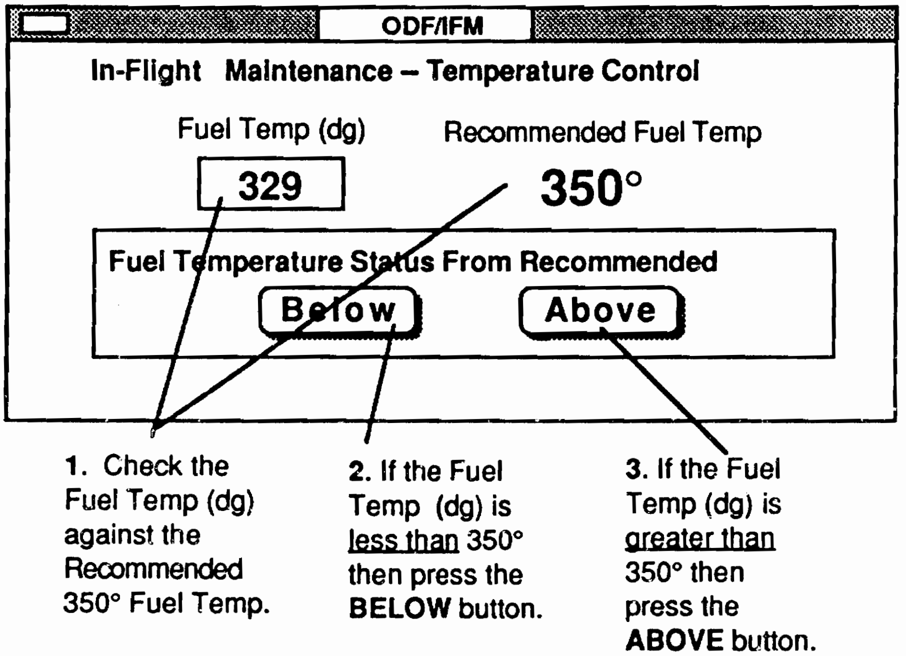


Figure 10. Operations Data File -- In-Flight Maintenance Task

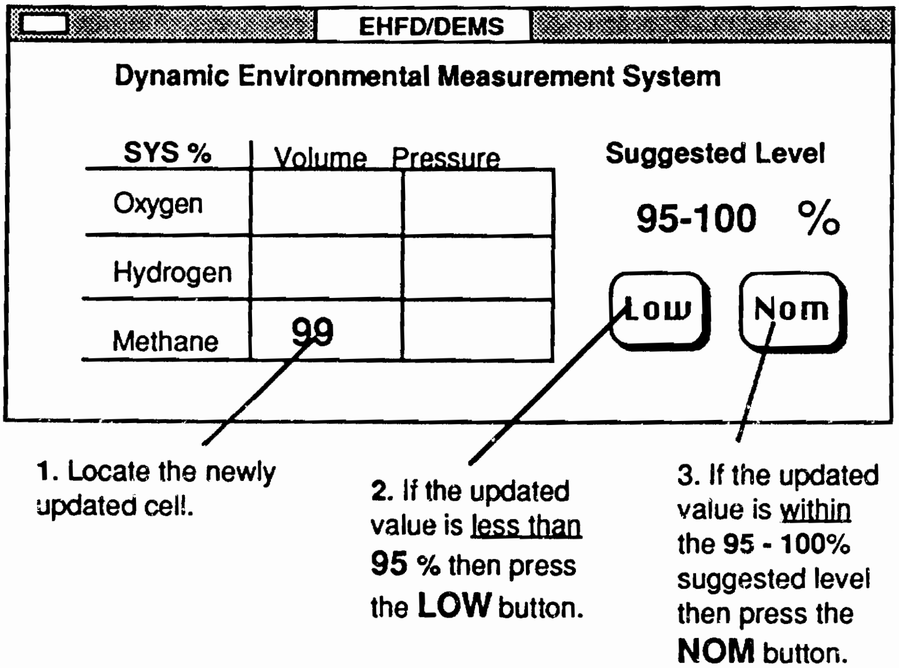


Figure 11. Environmental Health Facilities Database -- Dynamic Environmental Measurement System Task

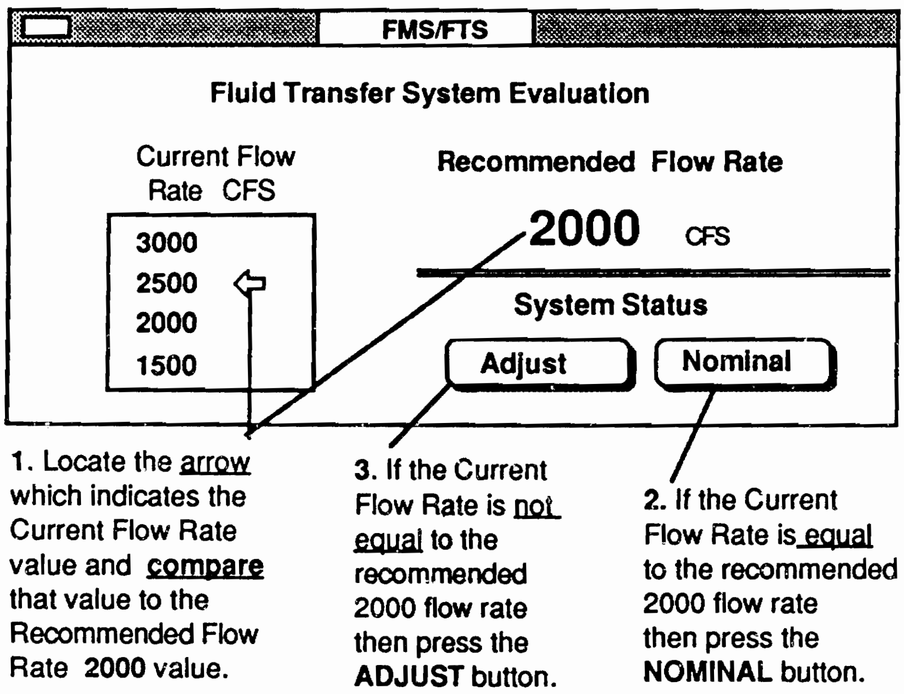
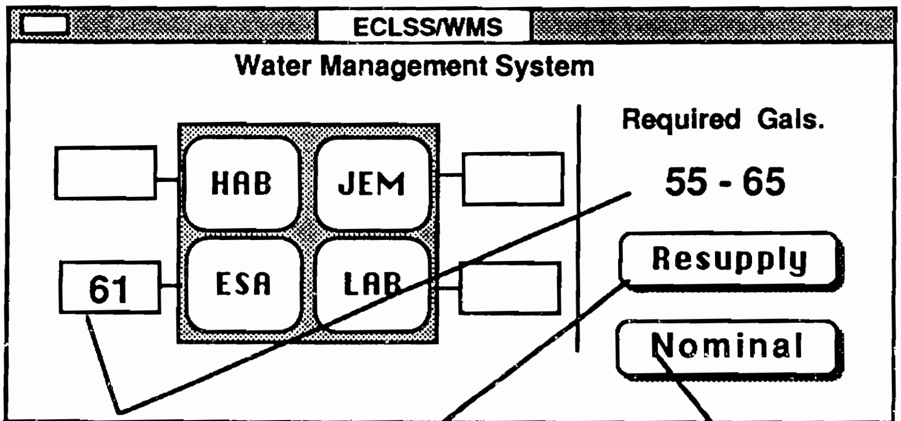


Figure 12. Fluids Management System -- Fluid Transfer System Evaluation Task



1. Locate the updated value that corresponds to one of the four SSF nodes and compare it to the Required Gals. **55-65** range.

2. If the updated value is not within the **55-65** range then press the **RESUPPLY** button.

3. If the updated value is within the required range then press the **NOMINAL** button.

Figure 13. Environmental Control and Life Support System -- Water Management Task

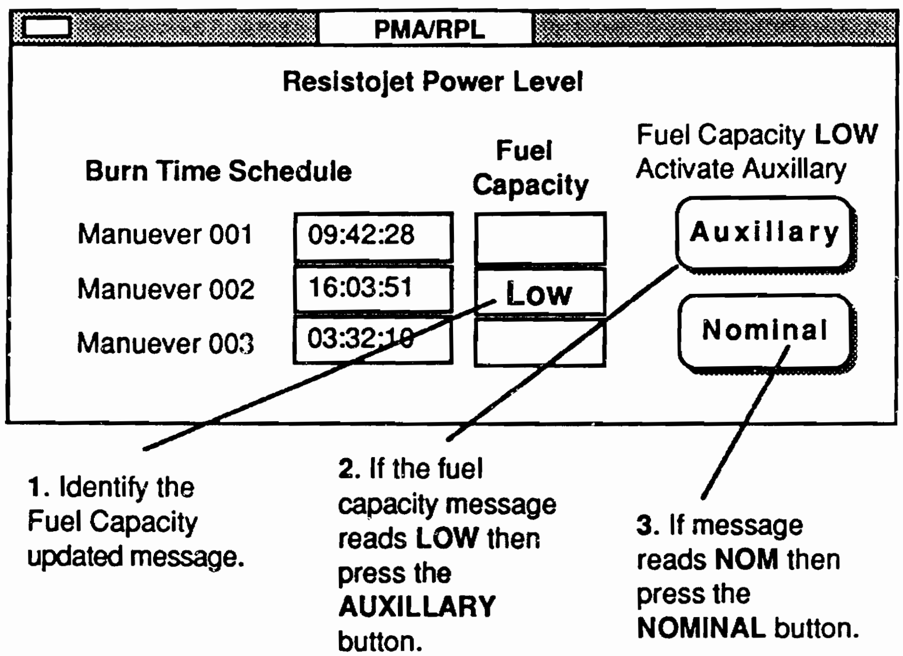


Figure 14. Propulsion Module Assembly -- Resistojet Power Level Task

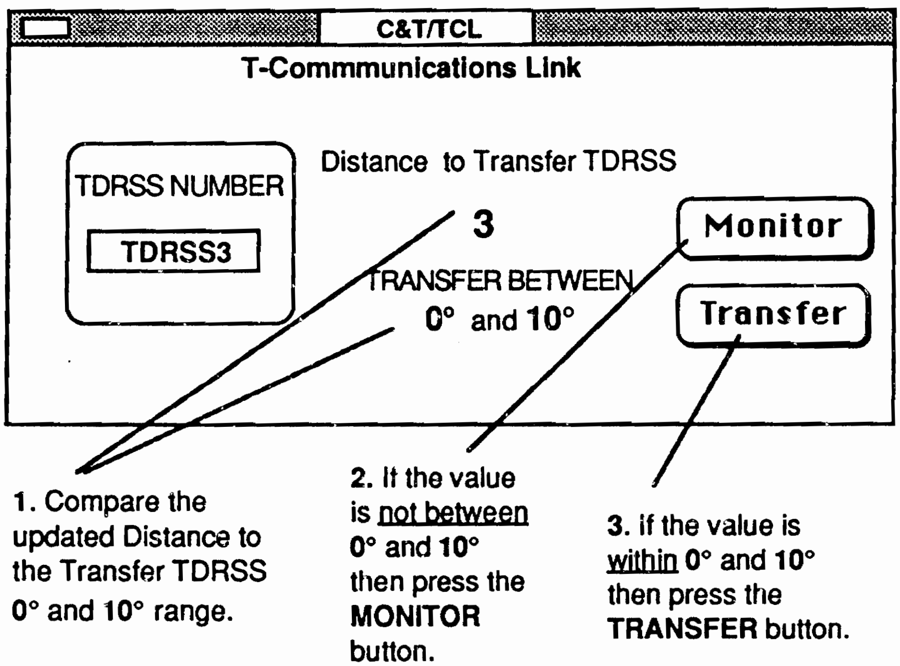


Figure 15. Communications and Tracking -- TDRSS Communications Link Task

box located in the upper left hand corner, and a title box in the center. Each title was unique and followed a specific format. The set of letters to the left of the slash identified the system (e.g., Thermal Control System) by abbreviation (e.g., TCS). The letters following the slash identified the task[1] within the system. In addition to the abbreviated system/task title, the complete and full task name was provided directly below the menu bar for each task window.

Apart from the overall window formatting features, each task was consistent in the type of monitoring and decision making strategy required of the subject. That is, the tasks required verbal information processing versus spatial processing. Most tasks required a simple response to be made between a changing value as compared to a recommended system level. Tasks depicted in Figures 10 and 12 required comparing one numerical value to a specific system value. The tasks shown in Figures 11, 13, and 15 required comparing one numerical value to a specified system range. Only one task, shown in Figure 14, deviated from numerical comparisons; however, the underlying verbal information processing paradigm remained the same. In all task situations, the specific system recommended range, value, or level remained constant. Moreover, these values were always visible to the subject. To make the distinction clear, the system recommended range, value, or level was in bold and typically larger type.

Finally, the choice of responses was limited to two. For example, with the task depicted in Figure 10 the choice was either "below" or "above," with the task depicted in Figure 13 the choice was between "resupply" or "nominal." The

---

[1] Recall some tasks were actually subsystems within the larger system. For all purposes, however, subsystems and subactivities are called tasks.

response choices were active buttons up to the point a decision or mouse click was made. Once a choice was made, both buttons would become disabled (greyed out). This prevented any further responses from occurring. To further confirm the selection, the value of comparison (not the system value) would disappear. Some tasks varied slightly. For example, in the FTS task (see Figure 12), the arrow would disappear. In the RPL task (see Figure 14), the word "Low" or "Nom" would disappear. At the appropriate time interval (approximately every 15 seconds), new values would appear for comparison.

Interruption Task

The interruption task was a simulated Guidance, Navigation, and Control (GN&C) emergency task. The GN&C emergency task was adapted from the actual system that will be operational on the SSF. The actual GN&C system will be responsible for monitoring and controlling the manned base orbit and space vehicle traffic around the SSF. Table 2 lists the individual GN&C system functions. In general, status data and trajectory information will be continuously validated, updated, and presented to the flight crew when necessary. Discrepancies in the GN&C data will cause a warning message to be transmitted to the crewmember on duty via crew workstations, who is then responsible for responding to the situation.

Like most of the procedural tasks described earlier, the GN&C task was hypothetical and was modified for the purposes of this multitasking study. The simulated interruption task involved monitoring GN&C data and performing simple decision making inputs similar to the functions described in the procedural tasks. The GN&C interruption task depicted in Figure 16 was a

Table 2. Guidance, Navigation, and Control System

Guidance Function Displays

* Targeting data
* Maneuver coordination data
* Maneuver command data
* Maneuver monitoring
* Crew override provisions

Navigation Function Displays

* Space Station state data
* Ground update data
* Ephemeris data (coordinates at a given time)
* Navigation fault detection and redundancy management data

Control Functions

* Support commands to intervene and override software and system activities in order to control or prevent undesirable events
* Permitting flight crew to override or reverse any automatic safing or switchover capability of functional paths
* Permit authorized crewmembers to command GN&C Management directly, and lower tier applications if applicable
* Accept the overriding or inhibiting of automated functions by crewmembers

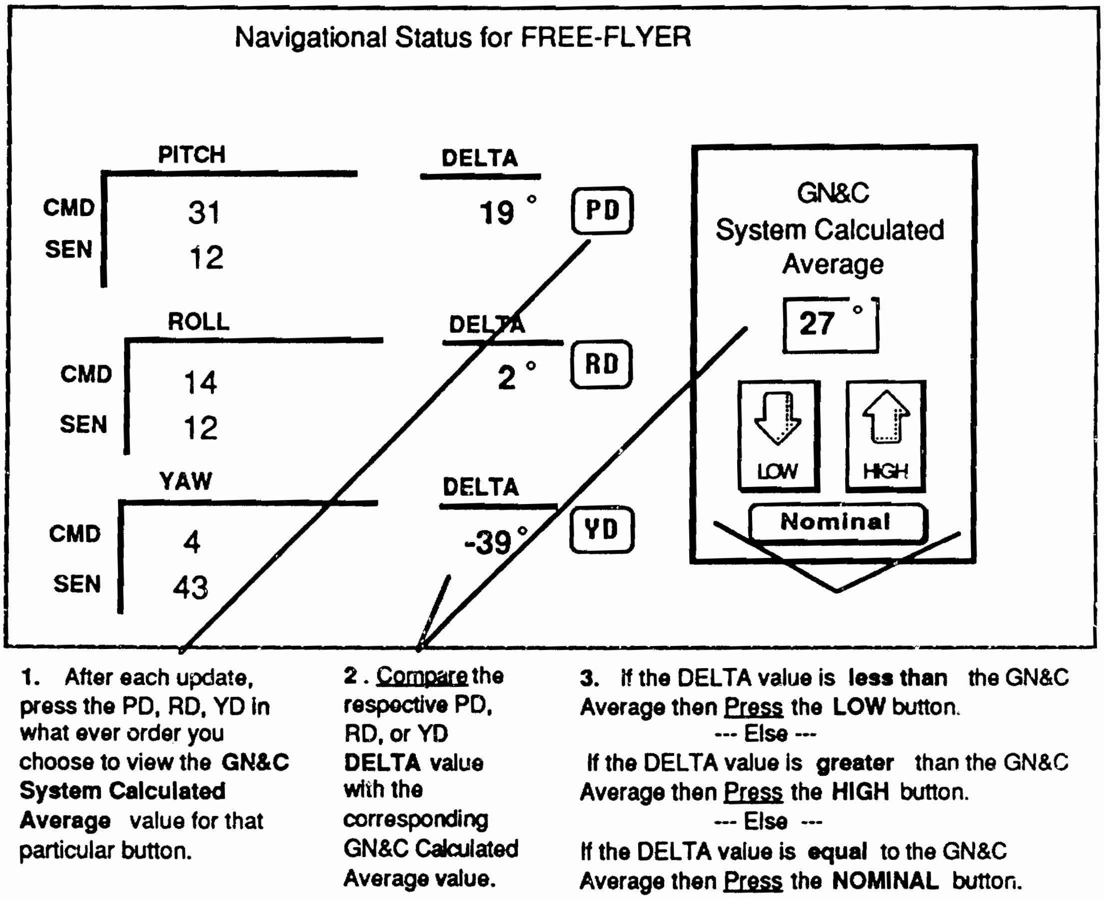


Figure 16. GN&C Emergency Interruption Task (not at actual scale)

modified version of a GN&C task designed for a multitasking experiment ongoing in the HCIL (Holden and O'Neal, 1990). This study's GN&C version was designed to be visually but not fundamentally different from the procedural tasks. That is, the interruption task required verbal processing versus spatial.

For this experiment, the GN&C interruption task required subjects to monitor an incoming unmanned space vehicle's pitch, roll, and yaw coordinates which was threatening the safety of the SSF (refer to Figure 16). CMD referred to the system's programmed and expected coordinate *Command* value. SEN, an abbreviation for *Sense*, represented the actual coordinate values as they were being received from the GN&C system. Delta referred to the difference between the CMD and SEN values for pitch (P-DLT), roll (R-DLT), and yaw (Y-DLT) coordinates.

The corresponding PD, RD, and YD buttons controlled what System Calculated Average number subjects would use for comparison. There were three different averages. Selecting one of these buttons caused the respective PD, RD, or YD System Calculated Average to appear in the smaller box. Subjects would then proceed to make a comparative decision between the respective P-DLT, R-DLT, and Y-DLT value, and the System Calculated Average. Upon making the comparison, subjects were required to click the "Low", "High", or "Nominal" button. Once a decision was made, the remaining two buttons would disable. Again, this was purposely done to prevent numerous responses to the same value.

As with the six procedural tasks, the GN&C values (except for the System Calculated Average) were changing at a constant rate. Unlike the 15 second update rate for the main tasks, the GN&C values updated every 10 seconds. In addition, the three corresponding PD, RD, and YD System Calculated Averages did change with every new treatment condition.

Display Menu

Three distinct menus were designed for the purposes of simulating a realistic system. These menus, shown in Figure 17, replaced the default Apple menu and all but one SuperCard™ menu. The four leading menus were Editor, Resource, Systems, and Documents. The Editor menu was actually part of the SuperCard™ program. It was included for demonstration purposes only. Both the Resource and Documents menu contained superfluous information related, but not critical, to the experiment.

The most important menu was the hierarchically designed Systems menu. This was the only menu with which subjects were required to interact during the experiment. It contained a total of nine SSF larger systems (C&T, DMS, ECLSS, EHFD, FMS, GN&C, ITA, OFD, and PMA). In turn, each system contained eight different menu items. Although there were a total of 72 submenu items, this experiment focused on six. These six corresponded to the procedural tasks illustrated in Figures 9-14.

The checkmark seen next to the RPL task under the PMA system (in Figure 17) was the computer-based reminder. This checkmark was only visible during the reminder treatment condition. In addition, the checkmark only appeared next to the task(s) subjects had been instructed to perform. Therefore, although there were a total of six tasks, only 1, 3, or 5 tasks would actually have a checkmark appearing next to its name. Due to programming

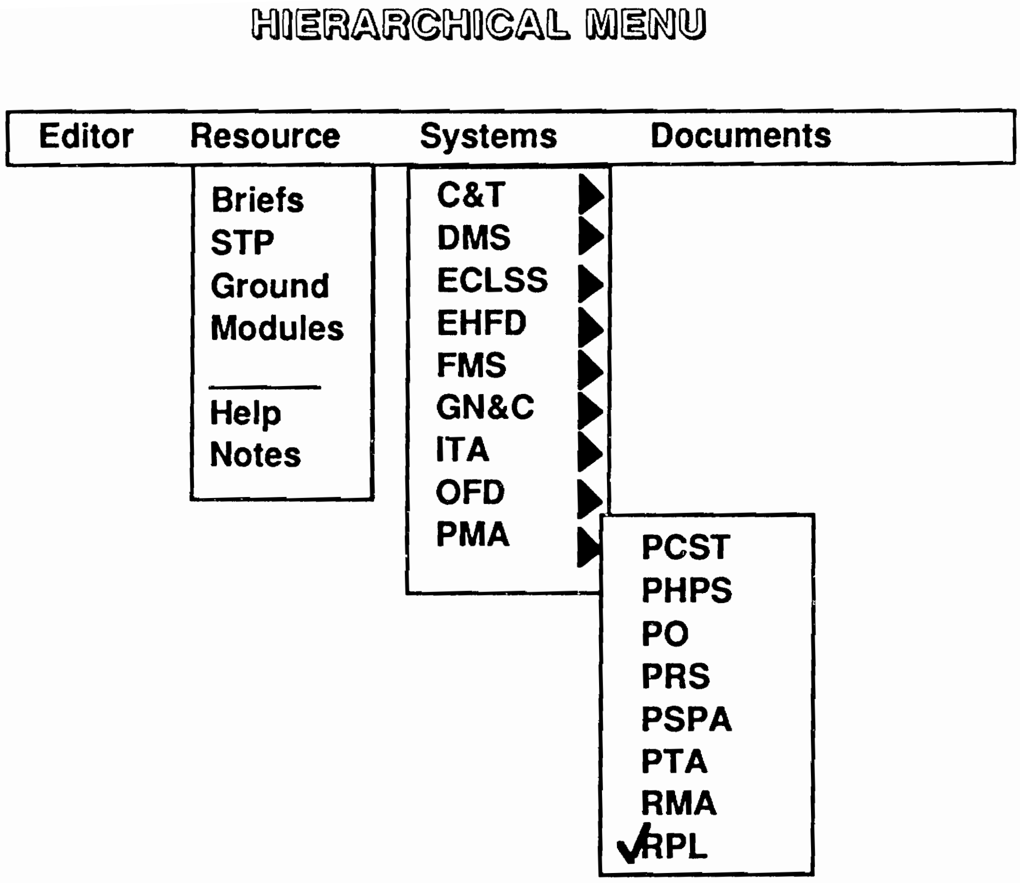


Figure 17. Resource, Systems, and Document Menu Bar

difficulties, the checkmark did not appear next to the C&T, DMS, ECLSS, EHFD, FMS, GN&C, ITA, OFD, and PMA system names.

EXPERIMENTAL PROCEDURE

Potential subjects were greeted and escorted to a test cubicle where they read the participants consent form shown in Appendix IV. In addition, the experimenter reviewed the information contained in the consent form with each candidate subject. Volunteers were told the study required a commitment of three consecutive days, the first of which was devoted to training. Upon signing the consent form, the questionnaire shown in Appendix III was given to each subject. The training session followed the completion of these prerequisites.

Training

Training was divided by a brief break into two comprehensive computer-based sessions. Subjects with little mouse experience were first allowed to practice clicking and moving the mouse. Once they felt comfortable with the device, the experimenter proceeded to explain the two-part training sessions. The average time to complete the two-part training phase was 1.5 hours.

Part I

The first training session focused on the individual procedural tasks and all components associated with those tasks. The training objectives were reviewed with each subject. These objectives were: (1) familiarizing each subject with the task window format and button features, (2) stepping through the instructions together with the experimenter, and (3) providing practice trial runs for each task separately. The goal was to acquaint subjects with all aspects of the experiment.

The six standard task window features were first reviewed. The window size, menu bar, close box, and title box were individually discussed. Next, the different button sizes and names were shown to the subjects. To help subjects understand the concept of a disabled button, a demonstration was provided. Subjects were instructed to click on one particular button which disabled half of the buttons. The functional and visible differences between the active and disabled buttons were illustrated.

Following this formatting and button overview, the instructions for the six task were reviewed separately. After discussing the basic concept and procedure for each task, subjects practiced performing the task. Error feedback was provided during each practice trial. For every incorrect response the computer would beep once. This gave the experimenter an opportunity to discuss the error with the subject and review the instructions. The same procedure was followed for all six tasks. At the end of the first training session, subjects were given a rest break.

Part II

The second training session focused on the GN&C interruption task, the Systems menu, and rating scale. At the end of the second training session, subjects performed two full practice trial runs for the 5-task condition.

As with the procedural tasks, the instructions for the GN&C task were reviewed and discussed together with the experimenter. Subjects were told the goal was to compare all three (PD, RD, and YD) values to a corresponding system calculated average before the next update occurred. After a period of questions and answers, subjects practiced performing the task. Once again, error feedback was provided in the form of a computer beep. When an error was committed, the experimenter discussed the error to reemphasize the correct task procedure.

The experimenter then proceeded to explain the use of the menu bar and illustrate the hierarchical Systems menu. Throughout training, subjects were told to note the system/task title name for each task window. Subjects were told this information would be needed after the interruption task in order to reestablish the tasks. Every subject was shown the location of each system/task name within the Systems menu as well as how to use the menu.

Subjects were also told that, once all the appropriate task(s) had been found and selected, the computer would automatically reinstate the remaining tasks.[2] Moreover, subjects were instructed to only open the task(s) they had been performing prior to the interruption. They were further told that, if they opened an incorrect window (one that was not assigned to them by the computer initially), they were to immediately close it. Failure to do so would be recorded as an error. Finally, subjects were told that the computer would end

---

[2] The choice for doing this was experimentally based. A comparison could not have been made between the pre-interruption and post-interruption performances if the display environment had changed. Reinstating the remaining tasks provided a consistent display environment as well as visual noise.

the trial after two minutes of unsuccessful searching and prepare them for the next trial.

Subjects were then presented with the rating scale shown in Appendix V. They were told that, at the end of each trial run (treatment condition), the computer would ask them to complete the rating scale. The experimenter explained the purpose of the scale and how it would be used to assess overall workload. At this point, the experimenter also encouraged subjects to write comments on the rating scale sheet regarding their impression of the trial run or other comments.

The final training step involved integrating all the training components from both training sessions. Up to this point, subjects had been trained on each task and the skills needed to perform them. Three points needed to be covered: (1) how would subjects know what task(s) to perform, (2) how would the interruption occur, and (3) how would they retrieve the GN&C emergency task. To integrate the components, subjects were shown the diagram depicted in Figure 18. This diagram represented the precise location of the six tasks. Based on this, a miniature representation of the display layout was designed as the means of identifying which task(s) to perform.

The miniature graphical icon of the display screen shown in Figure 19 was presented to each subject prior to every treatment condition. The small boxes would highlight the appropriate location along with the task name to show the subjects which task(s) were to be performed next. Hence, subjects had two forms of visual cues, the location of the highlighted box and the task name.

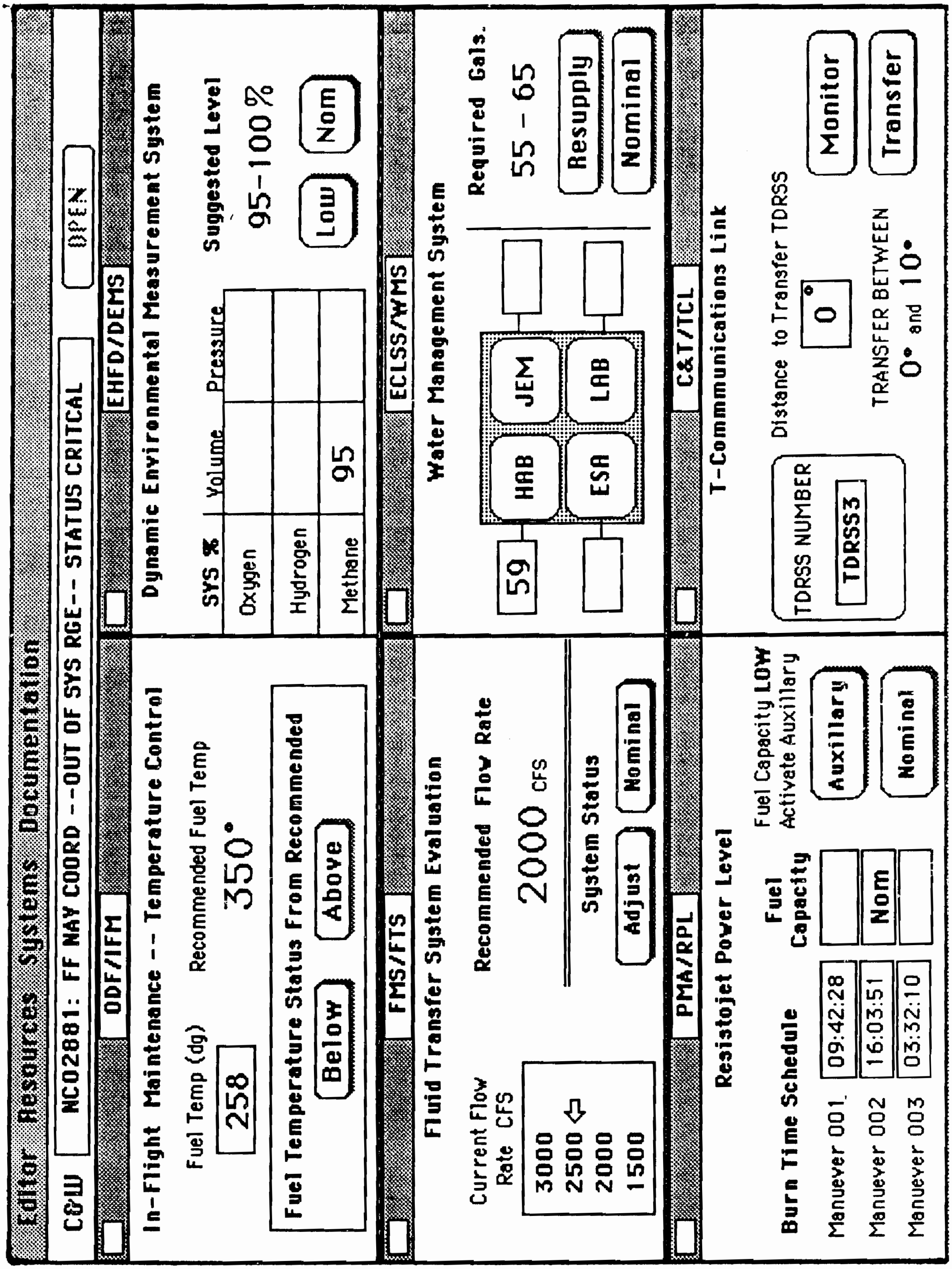


Figure 18. Six-Task Display Screen Layout (not at actual scale)

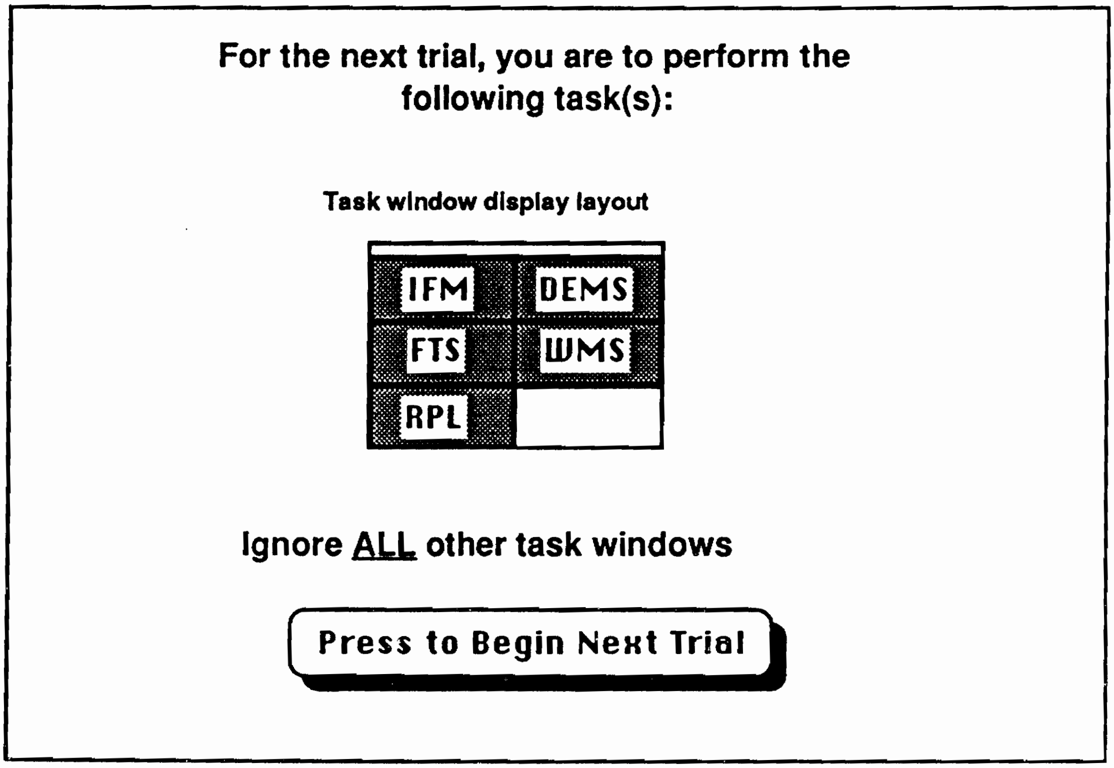


Figure 19. Miniature Task-Display Screen

For the interruption warning, subjects were told they would hear three beeps and that these beeps would occur at a random time as they performed the 1, 3, and 5 tasks. Having heard the beeps, subjects were instructed to look at the C&W (Caution and Warning) message box located below the menu bar, where a system message would appear (see Figure 18) in the C&W box regarding the GN&C emergency status. Subjects were told to make their last comparison decisions for the task(s) they had been instructed to perform. They were then told to close all six task windows. To retrieve the GN&C emergency task, subjects were told to click the "Open" button located in the C&W message box. Having clicked the button, the GN&C task would appear and the update numbers would begin automatically.

Having integrated all the components of the experiment, subjects were given two practice trial runs with the 5-task no-reminder treatment condition. These practice runs had all the features of an actual experimental testing condition with the exception that error feedback was still provided. However, subjects were cautioned they would not be receiving error feedback during the actual experiment.

At the completion of the training sessions, subjects were scheduled for two consecutive days of experimental testing. Subjects were told that, upon returning the following day, the procedures would be reviewed. It should be noted that subjects were never told about the computer-based reminder. As far as they were concerned, there would be no reminder checkmark on any trials.

Test Protocol

Subjects were randomly assigned to the reminder or no-reminder test condition the first day. This consequently affected which condition they would receive the second day as explained in the Experimental Design section.

Before testing began, subjects were given an experimental instruction review sheet (see Appendix VI). After reading the instructions, subjects were shown a display screen diagram similar to that illustrated in Figure 18. The experimenter then proceeded to review each task and the instructions. Subjects were reminded to make their response decisions as quickly and as accurately as possible because the computer would be recording their responses. They were not, however, told the amount of time they took to study the miniature display screen was being recorded.

Throughout training, subjects had been led to believe the GN&C task was more critical because the SSF safety was theoretically in jeopardy. Therefore, because this was a critical task, it would require undivided attention. This was the reason given to subjects for closing all six task windows. Subjects were told it was important they return to the unfinished tasks after the GN&C emergency was terminated because the tasks needed to be completed.

For subjects receiving the reminder condition, a semi-functional menu graphic was used to illustrate the checkmark. Subjects were told the checkmark was added to help them reestablish the task(s) after the interruption. In addition, subjects were reminded that the checkmark would only appear next to the task(s) they had been told to perform. Subjects in the no-reminder treatment condition simply proceeded with the experimental procedure after reviewing the instructions.

Regardless of the number of tasks to be performed, task updates occurred every 15 seconds. In all, subjects performed the procedural tasks for a total of 12 updates or three minutes (190 seconds). A separate time length and update rate was used for the GN&C interruption task. The GN&C numerical updates occurred every 10 seconds for a total of 90 seconds or nine updates.

The six rating scales were placed in a clipboard next to the terminal along with a pen. Subjects were reminded to complete the scale after each trial run and to write any comments on the blank space. Subjects were also given a five minute break after the third trial run. After the break, subjects continued to perform the remaining three trials until the computer displayed a screen informing them the experiment had ended.

The same experimental procedure was followed on the second day of testing. For those subjects who had received the reminder condition the first day, the experimenter informed them the reminder checkmark would not be available. The experimenter explained that, since the reminder was not part of the original experiment, it had to be removed. The explanation was accepted by subjects, although the majority expressed disappointment. At the end, subjects were debriefed (see Appendix VII) and thanked for having participated in the study.

DEPENDENT MEASURES

Objective

Nine performance measures were obtained for each subject: (1) reaction time pre-interruption, (2) pre-interruption response errors, (3) GN&C reaction time, (4) GN&C response error frequency, (5) reaction time post-interruption, (6) post-interruption response errors, (7) menu search time, (8) average viewing time, and (9) subjective mental workload ratings.

Reaction times were collected in order to assess the speed with which subjects could make simple decisions in a multitasking situation. Reaction time was recorded in milliseconds for each task response in every treatment condition. Traditionally, reaction time measurements reflect a simple choice response to a stimulus when it is first perceived by the subject.

Obtaining reaction times in a multitasking environment can be problematic. The very fact that subjects in this study used a mouse for all response inputs limited the number of responses subjects could make at one time. It is obvious then that, for the 3- and 5-task conditions, subjects would not be able to respond to all the tasks at the time of an update. To correct for this, reaction time measurements were taken to reflect the time interval between button clicks, with the exception of the very first update. For example, the initial update time was recorded, stored, and used to calculate the new reaction time once a response was made. A new time was then recorded and stored only after each button click response was made. For this study, reaction time is defined as the time interval between responses and not the initial update time.

Hence, consistent and valid reaction time measurements were obtained for all task conditions.

It should be noted that reaction time was obtained by using the internal computer tick count. One computer tick is equal to 1/60 second. Although the computer tick count is not the most preferred method for obtaining precise reaction time measurements, it was the method most suitable for the software design. This method provided a satisfactory tool for comparing multitasking reaction times given three and five simultaneous tasks.

Response errors frequency reflected the number of errors made when comparing an updated value to a recommended system value or range. Since the recommended system values were constants, this measure was easily obtained. In addition to incorrect responses, the error frequency count included all missed responses. If a subject failed to respond to an update, it was counted as an error. The reason for this was quite simple: if these procedures had been actual tasks on the Space Station Freedom, missing or failing to respond to information could have affected the overall system's performance.

The final measure of response error frequency was actually an average of the total number of responses possible plus the number of extra responses made. For the 5-task condition there were 60 possible responses, 36 possible responses for the 3-task condition, and 12 responses for the 1-task condition. The possibility of missing an update or responding to an unassigned task was taken into account. Responding to an unassigned task was counted as an error and included in the average, even if the response was technically correct.

As a result of missing an update, the reaction time was adjusted to reflect that error. Subjects were essentially penalized for failing to respond. For procedural tasks, a missed response resulted in a penalty of 15 seconds for that update, corresponding to the update time interval. In the event subjects responded to an unassigned task, the reaction time was reflected in the average time as well as in the response error average. For the GN&C interruption task, the penalty was 10 seconds per update which corresponded to the task's time update interval.

Menu search time reflected the average time it took subjects to find and reestablish the task(s). The average included the time for selecting incorrect menu items. This measure was used to assess the possible benefits of providing a simple checkmark reminder. Shorter menu search times would indicate the computer-based reminder was beneficial in helping identify which items to select.

The average viewing time dependent measure was a hidden measure obtained while the subjects studied the miniature display screen that told them what task(s) to perform, illustrated in Figure 19. The display screen had been designed so that subjects decided when to begin the experimental trial runs. Therefore, subjects had total control over the time they took to view and study this screen. This measure was seen as an important one because it had the potential of providing insights into the subjects' behavior when they knew a reminder would be available.

Subjective

Subjective mental workload measurements were obtained using the Modified Cooper-Harper (MCH) rating scale shown in Appendix V. This subjective measure has been shown to provide a global assessment of mental workload related to perception, monitoring, evaluation, communication, and problem solving tasks (Wierwille and Casali, 1983). These authors tested and demonstrated that the MCH rating scale was a valid and reliable measurement of global workload in three simulated aircraft experiments. In view of such support and general applications, the MCH rating scale was selected.

DATA ANALYSIS and RESULTS

Nine ANOVA's were conducted for each of the dependent measures: (1) pre-interruption reaction time, (2) response error pre-interruption, (3) GN&C reaction time, (4) GN&C error frequency, (5) post-interruption reaction time, (6) response error post-interruption, (7) menu search time, (8) average miniature task-display viewing time, and (9) subjective mental workload ratings.

Pre-Interruption

The pre-interruption reaction time analysis (Table 3) indicates significant main effects for Replication, $F(1,13) = 7.55$, $p < 0.0166$ and Number of Tasks, $F(2,26) = 54.74$, $p < 0.0001$. The post hoc analysis for the Number of Tasks main effect is presented in Table 4. The analysis shows a difference in reaction time performance between 1- and 5-tasks ($p<0.05$), 1- and 3-tasks ($p<0.05$), but not between 3 and 5 ($p<0.05$), distinguished by the letters A and B. In both instances, reaction time was significantly different when the single task performance was compared with each multitask performance measure. Multitask performance did not, however, significantly vary as the number of tasks increased from three to five. Futhermore, there was no significant interaction between the Replication and Number of Tasks factors. In addition, no significant main effects or interactions were observed for the response error pre-interruption measure (Table 5).

Table 3. ANOVA for Pre-Interruption Reaction Time

Source	*df*	*SS*	*F*	*Prob*
Between Subjects				
Subjects (Sub)	13	16.911		
Within Subject				
Reminder (REM)	1	0.019	0.20	0.6645
Sub x REM	13	1.313		
Tasks (TKS)	2	7.128	54.74	0.0001 *
Sub x TKS	26	1.6927		
Replication (REPL)	1	0.256	7.55	0.0166 *
Sub x REPL	13	0.440		
REM x REPL	1	0.109	1.20	0.2937
Sub x REM x REPL	13	1.186		
REM x TKS	2	0.190	3.04	0.0653
Sub x REM x TKS	26	0.812		
REPL x TKS	2	0.438	1.88	0.1722
Sub x REPL x TKS	26	3.024		
REM x REPL x TKS	2	0.266	1.68	0.2064
Sub x REM x REPL x TKS	26	2.063		
Total	167	35.851		

Reminder, $p = 2$
Tasks, $q = 3$
Replication, $r = 2$
Subjects per cell, $n = 14$

Table 4. Student-Newman-Keuls Test for Pre-Interruption Reaction Time

Alpha = 0.05

Number of Tasks	Mean		
3	1.494	A	
5	1.457	A	A
1	1.040		B

Note: Means with the same letters are not significantly different from one another at this Alpha level

Table 5. ANOVA for Response Error Pre-Interruption

Source	*df*	*SS*	*F*	*Prob*
Between Subjects				
Subjects (Sub)	13	0.086		
Within Subject				
Reminder (REM)	1	0.003	0.81	0.3847
Sub x REM	13	0.044		
Tasks (TKS)	2	0.002	0.51	0.6049
Sub x TKS	26	0.045		
Replication (REPL)	1	0.001	0.35	0.5631
Sub x REPL	13	0.037		
REM x REPL	1	0.001	1.97	0.1841
Sub x REM x REPL	13	0.008		
REM x TKS	2	0.000	0.09	0.9109
Sub x REM x TKS	26	0.056		
REPL x TKS	2	0.005	0.90	0.4187
Sub x REPL x TKS	26	0.075		
REM x REPL x TKS	2	0.011	1.71	0.1998
Sub x REM x REPL x TKS	26	0.082		
Total	167	0.457		

Reminder, p = 2
Tasks, q = 3
Replication, r = 2

Subjects per cell, n = 14

GN&C Interruption Task

The results obtained for the GN&C emergency interruption task are presented in Tables 6 and 7. A significant main effect with respect to reaction time is noted for the Replication factor, $F(1,13) = 12.65$, $p < 0.0035$. No significant effects or interactions were observed for the response error measure, reported in Table 7.

Post-Interruption

The results for the post-interruption reaction time and response error measures are shown in Tables 8 and 9 respectively. A significant main effect was found for the Number of Tasks, $F(2,26) = 10.81$, $p < 0.0004$. A Student-Newman-Keuls post hoc test indicated reaction time performance differed among the three conditions (Table 10). Reaction time performance was significantly faster with one task when compared to performance on both three and five tasks, which did not differ. As noted in the pre-interruption and GN&C response error results, no significant main effects or interactions were found for the post-interruption response error measure (Table 9).

Menu Search Time

As predicted, the main effect for the Reminder factor was significant $F(1,13) = 9.64$, $p < 0.0084$ (see Table 11). Menu search time was affected by the availability of the computerized checkmark reminder. There was a considerable increase in mean performance time when the reminder was not available (33.738) as opposed to when it was available(20.733). A significant effect was also obtained for the Number of Tasks, $F(2,26) = 45.64$, $p < 0.0001$

Table 6. ANOVA for GN&C Reaction Time (emergency interruption)

Source	*df*	*SS*	*F*	*Prob*
Between Subjects				
Subjects (Sub)	13	9.103		
Within Subject				
Reminder (REM)	1	0.002	0.02	0.8869
Sub x REM	13	1.251		
Tasks (TKS)	2	0.060	2.78	0.0802
Sub x TKS	26	0.279		
Replication (REPL)	1	0.441	12.65	0.0035 *
Sub x REPL	13	0.454		
REM x REPL	1	0.047	2.88	0.1133
Sub x REM x REPL	13	0.213		
REM x TKS	2	0.050	1.46	0.2496
Sub x REM x TKS	26	0.441		
REPL x TKS	2	0.008	0.28	0.7600
Sub x REPL x TKS	26	0.371		
REM x REPL x TKS	2	0.013	0.45	0.6395
Sub x REM x REPL x TKS	26	0.364		
Total	167	13.096		

Reminder, p = 2
Tasks, q = 3
Replication, r = 2
Subjects per cell, n = 14

Table 7. ANOVA for GN&C Response Error (emergency interruption)

Source	*df*	*SS*	*F*	*Prob*
Between Subjects				
Subjects (Sub)	13	0.019		
Within Subject				
Reminder (REM)	1	0.000	0.06	0.8069
Sub x REM	13	0.007		
Tasks (TKS)	2	0.000	0.40	0.6723
Sub x TKS	26	0.008		
Replication (REPL)	1	0.001	1.90	0.1913
Sub x REPL	13	0.006		
REM x REPL	1	0.000	0.00	1.0000
Sub x REM x REPL	13	0.005		
REM x TKS	2	0.001	0.86	0.4357
Sub x REM x TKS	26	0.123		
REPL x TKS	2	0.000	0.54	0.5882
Sub x REPL x TKS	26	0.011		
REM x REPL x TKS	2	0.001	0.89	0.4216
Sub x REM x REPL x TKS	26	0.009		
Total	167	0.080		

Reminder, p = 2
Tasks, q = 3
Replication, r = 2
Subjects per cell, n = 14

Table 8. ANOVA for Post-Interruption Reaction Time

Source	*df*	*SS*	*F*	*Prob*
Between Subjects				
Subjects (Sub)	13	21.000		
Within Subject				
Reminder (REM)	1	0.263	1.00	0.3353
Sub x REM	13	3.419		
Tasks (TKS)	2	6.058	10.81	0.0004 *
Sub x TKS	26	7.283		
Replication (REPL)	1	0.018	0.10	0.7539
Sub x REPL	13	0.006		
REM x REPL	1	0.752	0.94	0.3498
Sub x REM x REPL	13	10.395		
REM x TKS	2	1.179	0.94	0.4032
Sub x REM x TKS	26	16.296		
REPL x TKS	2	2.034	1.82	0.1820
Sub x REPL x TKS	26	14.524		
REM x REPL x TKS	2	0.311	0.57	0.5704
Sub x REM x REPL x TKS	26	7.040		
Total	167	92.826		

Reminder, p = 2
Tasks, q = 3
Replication, r = 2
Subjects per cell, n = 14

Table 9. ANOVA for Response Error Post-Interruption

Source	*df*	*SS*	*F*	*Prob*
Between Subjects				
Subjects (Sub)	13	0.344		
Within Subject				
Reminder (REM)	1	0.005	0.66	0.4300
Sub x REM	13	0.091		
Tasks (TKS)	2	0.000	0.01	0.9854
Sub x TKS	26	0.356		
Replication (REPL)	1	0.008	1.11	0.3107
Sub x REPL	13	0.095		
REM x REPL	1	0.034	1.12	0.3089
Sub x REM x REPL	13	0.395		
REM x TKS	2	0.054	1.14	0.3358
Sub x REM x TKS	26	0.621		
REPL x TKS	2	0.046	0.94	0.4031
Sub x REPL x TKS	26	0.642		
REM x REPL x TKS	2	0.000	0.00	0.9957
Sub x REM x REPL x TKS	26	0.337		
Total	167	3.030		

Reminder, $p = 2$
Tasks, $q = 3$
Replication, $r = 2$
Subjects per cell, $n = 14$

Table 10. Student-Newman-Keuls Test Post-Interruption Reaction Time

Alpha = 0.05

Tasks	Mean		
3	1.470	A	
5	1.385	A	A
1	1.032		B

Note: Means with the same letters are not significantly different from one another at this Alpha level

Table 11. ANOVA for Total Menu Search Time

Source	*df*	*SS*	*F*	*Prob*
Between Subjects				
Subjects (Sub)	13	12474.545		
Within Subject				
Reminder (REM)	1	7104.059	9.64	0.0084 *
Sub x REM	13	9582.898		
Tasks (TKS)	2	23222.645	45.64	0.0001 *
Sub x TKS	26	6615.026		
Replication (REPL)	1	1485.894	3.01	0.1062
Sub x REPL	13	6411.106		
REM x REPL	1	749.929	1.39	0.2597
Sub x REM x REPL	13	7019.002		
REM x TKS	2	875.977	1.32	0.2832
Sub x REM x TKS	26	8595.143		
REPL x TKS	2	343.949	0.79	0.4662
Sub x REPL x TKS	26	5688.597		
REM x REPL x TKS	2	342.840	0.93	0.4086
Sub x REM x REPL x TKS	26	4810.460		
Total	167	95322.070		

Reminder, p = 2
Tasks, q = 3
Replication, r = 2
Subjects per cell, n = 14

(see Table 12). However, the interaction of these tow factors was not significant ($p = 0.2832$). Further analysis showed a significant difference for all the means, reported in Table 12. The Student-Newman-Keuls post hoc analysis validates the increase in mean search time performance among the three Number of Tasks levels (see Table 12).

Viewing Task-Display

Due to initial software problems, data for the first two subjects was not used. Therefore, the results shown in Table 13 reflect the viewing time for 12 subjects only. As anticipated, a significant main effect was found for the Number of Tasks, $F(2,26) = 4.08$, $p < 0.0312$. This would indicate subjects viewed and studied the miniature task display screen according to the number of tasks they were assigned. The Student-Newman-Keuls post-hoc analysis indicates a significant difference in mean viewing time only between the 1-task condition as compared to the 5-task condition, with reaction time significantly longer with five tasks ($p<0.05$; see Table 14). The mean viewing times between 3- and 5-tasks and between 1- and 3-tasks were not significant.

Workload Ratings

The last dependent measurement obtained for each treatment condition was the subjective mental workload rating. The results shown in Table 15 indicate a significant main effect for the Number of Tasks variable, $F(2,26) = 20.12$, $p < 0.0001$). No other significant effects or interactions were found. The post hoc analysis showed significant differences among all three means (Table 16). As expected, subjects rated the mental workload requirements for 5-tasks higher than the other two treatment conditions. Likewise, the 3-task condition was rated higher than the single task condition.

Table 12. Student-Newman-Keuls Test for Total Menu Search Time

Alpha = 0.05		
Number of Tasks	Mean	
5	41.386	A
3	27.720	B
1	12.600	C

Note: Means with the same letters are not significantly different from one another at this Alpha level

Table 13. ANOVA for Time Spent Viewing the Miniature Task-Display Screen

Source	*df*	*SS*	*F*	*Prob*
Between Subjects				
Subjects (Sub)	11	19440.770		
Within Subject				
Reminder (REM)	1	233.896	0.23	0.6442
Sub x REM	11	11411.835		
Tasks (TKS)	2	10490.002	4.08	0.0312 *
Sub x TKS	22	28298.515		
Replication (REPL)	1	89.019	0.07	0.7935
Sub x REPL	11	13614.146		
REM x REPL	1	904.907	0.76	0.4028
Sub x REM x REPL	11	13144.482		
REM x TKS	2	1058.161	0.50	0.6155
Sub x REM x TKS	22	23460.280		
REPL x TKS	2	1110.566	0.44	0.6502
Sub x REPL x TKS	22	27826.556		
REM x REPL x TKS	2	1797.860	0.83	0.4511
Sub x REM x REPL x TKS	22	23954.953		
Total	143	176835.946		

Reminder, p = 2
Tasks, q = 3
Replication, r = 2
Subjects per cell, n = 14

Table 14. Student-Newman-Keuls Test for Time Spent Viewing the Miniature Task-Display Screen

Alpha = 0.05

Tasks	Mean	
5	25.467	A
3	11.491	A B
1	5.014	B

Note: Means with the same letters are not significantly different from one another at this Alpha level

Table 15. ANOVA for the Subjective Mental Workload Ratings

Source	*df*	*SS*	*F*	*Prob*
Between Subjects				
Subjects (Sub)	13	50.286		
Within Subject				
Reminder (REM)	1	0.381	0.25	0.6238
Sub x REM	13	19.619		
Tasks (TKS)	2	37.869	20.12	0.0001 *
Sub x TKS	26	24.464		
Replication (REPL)	1	0.595	2.52	0.1365
Sub x REPL	13	3.071		
REM x REPL	1	0.214	1.92	0.1894
Sub x REM x REPL	13	1.452		
REM x TKS	2	0.940	1.01	0.3767
Sub x REM x TKS	26	12.060		
REPL x TKS	2	1.583	3.05	0.0646
Sub x REPL x TKS	26	6.750		
REM x REPL x TKS	2	0.321	0.70	0.5081
Sub x REM x REPL x TKS	26	6.012		
Total	167	165.619		

Reminder, p = 2
Tasks, q = 3
Replication, r = 2
Subjects per cell, n = 14

Table 16. Student-Newman-Keuls Test for Subjective Mental Workload Ratings

Alpha = 0.05

Tasks	Mean	
5	2.554	A
3	1.911	B
1	1.393	C

Note: Means with the same letters are not significantly different from one another at this Alpha level

DISCUSSION

The ability to perform three or five multiple tasks simultaneously, a central theme of the study, presents three issues to evaluate: (1) would the level of response accuracy decrease with increases in the number of tasks to perform, (2) would reaction times differ among the tasks, and (3) would an interruption task greatly disrupt performance on the primary tasks. In addition to addressing basic multitasking and task interruption issues, this study examined the benefits of a computer-based reminder.

Multitasking Abilities

The multitasking response accuracy was at a higher level than anticipated. The results show that people perform three and five tasks simultaneously with no loss in performance accuracy. In addition, the pre- and post-interruptions did not impact the subject's level of response accuracy. Overall, the average error rate for all three task conditions resulted in a response accuracy of 98 percent.

The results further reveal that the number of tasks to be performed do not affect the level of response accuracy. Given such a high response accuracy, how was reaction time affected? Was this level of accuracy achieved by sacrificing reaction time--commonly referred to as the speed-accuracy paradigm?. Recall, reaction time was first calculated individually for each task to reflect the time between the most recent task response and the previous response (not the initial update time). The pre and post-interruption performance analysis showed the number of tasks subjects performed simultaneously affects reaction time. However, this difference occurred between the control baseline treatment condition (1-task) and the two multitasking conditions (3- and 5-tasks). Reaction time performance did not differ significantly between the two multitasking conditions. Subjects took approximately the same amount of time to respond to updates in the 3-task and 5-task multitasking treatment conditions as shown in Figures 20 and 21. In view of these findings, one may conclude that in the multitasking conditions subjects did not trade response speed for response accuracy. Therefore, the speed-accuracy paradigm explanation as discussed in the literature does not apply to the present multitasking findings.

This study postulates the hypothesis that increasing the number of tasks to be performed adversely affects reaction time and perhaps the number of errors. The underlying assumption was that additional tasks would require more attention and mental processing resources. This assumption presupposed that response time would be affected by the number of task-rules subjects needed to evaluate prior to making their response decision. Hence, the addition of more tasks to this mental evaluation/decision process would theoretically result in longer reaction times for the 5-task condition as opposed to the 3-task condition. It was therefore surprising to not find a significant difference between the reaction time performance for these two multitasking treatment conditions. This lack of significant difference may be attributed to the performance of the multiple tasks in sequence rather than in parallel. Serial or sequential reaction time has been shown to be affected by three factors: decision complexity, pacing, response complexity (Wickens, 1984).

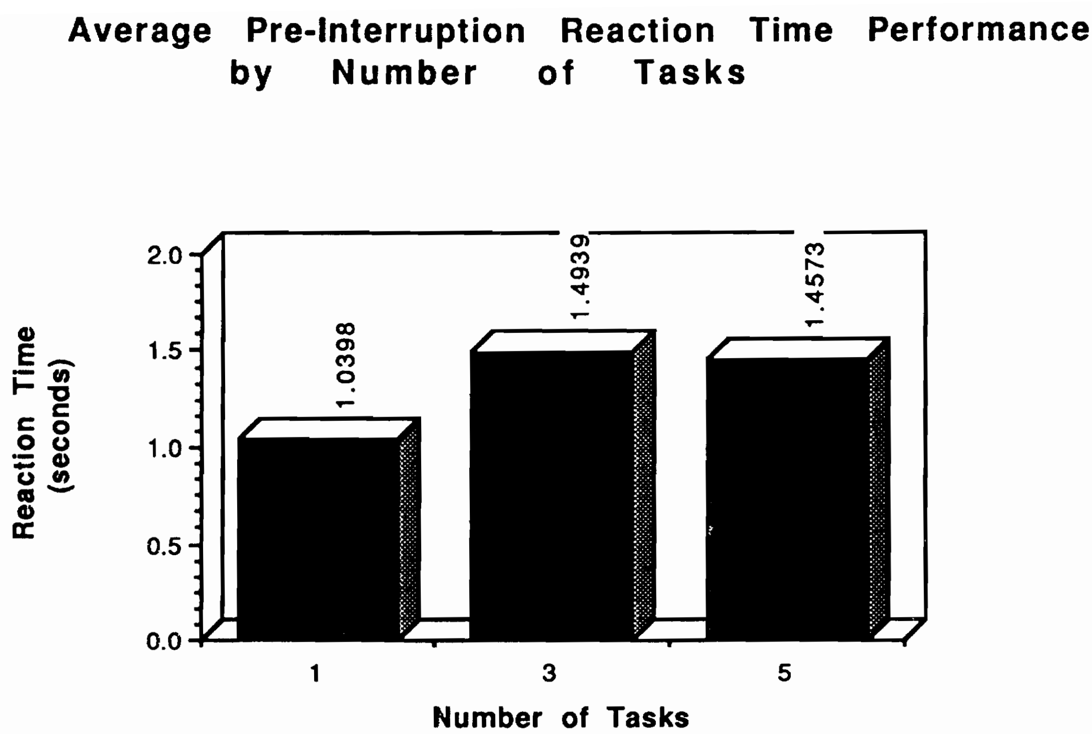


Figure 20. Average Pre-Interruption Reaction Time Performance by Number of Tasks

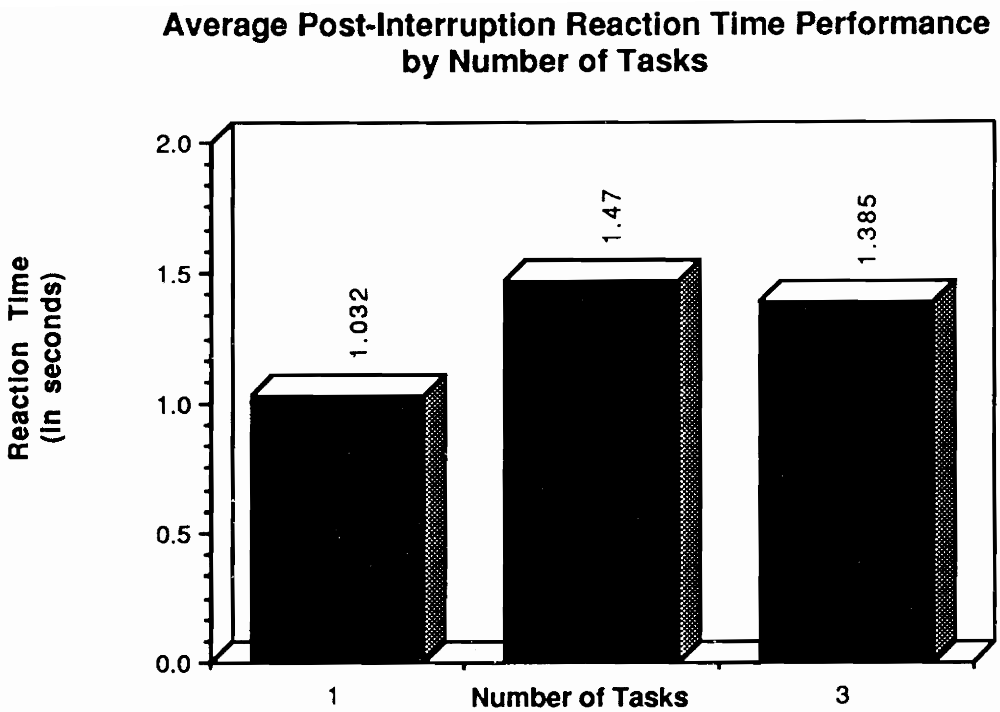


Figure 21. Average Post-Interruption Reaction Time Performance by Number of Tasks

Decision Complexity

In retrospect, the decision-making requirements for this study's multitasking tasks were quiet simple. Subjects were merely required to decide if an updated value was within a specified and constant system range. Therefore, decision complexity remained relatively unchanged despite the increase in the number of tasks being performed. Furthermore, the tasks were not memory intensive. Subjects did not have to rely heavily on their working memory to accomplish the tasks since the comparison system ranges were always visible. Perhaps greater reaction time differences would have been observed had memorization of all the comparison system values been required.

Pacing

The simple decision-making requirements and the constant 15 second update rate presented predictable tasks during the study. Subjects soon learned they could establish a pace that allowed all responses to be made before the next update occurred. After the experimental sessions, a number of subjects reported that the tasks became easier once they had learned to pace themselves. Along with pacing themselves, subjects also developed a response strategy or pattern of response. Most subjects employed a clockwise response pattern. Few subjects elected a counterclockwise pattern. Regardless of the pattern used, subjects reported they would position the cursor between two response buttons in anticipation of the next update. This clearly suggests pacing was a contributing factor to the similar reaction time performances among the two multitasking treatment conditions.

Response Complexity

Another factor that may have contributed to similar reaction times for the 3- and 5-task conditions is response complexity. The response requirements were simple. Subjects chose one of two possible responses (e.g., above versus below, nominal versus resupply, etc.). To select their response, subjects moved the cursor within the boundaries of a fairly large button, double clicked, and proceeded to the next response button. Since all the updates occurred at once, subjects had the opportunity to establish a response pattern and determine what response they were going to make as they moved the cursor across the display screen.

Interruption Effects

The results also indicated that the GN&C interruption task did not affect response errors or reaction time performance on any of the 1-, 3-, or 5-task conditions. There are two possible reasons that may help explain these findings. The first explanation is that the GN&C interruption task was too easy. That is, the interruption task did not interfere in any respect with the underlying processing of the main tasks as evident from the lack of significant findings in the post-interruption reaction time and response error performance measurements. The second explanation, and most viable, is that the GN&C interruption task was performed independent of the main tasks. Subjects were not required to perform the interruption task while they attempted to perform the standard tasks. This was clearly a departure from the typical primary and secondary experiments conducted in the past.

In traditional dual-task performance experiments, the secondary task is introduced during the performance of the primary task. These types of experiments have shown decrements in reaction time performance when two tasks are performed simultaneously. A technique known as "shadowing" is commonly used whereby subjects repeat a message out loud, purposely diverting attention away from the primary task. Much of the support for parallel processing and multiple resource theory has come from using this and similar techniques. However, the design of this study's multitasking tasks and the secondary interruption task departed from these traditional forms of testing, in order to accommodate the selected computer system. We therefore note that performance on a primary task is not necessarily hindered if the secondary task is performed independently. Another important factor that contributed to the findings in this study was practice.

Practice

Data analysis showed that practice (replication) was a significant factor in pre-interruption reaction time performance (Figure 22) and in the GN&C interruption task (Figure 23). In both instances, reaction time decreased from the first to the second replication. Apart from the mere exposure to these tasks, pacing and response pattern may have also contributed to the improved performance. A closer analysis of the raw data showed that response patterns had been clearly established prior to the interruption. Therefore, completing the primary tasks, post-interruption, was simply a matter of reestablishing the response pattern and pace. In essence, after the interruption, subjects knew

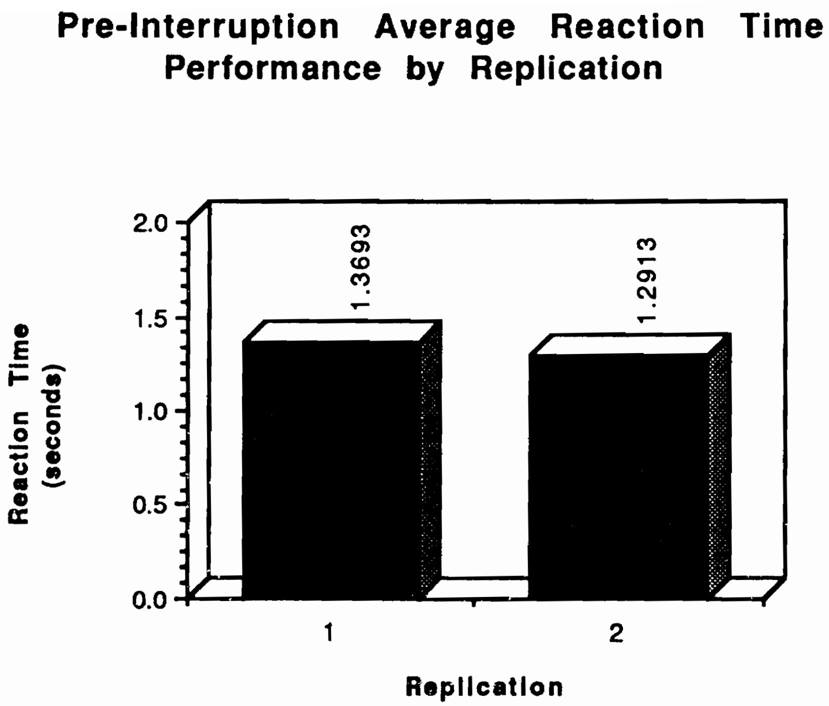


Figure 22. Pre-Interruption Average Reaction Time Performance by Replication

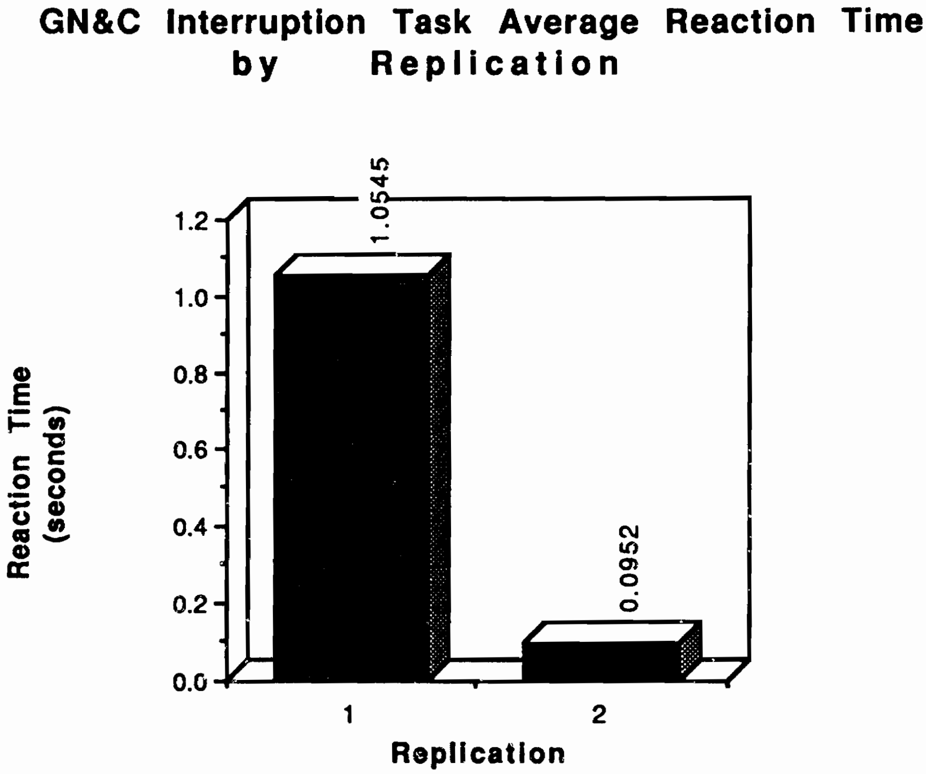


Figure 23. GN&C Interruption Task Average Reaction Time by Replication

what to expect. This would explain why post-interruption reaction time performance was not affected by replication.

Reminder Effects

The computerized checkmark reminder was shown to be beneficial regardless of how long subjects had studied, rehearsed, and attempted to memorize the system/task names. In addition, the benefit of the checkmark was independent of the number of system/tasks subjects had to locate in the hierarchical menu. The reminder was equally helpful for all treatment conditions. As shown in Figure 24, with the aid of the checkmark, total menu search time was decreased by 13 seconds. The increased search time for 3- and 5-tasks compared to that for 1-task (Figure 25) was naturally expected since there were more items to locate. However, the overall benefit of the checkmark reminder was evident in all three conditions.

Perhaps the most interesting discovery was that menu search time did not improve with exposure or practice. That is, search time did not decrease from the first to second replication. A parallel can be drawn between the no-reminder condition performance and the limited capacity of our working-memory to recall "chunks" of irrelevant items after a period of time. This limitation sometimes occurs even when rehearsal is allowed (Gillie and Broadbent, 1989).

An argument can be made that the system/task names were not irrelevant "chunks" of information since the meaning behind each task name had been explained to the subjects. Yet, with all the training and rehearsal, performance

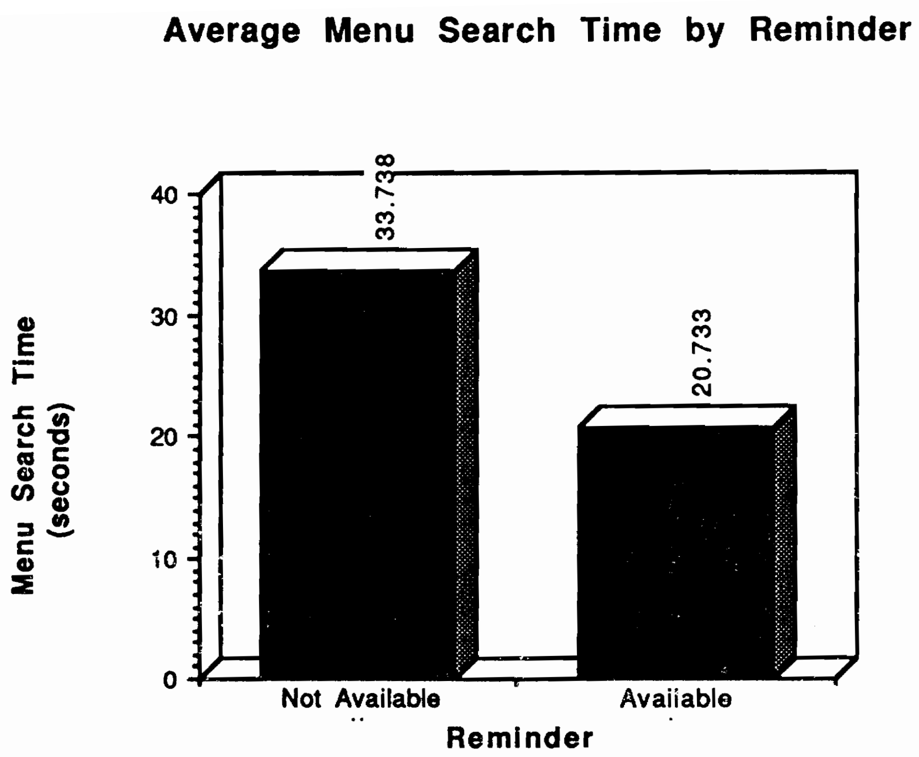


Figure 24. Average Menu Search Time by Reminder

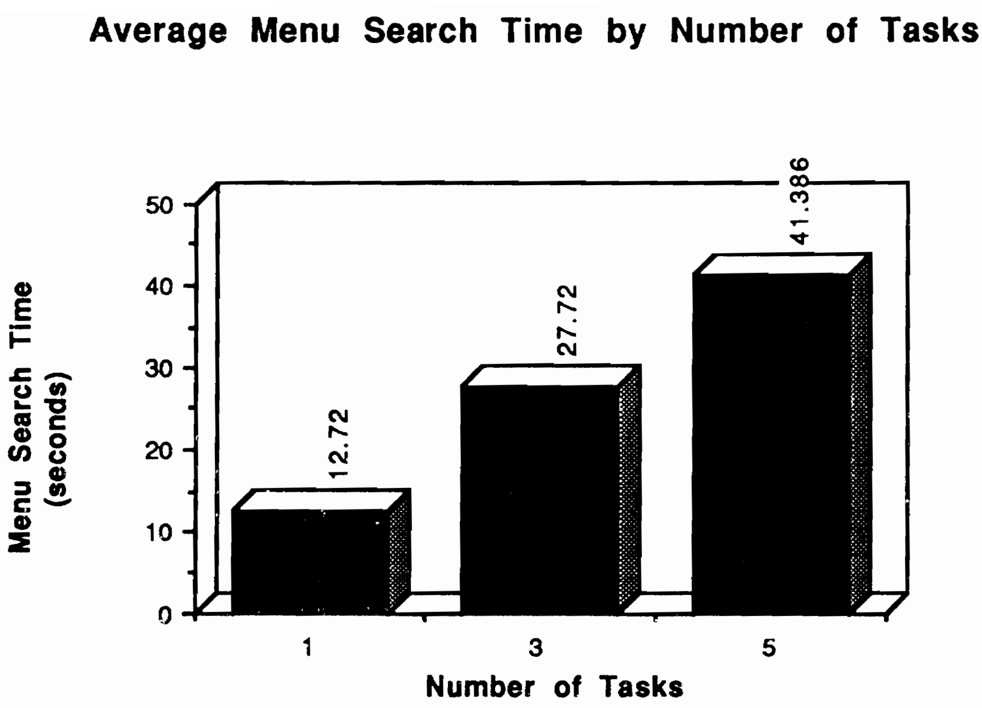


Figure 25. Average Menu Search Time by Number of Tasks

without the reminder was considerably poorer. These results coincide with Gillie and Broadbent's (1989) experimental findings. Although subjects were allowed to rehearse their position in the primary task in an effort to determine if rehearsal guaranteed immunity from disruption by an interruption, the authors found that rehearsal did not protect against the disruptive effects of an interruption. In the present study, the exposure to the system/task names and menu did not improve performance.

The externally-based computer checkmark was an effective tool in reducing the amount of time spent searching for menu items. In fact, two subjects in the no-reminder condition were never able to locate three system/task names in the menu. One subject was unable to locate the tasks in the first replication while the other subject was unable to locate the tasks in the second replication. Without the aid of the checkmark reminder, subjects reported trying to recall the system/task names by the number of letters in a particular system/task name, or a specific letter sequence (i.e., DEMS versus WMS). Hence, the computerized checkmark reminder alleviated some mental effort required by these other search strategies.

Perceived Workload

Despite exceptional performance with all three Number of Tasks conditions, mental workload was judged higher as the number of tasks increased. The 5-task condition was rated as the highest mental workload, followed by 3-tasks, and finally, the 1-task. These findings must be qualified by examining the meaning of the rating values. For the purposes of interpretation, the averages shown in Table 17 of the Data Analysis and Results section were rounded to the nearest whole number. Therefore, the respective task ratings become 1, 2, and 3. Referring to the scale in Appendix V, we see that these three ratings describe the task difficulty level as being "very easy" to "mildly difficult." Since the highest rating was 3 for the 5-task condition, we find that most subjects found the workload "acceptable" and the difficulty level "fair". Overall, the subjects in this study did not perceive the multitasking treatment conditions to be extremely difficult.

CONCLUSIONS

The results obtained in the present study offer new information into people's ability to perform several primary sequential tasks and an independent secondary task. First of all, multitasking performance was greatly influenced by practice. Other factors, such as decision complexity, response complexity, response pattern, and pacing, provided an alternative explanation to the multitasking performance findings. Second, if a secondary task is performed independent of the primary tasks, performance on the primary tasks will not likely deteriorate (given the task conditions in the present study). We can also infer that the increase in menu search time without the reminder indicates some limitation for storing and recalling abbreviated system/task names. Furthermore, computer-based reminders are beneficial (again, given the conditions described in the present study).

From the findings in this study, we can conclude that: (1) people can be trained to perform multiple tasks provided there is task-consistency, (2) a secondary task is not necessarily disruptive if it is performed independently from the primary tasks, and (3) the use of a simple computer-based reminder will reduce menu search time considerably. Further validation of these findings to evolving space applications would require more applied research.

The following study was not without its problems. Electing to use a Macintosh computer, which does not currently support internal multiprocessing, limited the methods available for investigating human multitasking capabilities.

In addition, the results of this study could not be readily compared to other multitasking studies nor explained by multiple resource theory for several reasons. First, the design and structure of this study differed considerably from those of past multitasking and dual-task experiments. In addition, the types of tasks were very different from those described in previous multitasking research. Second, subjects were not truly required to perform simultaneous tasks in the traditional sense. We have already noted the responses made to the tasks were serial in nature. Hence, parallel processing as described in multiple resource theory was not achieved. The vast majority of experiments described in support of multiple resource theory and parallel processing typically present simultaneous stimuli to different modalities (e.g., auditory and visual) and vary the response format (manual versus verbal). The tasks in this study were all verbal and the response format was always manual. Due to the differences in design of this study to other multitasking studies, a generalized comparison could not be made.

Several recommendations are suggested for future multitasking studies. First, a more accurate method of recording reaction times should be developed and used. Second, in order to truly simulate a more realistic multitasking environment, the tasks should be less predictable and perhaps more demanding. Third, a high memory load and complex interruption task should be selected and performed simultaneously as a secondary task--if it is to prove disruptive to the main tasks. Fourth, careful consideration should be given to the type of computer and software used for the investigation. If possible, a system that supports internal multiprocessing should be chosen. Fifth, various other forms of computer-based reminders should be investigated, such as history charts, icons, etc. Finally, all efforts should be made to obtain a subject's comments and observations.

REFERENCES

Adams, J. A. (1980). *Learning memory: An introduction.* Homewood, IL: Dorsey.

Baddeley, A. D., and Hitch, G. (1974). Working memory. In G. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8). New York: Academic Press.

Bannon, L., Cypher, A., Greenspan, S., and Monty, M. L. (1983, December). Evaluation and analysis of users' activity organization. In *Proceedings of CHI '83* (pp. 54-57). New York: Communications of the ACM.

Billingsley, P. A. (1988). Taking panes: Issues in the design of windowing systems. In M. Helander (Ed.), *Handbook of human-computer interaction* (Chapter 19). North-Holland: Elsevier.

Bly, S. A., and Rosenberg, J. K. (1986, April). A comparison of tiled and overlapping windows. In *Proceedings of CHI '86* (pp. 101-106). New York: Communications of the ACM.

Boff, K. R., and Lincoln, J. E. (Eds.). (1988). *Engineering data compendium: Human perception and performance* (Vol. II). (7.0 Attention and Allocation of Resources). New York: Wiley.

Broadbent, D. A. (1975). The magical number seven after fifteen years. In A. Kennedy and A. Wilkes (Eds.), *Studies in long-term memory.* New York: Wiley.

Bower, G. H. (1972). Mental imagery and associative learning. In L. W. Gregg (Ed.), *Cognition in learning and memory.* New York: Wiley.

Card, S. K., and Henderson, A. Jr. (1987). A multiple, virtual-workspace interface to support user task switching. In *Proceedings of CHI + GI* (pp. 53-59). New York: Communications of the ACM.

Card, S. K, Moran, T. P., and Newell, A. (1986). The model human processor: An engineering model of human performance. In K. R. Boff, L. Kaufman, and J. P. Thomas (Eds.), *Handbook of perception and human performance: Volume II cognitive processes and performance* (Chapter 45). New York: Wiley.

Card, S. K., Pavel, M., and Farrell, J. E. (1987). Window-based computer dialogues. In R. M. Baecker and W. A. S. Baxton (Eds.), *Readings in human-computer interaction: A multidisciplinary approach* (pp. 456-460). Los Altos, CA: Morgan Kaufmann.

Chase, W. G., and Ericsson, K. A. (1981). Skilled memory. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition.* Hillsdale, NJ: Lawrence Erlbaum.

Cypher, A. (1986). The structure of users' activities. In D. A. Norman and S. W. Draper (Eds.), *User centered system design* (Chapter 12). Hillsdale, NJ: Lawrence Erlbaum.

Damos, D. L., and Smist, T. E. (1983). Individual differences in multiple-task performance as a function of response strategy. *Human Factors, 25(2),* 215-226.

Dumais, S. T. (1988). Textual information retrieval. In M. Helander (Ed.), *Handbook of human-computer interaction* (Chapter 30). North-Holland: Elsevier.

Findlay, J. M., Davies, S. P., Kentridge, R., Lambert, A. J., and Kelly, J. (1988). Optimum display arrangement for presenting visual reminders. In D. M. Jones and R. Winder (Eds.), *People and computers IV: Proceedings of the fourth conference of the British Computer Society* (pp. 453-464). Cambridge, London: Cambridge University Press.

Gaylin, K. B. (1986, April). How are windows used? Some notes on creating an empirically-based windowing benchmark task. In *Proceedings of CHI '86* (pp. 96-100). New York: Communications of the ACM.

Gillie, T., and Broadbent, D. (1989). What makes interruptions disruptive? A study of length, similarity, and complexity. *Psychological Research, Vol. 50,* 243-250.

Gilmore, W. E., Gertman, D. I., and Blackman, H. S. (1989). *The user-computer interface in process control: A human factors engineering handbook.* San Diego, CA: Academic Press.

Greenstein, J. S., and Rouse, W. B. (1982, March/April). A model of human decisionmaking in multiple process monitoring situations. *IEEE Transactions of Systems, Man, and Cybernetics, 12* /2, 182-193.

Grief, I. (Ed.). (1988). *Computer-supported cooperative work: A book of readings.* San Mateo, CA: Morgan Kaufman.

Henneman, R. L., and Rouse, W. B. (1986, March/April). On measuring the complexity of monitoring and controlling large-scale systems. *IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-16*/2, 193-207.

Hirst, W., and Kalmar, D. (1987). Characterizing attentional resources. *Journal of Experimental Psychology: General, Vol. 116(1),* 68-81.

Hirst, W., and Spelke, E. S. (1980). Dividing attention without alternation or automaticity. *Journal of Experimental Psychology: General, Vol. 109(1),* 98-117.

Holden, T., and O'Neal, M. (1990). *Human-computer interaction for multitasking systems.* Unpublished manuscript, Lockheed Human Factors Department, Houston, TX.

Intons-Peterson, M. J, and Fournier, J. (1986). External and internal memory aids: When and how often do we use them? *Journal of Experimental Psychology: General, Vol. 115(3),* 267-280.

Klapp, S. T., Marshburn, E. A., and Lester, P. T. (1983). Short-term memory does not involve the "working memory" of infomation processing: The demise of a common assumption. *Journal of Experimental Psychology: General, Vol. 112(2),* 240-264.

Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81-97.

Miyata, Y., and Norman, D. A. (1986). Psychological issues in support of multiple activities. In D. A. Norman and S. W. Draper (Eds.), *User centered system design* (Chapter 13). Hillsdale, NJ: Lawrence Erlbaum.

Myers, B. A. (1984, December). The user interface for Sapphire. *IEEE CG & A*, pp. 13-23.

NASA. (1988, December). *Space Station Freedom Program human-computer interface guide version 2.1* (USE 1000, V. 2.1). Lyndon B. Johnson Space Center, TX: Human Computer Interaction Laboratory.

Newman, W., Stephens, N., and Sweetman, D. (1985). A window manager with a modular user interface. In P. Johnson and S. Cook (Eds.), *People and computers: Designing the interface* (pp. 415-426). Cambridge, London: Cambridge University Press.

Norman, D. A. (1990, January). *Distributed cognition in aviation.* Position paper presented at the CHI '90 Computer-Human Interaction in Aerospace Systems Workshop. Available from Donald A. Norman, Department of Cognitive Science D-015, University of California, San Diego; La Jolla, CA, 92093.

Norman, K. L., Weldon, L. J., and Shneiderman, B. (1986). Cognitive layouts of windows and multiple screens for user interfaces. *International Journal of Man-Machine Studies, 25*, 229-248.

Roger, Y. (1989, April). Icons at the interface: Their usefulness. *Interacting with Computer: The Interdisciplinary Journal of Human Computer Interaction, 1/1*, 105-177.

Ross, B. H. (1984). Remindings and their effects in learning a cognitive skill. *Cognitive Psychology, 16*, 371-416.

Rouse, W. B. (1977, May). Human-computer interaction in multitask situations. *IEEE Transactions on Systems, Man, and Cybernetics, 7*/ 5, 384-392.

Sheridan, T. B. (1988). Task allocation and supervisory control. In M. Helander (Ed.), *Handbook of human-computer interaction* (Chapter 8). North-Holland: Elsevier.

Shiffrin, R. M., and Schneider, W. (1977, March). Controlled and automatic human information processing: Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*/2, 127-190.

Treisman, A. M., and Davies, A. (1973). Divided attention to ear and eye. In S. Kornblum (Ed.), *Attention and performance IV.* New York: Academic Press.

Tulga, M. K., and Sheridan, T. B. (1980, May). Dynamic decisions and work load in multitask supervisory control. *IEEE Transactions on Systems, Man, and Cybernetics, 10* /5, 217-232.

User interface requirements document (DR SY-45.1). (1990, March). Work package no. 2, revision A, resubmittal one. McDonnell Douglas H4261.

Wierwille, W. W., and Casali, J. G. (1983). A validated rating scale for global mental workload measurement applications. In *Proceedings of Human Factors Society 27th Annual Meeting* (pp. 129-133). Santa Monica, CA: Human Factors Society.

Wickens, C. D. (1987). Information processing, decision-making, and cognition. In G. Salvendy (Ed.), *Handbook of human factors* (Chapter 2.2). New York: Wiley.

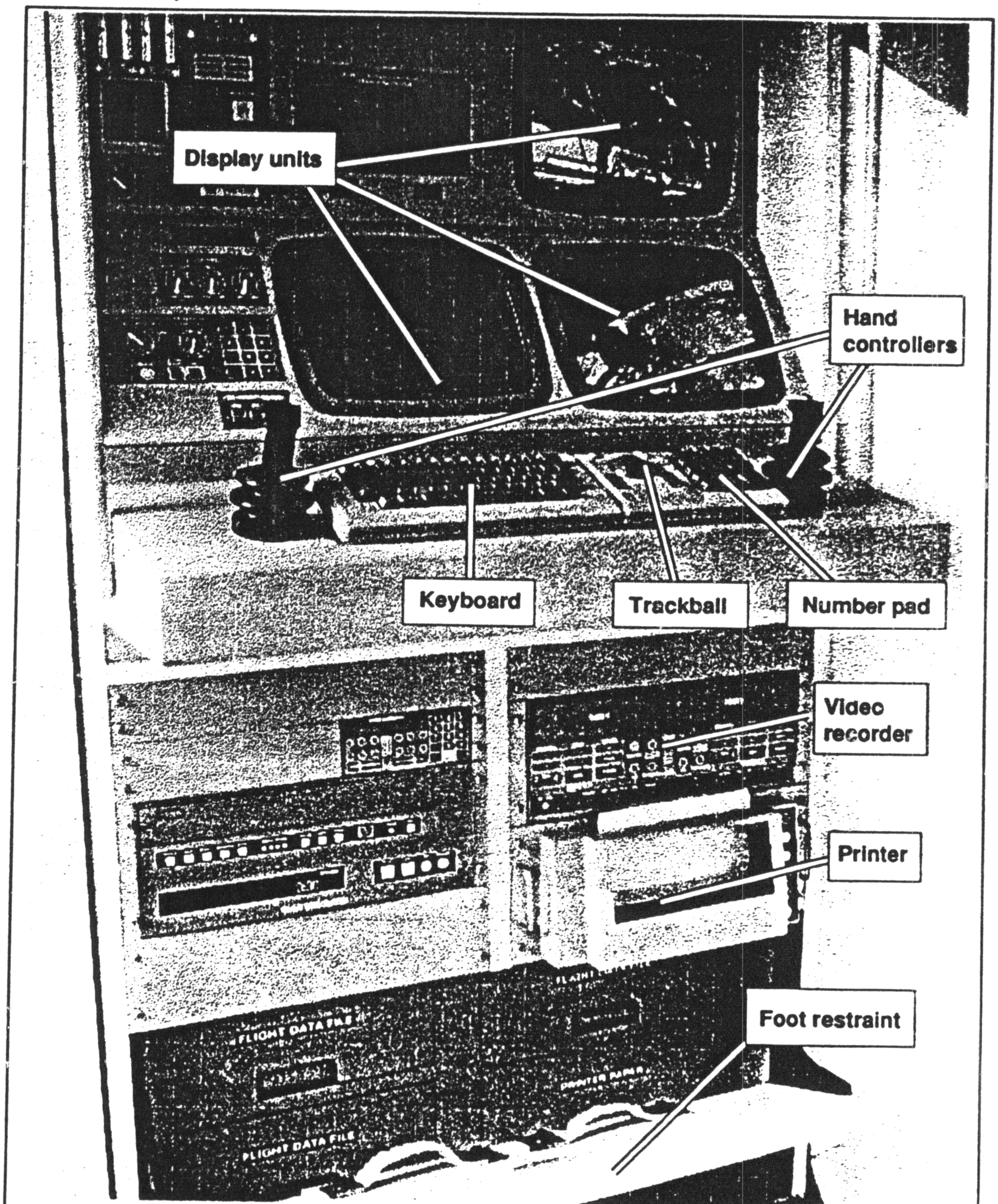
Wickens, C. D. (1984). *Engineering psychology and human performance.* Columbus, OH: Merrill.

Wiener, E. L., and Nagel, D. C. (Eds.) (1988). *Human factors in aviation.* San Diego, CA: Academic Press.

Yntema, D. B. (1963). Keeping track of several things at once. *Human Factors, 5* /1, 7-17.

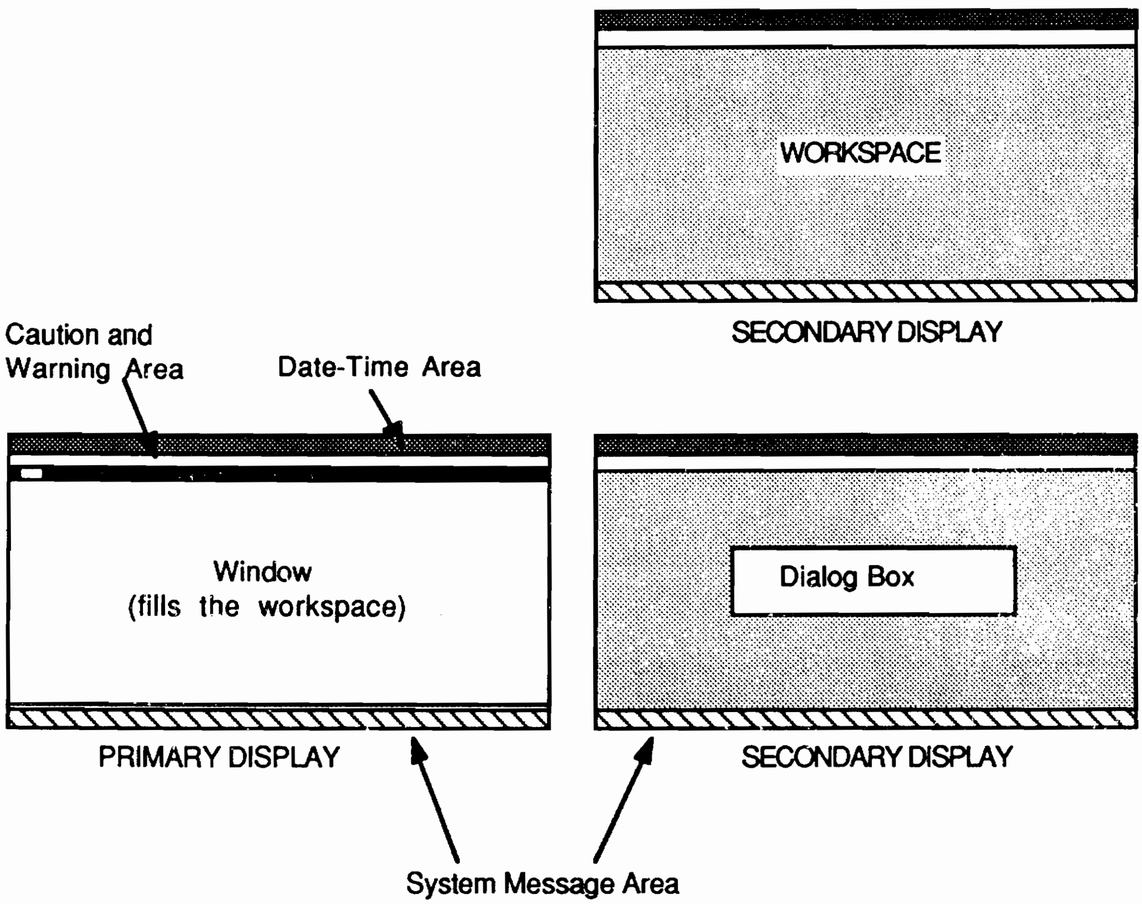
APPENDIX I

SPACE STATION FREEDOM CONCEPTUAL WORKSTATION MOCK-UP DESIGN


Display units
Hand controllers
Keyboard
Trackball
Number pad
Video recorder
Printer
Foot restraint
FLIGHT DATA FILE
PRINTER PAPER
FLIGHT DATA FILE

APPENDIX II

USER INTERFACE DISPLAY UNIT DESIGN REQUIREMENTS SPECIFICATION


WORKSPACE
SECONDARY DISPLAY
Caution and
Warning Area
Date-Time Area
Window
(fills the workspace)
Dialog Box
PRIMARY DISPLAY
SECONDARY DISPLAY
System Message Area

APPENDIX III

SUBJECT QUESTIONNAIRE

Sex ______

Age ______

Occupation ________________

Do you have normal or corrected vision? yes no

Do you have any motor impairments that prevent you from using a computer mouse? yes no

Have you worked on Macintosh computers? yes no

If **yes**, how often do you use the Macintosh? (circle one)

daily occasionally very little

If **no**, do you work with any other type of computer? yes no

Type ____
How Often? (circle one)

daily occasionally very little

For what purpose(s) do you use the Macintosh? (circle as many as needed)

a. word processing (some or mostly)
b. programming (little or intense)
c. graphics (some or mostly)
e. computer games
f. data analyses
g. other ________________________________

On the average, how many computer applications do you use simultaneously while you're working on a computer? ______

On the average, how many different tasks do you perform simultaneously on a computer? ______

On the average, how many windows do you have open on the computer screen at one time? _______

Do you play computer or video games? yes no

If **yes**, how often?

daily
often (3 to 4 times a week)
occasionally (< 3 times a week)
rarely (once a month)
hardly ever (once every several months)

How would you describe your perceived ability to keep track of several changing things at once (i.e., 3 or more different things)?

a. I would probably have no difficulty
b. I can do it if I have to, but prefer focusing my attention on fewer than 3 changing things
c. I would probably have some difficulty
d. I don't think I could keep track of several changing things at once

APPENDIX IV

PARTICIPANT CONSENT FORM

Participant's Informed Consent Form

The following experiment is a study concerning multiple activities and task interruptions. As a participant in this experiment, you have certain rights as explained below. The purpose of this document is to describe these rights and to obtain your written consent to participate in the experiment.

1. You have the right to discontinue your participation in the study at any time for any reason. If you decide to terminate the experiment, inform the researcher.

2. You have the right to inspect your data and withdraw it from the experiment if you feel that you should for any reason. In general, data are processed and analyzed after a subject has completed the experiment. At that time, all identification will be removed and the data treated with anonymity. Therefore, if you wish to withdraw your data, you must do so immediately after your participation is completed.

3. You have the right to be informed of the overall results of the experiment. If you wish to receive a synopsis of the results, include your address with your signature below.

This research is being supported by NASA's Office of Aeronautics and Exploration Technology. This particular study is part of a larger research program designed to understand multitasking for applications to the human-computer interface on spacecraft. Your cooperation is appreciated.

The principal investigator and researcher is Darlene Merced. The Human-Computer Interaction Laboratory (HCIL) manager is Dr. Marianne Rudisill. These persons may be contacted at SP34 extension 33703.

If you have any questions about the experiment or your rights as a participant, please do not hesitate to ask. The researcher will do her best to answer them provided it does not pre-bias the experiment.

____________________________ ____________________________

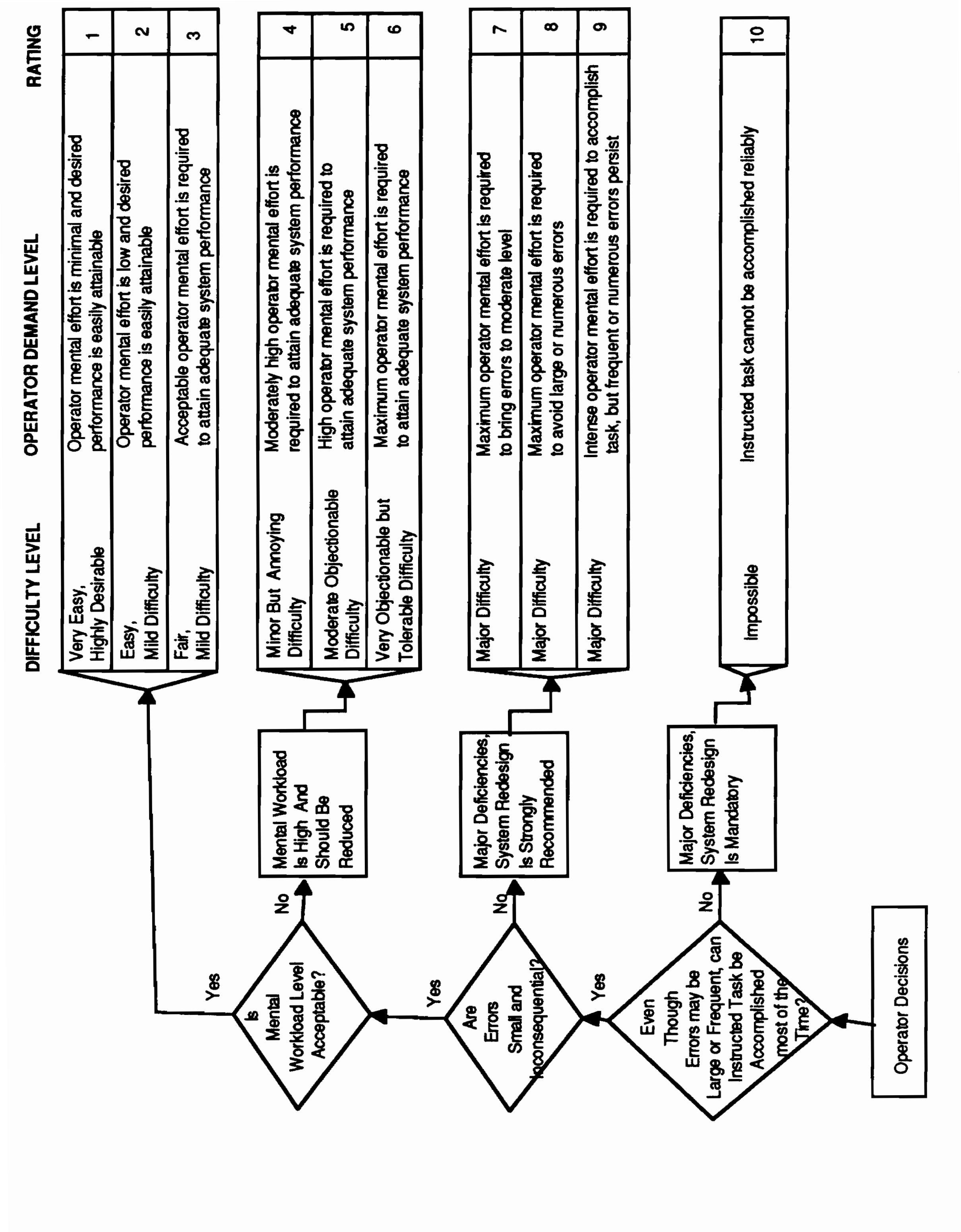
Participant Signature

____

Print address or mail code to receive summary of the results

APPENDIX V

MODIFIED COOPER-HARPER RATING SCALE


DIFFICULTY LEVEL
OPERATOR DEMAND LEVEL
RATING
Very Easy, Highly Desirable
Operator mental effort is minimal and desired performance is easily attainable
1
Easy, Mild Difficulty
Operator mental effort is low and desired performance is easily attainable
2
Fair, Mild Difficulty
Acceptable operator mental effort is required to attain adequate system performance
3
Minor But Annoying Difficulty
Moderately high operator mental effort is required to attain adequate system performance
4
Moderate Objectionable Difficulty
High operator mental effort is required to attain adequate system performance
5
Very Objectionable but Tolerable Difficulty
Maximum operator mental effort is required to attain adequate system performance
6
Major Difficulty
Maximum operator mental effort is required to bring errors to moderate level
7
Major Difficulty
Maximum operator mental effort is required to avoid large or numerous errors
8
Major Difficulty
Intense operator mental effort is required to accomplish task, but frequent or numerous errors persist
9
Impossible
Instructed task cannot be accomplished reliably
10
Mental Workload Is High And Should Be Reduced
Major Deficiencies, System Redesign Is Strongly Recommended
Major Deficiencies, System Redesign Is Mandatory
Is Mental Workload Level Acceptable?
Are Errors Small and Inconsequential?
Even Though Errors may be Large or Frequent, can Instructed Task be Accomplished most of the Time?
Yes
No
Yes
No
Yes
No
Operator Decisions

APPENDIX VI

EXPERIMENTAL INSTRUCTIONS

Instructions

You will be asked to perform two different types of hypothetical tasks. The first task involves monitoring 1, 3, or 5 simulated Space Station Freedom tasks. The second and more important task will be to perform an emergency Guidance, Navigational, and Control (GN&C) interruption task.

While performing the monitoring task, you will be interrupted by 1) three beeps and 2) an emergency GN&C textual message.

Each monitoring task will be presented as a separate display window. Your task will be to:

1. Detect a change (i.e., a field update) in the task window(s) you are asked to perform.
2. Make a comparative decision of the information that has been updated/presented against a specific system recommended criteria for every monitoring task (or window) you are performing.
3. Perform the emergency GN&C task when you receive the interruption message. Following the GN&C interruption auditory beep signal and textual message, your task will be to:

    a. Finish making your last comparative decisions for each task you are performing.

    b. Close **all** the monitoring/test display windows by clicking the upper left-hand "close box".

    c. Open the GN&C window by clicking the "Open" button located in the C&W (caution and warning) display area.

    d. Perform the GN&C navigational correction procedures (i.e., comparing incoming data to a specific system average value).

4. Reopen **ONLY** the monitoring task(s) you were performing before the interruption and complete them. Use the "Systems" menu to find the tasks to find the tasks.

5. Complete the Rating Scale after each trial.

Remember **SPEED** and **ACCURACY** are important

VITA

DARLENE MERCED

Permanent Address

4445 Nasa Road One, Apt. 378
Seabrook , Texas 77586
(713) 326-4746

EDUCATION

M.S. Industrial and Systems Engineering, February 1991
Human Factors Engineering concentration
Virginia Polytechnic Institute and State University, Blacksburg, Virginia

B.S. Liberal Arts and Sciences, October, 1985
Engineering Psychology concentration
University of Illinois at Urbana-Champaign, Urbana, Illinois

WORK EXPERIENCE

January 1991 - present
Computer Programmer
Econ, Inc., Houston, Texas

January 1990 - October 1990
Human Factors Engineer, Man-Systems Division
Cooperative Education Program
NASA Johnson Space Center, Houston, Texas

March 1989 - December 1989
Programs Support Technician
Management Systems Laboratories, Blacksburg, Virginia

September 1988 - December 1989
Laboratory Assistant
Human-Computer Interaction Laboratory
Virginia Polytechnic Institute and State University, Blacksburg, Virginia

August 1987 - August 1988
Claims Adjuster
Marshall Field's, Chicago, Illinois

PROFESSIONAL AFFILIATIONS

Human Factors Society
Society for Information Display
Alpha Pi Mu Industrial Engineering Honors Society

PERSONAL DATA

Born in Chicago, Illinois on September 16, 1963
Nationality: of Puerto Rican descent
Languages: Fluent in Spanish and English