To the Graduate Council:

I am submitting herewith a dissertation written by John V. Draper entitled "A Comparison of Manual, Voice and Automated Control of a Teleoperator Viewing System." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Industrial and Organizational Psychology.

John M. Larsen

John M. Larsen, Major Professor

We have read this dissertation
and recommend its acceptance:

Accepted for the Council:

Vice Provost
and Dean of The Graduate School

# A COMPARISON OF MANUAL, VOICE, AND AUTOMATED CONTROL OF A TELEOPERATOR VIEWING SYSTEM

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

John V. Draper

March 1987

## ACKNOWLEDGEMENTS

A number of persons have generously assisted me during the course of this research. I am particularly indebted to Professor John M. Larsen, who served as chairman of my dissertation committee; Professor Stephen J. Handel, who has provided a great deal of professional guidance and encouragement; Dr. Margaret M. Clarke of Rockwell International Corporation, who was also source of guidance and encouragement during the course of data collection; Dr. William R. Hamel of Martin Marietta Energy Systems, Inc., who provided the equipment and financial support for the research and who also served on my dissertation committee; and Joseph N. Herndon of Martin Marietta, who replaced Dr. Hamel as Task Leader of the Remote Control Engineering task and continued to support this work. With their help in the form of consultation, criticism, and encouragement, I was able to achieve results otherwise impossible. However, any errors or omissions that remain in my work are solely my responsibility.

I owe special thanks to Dr. George W. Baxter of King College, Bristol, Tennessee, for his guidance and supervision during my undergraduate preparation. I am also indebted to the management and staff of the Fuel Recycle Division of the Oak Ridge National Laboratory for their assistance during the course of this research. I have already mentioned W. R. Hamel and J. N. Herndon; I am also grateful to W. D. Burch, M. J. Feldman, D. E. Spangler, and others too numerous to mention, who directly or indirectly supported this investigation.

The study was sponsored by the Office of Facilities, Fuel Cycle, and Test Program, U.S. Department of Energy, under contract number DE-AC-05-84OR21400 with Martin Marietta Energy Systems, Inc.

ABSTRACT

Teleoperators are devices used to perform tasks in areas which are hazardous for humans to enter. The Oak Ridge National Laboratory's Consolidated Fuel Reprocessing Program is developing a teleoperator to perform maintenance and repair of future nuclear fuel reprocessing facilities. This effort involves designing optimal human-machine interfaces for the teleoperator cockpit. Current conceptual designs specify a two-operator team since it is thought that the demands of teleoperation are best divided into manipulative and sensory/locomotive areas of responsibility. However, in some circumstances it will be necessary to perform remote maintenance tasks with one operator. In this case, operation of manipulators and television cameras will interfere with one another to the detriment of overall teleoperator performance, since they place conflicting demands on the operator's finite information processing resources. This experiment investigated options for camera controls which may be less interfering with manipulator operations than conventional manual controls. Specifically, the experiment compared performance of remote maintenance tasks with control by computer word recognition and automation of camera aiming with performance using conventional, manually controlled pushbuttons and joysticks.

Operators worked more slowly with manual control, and were engaged in control motions for significantly smaller percentages of total task time than with either of the other control systems. There were no differences between voice control and automated control. There was no

difference between manual control and voice control in the number of camera changes made; the automated tracking system made significantly more changes than the operators did when using either of the other control systems. Camera movements were of longest duration when operators used voice control, and changes made by the automated control system were of longer duration than changes made with manual control.

The results indicate that either voice control or automated tracking is preferable to manual control in the solo-operator situation. There was little to distinguish between voice control and automated tracking; however, operators retain control over camera functions with voice control, so it is the preferred control system.

TABLE OF CONTENTS


| CHAPTER | | PAGE |
|---|---|---|
| 1. | INTRODUCTION | 1 |
| | 1. TELEOPERATORS | 1 |
| | 2. METHODS FOR IMPROVING THE HUMAN-MACHINE RELATIONSHIP | 3 |
| | 3. PURPOSE OF THE STUDY | 7 |
| | 4. ORGANIZATION OF THE PAPER | 10 |
| 2. | THE MULTIPLE RESOURCE MODEL | 11 |
| | 1. INTRODUCTION | 11 |
| | 2. THE MULTIPLE RESOURCE MODEL | 11 |
| | 3. WORKLOAD | 22 |
| | 4. TASKS AND CONTROLS | 27 |
| 3. | METHODS | 36 |
| | 1. SUBJECTS | 36 |
| | 2. APPARATUS | 37 |
| | 3. DEPENDENT VARIABLES | 48 |
| | 4. EXPERIMENTAL DESIGN | 54 |
| | 5. PROCEDURE | 55 |
| 4. | RESULTS | 58 |
| | 1. METHOD OF ANALYSIS | 58 |
| | 2. MANOVA RESULTS | 58 |
| | 3. ANOVA FOR REPEATED MEASURES DESIGNS | 60 |
| | 4. STRATEGY OF ANALYSIS | 64 |
| | 5. RESULTS | 66 |
| | 6. SUMMARY OF RESULTS | 99 |
| 5. | DISCUSSION AND RECOMMENDATIONS | 104 |
| | 1. DISCUSSION | 104 |
| | 2. RECOMMENDATIONS | 113 |
| REFERENCES | | 119 |
| APPENDIX | | 126 |
| VITA | | 131 |

LIST OF TABLES


| TABLE | | PAGE |
|---|---|---|
| 1. | Components of the Remote Systems Development Facility (RSDF) . . . . . | 39 |
| 2. | Control system presentation schedules . . . . . | 57 |
| 3. | Averages of within-operator correlations between dependent variables . . . . . | 60 |
| 4. | Expected values of mean squares and mean square ratios for F tests . . . . . | 65 |
| 5. | ANOVA results for logarithm to base 10 of task completion time. . . . . | 67 |
| 6. | Orthogonal, normalized contrasts for the effect of camera control type . . . . . | 69 |
| 7. | Mean square ratios for camera control type . . . . . | 69 |
| 8. | ANOVA results for the average rate of errors committed per minute during task completion . . . . . | 75 |
| 9. | ANOVA results for the percent of time performing slewing motions . . . . . | 78 |
| 10. | ANOVA results for percent of time performing fine-adjusting motions . . . . . | 85 |
| 11. | ANOVA results for number of camera changes . . . . . | 90 |
| 12. | ANOVA results for percent of time cameras were in motion . . . . . | 97 |

LIST OF FIGURES


| FIGURE | | PAGE |
|---|---|---|
| 1. | Hypothetical performance operating characteristic curves for the driving example . . . . . | 20 |
| 2. | The remote work cell at the Remote Systems Development Facility (RSDF) . . . . . | 40 |
| 3. | The TeleOperator Systems (TOS) SM-229 manipulators (slave portion) in the RSDF work cell . . . . . | 41 |
| 4. | The RSDF control room . . . . . | 42 |
| 5. | TOS SM-229 master controller handles . . . . . | 43 |
| 6. | Operator seated at the master controllers within the RSDF control room . . . . . | 44 |
| 7. | Thirty-inch flange task mock-up (at top with crane hook attached) in the RSDF remote work cell . . . . . | 49 |
| 8. | Mean of logarithm to base 10 of task completion time for each control system . . . . . | 70 |
| 9. | Mean of logarithm to base 10 of task completion time for each task . . . . . | 72 |
| 10. | Mean of logarithm to base 10 of task completion time for each block of testing . . . . . | 73 |
| 11. | Mean percent of time engaged in slewing motions for each task . . . . . | 81 |
| 12. | Mean percent of time engaged in slewing motions for each block of testing . . . . . | 82 |
| 13. | Interaction of control type, task performed, and block of testing for percent of time engaged in slewing motions . . . . . | 83 |
| 14. | Mean percent of time engaged in fine-adjusting motions for each camera control system . . . . . | 87 |
| 15. | Mean percent of time engaged in fine-adjusting motions for each task . . . . . | 89 |
| 16. | Mean number of camera changes per trial for each camera control type . . . . . | 92 |

FIGURE PAGE

17. Mean number of camera changes per trial for each task . . . . . 93

18. Mean number of camera changes per trial for each block of testing . . . . . 94

19. Interaction of control type and task for the mean number of camera changes per trial . . . . . 95

20. Mean percent of total task completion time that cameras were in motion for each camera control type . . . 98

21. Camera movement duration for each camera control system . . . . . 107

# CHAPTER 1

# INTRODUCTION

## 1. TELEOPERATORS

Technology extends human capabilities by providing machines that do things which people cannot. One example of such a machine is the teleoperator, a system which extends human sensory and motor abilities by coupling human operators with remote sensors and remotely controlled output devices. Corliss and Johnsen (1968) define a teleoperator as a "general purpose, dexterous, man-machine system that augments man by projecting his manipulatory and pedipulatory capabilities across distance and through physical barriers into hostile environments." Looking at teleoperators from another perspective, teleoperators are devices that extend the capabilities of robots by installing a human in the role of central processing unit, thereby increasing decision making power and flexibility.

Teleoperators do work in places people cannot enter without being injured or donning protective gear, such as outer space or radioactively contaminated areas. Teleoperators are safer than suited personnel because the human operator does not actually enter the dangerous area, and they are more effective than robots when tasks demand decision making power. They are also better at coping with unexpected or unusual events because teleoperators include human operators in the control loop; humans are better than machines at handling unanticipated events and

unstructured tasks (McCormick & Sanders, 1982; Swain, cited in Price, Maisano, & Van Cott, 1982).

The type of teleoperator that is the focus of this study is the bilateral master/slave manipulator (MSM). Manipulators are remote handling devices which mimic the function of the human arm and hand. Master/slave manipulators are coupled manipulator arms, one (the master) in the cockpit and one (the slave) in the hazardous area. The slave replicates the position of the master (hence the cockpit portion is sometimes called a replica master), so operators can move the slave by moving the master. The term bilateral means that the slave can also move the master (this is sometimes called back-driving); in this way the operator can receive force-feedback from the slave's environment, which gives a limited sense of touch to the operator.

The performance of the human operator is a critical determinant of overall teleoperator performance. The operator functions like a homunculus for the machine, receiving sensory data via television and other devices, processing this data and making decisions, and then causing responses to occur in the remote environment by manipulating controls in the teleoperator cockpit. The speed and accuracy with which the teleoperator does tasks depends on how speedily and accurately the operator can perceive, process, and respond to information transmitted by the machine, as well as on the speed of the machine itself. Information from the remote scene must flow as quickly and accurately as possible to the operator and control outputs must flow back with speed and precision; therefore, to achieve the highest levels of performance the relationship of human and machine must be optimized. Unfortunately,

teleoperators are very complex in terms of the input and output demands they place on operators. Master-slave manipulators (MSM) have become so demanding to operate that the current trend in the United States is toward multi-operator teams of up to seven persons.[1] Complexity results from efforts to make MSMs more flexible, i.e. able to do a wider variety of tasks, which requires the addition of subsidiary systems like cameras, remote tools, remotely controlled cranes, multi-degrees of freedom transporters, etc. These place additional demands on operators and make the human-machine relationship even more critical. The increased complexity of teleoperators makes it necessary to attend to how the operator will interact with the hardware. Otherwise, the system may not operate at its full potential due to a bottleneck which may develop in information flow between operator and machine.

## 2. METHODS FOR IMPROVING THE HUMAN-MACHINE RELATIONSHIP

There are four ways to improve the relationship between operator and hardware. First, operators may be selected to fit the system. The crew can be picked so that only those persons who fit the machine are included and persons who do not have the necessary dimensions or abilities are kept out of crews. Improvements in performance using this method result from making people fit the hardware system and are only

[1] Survey conducted by Dr. M. M. Clarke for the Oak Ridge National Laboratory's Fuel Recycle Division in 1982. The survey was done to help determine the optimal crew size for the cockpit of the Advanced Servomanipulator (ASM) developed by the Fuel Recycle Division's Remote Control Engineering task. For more information on the ASM and the ASM control room, see Clarke and Kreifeldt (1985a, b).

possible if the demands of the machine are within the limits of the characteristics of the user population. Effective selection also requires that the operator characteristics that produce good performance are known and can be accurately measured.

Second, operators may be trained to use the system. Training can improve system performance by making the behaviors of operators more compatible with the demands of the machine. Training has limitations, however. Certain characteristics of operators, like height and the visible spectrum, cannot be modified by training. It is impossible to train operators to use a machine which makes demands that surpass the physiological and psychological limitations of people. Like selection, training is aimed at modifying the crew to fit the system. It requires knowledge of the characteristics which result in good performance and how to improve these characteristics.

Third, the task may be changed to reduce demands on operators, either by increasing the size of the crew (dividing the demands of the task among several persons) or by rearranging task elements (making bolts easier to reach, for example). Suitable functions can also be allocated to machine control, with the same effect on operators as increasing team size. The problem with teams of operators is that multiple operator cockpits are more expensive in dollars and volume to build, require more operator-hours to run, and require the availability of a larger pool of trained persons. Team operated systems also function below par in the event of situations that preclude assembling a full crew, for example because of short preparation time in an emergency. This problem is particularly salient in handling radioactive waste

because mistakes can cause long lasting contamination, and is no less important in space where a mistake may prevent completion of a mission or cause damage to vehicles. Cockpit volume and crew size are also critical problems for space applications since the payload of modern space vehicles is limited.

Changes in tasks are not possible when task designs are dictated by factors beyond the control of the human-machine system designer. For example, a nuclear facility designed to be maintained by crane and impact wrench may be converted to maintenance by teleoperator. The designer of the teleoperator would have no control over the equipment already in the facility and so could not do much to change the tasks that the system would be required to do. Inability to specify the tasks and environments teleoperators work in is the reason for the proliferation of complexity; in order to be flexible enough to handle a wide range of tasks teleoperators must be designed to be highly dexterous and must be equipped with a wide variety of subsidiary systems. Increasing dexterity increases component complexity; increasing the number of components increases the complexity of the system as a whole.

Changing the crew size or the task uses a different intervention philosophy than either selection or training; it takes the human as a given and changes the system to fit the operator. The fourth form of intervention also takes the human as a given but makes changes in the machine rather than in the task. This approach is human-machine interface design, which modifies the machine to better accommodate people. This type of intervention can be of any scope, from making sure the knobs and dials on a control panel are the right size (see Chapter 5 in

MIL-STD-1472c, 1981; or Chapter 1 in MIL-HDBK-759a, 1981 for examples) to programmatic research into a particular design problem (for example see Bradford, Cooper, Garcia, Wildgen, and Spencer, 1984). Good interfaces promote good system performance and optimize the requirements on operators by altering the way machines transmit and receive information. Good human-machine interfaces provide the most appropriate match to the way humans receive and transmit information. At the same time selection and training requirements may be reduced, since constraints placed on the human operator by the machine are reduced. In addition, the knowledge gained by study of the human-machine interface can provide data pertinent for constructing appropriate selection devices, performance measures, and training procedures.

Ideally, the several interventions do not take place independently. Selection procedures, training programs, and hardware should be designed within a systems framework, since each of these subsystems affects each of the other subsystems. For example, suppose the hardware subsystem includes a 9 kilogram (20 pound) remotely controlled wrench. This may have an impact on selection since there is a significant section of the population of potential operators who lack sufficient strength to handle this load for long periods (Greenberg & Chafin, 1974). These people should not be included in crews if handling the wrench is an important part of most tasks. It affects training since operators will have to learn how to use the wrench correctly. It also affects other hardware design decisions: the bolts must be accessible to the bulk of the wrench, for example. Unfortunately, in equipment design future selection and training requirements are often neglected.

## 3. PURPOSE OF THE STUDY

In this study a problem in the design of an advanced teleoperator was addressed using task modification and the human- machine interface design approach. The problem was how best to control television cameras mounted on the teleoperator, and the solution was discovered through an experiment involving simulated remote maintenance tasks.

The Consolidated Fuel Reprocessing Program (CFRP) of the Oak Ridge National Laboratory is developing technology for future reprocessing of fuel from nuclear reactors. One part of CFRP, the Remote Control Engineering Task (RCE), is engaged in the design, fabrication, and testing of an advanced, electronic, bilateral master/slave manipulator for maintenance and repair of reprocessing equipment in high radiation environments. The basic philosophy of RCE, expressed by Hamel and Feldman (1984) and Herndon and Hamel (1984), is that operators using television as their primary sensory channel and bilateral MSMs as their motor channel can maintain and repair process equipment for the life of a reprocessing facility. RCE is responsible for developing teleoperator systems capable of performing this function.

The work at RCE includes designing optimal human-machine interfaces for the cockpit of the teleoperator. The basic conceptual design of the cockpit has been completed, and is outlined in Clarke and Kreifeldt (1984a, b). This design calls for a two person crew, one operator handling all the manipulator functions and the other operator controlling subsidiary equipment such as cameras, lights, tools, cranes, and a three degree of freedom transporter. This allocation of tasks

makes the RCE cockpit vulnerable to performance degradation in the event of operations with one crew member (for example in emergencies or for simple tasks assigned to one person) since the manipulator operator has no means for controlling remote cameras. Without camera controls manipulator operators cannot use the teleoperator to full advantage since they cannot optimize the information received from the remote environment.

In order to close this window of vulnerability the manipulator operator in the RCE cockpit must be given some means to control remote cameras. However, the task allocation in the RCE cockpit was based on careful study of the requirements of the hardware system and the capabilities of humans (Clarke & Kreifeldt, 1984a, b), and the separation of manipulator tasks from all other functions reflects an important truth about remote manipulation: it is a cognitively and physically taxing activity. The addition of camera controls to the manipulator operator's duties (though necessary in light of the potential for solo operations) may significantly increase the burden on these already heavily loaded persons. Therefore it is important to provide the best possible human-machine interface for controlling cameras.

Alternatives for camera control systems were initially evaluated in light of two criteria: first, the control systems were required to be available at the present time, i.e. futuristic concepts not yet built were excluded; second, systems that required that operators wear headgear were excluded because of the fatigue, discomfort, and cost associated with head mounted systems. Three camera control alternatives passed this preliminary culling: control by manual inputs from the

operator using joysticks, knobs and switches; control by verbal inputs from the operator, using a computer word recognition system; and machine control of the cameras using automatic tracking of the manipulator's end effectors. Hypotheses about the effects of the control systems were developed from the framework of the structured resource model described in Chapter 2; this model was also used to develop dependent measures for the study.

All three of the systems were installed in the Remote Systems Development Facility (RSDF), a facility devoted to development and demonstration of advanced remote maintenance technology. Four experienced operators performed a set of tasks typical of what might be encountered in reprocessing plant maintenance, and the time to task completion, number of errors, rate of control motions, and operators' ratings of subjective workload were recorded. Two experimental design problems demanded special attention in this study: small sample size and learning effects. To collect data that was reliable, the subjects had to be experienced manipulator operators. Unfortunately, the pool of experienced operators at CFRP is small; the sample of four used in this study represents 80% of the experienced operators. It was not practical to train subjects given the initial time constraint, so the experiment used a repeated measures design. This satisfies the requirements of internal validity and statistical conclusion validity (Cook & Campbell, 1979), but makes generalization of the results to a wider user population somewhat problematic.

Learning effects were handled by allowing subjects to practice with the control systems until their performance became fairly stable and by

counterbalancing the order of task and control system presentation during experimental trials. The operators completed each task six times per camera control system in two blocks of three testing sessions. A session consisted of three trials per task with a single camera control system. The order of presentation of camera controls within blocks of testing was selected so that no control system or task was favored over any other in terms of being consistently presented later than others.

## 4. ORGANIZATION OF THE PAPER

Chapter 2 describes a model of humans as information processors which was used to generate predictions about the effects of the three camera control systems and to develop dependent variables to measure these effects. Chapter 3 describes the methods and apparatus used in the study, and Chapter 4 describes the results of the data collection and statistical analyses. Chapter 5 discusses the results in terms of the model described in Chapter 2 and concludes with a set of recommendations for the RCE cockpit.

# CHAPTER 2

# THE MULTIPLE RESOURCE MODEL

## 1. INTRODUCTION

McCauley (1984) suggested that the best way to conceptualize the effects of word recognition control systems on system performance and human operators is in terms of a multiple resource information processing model, since there seems to be an important difference in the cognitive aspects of control with word recognition versus manual control. Multiple resource models describe the way in which persons manage the task of processing information inputs and converting them to outputs. This chapter will describe and apply such a model to the effects of all three of the camera control systems under study.

## 2. THE MULTIPLE RESOURCE MODEL

### Processing Resources

Miller (1956) proposed that people have a limited capacity for processing information. According to Miller, people can be considered as communication channels with an information input, a throughput (processing), and an information output. Up to a point, the amount of information transmitted increases with information input. However, beyond that point the human channel becomes saturated with information and further inputs are not transmitted. The point at which the human transmission channel becomes saturated is the channel capacity of the system, typically 5 to 9 bits of information according to Miller (1956).

Norman and Bobrow (1975) explained the limitations on the human information processing system by invoking the concept of resources. These authors say that humans as information processors have "programs and some mechanism for executing those programs." When programs are executed in order to process information inputs, resources are consumed. Resources are "such things as processing effort, . . . memory capacity, and communication channels" and are finite. The ability of persons to change inputs to outputs is limited by the finite nature of their resource pools since resources provide the means by which processing programs are performed.

Norman and Bobrow (1975) termed the combination of programs being executed to perform a function and the resources used to execute those programs a process. Processes may be resource-limited processes, which are affected by the amount of resources allocated to the performance of the process, or data-limited processes, which are not affected by resource allocation. Generally, for resource-limited processes information output increases with increased allotments of resources.

The nature of the task being performed by a process and/or the level at which the process is operating may determine whether it is resource-limited or data-limited. Some processes are by nature data-limited; Norman and Bobrow (1975) use the example of detection of a weak signal against noise, in which case performance depends on the signal to noise ratio and is unaffected by the quantity of resources allocated to detecting the signal. Resource-limited processes respond to the resources allocated to them, although below some minimum value performance is not improved by adding resources and at some higher level the curve

that describes the relationship of output to resources (the performance-resource function) asymptotes. Above the asymptotic value and below the minimum these processes are data-limited: in the portion of the function between the threshold and the asymptotic value they are resource-limited. This means that people can affect their performance of tasks through the allocation of resources, within limits imposed by the nature of the task and by their own processing capabilities.

Navon and Gopher (1979, 1980) extended these concepts by hypothesizing that the human information processing system is not a single channel but rather a set of channels, each with a specialized set of programs for processing specific types of information and each with a limited capacity for information transmission. Navon and Gopher (1979) opined that the channels of the human processing system are arranged in a structure which is specific to input/output modalities, information characteristics, and/or temporal sequence. Each channel has a pool of resources which can be committed in varying amounts to the processing of information inputs. The size of the pool of resources is finite and determines the capacity of the channel for processing information. Since the resources available to a channel are limited and cannot be expanded, the total amount of resources available to a person at any point in time is immutable. However, in the long term training may improve allocation skills or the efficiency of resources for handling information (Gopher, 1980).

According to Navon and Gopher (1979) task performance is the product of the resources committed to a task and the efficiency of the resources, or

$$p_x = r_i e_i \qquad (1.0)$$

where $p_x$ = performance on task x,

$r_i$ = the amount of resources allocated to task x from pool i, and

$e_i$ = the average efficiency of the resources from pool i.

Norman and Bobrow (1975) termed the function relating task performance to resource commitment (holding resource efficiency constant) the performance-resource function.

Resource efficiency ($e_i$ in formula 1.0) is the amount one unit of resources can raise the level of process performance and is determined by subject-task parameters. Subject-task parameters are the characteristics of the performer, the task, and the environment (Navon & Gopher, 1980). Performer skill, physical and cognitive requirements of the task, and characteristics of the task environment (like ventilation, lighting, or noise) are examples of subject-task parameters. Subject-task parameters place constraints on performance by the combination of task, environment, and performer characteristics.

One important aspect of the performance-resource function is that it is not a monotonically increasing function. This is due to the variability of resource efficiency through the range of the function. As performance increases, at some point the efficiency of resources for further improvement decreases, which causes the performance-resource function to follow the law of diminishing returns. Task performance level is itself a subject-task parameter, and as the performance level improves the requirements for further improvements constrain the system.

## Task Difficulty

The amount of resources required to perform a task at a criterion level of performance is termed the demand of a process by Navon and Gopher (1980). Task difficulty is synonymous with demand, and may be manipulated by decreasing resource efficiency, by raising the criterion performance level, by reducing the time available for processing, or by increasing the required frequency of responses.

Resource efficiency may be manipulated by changing subject-task parameters. The more constrained performance is by subject-task parameters the lower efficiency and the larger the amount of resources that must be committed to achieve a given level of task performance. In other words, the demand of a criterion performance level increases as resource efficiency decreases. Riley and Cochran (1984) afford an example of this type of manipulation: by reducing the temperature in an environmental chamber these authors were able to produce decrements in the performance of subjects on a set of dexterity tests. In this case the temperature decrease reduced the efficiency of the subjects for doing the tasks; this reduction was reflected in lower task performance time. Had the subjects been required to perform the tasks in some specified time, the efficiency reduction would have manifested itself as an increase in resource expenditure and concomitant increase in subjective workload.

Task difficulty may be increased by raising the task performance criteria, requiring the allocation of a larger number of resources to perform the task at the criterion level. Since performance is the

product of resource efficiency and resource amount (see formula 1.0), raising the level of criterion performance while maintaining resource efficiency requires an increase in the amount of resources expended in order to meet the new criteria and therefore, increases task difficulty.

Many tasks are not paced by the performer. In this case performance is usually measured by accuracy rather than by the speed with which the task is done. The amount of resources necessary to perform such a task may be raised by decreasing the time available to do the task, which increases difficulty by compressing the load (the amount of output needed to satisfy the criteria) of the task in time. Granaas, McKay, Laham, Hurt, and Juola (1984) describe a recent study that is an example of this type of manipulation. These authors investigated the effect of scrolling speed on the accuracy of recall of text displayed on a cathode-ray tube. Accuracy was significantly degraded by higher levels of speed, indicating that task difficulty increased as processing time was shortened. In this case subjects were unable or unwilling to increase their allotment of resources to task performance (no criterion performance level was prescribed) and increased difficulty caused a degradation in performance.

Finally, the difficulty of continuous tasks can be increased by increasing the criterion rate of processing outputs, i.e., by increasing the required frequency of responses. This is done in tracking and simulator tasks by increasing the frequency of the task, for example increasing the number of course disturbances per unit of time (see Wierwille & Connor, 1983, for example). This also increases the output load

on task performers, which requires allocation of more resources to reach a criterion level of performance.

## Task Interference and Resource Allocation

When a person performs two or more tasks which tap the same resource pools the tasks interfere with each other and the performer must make a decision about where to allocate resources (Norman & Bobrow, 1975). Think of this in terms of matrix algebra: a vector containing the processing resources associated with each resource pool describes the total processing capacity of the system, and a corresponding vector of processing needs describes the resource requirements of any task. In a multi-task situation a matrix exists which contains the vectors of all possible joint performance levels. Each set of performance levels (each performance vector in the matrix) has some valence for the performer, and the performer allocates resources in a manner which produces the performance outcomes with the highest positive valence.

Consider the task of driving an automobile while carrying on a conversation with a passenger, and suppose three joint performance levels are possible: good driving/poor conversation, medium driving/medium conversation, and poor driving/good conversation. On a long straight stretch of road with no traffic, poor driving performance is adequate and the driver's companion is telling an interesting story. Therefore, the vector [poor driving, good conversation] is the most attractive for the driver. Another driver who had witnessed a collision on the same stretch of road and who had already heard the story might assign higher valence to the vector [good driving, poor conversation]. The

performance attempted is the vector with the highest positive valence, and the performer allocates resources to satisfy the requirements of the tasks for the selected joint performance levels.

Performance level selection and resource allocations are generally automatic and unconscious: people pay attention to what seems important, often without being aware of making a decision. On the other hand, attention allocations may be conscious and require some effort, as in the case of the football fan trying to watch the big game while his wife is detailing the things that need to be done around the house. The fan tunes his wife out, i.e., he allocates his input resources to the task of watching the game.

As in the single task situation, for multi-task performance the amount of resources allocated to a task and the efficiency of resources determine level of performance. In the multi-task situation the formula can be expressed in matrix algebra as

$$\underset{txt}{\mathbf{P}} = \underset{txr}{\mathbf{E}} \quad \underset{rxt}{\mathbf{R'}} \quad \underset{txt}{\mathbf{I}} \qquad (2.0)$$

where **P** = the vector whose non-zero values are the performance levels for each of the tasks,

**E** = the set of vectors of resource efficiency for each resource given the subject-task parameters for each task,

**R'** = the r x t matrix of allocations of each resource for each task,

**I** = the t x t identity matrix,

t = the number of tasks, and

r = the number of resource pools available.

## Performance Operating Characteristic Curves

Joint performance of two tasks can be described by performance operating characteristic (POC) curves (Norman & Bobrow, 1975). The POC curve is the vector of all possible performance level matrices from formula (2.0) in the case where only two tasks are performed. For any dual task situation there is a curve that describes all possible levels of joint performance. Since manipulating difficulty affects the relationship of resources and performance, these manipulations also affect the slope of POC curves. Therefore, task difficulty manipulations can be represented by a series of POC curves (Navon & Gopher, 1980). Figure 1 illustrates the POC curves for the driving example given above. One line represents the vector of performances possible in the example; the other line represents a performance vector that might occur should resource efficiency be reduced by the manipulation of task parameters (perhaps rain falling on the stretch of road to be traversed).

## A Structured-Resource Model

Navon and Gopher (1979) stated that resources are probably organized into a structure of some sort but did not speculate about the precise nature of the structure. Wickens (1980) reviewed several studies that indicated that task performance is affected by the compatibility of the input mode, the coding of the task, and the output mode. From these studies he and his colleagues (Wickens, 1980; Wickens, Derrick,

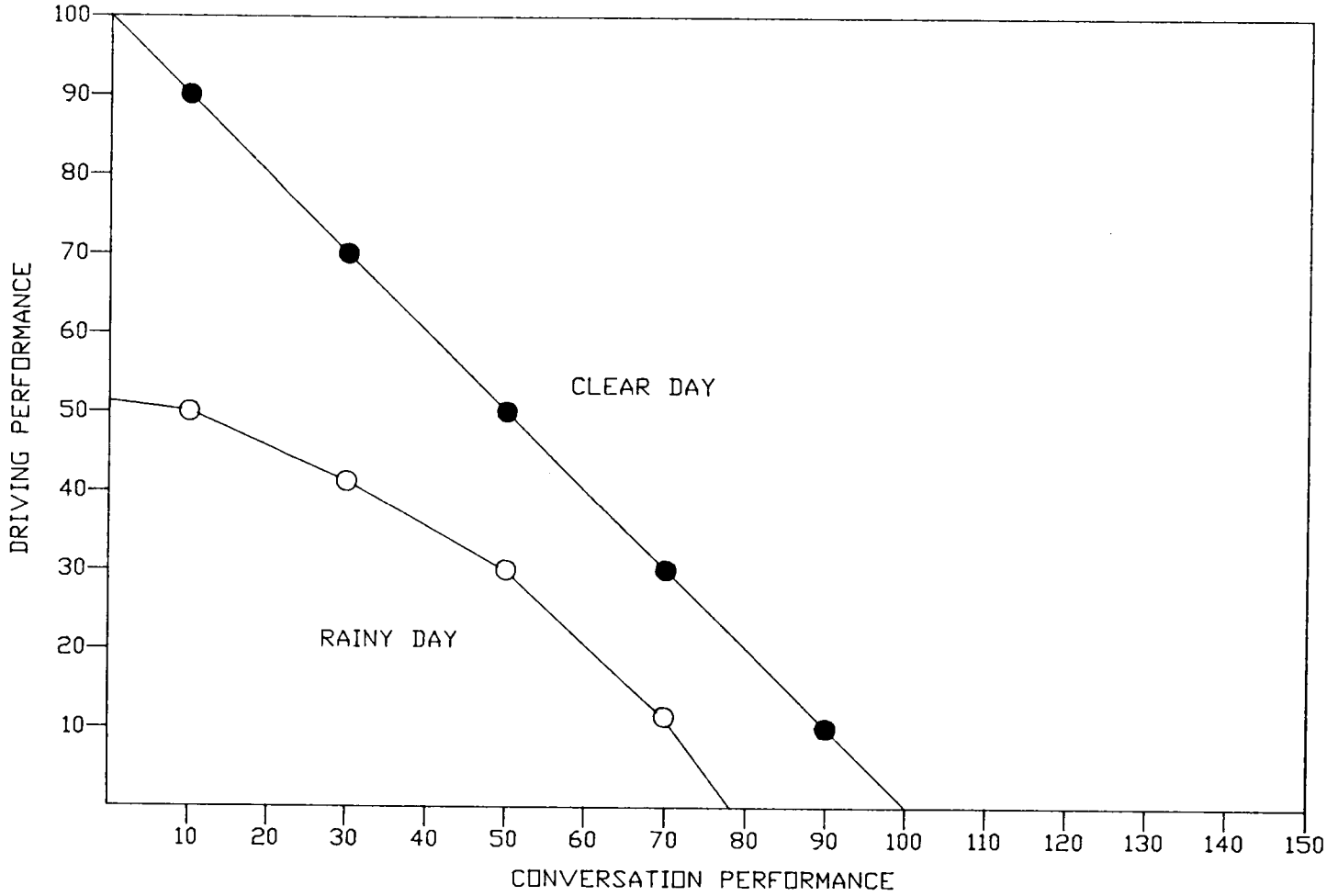



Figure 1. Hypothetical performance operating characteristic curves for the driving example.

Berringer, & Micalizzi, 1980; Wickens, Sandry, & Vidulich, 1983; and Ackerman, Schneider, & Wickens, 1984) developed a model of how resources might be structured. These authors organized resources into six pools corresponding to the poles of three dimensions, which were processing stages, processing modalities, and processing codes.

## Processing Stages

Processing stages are the steps through which information is processed, and include perceptual input, central processing, response (output) selection, and response execution. According to Wickens (1980) and Wickens et al. (1980, 1983) these stages define two resource pools, one for input/central processing and one for response selection/execution.

## Processing Modalities

Processing modalities are defined by the nature of the input and output. There are two pools, one for information flow through a visual input to a manual output and one for processing from an aural input to a verbal output. It seems obvious that other pools are possible; perhaps each of the sensory modalities has a unique pool of resources for inputs, or perhaps the "minor" senses like proprioception and the tactile sense have a common pool. Wickens and colleagues cite a great deal of literature in the areas of visual and aural perception to support their resource structure but do not mention senses other than sight and hearing.

## Processing Codes

Processing codes have to do with how information is presented. Again there are two pools, one for spatially coded information and one for symbolically coded information. Wickens, Sandry, Vidulich, and Schiflett (1981) defined spatially coded tasks as "any task requiring a judgment or integration concerning the 3 axes of translation or orientation." Symbolically coded tasks require "the use of language or some arbitrary, symbolic coding." Symbolically coded tasks have also been termed verbally coded tasks since the majority are coded by language. For an exception see Aretz (1983).

# 3. WORKLOAD

## Resources and Workload

Task performance is the product of the amount of resources committed and the efficiency of the resources. Since task performance can be maintained at a given level in the face of reduced resource efficiency by allocating more resources to the task, outcome measures alone do not adequately measure the impact of task difficulty on performers. For example, resource expenditure may affect time to task completion and error rate, with greater resource expenditure producing lower task time and error rate. Operators may try to maintain the level of performance they achieved with system A when they use system B, allocating more resources to the task if B is more difficult to use. A comparison on the basis of time and error rate would not indicate any differences between the systems even though system A is easier to use. Workload, the rate

of expenditure of resources, and subjective workload, an individual's perception of the expenditure of resources, must also be measured. Stave (1977) found this to be the case in an investigation of the performance of helicopter pilots in flight simulators. As task difficulty (manipulated via the task parameters of noise level and vibration) increased, Stave (1977) found no concomitant degradation in flying performance. However, the pilots reported extreme fatigue after tasks with high workload levels and exhibited short duration lapses in flying performance. These data indicate that performance can be maintained in the face of increasing task difficulty by expending additional resources, and also indicate that increased fatigue is a result of the expenditure of the additional resources. Had Stave (1977) failed to measure variables other than primary task performance he would not have detected any of these effects.

## Subjective Workload

In multiple resource model terms subjective workload can be expressed as

$$w = \underset{1xr}{\mathbf{S}} \quad \underset{rxt}{\mathbf{R}'} \quad \underset{txl}{\mathbf{I}'} \tag{3.0}$$

where w = the scalar of subjective workload,

**S** = the vector of sensitivity of the individual to the expenditure of resources from each pool,

**R** = the matrix of the amount of each resource expended performing each task,

**I** = the vector of the impacts of each task on the individual,

t = the number of tasks, and

r = the number of resource pools available.

The **S** vector represents the differences in operator sensitivity to expenditure of resources. Complete expenditure of a resource pool may make an individual more sensitive to further demands on that pool, and sensitivity may increase as certain critical levels of resources expended are passed. Some resource pools may have greater innate importance than others: the visual/manual channel is the primary modality for humans (Wickens, 1984) and may have greater salience than the auditory/verbal channel, for example.

The **I** vector represents possible differences in task importance which may affect performer sensitivity to resource expenditure. The weights in this vector combine the impacts of perceptions of task performance and the utility of each task for the performer. The product of the **S** and **R'** matrices represents the subjective workload for each task, and multiplication by the **I'** vector weights and sums the workload experienced for the separate tasks to arrive at the global experience of workload.

Since workload is a reaction to the expenditure of resources, and the amount of a resource required to achieve some intended level of performance depends on the efficiency and amount of resources used, workload reflects the difficulty associated with operating a particular system at a given performance level. Comparing control systems on workload and task performance indicates which is most difficult to use. Also, in the long run workload is positively related to operator fatigue and

systems producing high levels of workload for short duration tasks will produce high levels of fatigue for long duration tasks. Measurement of workload along with measurement of task performance is necessary for a complete understanding of the impact of the different control systems.

## Workload Assessment

### Subjective Measures

Wierwille and Williges (1978; see also Williges & Wierwille, 1979; and Wierwille, 1979) organized workload measures into four categories, which were subjective opinions, spare mental capacity measures, primary task measures, and physiological measures. Their extensive review examined more than 400 studies, from which they concluded that subjective opinions when gathered by well developed rating scales were the most reliable and sensitive workload measures. Hicks and Wierwille (1979) and Wierwille and Connor (1983) compared the sensitivity of workload measures from each category in simulator experiments and verified that conclusion. Hicks and Wierwille defined sensitivity as "the relative ability to discriminate statistically differences in workload levels" and evaluated two spare mental capacity measures (a digit detection task and visual occlusion), a physiological measure (cardiac arrhythmia), a two-item semantic differential scale, and primary task performance measures (lateral position, yaw deviation, and steering reversals) on sensitivity to three levels of simulated wind gust disturbance severity in a driving simulator. Only steering reversals, yaw deviation, and the semantic differential were sensitive to all three levels of disturbance

severity. Wierwille and Connor (1983) compared 20 different workload measures for sensitivity to manipulations of instability and random disturbances in a flight simulator, and found that only subjective rating scales and an index of the rate of control motions were sensitive to all levels of task difficulty under the Hicks and Wierwille (1979) definition. Two measures of operator errors were not sensitive to the task difficulty manipulation. Casali and Wierwille (1983) also found that subjective measures were sensitive to workload manipulation, while some primary task measures (e.g. rate of control motions in a moving-base aircraft simulator) were not. These studies indicate that operator perceptions of workload accurately reflect task difficulty, and also indicate that workload is not accurately measured by all performance indices.

## Primary and Secondary Task Measures

Primary and secondary task measures are also useful measures of workload, especially when used in conjunction with measures of subjective workload. Wierwille and Williges (1978) cited 22 studies of workload that used a single primary task measure, and concluded that these measures are of limited value. Single primary task measures have not been found to be sensitive to low levels of workload, perhaps because operators have sufficient spare capacity at low levels to maintain performance in the face of rising workload (Wierwille & Williges, 1978). These measures have been able to detect high levels of workload.

Wierwille and Williges (1978) reported that multiple measures of primary task performance have been more successful in detecting the

entire range of workload, although fewer studies using multiple measures have been conducted (they cite 12). Since their review more data have become available. Stave (1977) used two measures of task performance (root mean square error off navigational course and degrees off course on final approach) and found that pilot load was not reflected in scores on these measures. However, his variables are very similar, essentially the same operationalization applied to two different subtasks. Therefore, his results are not surprising in light of the results Wierwille and Williges (1978) report for single measures. Multiple measures should include indices of performance and of variables that are sensitive to changes in strategy which reflect workload. Hicks and Wierwille (1979) used steering reversals and yaw deviation as measures of primary task performance and found them sensitive to all levels of workload, for example.

## 4. TASKS AND CONTROLS

### Tasks

Teleoperator tasks can be divided into manipulative tasks, sensor tasks, and transportation tasks. Manipulative tasks are performed by handling objects in the remote environment using the remote control arms; sensor tasks are orientation and adjustment of viewing, touch, or sound systems; and transportation tasks are movement of the teleoperator from point to point within the remote area. Transportation usually occurs at the beginning and end of a job and is not often concurrent with other types of tasks, but sensor tasks and manipulative tasks do require

concurrent performance. Therefore, manipulator control demands and sensor control demands often conflict, especially when both control systems depend on manual inputs from the same person. For example, operators sometimes desire a different view of the remote work area while performing manipulative tasks, perhaps panning a camera to follow the path of an end effector reaching for a tool. In a multi-operator cockpit a camera operator can adjust the cameras as the manipulator operator works, either in anticipation of the manipulator operator's needs or on command. When one operator runs the system, that person must take his hands and attention away from the manipulator controls and transfer them to camera controls in order to change views. Because of the competing demands for finite resources in this situation, operators tend to adopt either a strategy of manipulate-adjust cameras-manipulate, or they set up the cameras in a position that offers a compromise between the viewing needs of different stages of the work. Neither strategy is optimal since the former produces a stop and go work pattern with tasks performed in sequence rather than in parallel while the latter may require operation with inadequate visual information.

Since joint performance of manipulative tasks and sensor tasks (mainly television camera control in this system) imposes the greatest degradation on solo performance it is the area to concentrate on in order to improve performance by one person crews. Manipulation is the primary task of teleoperators and viewing system control is a subsidiary activity, so solo operators need viewing system controls which do not interfere with manipulator controls, i.e., sensor controls should be

designed for minimal resource competition with manipulator controls. Manipulation is the primary task and camera control is the secondary task.

## Controls

There are two control systems with the potential to be less interfering than manual control and which meet the criteria discussed earlier (that is current availability and no requirement for headgear). These are control by computer word recognition and machine control of the viewing system. The machine control system studied here has been discussed in detail by Stoughton, Martin, and Bentz (1984). In this system the teleoperator's supervisory computer automatically calculates the position of a slave end-effector and aims the cameras at a point near it. As the end-effector moves while in use, the computer recalculates its position and adjusts the cameras. During installation of the tracking system some operators complained about the constant motion of the system, so a deadband area, within which the slave could move without provoking a camera movement, was calculated for the slave tong.

### Manual Control

When sensor and manipulative systems both depend on visual inputs and manual outputs, concurrent task resource demands may be greater than the resources available. This is the case when camera controls and manipulator controls are both operated manually. However, manual control does have the advantage of using the hands, which are the most precise human output mode for control purposes. Hands are also the most

compatible output mode for visual inputs according to Wickens, Sandry, & Vidulich (1983). These authors reviewed several studies which indicate that response times are lower when visual inputs are coupled with manual responses than when the same inputs require verbal responses. They interpret these findings as evidence that there are distinct processing channels for visual to manual processing and for auditory to verbal processing. Optimal processing in single task situations occurs when auditory inputs are processed to verbal outputs, and when visual inputs are processed to manual outputs.

## Computer Word Recognition

Control by computer word recognition is one alternative to manual control of camera degrees of freedom. A computer can recognize a limited vocabulary by comparing the electrical signals generated by sound waves striking a microphone to patterns of signals in the computer's memory. Incoming signals which match stored signals can trigger outputs associated with the stored signal.

Some literature about voice technology exists, but very little of it reports controlled experiments comparing voice input with other forms of control. In McCauley's (1984) review of human factors and voice technology he cites only four references (Mountford & North, 1980; Poock, 1980; Skriver, 1979; and Welch, 1977) to experiments of this nature. All of the experiments used word recognition as an alternative to manual control via keyboard for a symbolically coded task. According to McCauley (1984) the results of these studies were similar: voice control was no better than keyboard entry for simple tasks, but under dual task

conditions (where the other was a spatially coded, visual input-manual response task, typically some form of tracking) voice control was superior. Mountford and North (1980) used two dimensional compensatory tracking and digit selection as tasks, and found that tracking performance was poorer when digit selection was by keyboard than when it was by word recognition. Data entry took slightly longer with voice than with keyboard in single task performance, but voice entry took as little as 57% of the time for keyboard entry in the dual task situation. Tsang (1982) and Aretz (1983) found similar effects. Tsang used a tracking task and a symbolically coded reaction time task, and found that for both visual and aural inputs the reaction time task interfered less with the tracking task when responses were by voice than when they were by keyboard input. Aretz (1983) found similar outcomes in a less esoteric setting. His primary task was "flying" a flight simulator and his secondary task was control of subsidiary systems by keyboard (in this case an array of dedicated push buttons) or by computer word recognition. Flying performance was better in the dual task situation when subsidiary systems were controlled by voice, but reaction times on the secondary task were better for keyboard entry. However, Aretz notes that when the delay inherent in the word recognition system he used is taken into account word recognition reaction times were actually somewhat shorter than keyboard reaction times.

In two studies which compared manual control to voice control of a spatially coded task, Vidulich and Wickens (1982) and Murray, Gilfoil, and Van Praag (1983) found data that support the dual task literature. Vidulich and Wickens attempted to duplicate a spatially coded flight

simulator task with voice input and found that subjects were much worse with voice control than with manual control in single task performance. However, these authors admit that they were unable to equate the difficulty level on the two tasks, which makes inferences about performance of spatially coded tasks with voice control from their data impossible. Their analysis of dual task interference tends to support the conclusions reported in the literature using symbolically coded voice controlled tasks.

Murray et al. (1983) report that moving a cursor across a cathode-ray tube (CRT) takes longer when operators use word recognition than when they use a spatially coded keyboard. This finding is not surprising in light of the evidence from symbolically coded tasks and the predictions of the structured-resource model; in the absence of dual task resource competition the compatibility of input modality, coding, and output modality determines performance. Since the input was spatially coded and visually presented, performance should have been best when output was spatially coded and delivered via the manual output mode.

Funk and McDowell (1982) suggested that voice control is not appropriate for control of continuous functions because the structure of voice communication is based on discrete phonemes. These authors feel that this limits the number of dimensions that can be controlled through voice input without increasing message length. Control of multidimensional, continuous functions would increase message length beyond acceptable levels. The Murray et al. (1983) results support the Funk and McDowell (1982) position.

## Automatic End-effector Tracking

Automatic tracking of end effectors (machine control) is a second alternative to manual control. This method uses electrical signals from manipulator servomotors to compute the position of one of the end effectors relative to the cameras and then automatically aims the cameras at a point slightly beyond the end effector. Calculation of the end effector position and appropriate camera pan/tilt angles takes no perceptible time so response time and tracking speed depend on the speed of the camera mounts. In effect automatic end-effector tracking relieves the manipulator operator of control of cameras; it is a modification of the task rather than of the human-machine interface. This approach has been investigated for tasks in process control (monitoring and resetting monitoring equipment) by Greenstein and Revesman (1981) and Rieger and Greenstein (1982) among others. The general conclusion from these studies seems to be that when computers and humans share tasks, human workload can be reduced without concomitant degradation of task performance, provided an appropriate task allocation is made and the human can communicate adequately with the computer.

## Specific Hypotheses

Since manipulator control also uses visual to manual processing, manual control of cameras competes for resources with manipulator control. Resource competition will result in higher levels of workload for manual control than for the other control systems and may also result in poorer performance, although operator effort and task strategies may

ameliorate the effect. The skill with which operators queue inputs and outputs in particular will affect the impact of manual control.

Control by word recognition is similar to having a two-operator crew since the cameras will respond to requests by the manipulator operator. Performance and workload should be better (i.e., lower task times, error rate, and workload) for word recognition than for manual control since word recognition does not compete with manipulator control for as many resources. Voice control will require more resources for the task of camera control because of the extra processing required to make a verbal response to a visual input, but reduced competition for resources will result in improved dual task performance and reduced workload.

End effector tracking removes all operator responsibility for camera control. Camera-related resource demands are eliminated, which should result in low workload and good performance in comparison to the other control alternatives. Vertut, Hugon, Fournier, Espiau, and Coiffet (1983) reported that they were able to improve task performance times from 9 times hands-on speed to 4 times hands-on speed by adding an end effector tracking system to a teleoperator. However, these authors do not report the details of the tasks nor do they mention any use of an experimental design and statistical procedures to insure that these results are due to anything other than chance. Therefore, these data can not be considered evidence that automatic tracking improves teleoperator performance of remote handling tasks.

Automatic tracking causes an unusual television scene since it shows the background streaming past the end effectors instead of the end

effectors moving against a stable background. This may reduce the efficiency of perceptual and central processing resources by requiring adaptation to different figure-ground relationships than are normally experienced. Since the tracking system will employ a deadband, an area in the center of the scene within which the end effectors can move without provoking a camera response, operators will need to adjust to changing figure-ground movement relationships during trials. This may reduce the efficiency of resources in the visual/manual channel and cause an increase in workload or a decrease in performance.

To summarize, the hypotheses are (1) manual control of cameras will produce relatively high levels of workload and poor performance, due to resource competition; (2) control by computer word recognition will produce lower workload and better performance than manual control, due to a relative reduction in resource competition; and (3) automated end-effector tracking will produce the lowest levels of workload and the best performance by removing camera control demands from the task.

# CHAPTER 3

# METHODS

## 1. SUBJECTS

The subjects who participated in the study were experienced manipulator operators assigned to the Consolidated Fuel Reprocessing Program (CFRP). These persons use a teleoperator regularly for testing and demonstrations. Teleoperations require a high level of psychomotor skill, so operators must be trained to use the equipment before they can do effective work; therefore, inexperienced persons were not acceptable subjects. Inexperienced operators' data tends to be unstable. Inexperienced persons make mistakes that inflate their task performance times, and these mistakes seem to occur at random intervals. Mistakes are due to errors in interpretation of televised scenes, lack of familiarity with task configurations and procedures, inadequate experience with manipulator operating procedures, and clumsy handling of the manipulators. As operators gain experience they become more adept at planning and executing task strategies because of familiarity with tasks and teleoperator capabilities and because of improved perceptual and motor skills. Learning to use a teleoperator effectively takes time; during an earlier demonstration some operators took weeks to achieve stable performance of some very simple tasks.

Since operators must be experienced to produce meaningful data, subjects must come from the pool of persons already familiar with teleoperators or they must go through a long training program.

Unfortunately, the manipulators used in the RSDF were on loan from another facility; the conditions of the loan limited the time that they would be available. An adequate training program could not be conducted in the time available. Therefore, the experienced operators assigned to the CFRP were used. All available operators were included. This produced a sample of four experienced operators (a fifth operator assigned to the CFRP was not available during testing). All of the subjects were male, with stature within the $5^{th}$ percentile to $95^{th}$ percentile range of the male population in the United States. All of the subjects indicated that they were right-handed.

## 2. APPARATUS

### The Remote Systems Development Facility

The venue for this study was the Remote Systems Development Facility (RSDF), a facility devoted to the development and testing of advanced remote handling concepts. Since the completion of the testing reported here, the RSDF has been dismantled and its place taken by the Remote Operations and Maintenance Development (ROMD) facility. Clarke, Hamel, and Draper (1983) and Stoughton (1985) described the RSDF in detail; it consisted of a set of TeleOperator Systems SM-229 bilateral MSM slave arms, a television viewing system, and power and signal transmission equipment installed in an unused radioactive materials handling cell; a remotely located control room with replica master arms, television monitors, and other controls necessary to operate the system; and tool and task mock-ups located in the cell. The components of the RSDF

are listed in Table 1. The cell and manipulator arms are depicted in Figure 2 and Figure 3; the remote control room is shown in Figure 4.

### TeleOperator Systems SM-229

The TeleOperator Systems SM-229 master/slave manipulators were designed for use in handling radioactive materials in hot cells and accelerators (Kohler, 1981). The control system is bilateral position-position, meaning that the control system reads the position of the master and moves the slave to match it; to achieve force reflection the control system also reads the position of the slave and attempts to drive the master to its position. Movements are powered by 28-volt D. C. electric motors with metal tape transmission of motor revolutions to master/slave motions. Lifting capacity is about 20 pounds.

The force reflection ratio (the ratio of force on the slave arm to the force output at the master) is listed in the operating manual as 1:1 (1 unit of force at the slave causes 1 unit of force at the master) in the low end of the weight capacity range for these manipulators, increasing to about 2.2:1 at the high end of the range. However, other research with these manipulators (Dumbreck, unpublished data) indicates that the force reflection ratios were not as advertised, and indeed could vary with the state of repair and "fine tuning" of the arm. Although this may have affected task difficulty, the fluctuations in feedback level were probably neither large enough nor consistent enough to have had an impact on the effects studied in this experiment.

The slave end effector is a two-finger tong, operated by a caliper on the master handle. Figure 5 shows the master handle, and Figure 6

Table 1. Components of the Remote Systems Development Facility (RSDF).

1. Hot Cell Size: Length = 6.1 m (20 ft)
   Width = 6.1 m (20 ft)
   Height = 6.1 m (20 ft)
   Wall thickness = 1.7 m (5.5 ft)

2. Remote handling equipment:

   2.1 Overhead transporter
   - Bridge, carriage travel = 6.1 m (20 ft)
   - Telescoping tube hoist = 5.2 m (17 ft)
   - Microprocessor (Intel 8080)-based 3-axis servocontrol
   - DC servo drives on all 3 axes

   2.2 Servomanipulators
   - Dual TeleOperator Systems SM-229s[a]
   - Hoist mounted
   - 7 degrees of freedom, force reflecting
   - 9 kg (20 lb) capacity; 1 kg (2 lb) force threshold
   - Analog electronics

   2.3 TV viewing
   - 2 wall-mounted color cameras with pan, tilt, and zoom
   - Dual, 4 degree of freedom (pan, tilt, r, θ), tube hoist-mounted color cameras
   - Subsea Stereo viewing system

3. Remote operations control room: (2nd floor, about 20 m (65 ft) cabling separation)

   3.1 Servomanipulator master controllers
   - TOS SM-229 replica masters, dual

   3.2 Overhead transporter and TV control panel
   - Transporter joystick controls
   - TV monitor/camera select
   - Remote lighting controls
   - Camera lens controls (zoom, iris, focus)
   - Camera position controls (pan, tilt, r, θ)

   3.3 Computer-based collision avoidance monitor
   - Color graphics display of overhead transfer with respect to cell equipment map
   - Tektronix Model 4027

   3.4 TV displays
   - 2 x 2 cluster monitor, 2-in. black and white displays
   - 25-in. color monitor
   - 48-in. wide screen color projection monitor
   - Complete monitor/camera selection switching

4. Supervisory computer:

   Digital Equipment Corporation PDP 11/34
   - 128 K memory
   - 2 RL01 disc storage units

   RSX-11M operating system
   Primarily FORTRAN software
   Functions:
   (1) collision avoidance
   (2) transporter automatic positioning functions:
   - -absolute commands
   - -teach/playback
   - -terminal-based control with cursors

   (3) servomanipulator duty cycle data storage and analysis

5. Tooling and generic task setups:

   5.1 Miscellaneous hand tools
   - Electric impact wrench

   5.2 Generic work task mockups
   - In-line instrument package (fluid-line and electrical connectors)
   - Horizontal shaft motor (190 W, 1/4 hp)
   - Diaphragm-actuated control valve (1.2 cm, 1/2 in.)
   - 76 cm (30-in.) bolted flange
   - 15 cm (6-in.) bolted flange
   - TRU tubing service jumpers

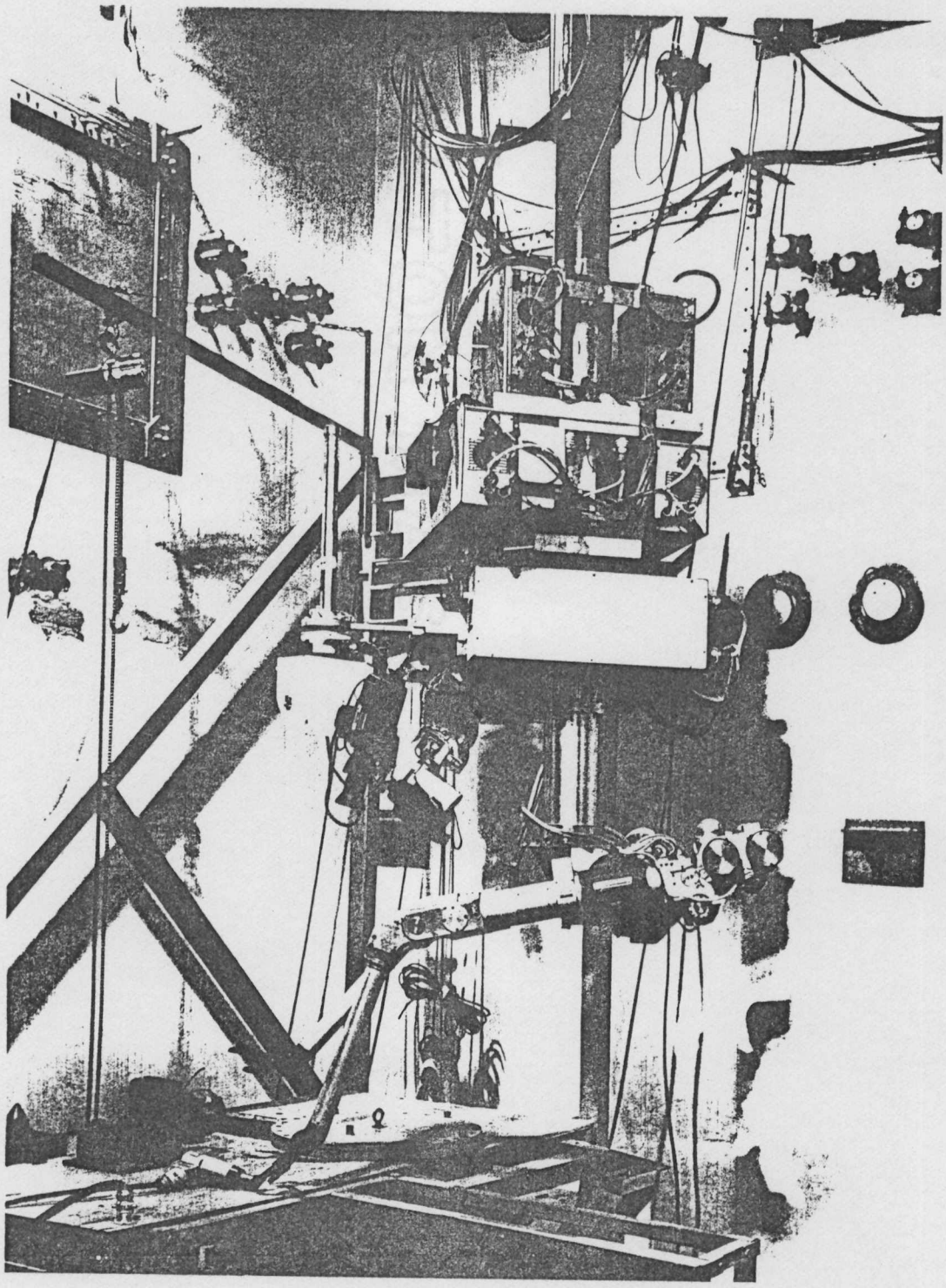


Figure 2. The remote work cell at the Remote Systems Development Facility (RSDF).

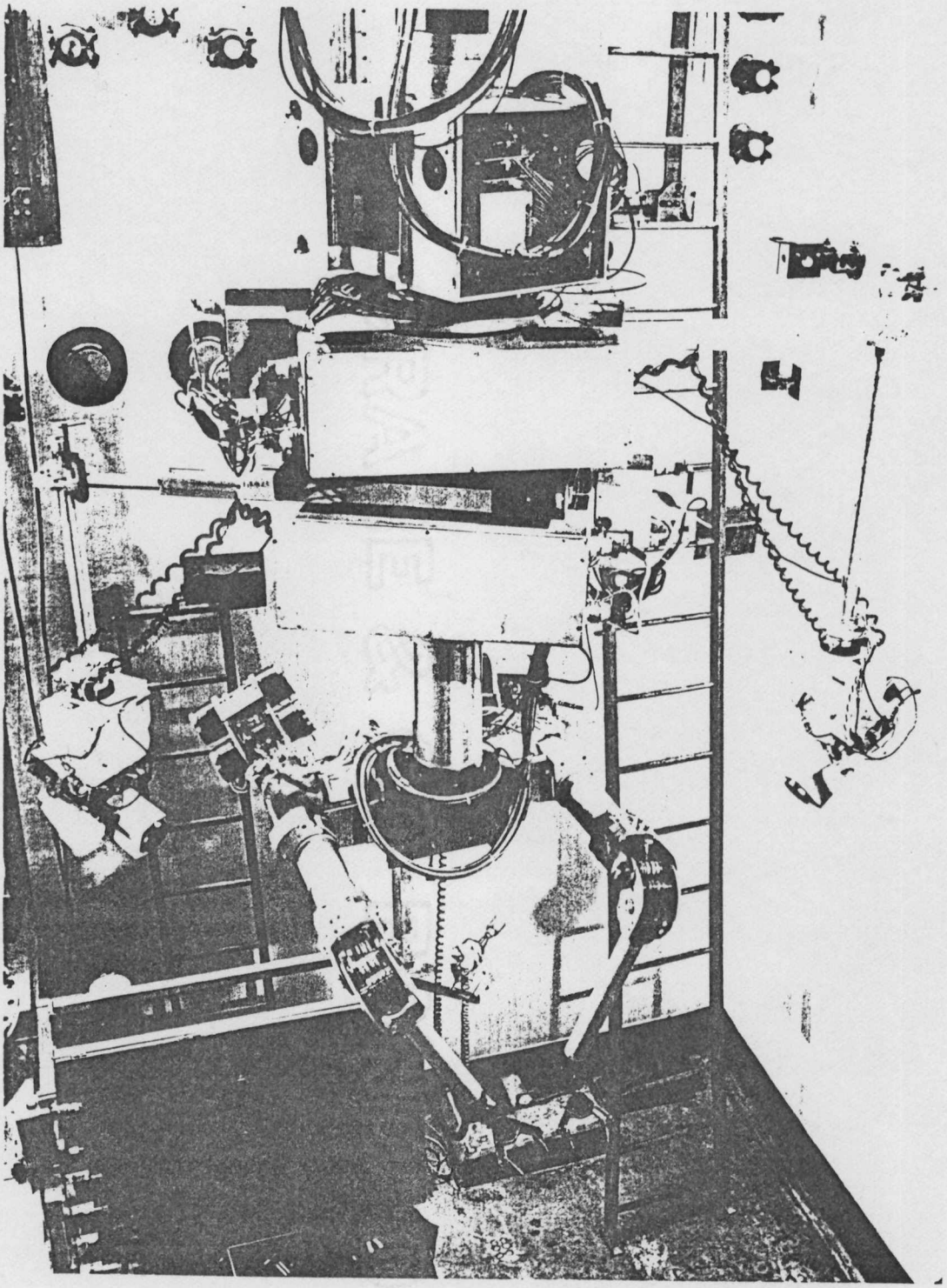


Figure 3. The TeleOperator Systems (TOS) SM-229 manipulators (slave portion) in the RSDF work cell.

Declassified

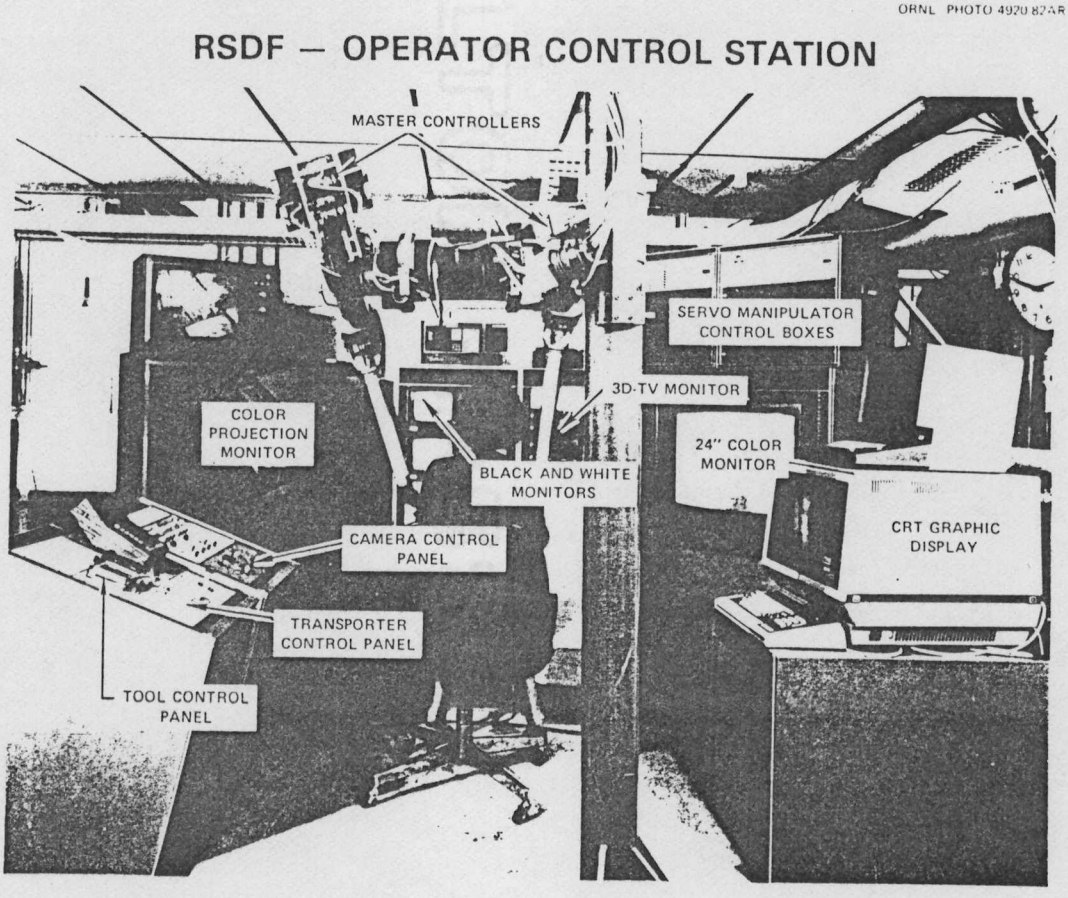



Figure 4. The RSDF control room.

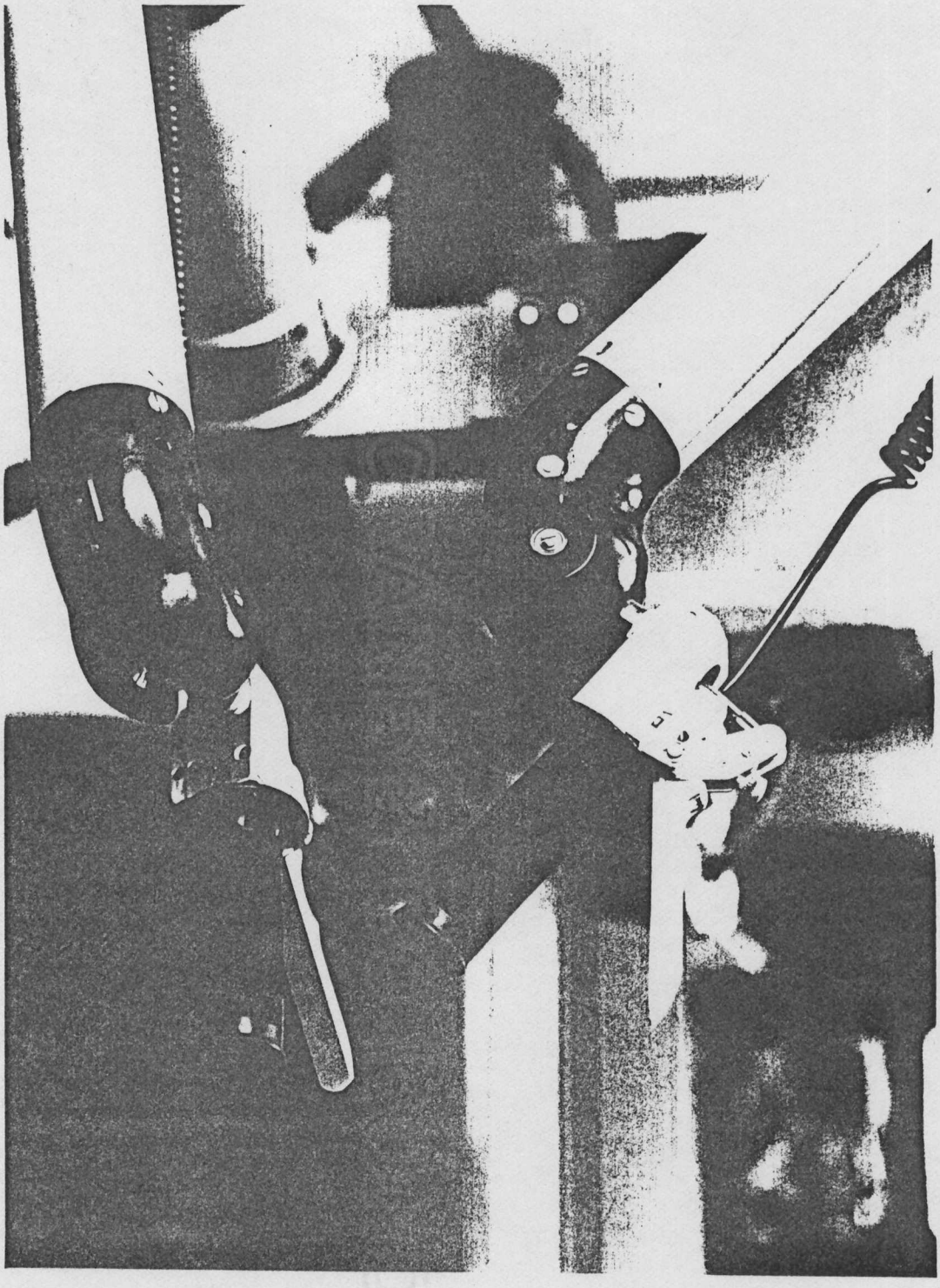


Figure 5. TOS SM-229 master controller handles.

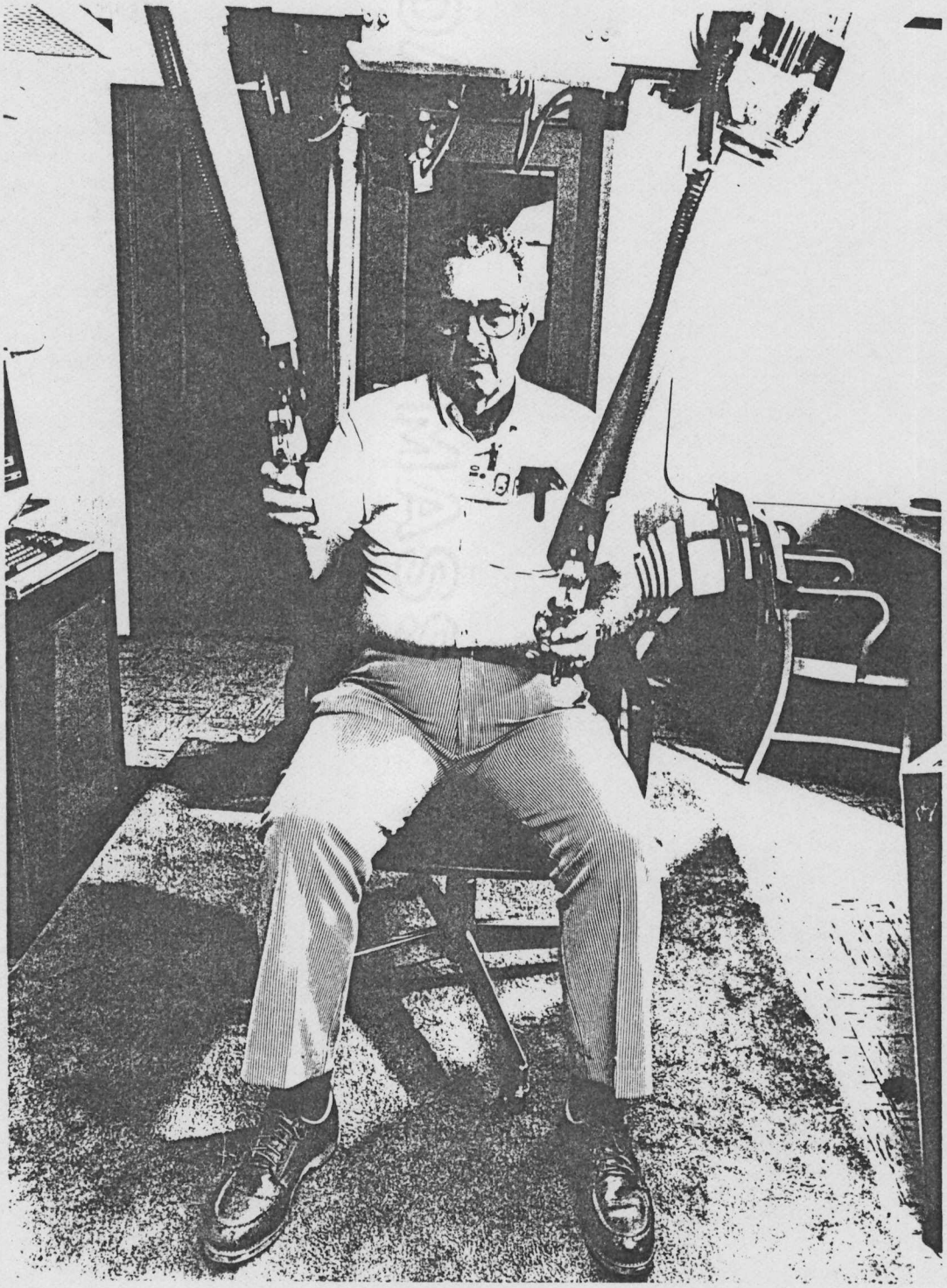


Figure 6. Operator seated at the master controllers within the RSDF control room.

depicts an operator seated at the manipulators. Push button switches for locking the position of the arm and locking the tong independently of the arm are located just above the gripping surface on the front (the side facing the operator) of the handle.

### Closed-Circuit Television System

Views from the cell were provided by a closed-circuit television system. Two television cameras attached to four-degrees-of-freedom mounts on the transporter were used for views of the task area. Two cameras on two-degrees-of-freedom wall mounts were used for views of the whole teleoperator. The transporter mounted cameras could be panned, tilted, rotated around the vertical axis of the transporter, and moved towards and away from the transporter by extending their mounting arms.

### Camera Controls

The manual controls were hard-wired switches and pushbuttons located on a panel to the left of the operator seat. The placement of the controls within the panel was optimized by the author, using standard human factors guidelines (Van Cott & Kinkade, 1972; McCormick & Sanders, 1983). The control panel was placed within easy reach to the left of the operator's chair.

Voice input was via an Interstate Electronics Corporation SYS300 Voice Recognition System. Operators wore a light weight headset which mounted a microphone connected to the system. Before the start of

trials and periodically during testing, operators stored a reference set of commands. When an operator uttered a command, the microphone transmitted it to the voice recognition system. There it was compared to the current operator's stored voice file. Voice files were updated periodically to insure recognition accuracy; on average, files were updated every two weeks. If the signal from the microphone matched one of the stored commands, that command was executed. The command was also displayed on a CRT mounted in front of the operator. If it did not match, the system responded with an aural error signal (a beep) and displayed an error message on the CRT. Voice commands were entered in two steps. First, the operator indicated which camera was to be controlled (camera 1, camera 2, or both). Second, the operator indicated the function to be performed by the selected camera. Available selections included left (pan to the left), right (pan to the right), up (tilt up), down (tilt down), in (zoom in), out (zoom out), near (focus), far (focus), open (increase the size of the camera's lens aperture), close (decrease the size of the aperture), and track. The track command caused the cameras to center the right hand end-effector in the camera view. Voice-controlled tracking was limited to a 3-second interval; at the end of the interval the cameras stopped moving. While the operators were free to develop their own vocabularies, the only change they made in the vocabulary provided was to modify the camera selection command by shortening the camera names to "cam 1" and "cam 2." All commands were executed as soon as the voice control system received them, with no additional verification of inputs required. Command execution continued until the

operator said "stop." When the stop command was entered, active voice commands were cancelled, camera movement stopped, and the system reset.

Stoughton, Martin, and Bentz (1984) described the automatic camera tracking system. A PDP-11/34 computer calculated the positions of the slave end-effectors by finding the position of each joint of the manipulator. It then calculated the pan and tilt commands necessary to aim the onboard cameras at a point just ahead of the right-hand end-effector. A deadband of plus or minus 2 degrees around the end-effector was also calculated. Within this area, movement of the end-effector did not trigger camera movement. This prevented continuous camera motion, which some operators found objectionable during earlier operations.

## Tasks

The subjects performed two tasks: disassembly and assembly of a 30-inch flange. The 30-inch flange was a round flange, 30-inches in diameter and made of plastic. It was bolted in position with three 3/4-inch steel hex-head bolts equally spaced around the outer edge of the flange. A lifting ring was bolted to the center of the flange. Normal disassembly procedure was to loosen and remove the hex-head bolts using an impact wrench (placed in a holster to the side of the flange), swing the overhead crane into position above the flange, attach the crane-hook to the lifting ring, remove the flange, and place the flange on the mock-up framework next to the flange's mount. Assembly of the flange required attaching the crane-hook to the flange, moving it into position above the mount, lowering the flange onto the mount while

maintaining the proper alignment of the flange with the mount, inserting the hex-head bolts, and tightening the bolts with the impact wrench. Figure 7 is a photograph of the flange task.

## 3. DEPENDENT VARIABLES

### Task Performance Time

The time each subject required to complete each task was measured with an electric stop-clock. Task completion time was defined as the time in seconds from the point in each trial when the experimenter said "start" to the point when the task is completed. Time spent retrieving dropped objects and recovering from equipment failures was not included in task time, even though these events may have been due to mistakes caused by the viewing system in use at the time. Time spent recovering from events of this sort depends on the severity of the consequences of the mistake. This was a random function of what happened after the operator made his mistake; if the wrench dropped to the floor the recovery time was long, but if it landed on the flange it was short. Recovery time in this example depended on where the wrench landed and was unaffected by the task or the control system in use. Recovery time was not included in measurements of task completion time in order to prevent confounding variance in performance time due to the independent variables in the experiment with variance due to the randomly distributed consequences of mistakes.

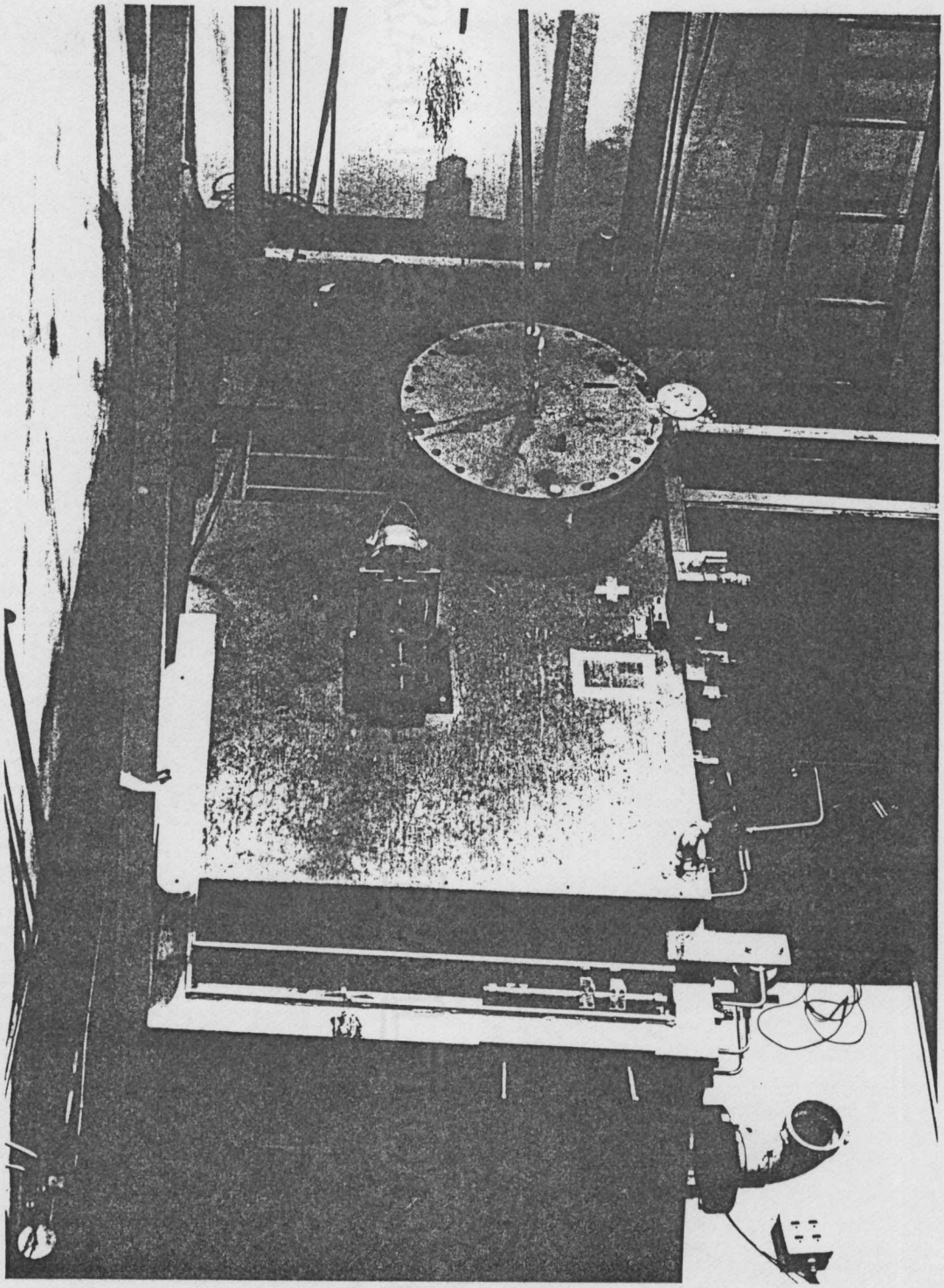


Figure 7. Thirty-inch flange task mock-up (at top with crane hook attached) in the RSDF remote work cell.

## Error Rate

Error rate is the ratio of the number of errors committed during completion of a trial to the time (in minutes) spent on the trial. Errors were counted by an observer seated in the control room during task completion and by another observer who viewed videotapes of trials.

Errors were defined as (1) accidental contact of a manipulator or an object in the grasp of the manipulator with any part of the mock-up frame or with any other object in the remote cell; (2) loss of control of any object in the grasp of the manipulator, as indicated by unintended rotation, translation, or dropping of the object; (3) misalignment of the manipulator or grasped object with another object, as when an operator attempting to insert a bolt into a hole misses the hole and makes contact with the mock-up frame instead; and (4) an action which causes a torque overload on a joint, which in turn triggers the manipulator motor brakes. These error categories were not mutually exclusive: the example of a misalignment could have been classified as an accidental contact. However, the observers only tallied the errors, and did not break error frequency down by category of error, so exclusivity is not important. Observers were instructed to count multiple error category events as one error event.

Observers were required to become familiar with error definitions and examples of errors prior to actually recording data. The observers also practiced recording errors by counting error events in videotapes of a system demonstration conducted prior to this experiment. The observers co-tallied data from videotapes twice during the experiment to

prevent observer drift of the type described by Taplin and Reid (1973). These authors found that observers assigned to record behaviors tended to make observations with good inter-rater reliability in experimental sessions immediately following observer training. However, as experimental sessions continued, inter-rater reliability declined steadily. The co-tallying sessions were designed to insure that the observers were recording errors in a consistent and uniform fashion.

## Subjective Workload

Subjective workload was measured with the Modified Cooper-Harper scale (MCH) developed by Wierwille and Casali (1983). The MCH is a refinement of Cooper's (1957) adjective rating scale and the Cooper-Harper Scale (Cooper & Harper, 1969). Both of these scales were developed to quantify pilot opinions of aircraft handling characteristics. Wierwille and Williges (1978) and Willeges and Wierwille (1979) hypothesized that the unmodified Cooper-Harper Scale would not adequately measure workload because items on the scale refer to aircraft handling characteristics. However, Wierwille and Connor (1983) found that the the scale can be a sensitive measure of workload. These authors compared the sensitivity of 20 workload measurements to manipulations of task difficulty through stability changes and random disturbances in an aircraft simulator. Only three measures were able to distinguish all three levels of the task difficulty manipulation. These measures were rate of control motions, ratings on Donnell's (1979) Workload-Compensation-Interference/Technical Effectiveness (WCI/TE) scale, and ratings on the Cooper-Harper Scale.

Wierwille and Casali (1983) introduced the modified version of the Cooper-Harper Scale to broaden its applicability. According to these authors the original scale, although successful as a workload measuring tool, was not readily applicable to situations in which operators performed perceptual, cognitive, and communication activities as well as motor activities. The MCH was used in this study rather than the original Cooper-Harper Scale because of its greater generality. Manipulator operators often perform tasks which could be categorized as perceptual, cognitive, or communicative and the MCH is therefore more appropriate in this situation. Examples of perceptual activities include spotting objects in televised scenes and distance estimation; cognitive activities include generation of task strategies, including schemes for locating and aiming cameras; and communication activities include giving voice commands to the camera control activities.

The MCH may be prone to memory effects (with operators answering MCH questions on the basis of previous responses rather than on their experiences during a trial). To combat this operators were instructed to match their responses on the MCH to their experiences during trials, and were asked not to let prior responses affect their judgments. In addition a similar form that required a slightly different response was alternated with the original MCH.

## Control Motions

The percent of total time to complete trials spent moving the right-hand master controller was measured as an indicator of workload. Scores on the percent of time spent moving manipulators variables are

indicative of task difficulty. The greater the percent of time operators spend moving manipulators, the lower the portion of time they spend moving cameras, operating tools, thinking about strategy, etc. In other words, the greater the time spent moving manipulators, the lower the demands of the task. As demands on operators increase, scores on percent of time for manipulator motions decrease. Since task difficulty is defined as the sum of the demands placed on an operator by a task (Navon and Gopher, 1980), the percent of time manipulators are in motion is a measure of task difficulty. The greater the difficulty of a task, the lower the percent of time manipulators are kept in motion. Wierwille and Connor (1983) found that the rate of control motions in an aircraft simulator accurately reflected differences in levels of workload; percent of time moving manipulator controls should also measure workload.

The manipulator motion measurement was done by automatic recording of electrical impulses at the motors which controlled the right-hand master controller. Recording were taken by the Digital Electronics Corporation PDP-11/34 computer which served as the teleoperator's master computer. Recordings of the left-hand master controller were not possible due to equipment limitations. However, operational and maintenance experience with the RSDF manipulators indicated that the operators prefer use the right-hand master during the majority of operations. Earlier research at RSDF found that operators used right-hand master or controlled the left-hand master with their right hand (relinquishing control of the right-hand master) 83% of the time (Handel & Draper, 1983). These data include periods when the right-hand manipulator was malfunctioning or inoperable. Given the operator preference for the

right hand master motor activity of the right hand master controller should have provided and adequate measure of workload. All the operators who participated in previous research at the RSDF were right-handed, as were all the subjects of this study.

### Number of Camera Changes

The number of times operators made camera changes was measured. The computer data logging system monitored the movement of cameras 10 times per second (10 hz). A camera change was recorded during a sample if the camera was in motion and it had not been moving in the previous sample. This variable was included to determine how frequently cameras were moved and whether movement frequency differed according to the type of control in use.

### Percent of Time In Motion for Cameras

The computer data logging system (sampling 10 times per second) counted the number of times that it found the cameras in motion during a trial. It divided this number by the total number of samples made in a trial to find the percent of time the cameras were in motion during the trial. This variable was included to determine the duration of camera motions and whether movement duration differed according to the type of control in use.

## 4. EXPERIMENTAL DESIGN

The experiment used a repeated-measures, randomized block design with 3 factors. The factors were camera control type, task performed,

and operator. The blocking variable was based on the order of administration of the experimental conditions. Each operator performed both tasks three times with each camera control system in a randomly determined sequence in the first block, and then repeated the tasks three times each per camera control system in the second block. The tasks were performed in the same order each time, with disassembly completed first, followed by assembly. In the second block, the order of presentation of camera control type was randomly determined, with the constraint that no operator was administered the same sequence as in the first block. A set of three trials was completed in each experimental session when possible; on two occasions equipment failures required extending a session across two days. Analyses of data were based on the average score per treatment combination per operator per block. The average was computed from three trials per block per operator in each of the treatment combinations.

## 5. PROCEDURE

The subjects were briefed about the purpose and plan of the experiment in a meeting held before the start of trials. The briefing included an outline of the goals and procedures of the experiment (without explicitly stating the experimental hypotheses) and a question and answer period. During an earlier experiment that some of the prospective subjects participated in, a subject reported that he was uneasy about being "analyzed" by psychologists. The pre-briefing was planned to inform subjects about the measures taken during the study and the reasons for each measurement. This insured the informed consent of participants

and allayed any anxieties the subjects might have had about the testing.

After the initial meeting subjects reported to the RSDF singly, twice or thrice a week for one-hour testing sessions. Operators were introduced to the equipment by performing some simple remote tasks during their first three experimental sessions. This helped them become accustomed to the TOS SM-229 manipulators and polished their skills. Before starting the fourth experimental session operators walked through the flange task with the experimenter. The walk-through was done under fully direct-contact conditions, i.e., the subject was physically present at the task site and was able to see and touch the task components. This allowed them to become familiar with the task and begin planning how to complete it with the manipulators. After the walk-through subjects began performing the flange task. They completed each task once with all of the camera controls before data collection started.

The subjects performed each task six times with each control system, in blocks of three trials. The order of control presentation was determined by the schedules listed in Table 2. The presentation schedules are composed of two of the six possible orders of control system presentation; since there were four subjects and six possible orders, with each control system used in two separate sets of trials, two orders appear twice. The schedules were constructed to randomize the order of control system presentation. Each subject was assigned at random to one of the schedules.

Table 2. Control system presentation schedules.

| Operator | Experimental Session 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | Auto | Voice | Manual | Voice | Manual | Auto |
| 2 | Manual | Auto | Voice | Manual | Auto | Voice |
| 3 | Voice | Auto | Manual | Auto | Manual | Voice |
| 4 | Manual | Voice | Auto | Voice | Auto | Manual |

# CHAPTER 4

# RESULTS

## 1. METHOD OF ANALYSIS

Analysis of the data followed the strategy suggested by Finkelman and Wolf (1977). The data for the dependent variables (logarithm to base 10 of time, percent of time performing slewing motions, percent of time performing fine-adjusting motions, errors committed per minute, percent of time cameras were in motion, and number of camera changes) were subjected to a multivariate analysis of variance (MANOVA) and then to analysis of variance (ANOVA). The section following describes the results of the MANOVA and the remaining sections in this chapter describe the results of the ANOVAs conducted for each variable.

## 2. MANOVA RESULTS

A multivariate analysis of variance (MANOVA) was performed on six of the seven dependent variables discussed in Chapter 3. The ratings of workload were not included. During the experiment a disturbing trend in the Cooper-Harper scale workload ratings appeared. It seemed that due to the relatively short length of time required to complete trials (approximately 299 seconds on average) operators' ratings of workload were not made independently from trial to trial. In any given trial, the operators were able to remember the rating they gave on the preceeding trial. Given the simplicity of the rating form, it is also likely that they were able to remember ratings from testing session to testing

session. This may have made the workload instrument unresponsive to differences between levels of the factors included in the experiment. A preliminary ANOVA conducted on these data confirmed that intuition: no significant differences were found in the liberally biased ANOVA. In fact, the majority of the mean square ratios were smaller than 0.67. Therefore, the operator's ratings of workload were not subjected to further analysis. It is impossible to tell whether there were in fact no effects on workload for any of the factors included in the experiment or the measure was simply insensitive to workload differences.

The remaining dependent variables were subjected to a three-factor, fully crossed, mixed-model MANOVA, using the Statistical Analysis System (SAS Institute, 1982) General Linear Models procedure (PROC GLM). The factors were type of camera control, task performed, block of testing, and operator. The operator variable was considered to be a random factor; all the others were treated as fixed factors. Wilk's lambda criterion was computed for each factor and converted to F. The camera controls effect produced the only significant F ($F[8,6] = 5.87$, $\alpha \leq 0.04$); however, the error matrices for the task performed and block of testing factors were singular and no tests could be made.

Correlations between the dependent variables were calculated by computing Pearson product-moment correlations separately for each operator and averaging across operators. Table 3 lists the averages of the correlations.

Table 3. Averages of within-operator correlations between dependent variables.

| | Time | Slew M. | Fine M. | Changes | Cam. M. | Errors |
|---|---|---|---|---|---|---|
| Time | | -0.35 | -0.32 | -0.08 | -0.18 | -0.15 |
| Slew M. | | | 0.73 | 0.02 | 0.29 | 0.25 |
| Fine M. | | | | -0.01 | 0.20 | 0.39 |
| Changes | | | | | 0.69 | 0.02 |
| Cam. M. | | | | | | - 0.01 |

## 3. ANOVA FOR REPEATED MEASURES DESIGNS

Repeated-measures data require special handling to produce valid results with analysis of variance (ANOVA). According to Huynh and Mandeville (1979), mean square ratios in repeated measures designs are distributed as F when the data satisfy the following requirements: (1) the covariance matrices for different levels of between-subjects factors (if any are present) are identical across all levels of the factor; and (2) the covariance matrix for the levels within each factor has equal variances and zero covariances. If these conditions do not apply, mean square ratios are not distributed according to F at their degrees of freedom. The criterion value of F for mean squares from data sets which

violate the assumptions concerning the covariance matrices is lower than it should be at the nominal $\alpha$ level. Tests using classical ANOVA mean square ratios for repeated measures designs risk being liberally biased, rejecting the null hypothesis more often than is appropriate at the nominal $\alpha$ level.

To insure valid tests of hypotheses are made in repeated measures designs, two strategies are available. The data may be tested using statistics which do not require restrictive assumptions regarding the covariance matrices, or the degrees of freedom of the F tests may be altered to adjust for the degree of departure of the covariance matrix from the assumed pattern. The former procedure produces an exact test and the latter procedure provides an approximate test. The approximate tests are conservatively biased, that is they tend to reject the null hypothesis less often than they should at the nominal $\alpha$ level and power of the test. The degree of bias depends on the method by which the degrees of freedom are adjusted.

Two methods exist for performing exact tests of hypotheses. Exact tests may be calculated using the mean squares for orthogonal, normalized contrasts on the effect in question as the numerator, and the mean squares for the specific error terms for the contrasts as the denominator of an F ratio (O'Brien and Kaiser, 1985). The mean square for the contrast is divided by the mean square for the contrast by operator interaction (the specific error term) to find the exact F ratio for the contrast. The F calculated using specific error terms requires no assumptions concerning covariance matrices (O'Brien and Kaiser, 1985).

Hotelling's $T^2$ also provides an exact test (Winer, 1968). Hotelling's $T^2$ is the test of choice for repeated measures data since it is relatively easy to calculate and makes use of all the information in the data, including the covariance matrix for the factor tested. Unfortunately, the power of $T^2$ is very poor when the number of subjects is close to the number of levels in the factor being tested (Imhof, 1962). In the design used in this experiment, $T^2$ would not be appropriate since only four subjects participated and three levels were included in the factor of primary interest (camera control type).

The alternative method of analysis is adjustment of degrees of freedom. A procedure described by Greenhouse and Geisser (1959) consists of three stages: first the ANOVA using full degrees of freedom is conducted, just as if the data fulfill the repeated measures assumptions. This produces a set of liberally biased (actual $\alpha$ is larger than nominal $\alpha$) tests. Factors not found significant at this stage are dropped from further analysis. Second, factors found significant in stage 1 are tested using a conservatively biased test. This is constructed by modifying degrees of freedom to reflect worst-case violation of the covariance assumptions. In the worst case scenario, degrees of freedom are equal to 1 and n-1, where n is the number of subjects. Any factor which has an F ratio greater than the criterion value of F with the modified degrees of freedom is considered significant, since the test is conservatively biased. Factors that exceed the criterion in the first stage but which are not found significant in the second stage are carried over into stage 3. In stage 3, factors whose F ratios exceeded the criterion value of F using the liberally biased

tests but not in the conservatively biased tests are re-evaluated using a more precise adjustment to degrees of freedom. An adjustment index called $\theta$ is calculated from the sample covariance matrix and the matrices which define orthogonal normalized variables for each factor. According to Huynh (1978) the procedure for calculating $\theta$ is:

1. calculate the sums of squares and cross products matrix **SSCP** for the raw data for each level of the between subjects factor and then sum across between subjects factors to obtain the matrix **Z** (if no between-subjects factors exist, **Z = SSCP**);
2. select a set of matrices **M** which define the r orthogonal normalized variables for each factor and for each interaction to be tested;
3. compute the r x r matrices **M x Z x M'** for each factor and interaction to be tested;
4. compute $\theta$-hat

$$\theta\text{-hat} = (\mathbf{E}s_{ii})^2/r\mathbf{E}s_{ij}^2 \qquad (4.00)$$

where $s_{ij}$ is an element of the **M x Z x M'** matrix and r is the number of orthogonal normalized variables associated with the factor (equal to the degrees of freedom of the factor).

5. compute $\theta$

$$\theta = Nr(\theta\text{-hat}-2)/r(N-1-r(\theta\text{-hat}) \qquad (5.00)$$

where N is the total number of subjects and r is the number of orthogonal normalized variables for each factor.

The degrees of freedom for each F in stage three are modified by multiplying by the appropriate $\theta$, and a new criterion level of F is found. Any F exceeding this criterion level is considered significant.

## 4. STRATEGY OF ANALYSIS

The analysis for these data followed the strategy suggested by Greenhouse and Geisser (1959), with one exception. Instead of calculating and adjusting the degrees of freedom of the mean square ratios, exact tests were made on factors carried over into the third stage of testing. These tests were done by calculating orthogonal, normalized contrasts for these factors and using these contrasts along with their specific operator by contrast error terms to construct mean square ratios.

Initially the data were run through an analysis of variance using a mixed effects model with one case per cell. Control method, task, and block of trials were considered fixed factors, and subjects entered the model as levels of a random factor. Mean square ratios were calculated in the normal fashion for this type of model, which produced a set of liberally biased F tests. Expected values of the mean squares and the appropriate test for each factor are listed in Table 4. Factors not found significant in this wave of the analysis were not included in further analyses. Factors that produced significant F ratios had their degrees of freedom adjusted as for worst-case violation of the covariance assumption, and their F ratios were compared to the criterion value of F with the adjusted degrees of freedom. Factors whose F ratios exceeded

Table 4. Expected values of mean squares and mean square ratios for F tests.

| Source | E(MS) | Appropriate Test |
|---|---|---|
| Control ($\alpha$) | $qrsn\ \sigma^2_{\alpha} + qrln\ \sigma2_{\alpha\pi} + \sigma2_e$ | $MS\alpha$ / $MS\alpha\pi$ |
| Task ($\tau$) | $prsn\ \sigma^2_{\tau} + prln\ \sigma2_{\tau\pi} + \sigma2_e$ | $MS\tau$ / $MS\tau\pi$ |
| Block ($\beta$) | $pqsn\ \sigma^2_{\beta} + pqln\ \sigma2_{\beta\pi} + \sigma2_e$ | $MS\beta$ / $MS\beta\pi$ |
| Control X Task ($\alpha\tau$) | $rsn\ \sigma^2_{\alpha\tau} + rln\ \sigma2_{\alpha\tau\pi} + \sigma2_e$ | $MS\alpha\tau$ / $MS\alpha\tau\pi$ |
| Control X Block ($\alpha\beta$) | $qsn\ \sigma^2_{\alpha\beta} + qln\ \sigma2_{\alpha\beta\pi} + \sigma2_e$ | $MS\alpha\beta$ / $MS\alpha\beta\pi$ |
| Task X Block ($\tau\beta$) | $psn\ \sigma^2_{\tau\beta} + pln\ \sigma2_{\tau\beta\pi} + \sigma2_e$ | $MS\tau\beta$ / $MS\tau\beta\pi$ |
| Ct. X Tk. X Bl. ($\alpha\tau\beta$) | $sn\ \sigma^2_{\alpha\tau\beta} + ln\ \sigma2_{\alpha\tau\beta\pi} + \sigma2_e$ | $MS\alpha\tau\beta$ / $MS\alpha\tau\beta\pi$ |
| Operator ($\pi$) | $pqrn\ \sigma^2_{\pi} + \sigma2_e$ | $MS\pi$ / MSE |
| Op. X Control ($\alpha\pi$) | $qrn\ \sigma^2_{\alpha\pi} + \sigma2_e$ | $MS\alpha\pi$ / MSE |
| Op. X Task ($\tau\pi$) | $prn\ \sigma^2_{\tau\pi} + \sigma2_e$ | $MS\tau\pi$ / MSE |
| Op. X Block ($\beta\pi$) | $pqn\ \sigma^2_{\beta\pi} + \sigma2_e$ | $MS\beta\pi$ / MSE |
| Op. X Ct. X Tk. ($\alpha\tau\pi$) | $rn\ \sigma^2_{\alpha\tau\pi} + \sigma2_e$ | $MS\alpha\tau\pi$ / MSE |
| Op X Ct. X Bl. ($\alpha\beta\pi$) | $qn\ \sigma^2_{\alpha\beta\pi} + \sigma2_e$ | $MS\alpha\beta\pi$ / MSE |
| Op. X Tk. X Bl. ($\tau\beta\pi$) | $pn\ \sigma^2_{\tau\beta\pi} + \sigma2_e$ | $MS\tau\beta\pi$ / MSE |
| Op. X Ct. X Tk. X Bl. ($\alpha\tau\beta\pi$) | $n\ \sigma^2_{\alpha\tau\beta\pi} + \sigma2_e$ | $MS\alpha\tau\beta\pi$ / MSE |

The model for the analysis is

$$x_{ijklm} = \tau + \alpha_i + \tau_j + \beta_k + \alpha\tau_{ij} + \alpha\beta_{ik} + \tau\beta_{jk} + \alpha\beta_{ik} + \pi_l + \alpha\pi_{il} + \tau\pi_{jl} + \beta\pi_{kl} + \alpha\tau\pi_{ijl} + \alpha\beta\pi_{ikl} + \tau\beta\pi_{jkl} + \alpha\tau\beta\pi_{ijkl} + e_{m(ijkl)}$$

the criterion value of F were declared significant and were not submitted to further analyses.

Factors found significant in the preliminary (liberally biased) testing but not in the secondary (conservatively biased) testing had orthogonal, normalized contrasts calculated, and exact F tests for these factors were conducted. These tests used the specific error terms (the contrast by operator interaction) rather than the omnibus terms (the factor by operator interaction). Tests using specific error terms require no assumptions concerning covariance matrices (O'Brien & Kaiser, 1985).

## 5. RESULTS

The remainder of this chapter describes the results of the analysis conducted for each dependent variable. The data for all of the variables were submitted to a repeated- measures ANOVA. The factors were control type (3 levels), task (2 levels), and block (2 levels). Control, task, and block were considered fixed factors. The criterion level of the F tests was at 0.10. All of the ANOVAs were conducted using the General Linear Models (PROC GLM) procedure of release 82.4 of the Statistical Analysis System (SAS Institute, 1982). Where necessary, the sums of squares for orthogonal, normalized contrasts were also calculated using SAS PROC GLM. Specific error terms were calculated using a custom-written, BASIC-language computer program. The custom-written program used a formula for specific error terms found in Kirk (1982). Appendix A contains the formula for specific error terms and the BASIC program.

## Task Completion Time

The task completion time data were transformed into the logarithm to base 10 of task time in seconds. Table 5 lists results of the preliminary (liberally biased) ANOVA for the logarithm to base 10 of task time. A measure of the proportion of variance accounted for by the

Table 5. ANOVA results for logarithm to base 10 of task completion time.

| Source | df | Mean Square | F | α |
|---|---|---|---|---|
| Within Subjects | 11 | | | |
| Control | 2 | 0.0783 | 5.27 | 0.05 |
| Task | 1 | 0.5556 | 25.23 | 0.02 |
| Block | 1 | 0.2418 | 12.17 | 0.04 |
| Control X Task | 2 | 0.0045 | 1.04 | 0.41 |
| Control X Block | 2 | 0.0122 | 2.68 | 0.15 |
| Task X Block | 1 | 0.0063 | 0.93 | 0.41 |
| Control X Task X Block | 2 | 0.0044 | 0.32 | 0.73 |
| Between Subjects | 36 | | | |
| Operator | 3 | 0.1811 | | |
| Operator X Control | 6 | 0.0149 | | |
| Operator X Task | 3 | 0.0022 | | |
| Operator X Block | 3 | 0.0198 | | |
| Operator X Control X Task | 6 | 0.0043 | | |
| Operator X Control X Block | 6 | 0.0045 | | |
| Operator X Task X Block | 3 | 0.0068 | | |
| Oper X Cont X Tsk X Blk | 6 | 0.0068 | | |
| Total | 47 | | | |

model was calculated by the intraclass correlation (Winer, 1971). For this model, all of the within-groups effects and the operator effect were considered treatment effects, and the interactions of the operator effect and the within-groups effects were considered error effects. For

this variable, the intraclass correlation was 0.71. In the initial tests, the effects of camera control type ($F[2,6] = 5.27$, $\alpha \leq .05$), task performed ($F[1,3] = 25.23$, $\alpha \leq .02$), and block of trials ($F[1,3] = 12.17$, $\alpha \leq .04$) were significant. The adjustment for worst case violation of the covariance assumption uses a criterion level of F with 1 and n-1 degrees of freedom, where n is the number of subjects. For these data, $n = 4$ and the criterion value of F is 5.54 ($F[1,3] = 5.54$ at $\alpha \leq 0.10$). Therefore, the camera control effect is not significant using the conservative test. The F ratios for task performed and block of testing exceeded the conservative criterion. These factors may be considered to have a significant effect on the task time variable.

To perform an exact test of the camera control type factor, a set of orthogonal, normalized contrasts was constructed and the mean square for each contrast was computed. The mean squares for the specific error terms (contrast by operator, in this case) were also computed, using a formula given in Kirk (1982). Table 6 presents the contrast matrix for the orthogonal, normalized contrasts and Table 7 presents the results of the analysis. The mean square ratio for the contrast of voice control versus automated tracking failed to reach the criterion value of F with 1 and 3 degrees of freedom. The contrast of manual control versus the average of voice control and automated tracking did exceed the criterion value of F, indicating a significant difference between the average for the manual control system and the average of the averages for voice control and automated tracking.

To summarize the results of the analysis of the task time variable, a significant F ratio was found in the liberally biased phase of testing

Table 6. Orthogonal, normalized contrasts for the effect of camera control type.

| | Control Type | | |
|---|---|---|---|
| Contrast | Manual | Voice | Automated |
| 1 | 0.00 | 0.707106 | -0.707106 |
| 2 | 0.816496 | -0.408248 | -0.408248 |

Table 7. Mean square ratios for camera control type.

| Source | SS | df | MS | F | α |
|---|---|---|---|---|---|
| C1 | 0.02027 | 1 | 0.02027 | 2.96 | 0.20 |
| C2 | 0.13630 | 1 | 0.13630 | 5.96 | 0.10 |
| C1xOP | 0.02052 | 3 | 0.00684 | | |
| C2xOP | 0.06863 | 3 | 0.02288 | | |

for the effect of camera control type. When exact tests on orthogonal, normalized contrasts were performed, manual control and the average of voice control and automated tracking were significantly different.

No significant difference was found between the voice control level and the automated tracking level. Figure 8 illustrates the mean differences in log 10 task time among camera control systems. Operators worked most rapidly with the automated tracking system (log 10 time = 2.406; mean time = 254.68 seconds). They were somewhat slower with the voice control system (mean log 10 time = 2.435; mean time = 272.27 seconds) and even slower with the manual control system (mean log 10 time = 2.485; mean time = 305.49 seconds). The results indicate that automated tracking allows operators to work more efficiently than manual control of cameras.

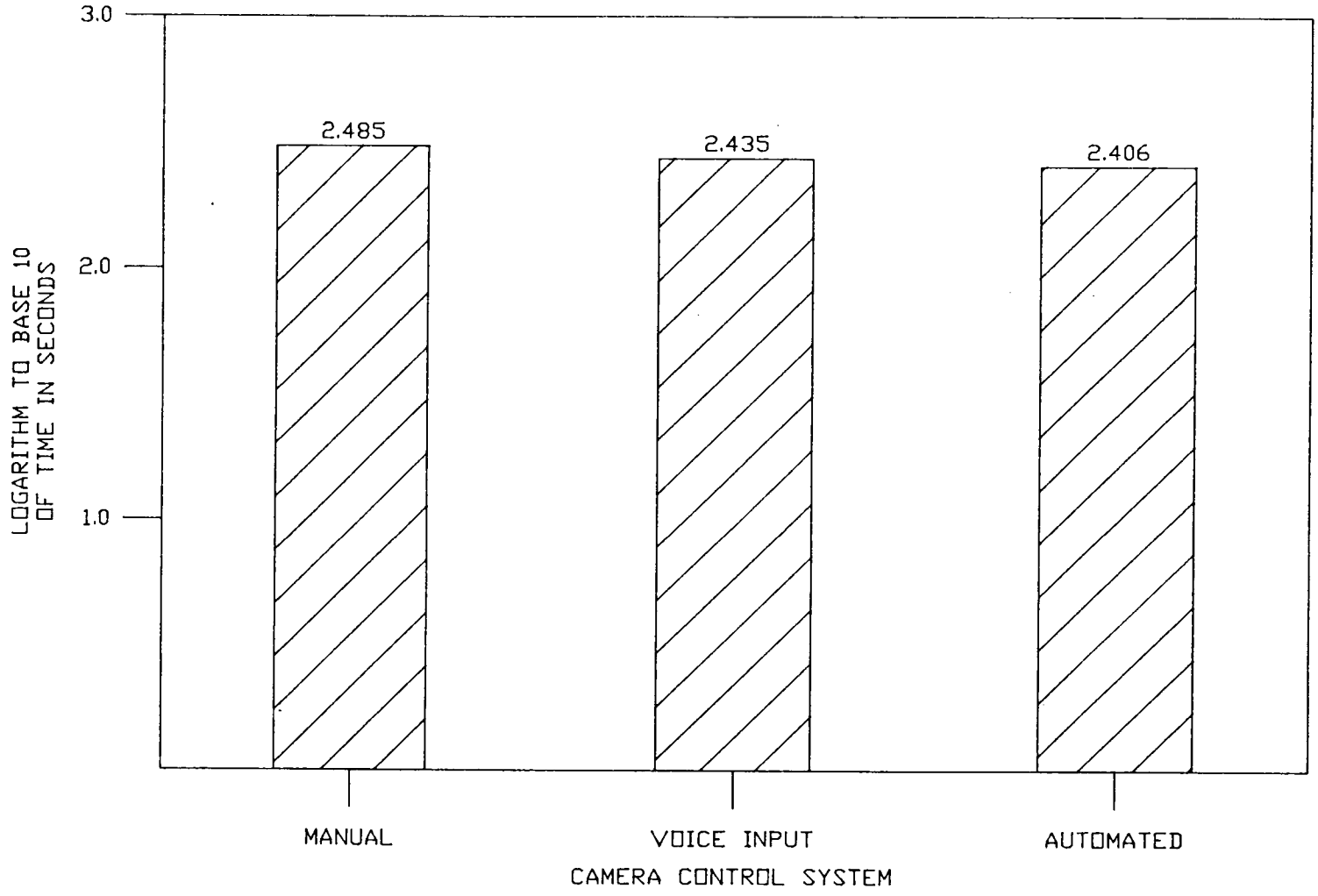



Figure 8. Mean of logarithm to base 10 of task completion time for each control system.

There were also significant differences for the task performed and block of trials, in both the liberally biased and conservative phases of testing, indicating that the means for the levels of these factors differ significantly from each other. Figure 9 illustrates the difference in the mean of the logarithm to base 10 of task time between tasks. Operators required longer times to perform the assembly task (log 10 time = 2.504; mean time = 319.15 seconds) than they did to perform the disassembly task (log 10 time = 2.380; mean time = 239.88 seconds). The difference between the two tasks in time to completion is due to the greater dexterity and finer touch required to complete the assembly task. Assembly of the flange required careful alignment of the flange cover and the flange base. The need for precise alignment caused the operators to work more slowly than they did when disassembling the flange, which did not require precise alignment of task components.

Figure 10 presents the mean difference for the block of trials. Operators required more time to complete tasks in the first block of trials than they did in the second block of trials, which probably reflects the effect of practice on performance. As operators became more familiar with the tasks and with the manipulators they became more proficient. In block 1, the mean log 10 time was 2.483 (mean time = 304.09 seconds) and in block 2 the mean log 10 time was 2.401 (mean time = 251.77 seconds).

## Error Rate

The error rate variable was the average of two observers' error tallies for each trial divided by the number of minutes required to

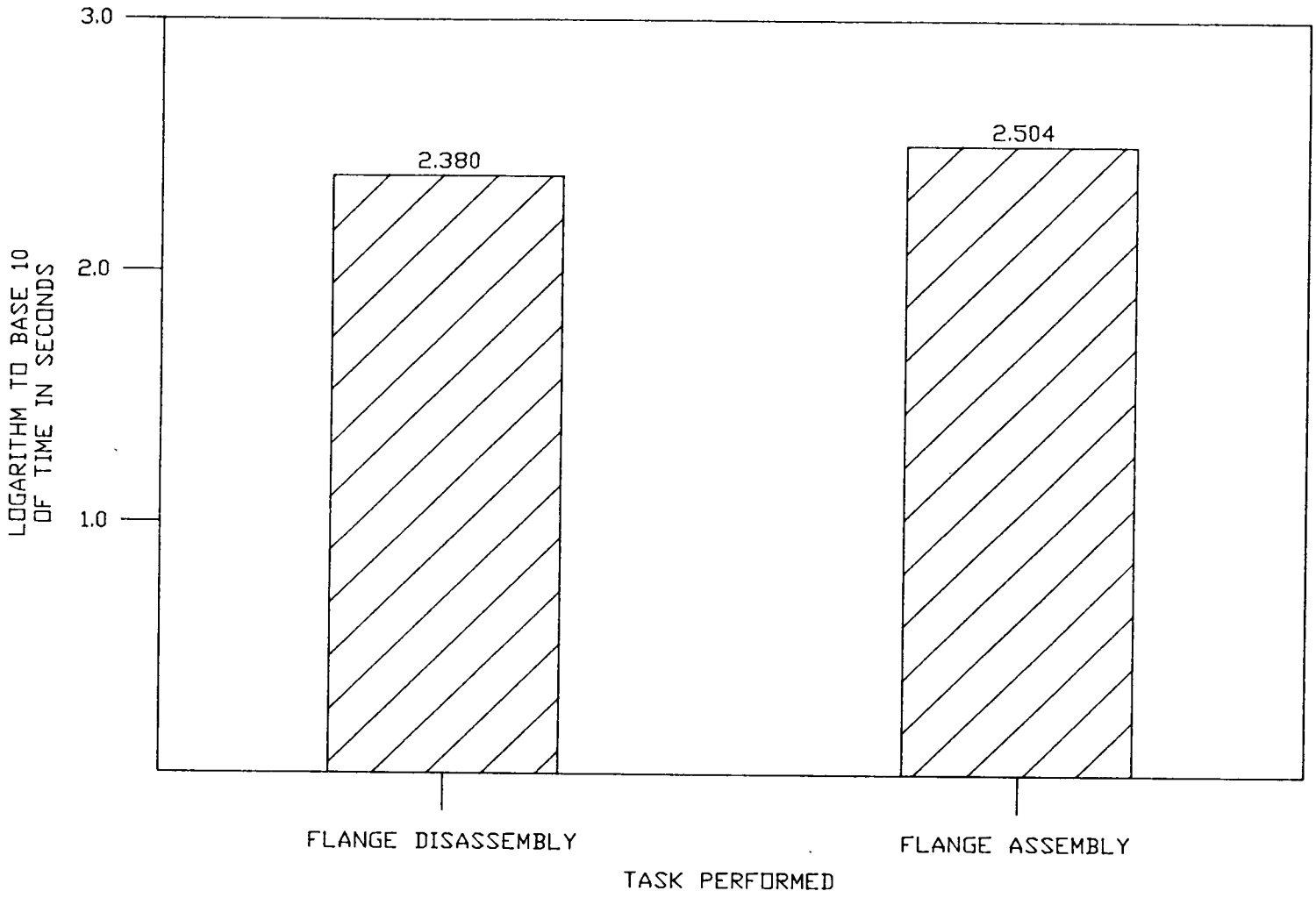



Figure 9. Mean of logarithm to base 10 of task completion time for each task.

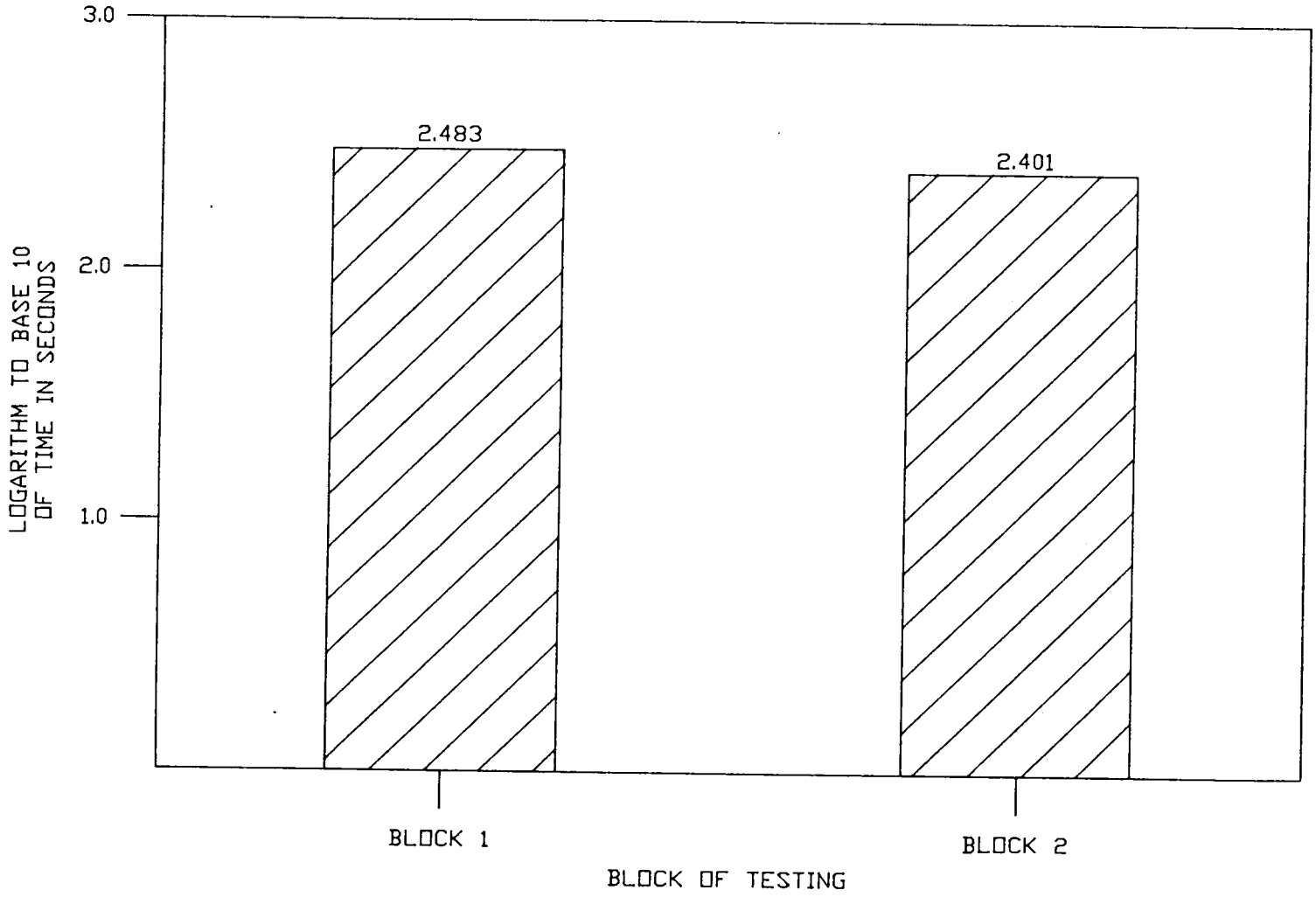



Figure 10. Mean of logarithm to base 10 of task time for each block of testing.

complete the trial. One of the observers recorded errors during testing; the other viewed videotapes of the trials and made tallies. The next section reports on the reliability of the error tallies and the section following reports the results of the analysis of variance for this variable.

## Interobserver Reliability

The reliability of each observer's error tallies was calculated by computing the product-moment correlation of the two sets of tallies. The correlation between the two sets was 0.595. To improve the reliability of this variable, the average of the two error tallies for each trial was used in the analysis. The reliability of the average was computed by adjusting the correlation as per the Spearman-Brown prophesy formula (Magnusson, 1967). This formula computes the reliability advantage for increasing the length of a multi-item test. In this case, the length of the test is doubled since the average of the two ratings represents two measurements (one by each of observer) of the number of errors which occurred in a single trial. According to the Spearman-Brown formula, reliability for the average of the two tallies was 0.746, which was deemed adequate to continue with analysis of this variable.

## ANOVA for Error Rate

None of the factors included in the model produced a significant F ratio for this variable. Table 8 lists the means squares and F ratios calculated in the ANOVA for the error rate variable. A measure of the proportion of variance accounted for by the model was calculated by the

intraclass correlation (Winer, 1971). For this model, all of the within-groups effects and the operator effect were considered treatment effects, and the interactions of the operator effect and the within-groups effects were considered error effects. For this variable, the intraclass correlation was 0.50.

Table 8. ANOVA results for the average rate of errors committed per minute during task completion.

| Source | df | Mean Square | F | α |
|---|---|---|---|---|
| Within Subjects | 11 | | | |
| Control | 2 | 0.0595 | 0.23 | 0.80 |
| Task | 1 | 0.0004 | 0.01 | 0.93 |
| Block | 1 | 0.2466 | 4.48 | 0.13 |
| Control X Task | 2 | 0.0644 | 0.96 | 0.43 |
| Control X Block | 2 | 0.1073 | 1.13 | 0.38 |
| Task X Block | 1 | 0.0058 | 0.37 | 0.59 |
| Control X Task X Block | 2 | 0.1373 | 0.62 | 0.57 |
| Between Subjects | 36 | | | |
| Operator | 3 | 4.1724 | | |
| Operator X Control | 6 | 0.7663 | | |
| Operator X Task | 3 | 0.1227 | | |
| Operator X Block | 3 | 0.1651 | | |
| Operator X Control X Task | 6 | 0.2011 | | |
| Operator X Control X Block | 6 | 0.2837 | | |
| Operator X Task X Block | 3 | 0.0468 | | |
| Oper X Cont X Tsk X Blk | 6 | 0.6618 | | |
| Total | 47 | | | |

Examining the mean squares in Table 8, it appears that the majority of the variance in this factor was due to differences between operators, and to differences in the ways operators reacted to the camera control conditions. Intra-operator differences were relatively weak. The wide variation between operators may have made this variable insensitive to

differences between levels of the repeated measures factors, if any existed.

## Control Motions

The computer recordings of control motions were subdivided into a slewing motions variable and a fine-adjusting motions variable on the basis of results reported in Draper and Sundstrom (Draper & Sundstrom, unpublished manuscript). These authors found that in this data set, principle components analysis indicates that certain joints were used with similar frequency while performing the remote handling tasks. To assess the existence of clusters in the operators' use of the different types of movements, Draper and Sundstrom (1986) conducted a principal components analysis for each of the four operators. Orthogonal factors were extracted by the SAS PROC FACTOR routine using oblique rotation and a minimum eigenvalue of 1.0 (SAS Institute, 1982). The analyses yielded two orthogonal factors for all four of the operators, and the factors were consistent within three of the four operators. Three of the operators showed a clear first factor that included the four degrees of freedom closest to the end-effector and a second factor that included the degrees of freedom closest to the shoulder. The authors conclude that these data indicate that skilled remote manipulator operators used two clusters of movements described by Chapanis and Kinkade (1972) in the use of a servomanipulator with seven degrees of freedom: fine--adjusting motions and slewing motions. Slewing motions are relatively large scale, imprecise adjustments of control position. Fine

adjusting-motions are relatively small scale adjustments to position which typically follow slewing motions (Chapanis & Kinkade, 1972).

In the current experiment, the slewing motion variable was defined as the percent of time the first three joints (shoulder pitch, shoulder roll, and elbow pitch) were in motion. Since these three joints control the motions of relatively long arm-links, their motions can be considered representative of slewing motions. The fine-adjusting motion variable was defined as the percent of time the remaining joints (end-effector yaw, end-effector pitch, end-effector roll, and tong open/close) were in motion. These joints control relatively small arm-links and are used for precise positioning.

Scores on both types of motion variables are indicative of task difficulty. The greater the percent of time operators spend moving manipulators, the lower the portion of time they spend moving cameras, operating tools, thinking about strategy, etc. In other words, the greater the time spent moving manipulators, the lower the demands of the task. As demands on operators increase, scores on percent of time for manipulator motions decrease. Since task difficulty is defined as the sum of the demands placed on an operator by a task (Navon & Gopher, 1980), the percent of time manipulators are in motion is a measure of task difficulty. The greater the difficulty of a task, the lower the percent of time manipulators are kept in motion.

### Slewing Motions

The slewing motions variable consisted of the percent of time joints 1 through 3 (shoulder pitch, shoulder roll, and elbow pitch) of

the right-hand manipulator were in motion. Table 9 lists mean squares and F ratios for the preliminary (liberally biased) ANOVA conducted on this variable. A measure of the proportion of variance accounted for by

Table 9. ANOVA results for the percent of time performing slewing motions.

| Source | df | Mean Square | F | α |
|---|---|---|---|---|
| Within Subjects | 11 | | | |
| Control | 2 | 121.315 | 3.22 | 0.11 |
| Task | 1 | 58.2062 | 0.10 | 0.02 |
| Block | 1 | 81.566 | 6.40 | 0.09 |
| Control X Task | 2 | 24.149 | 1.99 | 0.22 |
| Control X Block | 2 | 6.669 | 0.77 | 0.50 |
| Task X Block | 1 | 9.566 | 0.92 | 0.41 |
| Control X Task X Block | 2 | 10.067 | 3.48 | 0.09 |
| Between Subjects | 36 | | | |
| Operator | 3 | 774.053 | | |
| Operator X Control | 6 | 37.679 | | |
| Operator X Task | 3 | 2.896 | | |
| Operator X Block | 3 | 12.753 | | |
| Operator X Control X Task | 6 | 12.162 | | |
| Operator X Control X Block | 6 | 8.673 | | |
| Operator X Task X Block | 3 | 10.352 | | |
| Oper X Cont X Tsk X Blk | 6 | 2.895 | | |
| Total | 47 | | | |

the model was calculated by the intraclass correlation (Winer, 1971). For this model, all of the within-groups effects and the operator effect were considered treatment effects, and the interactions of the operator effect and the within-groups effects were considered error effects. For this variable, the intraclass correlation was 0.78. In the preliminary ANOVA significant differences were found for the task performed ($F[1,3] = 20.10$), the block of testing ($F[1,3] = 6.40$), and the

three-way interaction of camera control type, task performed, and block of testing ($F[2,6] = 3.48$). The mean square ratios for the task performed and for the block of testing also exceeded the criterion of the conservative test. The mean square ratio for the interaction did not reach the conservative criterion. Unfortunately, SAS PROC GLM declared the orthonormal contrast for this factor unestimable. Therefore, $\theta$-hat was calculated for the interaction instead of the mean square ratio based on the specific error term. The resulting $\theta$ was 1, indicating that no adjustment for violation of covariance assumptions was necessary for this test. In other words, the mean square ratio for the interaction is distributed as F with 2 and 6 degrees of freedom. In response to the significant interaction, a set of simple effects tests was conducted. Since (1) plots of the interaction indicated a difference in the relationship of the camera controls based on the task and block of testing; (2) the effects of task and block were significant in liberally-biased and conservative tests; and (3) the camera control factor is of primary interest, the simple effects tests concentrated on the effect of camera controls at each task and block of testing combination. The orthogonal, normalized contrasts for the camera controls listed in Table 6 (page 69) were tested using specific error terms for the control by operator interaction (calculated using the BASIC-language program found in the Appendix) at each combination of task performed and block of testing. For the (Disassembly, Block 1) combination, the contrast for voice versus automated control was significant [$F(1,3) = 16.71$, 0.05]; for the (Disassembly, Block 2) combination, the manual versus voice and automated tracking contrast was significant [$F(1,3) =$

7.01, $\alpha \leq 0.10$]; for the (Assembly, Block 1) combination, the manual versus voice and automated tracking contrast was significant [$F(1,3) = 6.50$, $\alpha \leq 0.10$]; and for the (Assembly, Block 2) combination, neither contrast was significant.

To summarize the results of the ANOVA for this variable, the difference between the assembly task and the disassembly task was significant. Figure 11 illustrates the difference.

Slewing motions occupied 18.75% of the time required to complete the disassembly task and 17.39% of the time for the assembly task. This indicates that the disassembly task was the easier of the two tasks for the operators to complete. This result is consistent with the findings for the task completion time variable.

There was also a significant difference between the blocks of testing, which Figure 12 depicts. In the second block of testing operators spent more time in slewing motions (18.85%) than they did in the first block of testing (17.31%). This reflects the improving proficiency of the operators with practice; as the operators became more familiar with the tasks and the manipulators, they were better able to handle the demands of the task. This is consistent with the findings for task completion time.

A significant control type by task by block interaction was also found on the slewing motions variable. Figure 13 illustrates the interaction. Simple effects tests performed within the disassembly task, block 1 combination indicate that operators spent a higher proportion of their time performing slewing motions with the voice control system than with either of the other systems. In block 2, the contrast between

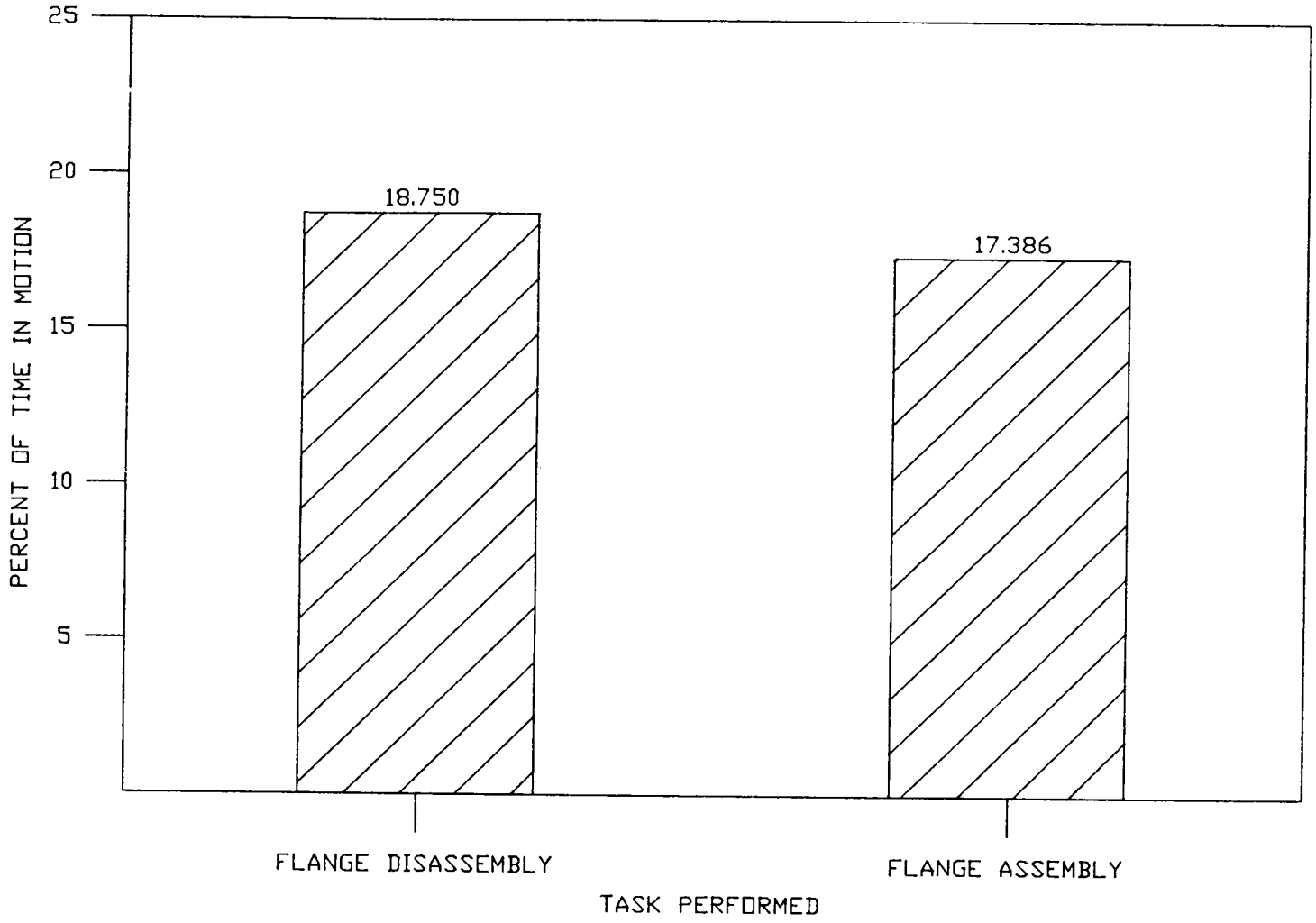



Figure 11. Mean percent of time engaged in slewing motions for each task.

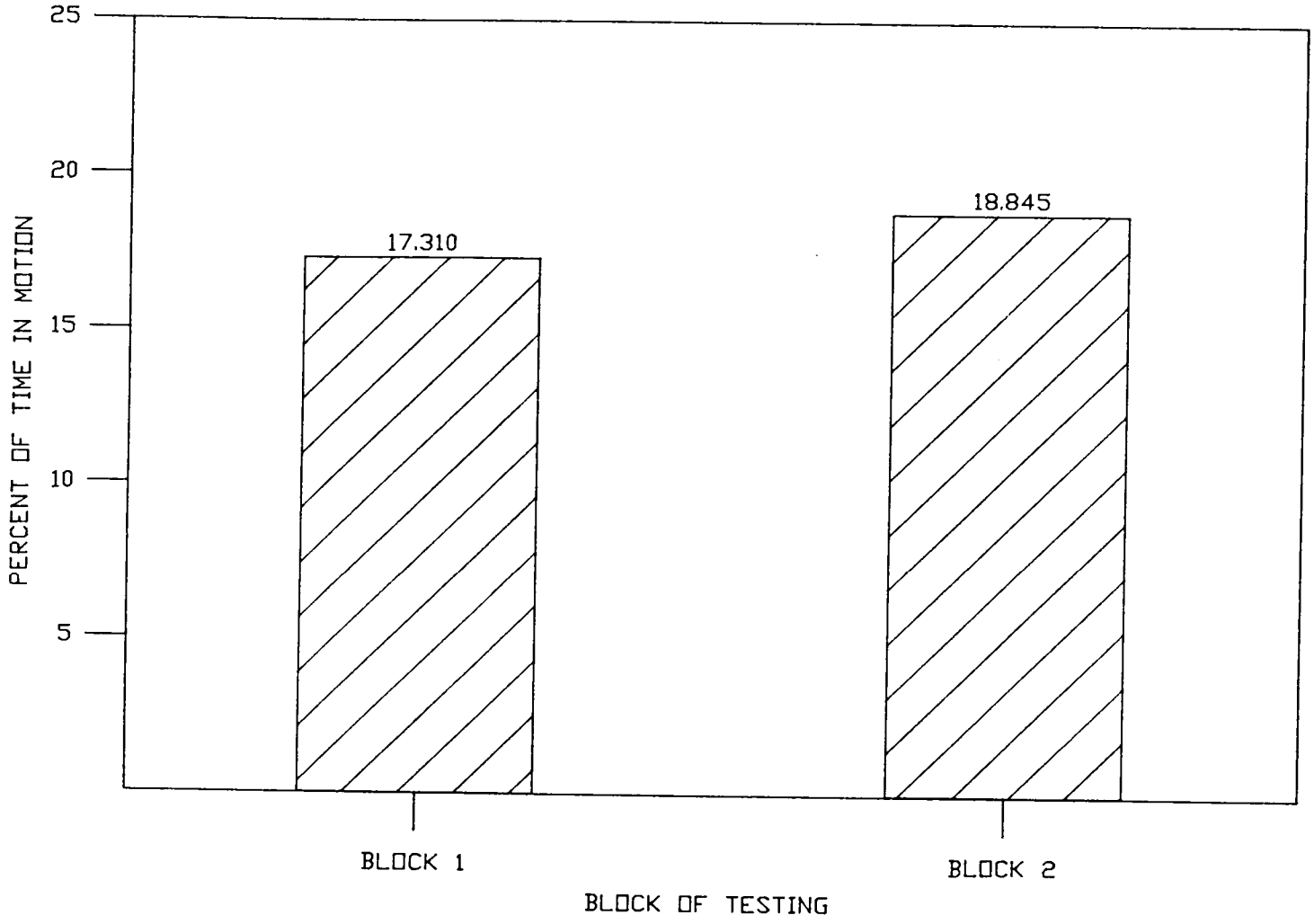



Figure 12. Mean percent of time engaged in slewing motions for each block of testing.

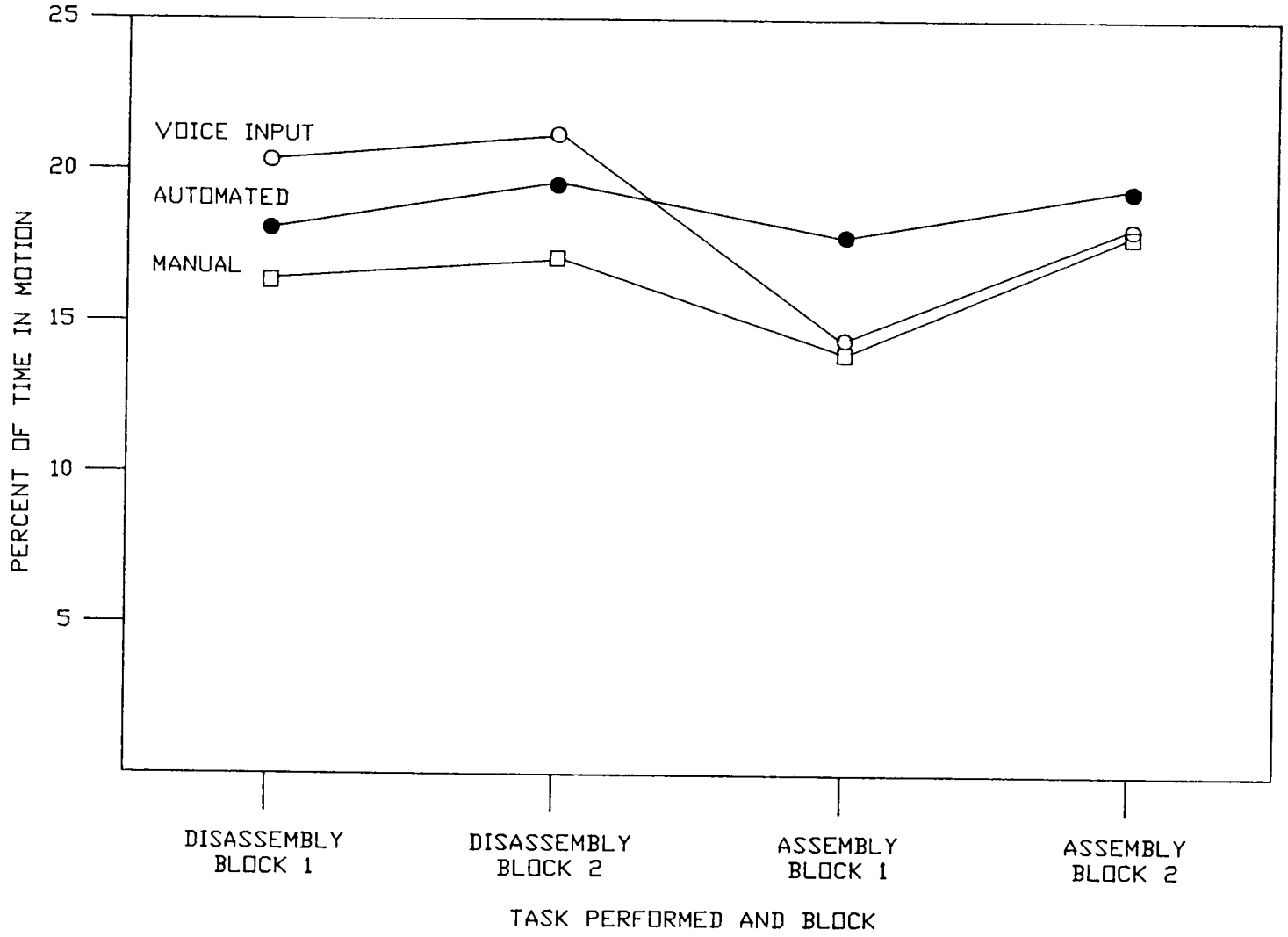



Figure 13. Interaction of control type, task performed, and block of testing for percent of time engaged in slewing motions.

voice and automated tracking was not significant, but the manual versus voice and automated tracking contrast was. This indicates that, on this task, the operator's performance with automated tracking improved more from block 1 to block 2 than it did with the other camera control systems. Initially there was an advantage for voice control over the other systems; as operators practiced with the systems, performance with automated tracking improved to the point of equivalence with voice control. On the disassembly task, there was a significant contrast for manual control versus voice and automated tracking in block 1; however, as operators practiced with the manual control system, their performance improved to the point that there were no differences between the systems.

There seem to be three causes for the significant interaction. First, the difficulty of the task performed makes no difference when using automated tracking. Fine-adjusting motion percent seems equivalent across tasks within the automated tracking condition. Second, the task difficulty affects the performance difference between voice control and manual control. There is a larger performance advantage for voice control on the easier task. Third, the effect of practice within the manual control condition differs between tasks. For the disassembly task, practice effects seem the same as for the other control systems; however, the effect of practice is much more pronounced for the assembly task.

### Fine-Adjusting Motions

The fine-adjusting motions variable consisted of recordings of the percentage of time joints 4 through 7 (wrist yaw, wrist roll, wrist

pitch, and end-effector open/close) of the manipulator were in motion. In the liberally biased ANOVA (reported in Table 10), two factors exhibited F ratios beyond the criterion value: the camera control factor (F[2,6] = 3.49) and the task completed factor (F[1,3] = 330.07).

Table 10. ANOVA results for percent of time performing fine-adjusting motions.

| Source | df | Mean Square | F | α |
|---|---|---|---|---|
| Within Subjects | 11 | | | |
| Control | 2 | 116.84 | 3.49 | 0.10 |
| Task | 1 | 146.5133 | 0.07 | 0.01 |
| Block | 1 | 55.78 | 1.34 | 0.34 |
| Control X Task | 2 | 73.20 | 2.63 | 0.16 |
| Control X Block | 2 | 34.77 | 2.97 | 0.13 |
| Task X Block | 1 | 20.73 | 1.06 | 0.38 |
| Control X Task X Block | 2 | 3.38 | 0.19 | 0.83 |
| Between Subjects | 36 | | | |
| Operator | 3 | 2635.82 | | |
| Operator X Control | 6 | 33.50 | | |
| Operator X Task | 3 | 0.44 | | |
| Operator X Block | 3 | 41.67 | | |
| Operator X Control X Task | 6 | 27.87 | | |
| Operator X Control X Block | 6 | 11.71 | | |
| Operator X Task X Block | 3 | 19.55 | | |
| Oper X Cont X Tsk X Blk | 6 | 17.81 | | |
| Total | 47 | | | |

A measure of the proportion of variance accounted for by the model was calculated by the intraclass correlation (Winer, 1971). For this model, all of the within-groups effects and the operator effect were considered treatment effects, and the interactions of the operator effect and the within-groups effects were considered error effects. For this variable, the intraclass correlation was 0.86.

The F ratio for the task completed factor was also significant when its degrees of freedom were adjusted for worst case violation of the covariance assumptions. The F ratio for the camera control factor was less than the criterion value for the conservative test. A set of orthogonal contrasts was computed for the camera control factor using the coefficients listed in Table 6 (page 69), and an exact F ratio calculated for each of the contrasts. Mean squares for the contrasts were calculated using the SAS GLM procedure. Mean squares for the specific error terms were calculated using the computer program described earlier. The first contrast tested the difference between voice control and automated end-effector tracking and was not significant. The contrast between manual control and the average of the averages of voice control and automated tracking was significant ($F[1,3] = 8.35$, $\alpha \leq 0.10$).

To summarize the results of the analysis for the fine-adjusting motion variable, there was a significant difference between the manual control system and the average of the voice control system and automated tracking system. The difference between the voice control system and the automated tracking system was not significant. Figure 14 illustrates the mean differences among the control systems in the percent of time spent performing fine-adjusting motions. Operators using the voice control system spent the greatest percentage of their time performing fine-adjusting motions (38.66%). The same operators performed fine-adjusting motions 37.89% of the time when using automated tracking and 35.65% of the time when using manual controls. This indicates that

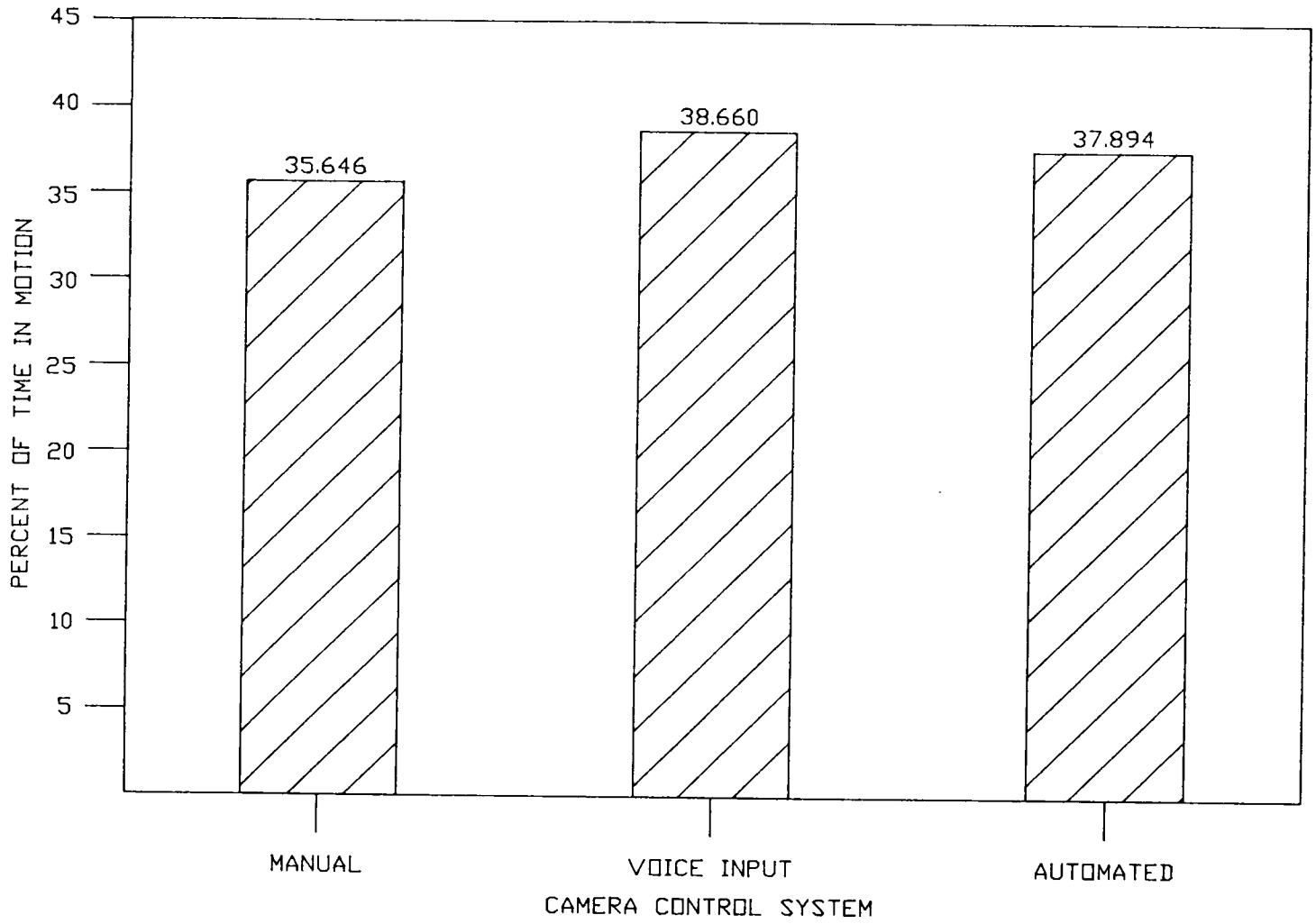



Figure 14. Mean percent of time engaged in fine-adjusting motions for each camera control system.

operators found the tasks easier to perform with voice control than with manual control.

There was a significant difference between tasks in the percent of time spent performing fine-adjusting motions. Figure 15 presents the means for the two tasks on this variable. Operators performing the disassembly task spent 38.39% of their time in fine-adjusting motions while operators performing the assembly task spent only 36.36% of their time performing fine adjusting motions. This is consistent with the findings for task completion time and slewing motions; the assembly task was more difficult to complete than the disassembly task.

## Number of Camera Changes

The camera changes variable was the number of times camera positions were changed during the course of a trial. The computer data logging system monitored the movement of cameras 10 times per second (10 hz). A camera change was recorded during a sample if the camera was in motion and it had not been moving in the previous sample.

Table 11 lists the results of the ANOVA for this variable, including degrees of freedom, mean squares, F ratios, and significance levels. A measure of the proportion of variance accounted for by the model was calculated by the intraclass correlation (Winer, 1971). For this model, all of the within-groups effects and the operator effect were considered treatment effects, and the interactions of the operator effect and the within-groups effects were considered error effects. For this variable, the intraclass correlation was 0.85. The camera control used ($F[2,6]$ =

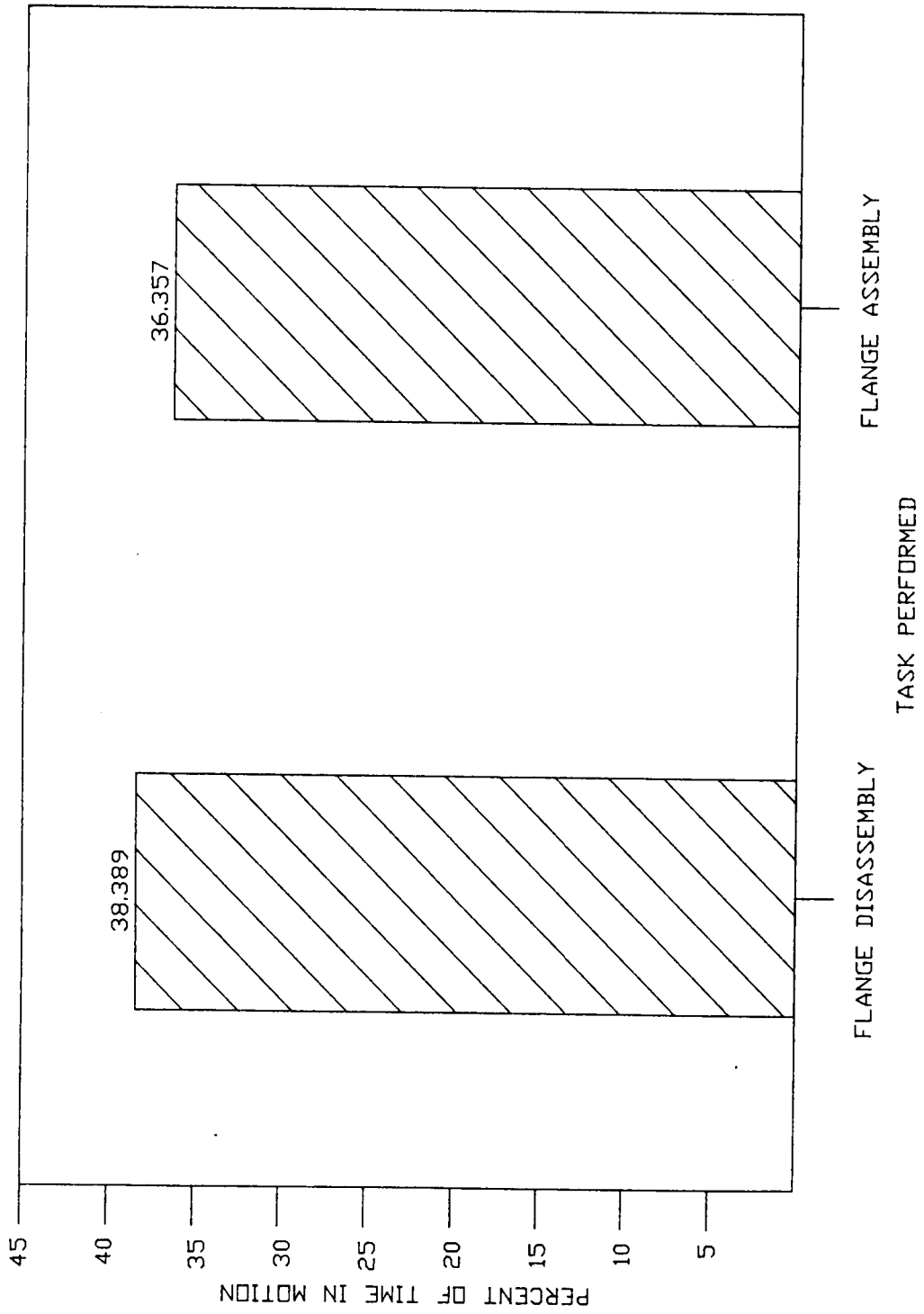



Figure 15. Mean percent of time engaged in fine-adjusting motions for each task.

Table 11. ANOVA results for number of camera changes.

| Source | df | Mean Square | F | α |
|---|---|---|---|---|
| Within Subjects | 11 | | | |
| Control | 2 | 6748.10 | 63.26 | 0.01 |
| Task | 1 | 500.52 | 26.88 | 0.02 |
| Block | 1 | 196.02 | 8.43 | 0.07 |
| Control X Task | 2 | 349.12 | 17.53 | 0.01 |
| Control X Block | 2 | 95.43 | 2.17 | 0.20 |
| Task X Block | 1 | 0.84 | 0.08 | 0.80 |
| Control X Task X Block | 2 | 13.42 | 0.35 | 0.72 |
| Between Subjects | 36 | | | |
| Operator | 3 | 238.75 | | |
| Operator X Control | 6 | 320.00 | | |
| Operator X Task | 3 | 55.86 | | |
| Operator X Block | 3 | 69.73 | | |
| Operator X Control X Task | 6 | 59.74 | | |
| Operator X Control X Block | 6 | 131.85 | | |
| Operator X Task X Block | 3 | 31.88 | | |
| Oper X Cont X Tsk X Blk | 6 | 115.89 | | |
| Total | 47 | | | |

63.26, $\alpha \leq 0.01$), the task performed ($F[1,3] = 26.88$, $\alpha \leq 0.02$), the block of testing ($F[1,3] = 8.43$, $\alpha \leq 0.07$), and the control by task interaction ($F[2,6] = 17.53$, $\alpha \leq 0.01$) produced significant F ratios in the liberally biased phase of testing. All of these F ratios exceeded the criterion value of F used in the conservative phase of testing. The significant interaction observed on this variable seemed to be due to a magnification of the impact of automated tracking on camera changes within the assembly task. On the disassembly task, the automated tracking system made more than twice as many changes as did the other systems; on the assembly task, the tracking system made more than four times as many changes. While the interaction for the slewing motions variable led to a series of simple effects tests on main effects within

levels of the interacting factors, the magnitude of the simple effect seemed to make such tests superfluous for this variable.

Figure 16 illustrates the means for the camera control types on the number of changes per trial. The operators made 13.19 changes per trial with manual control and 11.21 changes per trial with voice control; the automated tracking system made 37.29 changes per trial for the operators. Post hoc comparisons made using Scheffé's S test (Kirk, 1982) found no significant difference between manual control and voice control. There was a significant difference between automated tracking and both of the other camera control systems.

Figure 17 shows the mean number of camera changes per trial for each task. Operators made 17.33 changes on the disassembly task and 23.79 changes on the assembly task. The assembly task was more difficult than the disassembly task. It required accurate alignment of task components and of the impact wrench with the bolts securing the flange. In response to the requirement for accurate alignments by the assembly task, the operators made more frequent adjustments to the zoom, focus, pan, and tilt of the cameras during performance of this task.

Figure 18 shows the mean number of changes for each block of testing. In the first block, operators made 22.58 changes per trial, and in the second block they made 18.54 changes. This reflects the effect of practice on operator performance. With practice, operators were able to position the cameras so that fewer changes were necessary.

Figure 19 illustrates the interaction of control type and task for the number of camera changes per trial. The averages for the manual control and voice control systems are very close, and the differences

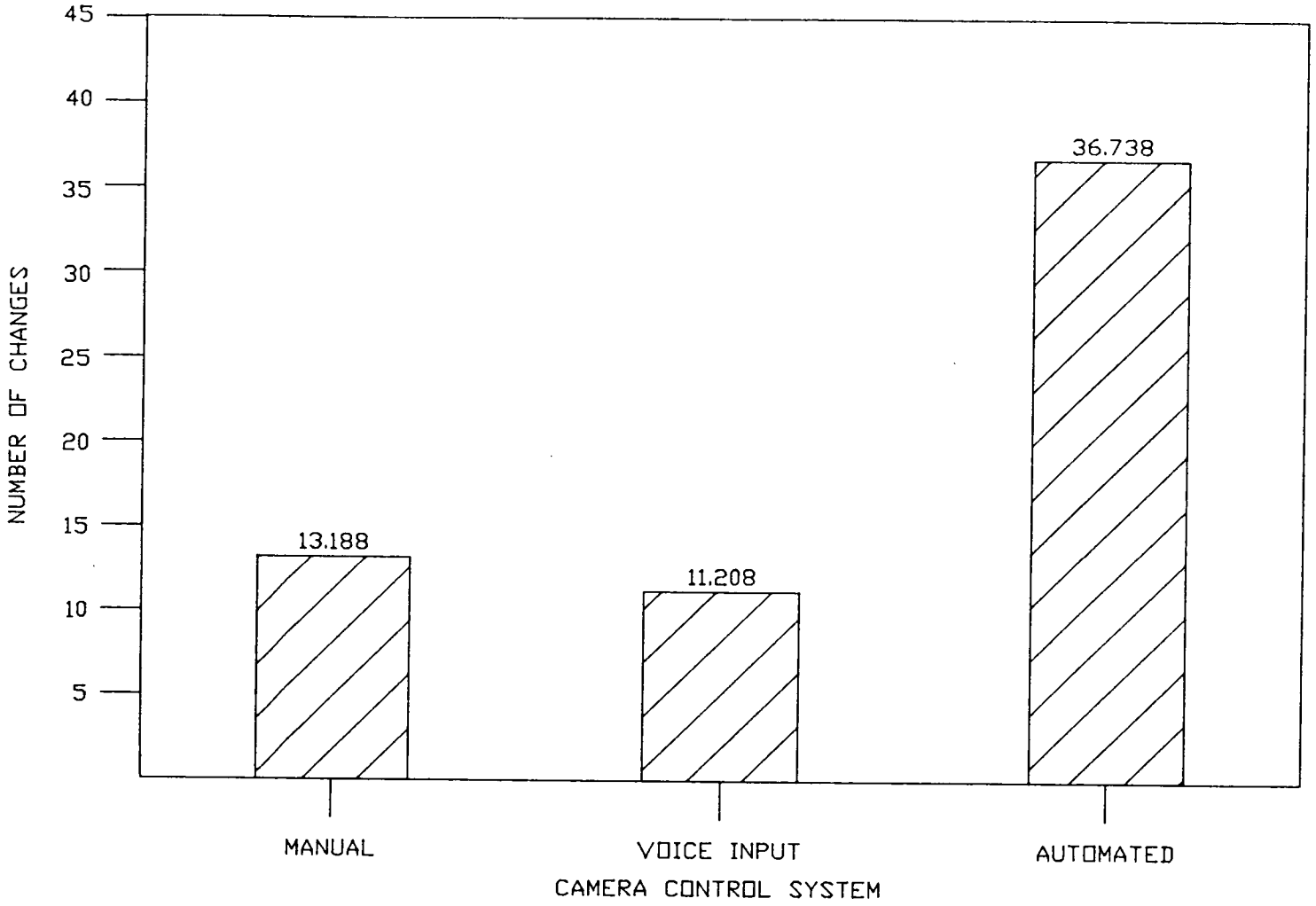



Figure 16. Mean number of camera changes per trial for each camera control type.

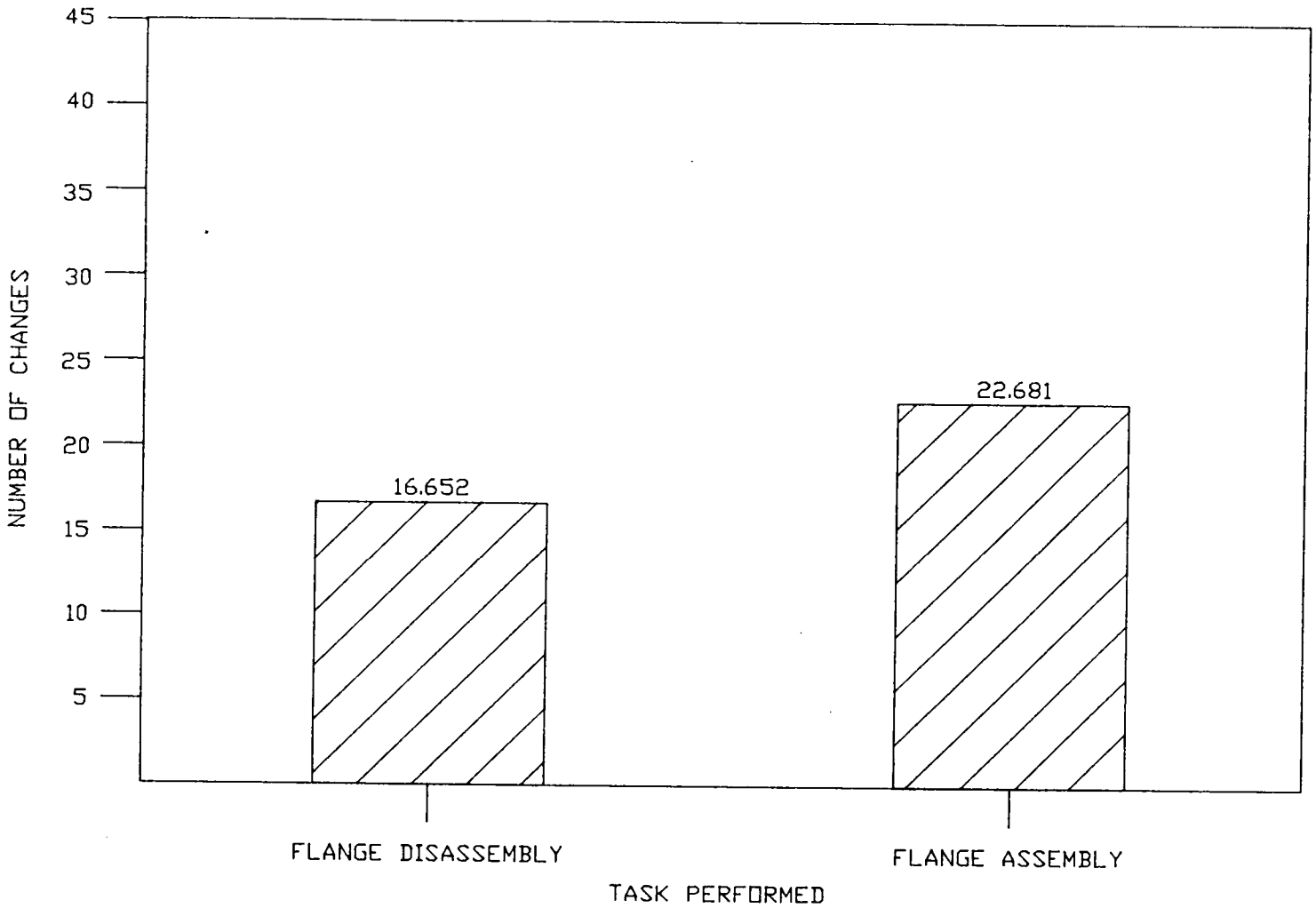



Figure 17. Mean number of camera changes per trial for each task.

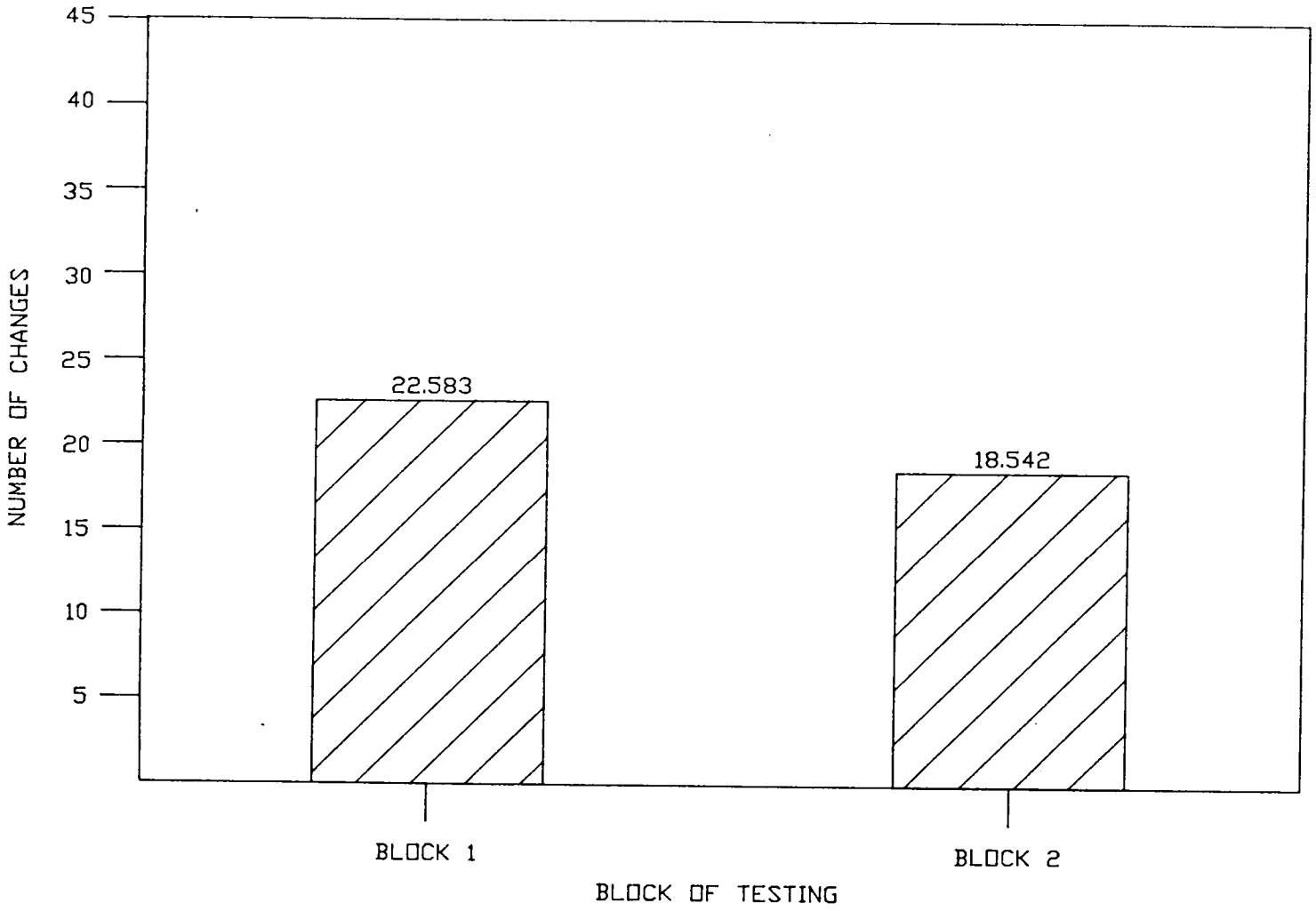



Figure 18. Mean number of camera changes per trial for each block of testing.

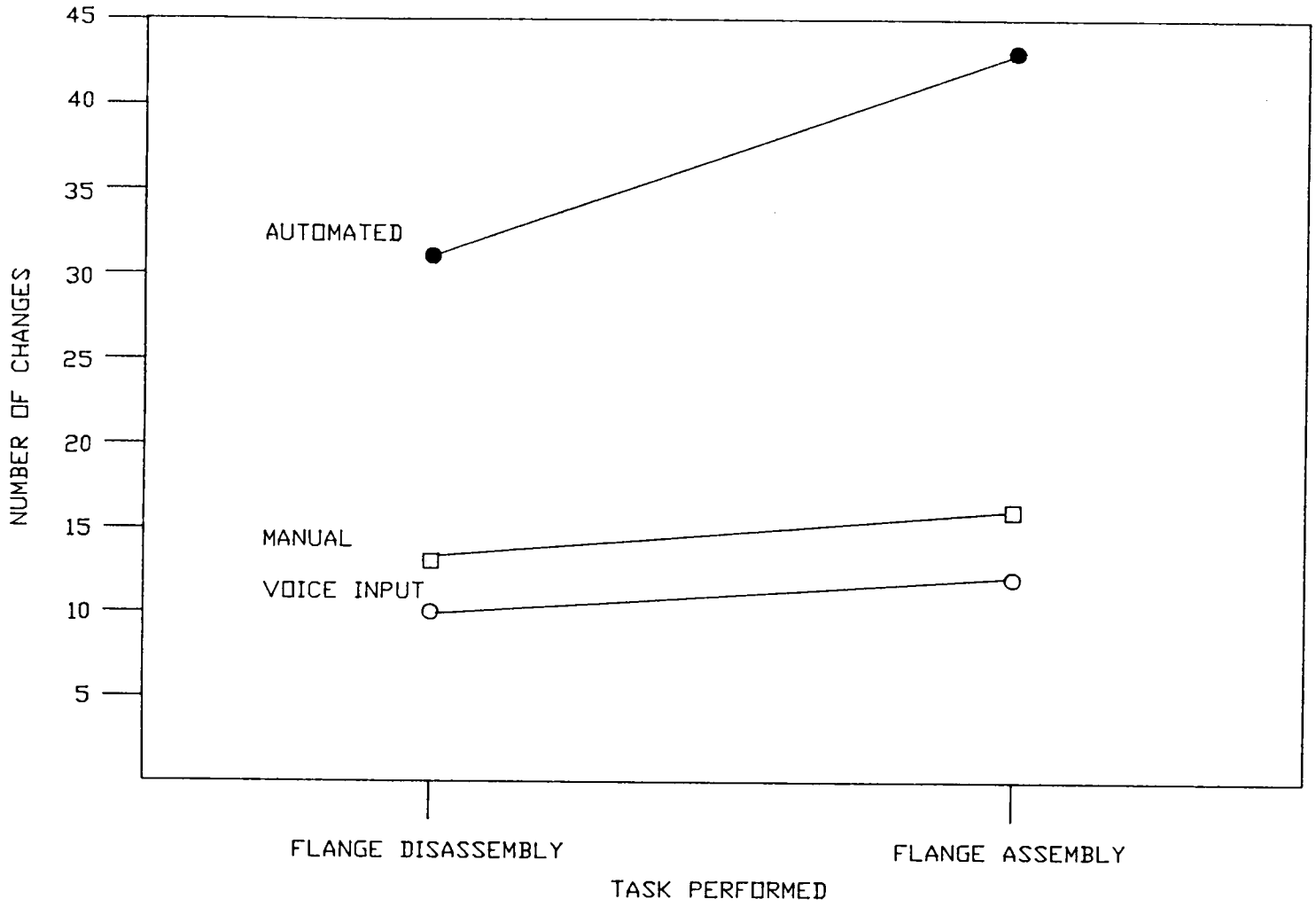



Figure 19. Interaction of control type and task for the mean number of camera changes per trial.

are uniform across levels of the task performed. The average for automated tracking is much higher than the averages for the other control systems, and the difference is greater on the assembly task than on the disassembly task. The difference in the effect size between tasks is probably due to the larger number of small scale adjustments necessary during the assembly task. This task required precise alignment of task components, which in turn necessitated more adjustments to manipulator position, causing more frequent changes in camera position.

## Percent of Time in Motion

The computer data logging system (sampling 10 times per second) counted the number of times that it found the cameras in motion during a trial. It divided this number by the total number of samples made in a trial to find the percent of time the cameras were in motion during the trial.

In the ANOVA for this variable, only the camera control factor ($F[2,6] = 80.67$, $\alpha \leq 0.01$) produced a significant F ratio in the liberally biased phase of testing; the F for this factor also exceeded the criterion value in the conservative phase of testing. Table 12 lists the degrees of freedom, mean squares, F ratios, and significance levels for this variable. A measure of the proportion of variance accounted for by the model was calculated by the intraclass correlation (Winer, 1971). For this model, all of the within-groups effects and the operator effect were considered treatment effects, and the interactions of the operator effect and the within-groups effects were considered error effects. For this variable, the intraclass correlation was 0.85.

Table 12. ANOVA results for percent of time cameras were in motion.

| Source | df | Mean Square | F | α |
|---|---|---|---|---|
| Within Subjects | 11 | | | |
| Control | 2 | 5452.64 | 80.67 | 0.01 |
| Task | 1 | 130.02 | 3.18 | 0.18 |
| Block | 1 | 20.02 | 0.58 | 0.51 |
| Control X Task | 2 | 24.01 | 1.90 | 0.23 |
| Control X Block | 2 | 0.68 | 0.04 | 0.96 |
| Task X Block | 1 | 6.02 | 0.31 | 0.62 |
| Control X Task X Block | 2 | 0.51 | 0.08 | 0.92 |
| Between Subjects | 36 | | | |
| Operator | 3 | 360.62 | | |
| Operator X Control | 6 | 202.76 | | |
| Operator X Task | 3 | 122.71 | | |
| Operator X Block | 3 | 104.12 | | |
| Operator X Control X Task | 6 | 37.95 | | |
| Operator X Control X Block | 6 | 50.54 | | |
| Operator X Task X Block | 3 | 59.16 | | |
| Oper X Cont X Tsk X Blk | 6 | 18.89 | | |
| Total | 47 | | | |

Figure 20 shows the means for each level of the camera control factor for this variable. Using manual control, the cameras were in motion an average of 8.23% of the time; for voice control, they were in motion 14.88% of the time; and for automated tracking 33.42% of the time. Post hoc tests conducted using Scheffé's S test (Kirk, 1982) found significant differences between all three of the camera control systems.

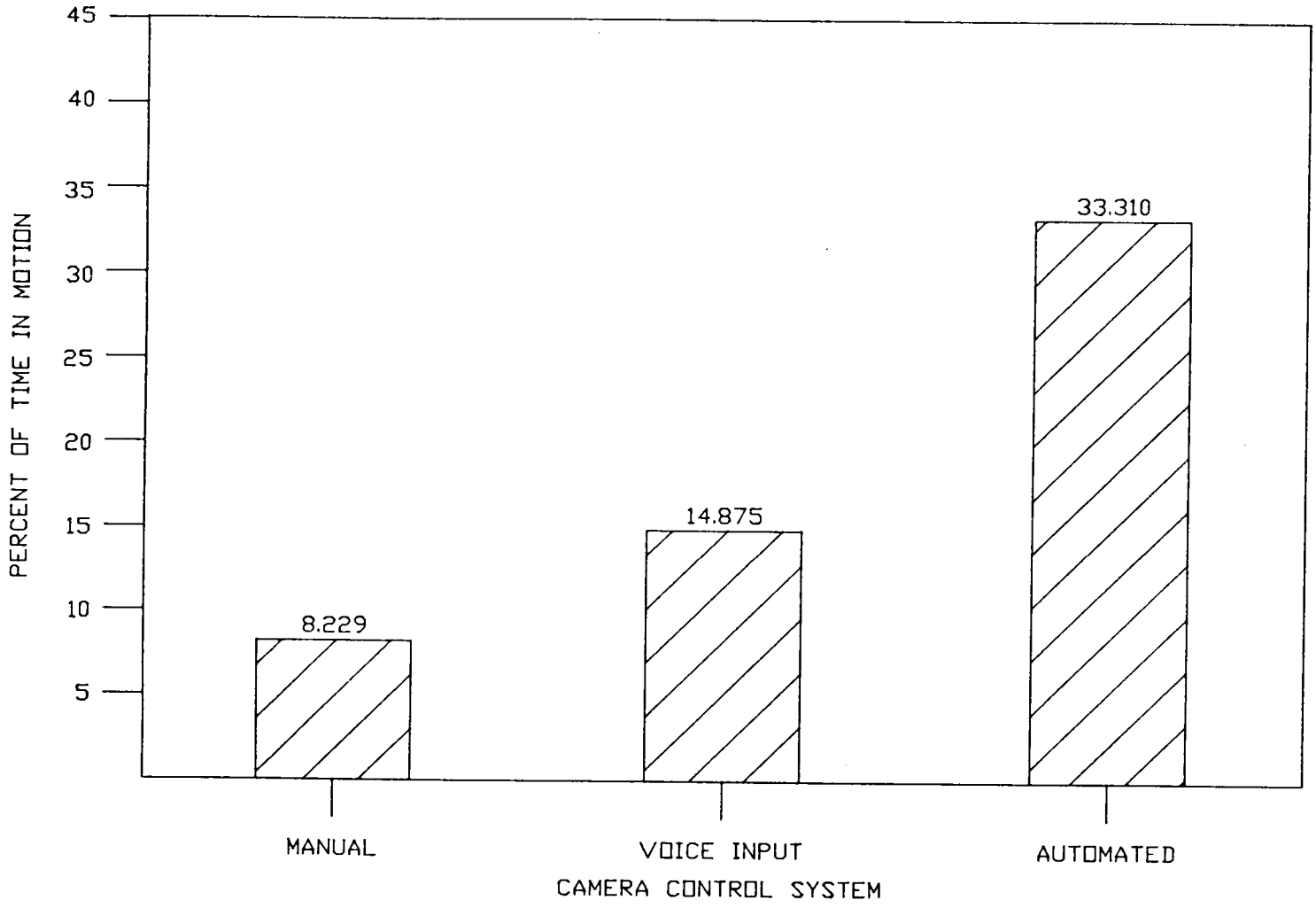



Figure 20. Mean percent of total task completion time that cameras were in motion for each camera control type.

# 6. SUMMARY OF RESULTS

## Camera Controls

The MANOVA discovered a significant difference between camera control conditions. The ANOVAs for four of the dependent variables found significant differences among the levels of this factor. The variables that were affected by camera control type were the logarithm to base 10 of task completion time in seconds, the percent of time operators were engaged in fine-adjusting motions, the number of camera changes made per trial, and the percent of time per trial that cameras were in motion. No effect of camera controls was found for the rate at which errors were committed during task completion. A significant effect of camera controls was found within the slewing motions variable during completion of the disassembly task and during completion of the assembly task in block 1 of testing; no effect was found during performance of the assembly task in block 2.

For the logarithm to base 10 of time in seconds, operators completed tasks more slowly with manual control than they did with voice control and automated tracking. There was not a significant difference between automated tracking and voice control, although operators completed tasks in less time with automated tracking than with voice control. Operators spent less time engaged in fine-adjusting motions while using the manual controls than they did when using the voice controls or the automated tracking system. There was no significant difference between voice control and automated tracking, although the average percent of time was higher when operators used the voice control system. For

the disassembly task, operators were engaged in slewing motions more with voice control when the completed the disassembly task in block 1 than with any other system; during block 2 they were engaged is slewing motions for less time with manual control, but there was no difference between voice control and automated tracking. While completing the assembly task, the slewing motion percent was lowest for operator using manual control during block 1 but there were no differences between control systems during block 2. More camera changes occurred during trials with automated tracking than during trials with manual control or voice control. There was no significant difference between manual control and voice control, although more changes were made with manual control. The cameras were in motion for a larger percentage of time when the operators used the automated tracking system than when they used either of the other systems, and the cameras were in motion for a larger percentage of time when the operators used voice control than when they used manual control.

These data indicate that the voice control system and the automated tracking system were superior to manual control for these tasks. Manual control led to the longest task completion times and the highest levels of task difficulty (lowest proportion of time spent in slewing motions and fine-adjusting motions). The overall pattern of results suggests that there is little to separate voice control and automated tracking in terms of the performance variables; however, in terms of the use of cameras during task performance, automated tracking caused much more frequent movement of the cameras to occur than did voice control. Operators using voice control made roughly the same number of camera

changes as with manual control, but these changes were of longer duration.

## Task Performed

The error matrix for the MANOVA test of the task performed factor was singular, so no test could be performed. However, results of the ANOVAs for the individual dependent variables are quite consistent. Significant differences were found for the logarithm to base 10 of task completion time, the percent of time engaged in slewing motions, the percent of time engaged in fine-adjusting motions, and the number of camera changes per trial. No significant effect was found for two variables: the rate at which errors were committed during task completion and the percent of time cameras were in motion. Operators completed the disassembly task more quickly than the assembly task. The operators spent a larger percentage of total task time engaged in slewing motions and fine-adjusting motions while completing the disassembly task than they did while completing the assembly task. More camera changes were made during completion of the assembly task.

## Block of Testing

The error matrix for the MANOVA test of the block of testing factor was singular, so no test could be performed. However, the results of the ANOVAs are consistent. The blocks of testing were significantly different for the logarithm to base 10 of task completion time in seconds, the percent of total task time engaged in slewing motions, and the number of camera changes made per trial. Operators required longer to complete

tasks during the first block of testing, spent a lower percentage of total task time engaged in slewing motions, and made a larger number of camera changes than in the second block of testing.

## Interactions

No interaction was significant in the MANOVA. However, two separate ANOVAs found significant interactions. There was a significant control type by task performed by block of testing interaction on the percent of time engaged in slewing motions, and a significant control type by task performed interaction on the number of camera changes per trial. For the slewing motions variable, simple effects tests performed within the disassembly task, block 1 combination indicate that operators spent a higher proportion of their time performing slewing motions with the voice control system than with either of the other systems. In block 2, the contrast between voice and automated tracking was not significant, but the manual versus voice and automated tracking contrast was. This indicates that, on this task, the operator's performance with automated tracking improved more from block 1 to block 2 than it did with the other camera control systems. Initially there was an advantage for voice control over the other systems; as operators practiced with the systems, performance with automated tracking improved to the point of equivalence with voice control. On the disassembly task, there was a significant contrast for manual control versus voice and automated tracking in block 1; however, as operators practiced with the manual control system, their performance improved to the point that there were no differences between the systems. For the camera changes variable,

the interaction was due to a difference in the response of the automated tracking system to the two tasks. On the disassembly task, the automated tracking system made more than twice as many changes as did the other systems; on the assembly task, the tracking system made more than four times as many changes.

# CHAPTER 5

# DISCUSSION AND RECOMMENDATIONS

## 1. DISCUSSION

The hypotheses stated at the end of Chapter 2 were: (1) manual control of cameras will produce relatively high levels of workload and poor performance, due to resource competition; (2) control by computer word recognition will produce lower workload and better performance than manual control, due to a relative reduction in resource competition; and (3) automated end-effector tracking will produce the lowest levels of workload and the best performance by removing camera control demands from the task. As expected, these data show a consistent performance disadvantage for manual control. For workload (measured by the percent of time operators were engaged in slewing and fine-adjusting motions), operators found the tasks most difficult when completing them with manual control. In terms of performance, the operators required more time to complete the tasks with manual control than with the other camera control systems, but there was no difference in the number of errors committed during task completion. There was not a clear difference between control by computer word recognition and automated end-effector tracking. For the measures of workload, there were no significant differences between voice control and automated tracking, although trials completed with computer word recognition produced lower average scores on the workload measures. For the measures of task performance, operators were able to complete the tasks in less time on average when

using automated tracking although the difference between automated tracking and word recognition was not significant. There was not a significant difference between the two systems on the error rate variable.

For these tasks, the data support the conclusions of Vertut et al. (1983) in that the automated tracking system allowed operators to complete tasks more quickly than with the manual control system. However, the difference in these data was not as dramatic as the performance difference found by Vertut et al. (1983). They report a 2.25 to 1 task completion time advantage for automated tracking; in these data, automated tracking allowed operators to complete the tasks in 83% of the time required to perform them with manual control. The difference in the findings of this experiment and Vertut et al. (1983) may be due to a number of experimental details, including the tasks performed, operator skill and training, practice effects, and the design of the manual control system's human-machine interface.

The analysis of the camera motions variables (percent of time cameras were in motion and the number of changes in camera position made per trial) yields insights into the ways that the operators used the camera controls. For the manual control system, the operators made relatively few camera changes, averaging slightly more than 13 per trial. For the voice control system, the made even fewer (although the difference between manual control and voice control was not significant), averaging slightly more than 11 per trial. The automated tracking system made more than 36 changes per trial for the operators, in spite of the fact that the trials were (on average) of shorter duration. The

percent of time cameras were in motion shows that cameras were moving more than 1/3 of the time during trials with automated tracking, compared to approximately 15% of the time for voice control and 8% for manual control. Since there were no performance or workload differences between voice control and automated tracking, it seems that the tracking system over-adjusted cameras by a factor of 2; perhaps the deadband area around the end-effector was too small.

Figure 21 illustrates the average length of camera adjustments (in terms of the percent of time required for camera changes) for each of the camera control systems. For manual control, each camera change required approximately 0.62% of total task time; for voice control, each change required 1.33%; and for automated tracking, each change required 0.91% of total task time. The data for the number of changes per trial and the length of each change indicate that operators using manual control tended to make a small number of short duration camera changes. This seems to support the speculation that operators using manual control tend to avoid making camera changes. They attempt to establish the best camera positions for performance of the task and then make minor small-scale adjustments when necessary. Operators using voice control made no more changes than when using manual control, but the changes required more than twice as long to complete. This indicates that operators using voice control were more willing to make relatively large adjustments in camera position during trials. Operators using voice control were able to establish optimal camera positions concurrently with manipulator operation and were less dependent on camera positions established at the beginning of trials.

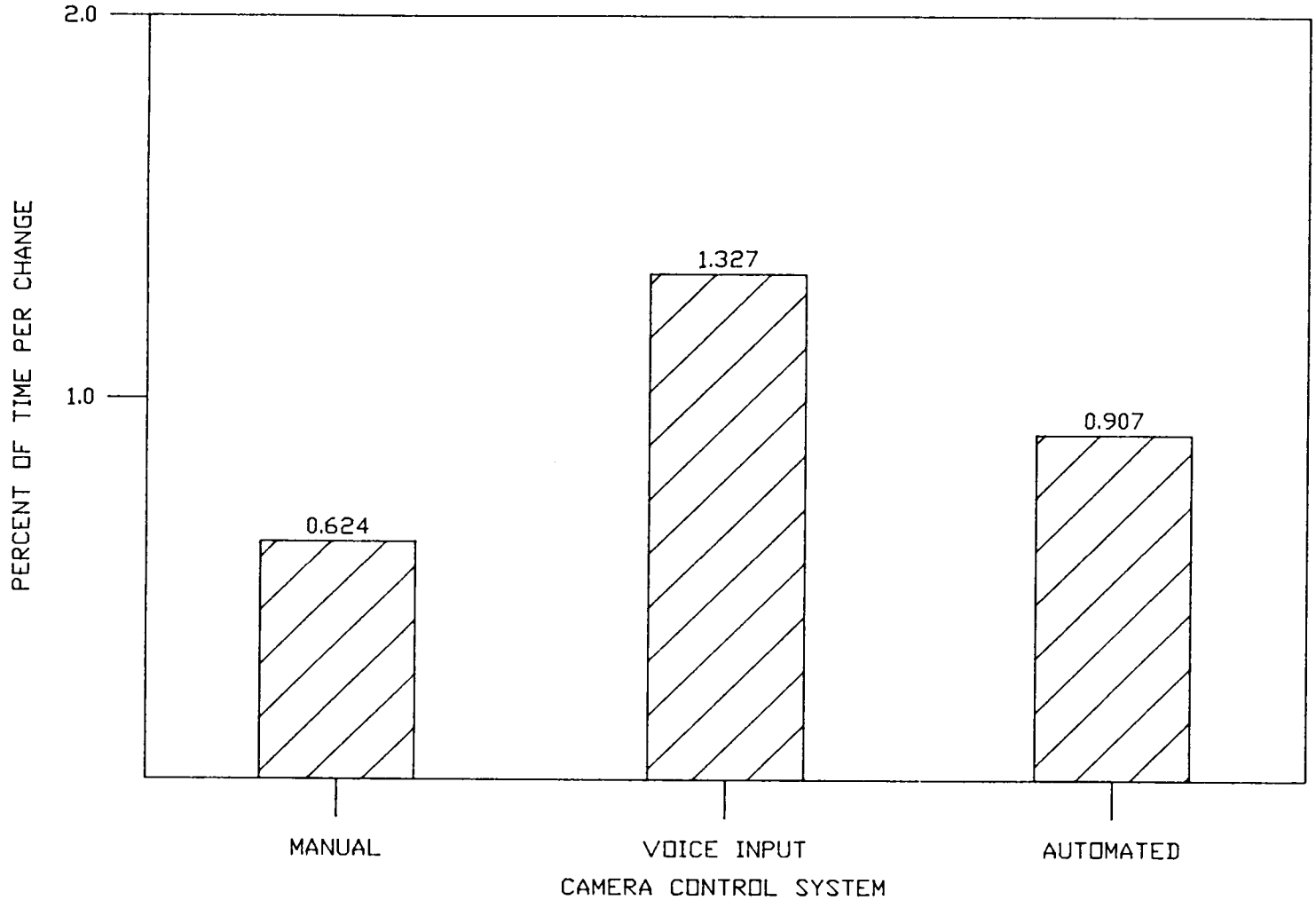



Figure 21. Camera movement duration for each camera control system.

The performance of two variables was disappointing. The errors per minute variable, which was intended to measure the quality of operators' performance, was not sensitive to any of the effects included in the study. Given the consistent results for the other variables (especially in regard to the task performed factor), this is surprising. It may be that the errors included in the tally are not appropriate for use in experiments involving teleoperation. A critical incidents checklist, which could be used to tally occurences of individual events (including the ones tallied as errors in this experiment) might be more sensitive. Another reason for the insensitivity of the errors per minute variable was the wide intra-operator and inter-operator variance exhibited by the subjects in this experiment. A larger operator sample, more repetitions per condition, or more pre-training may have reduced this variance and made the errors per minute data more sensitive.

The other disappointing variable was the measure of subjective workload. In this experiment, the modified Cooper-Harper scale did not provide an adequate measure of subjective workload, probably due to the relatively short duration of the trials. In spite of efforts to reduce the effect of operator memory effects (the special instructions and alternating forms mentioned in Chapter 3), the scale seemed to be affected by the high frequency of administration and short time span between administrations. The scale might have worked better had more complex, longer duration tasks been used or if the operators had been asked to make one rating per experimental session, rather than one per trial. With failure of the Cooper-Harper scale to provide a measure of subjective workload, the evaluation of workload for these tasks must depend

on the control motions variables. Other authors have found control motions to be adequate measures of workload (Wierwille & Connor, 1983); however, these authors measured the rate of control motions. Rate of control motions makes the assumption that workload levels are related to the speed with which operators can make motor outputs in response to display changes. Percentage of total task time that controls were in motion is a direct measure of task demand, and therefore of task difficulty. One may assume that more difficult tasks require greater workload to complete in a given span of time, if performance levels are held constant. Therefore, percentage of time that the manipulator was in motion may serve as evidence of the workload associated with a given combination of control type, task performed, and block of testing.

The experimental design of the study is weak in two respects, both related to the small operator sample included in the study. With four operators, the power of statistical tests employed was low; therefore, the criterion significance level for the ANOVAs was set at 0.10. Fortunately, the MANOVA for the camera controls factor (the effect of primary interest) found a robust effect; the F ratio for this factor exceeded the criterion value with a probability of Type I error of only 0.04. While no MANOVA tests could be performed on the other main effects, the ANOVAs for the dependent variables were consistent. The interaction effects that were found in two of the ANOVAs were not replicated in any of the other variables, and no significant effects were found in the MANOVA for the interactions. Less confidence may be placed in the interactions than in the main effects.

The other problem related to the small sample size has to do with generalizability. With such a small operator sample, it is possible that the results of the experiment are due to a subjects by treatments interaction. If this is the case, then these findings might not be replicable in the general operator population. However, in defense of the experiment, the sample in the experiment represents 80% of the trained operators of electrical master/slave manipulators available at the Oak Ridge National Laboratory at the time of testing.

The performance of the multiple resource model in generating hypotheses about the various control systems was good. The three central hypotheses, detailed earlier in this section, were derived directly from the multiple resource model. The results of the experiment are for the most part consistent with the hypotheses, indicating that the model was adequate for generating predictions of operator behavior in this case. This supports McCauley's (1984) contention that the effects of voice recognition are best understood within the framework of the multiple resource model. The single deviation from the multiple resource model hypotheses was the inability of the experiment to detect differences between voice control and automated tracking. Since automated tracking removed all camera control responsibility from the operators, it should have reduced the task demands on operators from those present when using voice control. However, the performance indices show no consistent advantage for the automated tracking system. Two explanations for this observation seem possible. First, it may be that the limitations of the tracking system may have raised task demands over the expected levels. The data show that the automated tracking system kept cameras in motion

for a much higher percentage of total task time than the other systems did. The frequent motion may have made the automated tracking system more demanding to use than if it had operated in a fashion similar to that of the other systems. A more intelligent tracking system, which could take advantage of information about the progress of the task, the manipulator movements necessary to complete the task, or other data, could make tracking performance better than that observed.

Second, the small sample size used in the study may have made the statistics insensitive to differences which may exist in the population. It is also possible that the small number of repetition of each task with each system may have been insufficient to provide stable data for analysis. A larger sample, in terms of operators participating in the study or the number of repetitions of each task by each operator could improve the performance of the statistical tests and lead to the discovery of performance differences between voice control and automated tracking.

Small sample sizes are a constant problem in remote handling research; it seems necessary to use skilled operators in order to generate meaningful data, but the skilled operator population is small due to the relative expense and youth of the technology. Statistics which derive their power from the number of experimental units employed in a study suffer from low power in this situation. There are a limited number of methods for coping with this problem. The use of repeated measures designs is an effective strategy, but these designs necessitate the use of unusual strategies for constructing mean square ratios in order to produce tests which are meaningful, as demonstrated by the analysis

employed in this study. Repeated measures designs also require careful counterbalancing of treatment presentation to prevent practice effects and transfer from becoming confounded with effects under study. Strategies for stabilizing error variance can also improve the power of tests with small sample sizes; conversion of dependent variables to more stable indices, as in conversion of task time in seconds to it logarithm to base 10, is an example. The use of medians rather than averages is another example. Averages tend to be unduly influenced by outlying scores and are sometimes less stable than medians. Another strategy for improving the power of experiments with small sample sizes is the use of multiple performance criteria. In this study, six dependent measures were taken. Each served to operationalize a different aspect of performance; in the aggregate, they were more powerful than any one of them would have been alone.

To summarize the results of the experiment, (1) tasks performed using manual control required longer to perform than tasks completed using the other control systems; (2) operators used their primary controls (the manipulator master controllers) with more frequency when tasks were completed with voice control and automated tracking, as indicated by the results on the fine-adjusting motions and slewing motions variables; and (3) cameras were in motion for a larger percent of the time during task completion and made more frequent changes when automated tracking was engaged than when the other control systems were in use. These results seem to support two conclusions about the effects of the three camera control systems: (1) performance with voice control and automated tracking is superior to performance with manual control, as demonstrated

by the results for the task completion time variable; and (2) voice control and automated tracking reduce the workload associated with teleoperations, compared to the workload experienced using manual control, as demonstrated by the results for the manipulator motions variables. A third conclusion which seems appropriate is that the automated tracking system used in this experiment made camera changes with unnecessary frequency during the course of task completion.

## 2. RECOMMENDATIONS

While there was little to separate voice control and automated tracking of end-effectors in these data, voice control has the advantage of placing cameras under operator control. This increased the demands placed on the operator by the task, but the increased demands were not enough to cause a degradation in performance nor did they cause an increase in workload for these tasks. On the other hand, automated tracking limits the viewing options available to an operator. For example, the optimal camera aiming strategy for these tasks may have been to aim one camera at the left-hand side of the task volume and another at the right-hand side. With automated tracking it was not possible for an operator to aim the cameras at different points within the task area. Control by computer word recognition allows greater camera aiming flexibility, and does not increase workload or degrade performance when compared to automated end-effector tracking. Teleoperator cockpits which are expected to be used by one operator should be equipped with a system for voice control of television cameras.

Control systems not evaluated during this testing may be effective in some situations. For example, this experiment did not involve a system which allowed users to control cameras by means of foot pedals. Foot pedals have been used for many years in automobiles and aircraft. These devices have served reliably and with sufficient accuracy in these roles. A recent Japanese manipulator design (the Advanced Servomanipulator Prototype I, installed in the EDF-1 facility) includes foot pedals for control of some functions. This manipulator has an unusual and somewhat restrictive (to the operator) stance, with the upper and lower links horizontal and the elbows pointing out. The workstation is limited to seated operations. Foot pedals were not considered in this study because of the tendency of manipulator operators in the through-the-wall stance to move around under the master controller handle during task completion. Movement within the workstation makes foot pedals difficult to use, since operators must search for the location of the pedals after each movement. Manual controls suffer from the same problem, but the impact is not as great. Operators searching for manual controls are more likely to correctly locate a control and less likely to accidentally activate it than operators searching for foot pedals, since the hand is so much more sensitive than the foot. Another reason foot pedals were not considered was the potential for developing future, advanced servomanipulator control stations as sit/stand workstations. While standing the operators' feet would not be free to use pedals.

The automated tracking system was in motion much more of the time than either the manual system or the word recognition system. It was

probably in motion more than was necessary. Performance with this system could be improved by more sophisticated algorithms for automated motion. The simple deadband system should be replaced with a more intelligent system involving teach/playback of camera motions, consideration of task characteristics (e.g. volume, accessibility), knowledge of the structure and sequence of subtasks, adaptation to the type of motion (slewing versus fine-adjusting) being conducted, and adaptation to the level of operator skill. A voice override used in conjunction with the automated tracking system would also be helpful; it would retain the advantage of automated tracking (removal of responsibility for camera control from the manipulator operator) and yet leave the operator with some control over camera placement and motion. The ideal camera control system would normally operate in automated tracking mode (although more intelligently than in this study) and could be switched to voice control quickly and easily during task performance. During normal operation this system would serve as an automated assistant to the manipulator operator, and during unusual events the operator could take control of the system. The teach/playback function could be invoked during unusual events so that in later occurences of the event the operator could allow control to remain in the hands of the automation system.

While these data provide valuable insights into the impact of voice control and automated tracking, they should not be interpreted as providing the final solution to the problem. The size of the sample (4 operators) makes these data somewhat questionable, in terms of generalizing to the entire population of potential users. It also limits the sensitivity of the study to potential differences between the control

systems. Real differences between voice control and automated tracking may exist in terms of time to completion, errors, and workload, but the sample size problem limited the power of the statistics employed to detect differences. To provide definitive data, further testing should employ a larger group of operators. Future research should also examine the role of operator practice on the effects of the three control systems on performance. Previous experience with teleoperators may have affected the performance of the operators in this study. Each had prior manipulator experience, exclusively with manual control of viewing systems. Habits acquired with these systems may have affected performance in this study. For example, naive operators might find automated tracking even easier to use than these operators did, while operators with more experience with computer word recognition may be better with voice control than these operators.

Another consideration not addressed by these data is the rate of training with each system. Automated tracking may make it easier for naive operators to learn to use the manipulator system than manual control or voice control. Naive operators training with the latter systems must simultaneously acquire expertise with two complex systems, the manipulator and the camera control system. Initial training with automated tracking might break the training into more easily digestable chunks.

A larger sample of remote handling tasks should be included in future testing. The tasks performed in this study include a number of critical manipulator tasks (insertion, fine alignment, rotations, movement and positioning of heavy objects), but they represent a small subset of the tasks a teleoperator may be asked to perform in maintenance

operations. It may be more effective to sample a wider variety of less complex tasks in future testing, in order to provide better coverage of important remote handling task elements.

Future testing should also use a more effective measure of subjective workload. This experiment attempted to use the Modified Cooper-Harper (MCH) scale (Wierwille & Casali, 1983), a modified version of the Cooper-Harper scale (Cooper & Harper, 1969). The scale failed to produce an effective measurement of subjective workload. The reason for the failure seems to be the length of the tasks. The MCH and the Cooper-Harper scale were developed to measure handling characteristics and workload in aircraft. Trials involving aircraft flight can be expected to be much longer than the trials in this study; the average trial length for these data was about 4 minutes. Under this condition, the MCH scale was more responsive to operator memory of previous measurements than it was to subjective workload. It is also possible that the MCH scale failed to measure subjective workload because the operators were unable to judge workload on the basis of short work sessions. Operators may associate subjective workload with fatigue, and short sessions may not produce perceptible fatigue. The short duration tasks may have had insufficient impact on the operators to allow them to serve as effective judges of workload.

In summary, the data reported here provide valuable insights into the performance of three camera control systems on teleoperator performance. Advanced control systems currently available were better in terms of task completion time, errors committed during task performance, and workload than manual control, but the advanced systems are in need

of improvement. More intelligent systems and a combination of computer word recognition and automation may be the most effective control method. Limitations to this study stemming from the small sample size, the small number of remote handling tasks included, and the failure to effectively measure subjective workload make conclusions based on these data somewhat questionable without further testing.

# REFERENCES

## REFERENCES


Ackerman, P. L., Schneider, W., & Wickens, C. D. Deciding the existence of a time-sharing ability: a combined methodological and theoretical approach. Human Factors, 1984, 26(1), 71-82.

Aretz, A. J. A comparison of manual and vocal response modes for the control of aircraft systems. Proceedings of the Human Factors Society 27th Annual Meeting, October 1983, 97-101.

Bradford, K. Z., Cooper, T. W., Garcia, K. D., Wildgen, A. K., & Spencer, R. A. Teleoperator human factors study: Human factors research plan (MCR 84-512). Denver, Colorado: Denver Aerospace, Martin Marietta Aerospace, 1984.

Casali, J. G., & Wierwille, W. W. A comparison of rating scale, secondary-task, physiological, and primary-task workload estimation techniques in a simulated flight task emphasizing communications load. Human Factors, 1983, 25(6), 623-642.

Chapanis, A., & Kinkade, R. G. Design of controls. In H. P. Van Cott and R. G. Kinkade (Eds.), Human Engineering Guide to Equipment Design. Washington, D.C.: American Institutes of Research, 1972, 345-379.

Clarke, M. M., Hamel, W. R., & Draper, J. V. Human factors and remote control engineering development activities. Proceedings of the 31st Conference on Remote Systems Technology, Vol. 1, 1983, 8-16.

Clarke, M. M., & Kreifeldt, J. G. A control room concept for remote maintenance in high-radiation areas. Proceedings of the Human Factors Society 28th Annual Meeting, October 1984a,

Clarke, M. M., & Kreifeldt, J. G. Elements of an advanced integrated operator control station. Proceedings of the 1984 National Topical Meeting on Robotics and Remote Handling in Hostile Environments, 1984b, 425-432.

Cook, T. D., & Campbell, D. T. Quasi-experimentation: Design and Analysis Issues for Field Settings. Boston: Houghton Mifflin Company, 1979.

Cooper, G. E. Understanding and interpreting pilot opinion. Aeronautical Engineering Review, March 1957, 47-52.

Cooper, G. E., & Harper, R. P. The use of pilot rating in the evaluation of aircraft handling qualities. (NASA TN D-5153) National Aeronautics and Space Administration, April 1969.

Corliss, W. R., & Johnsen, E. G. Teleoperator controls (NASA SP-5070). Washington, D. C.: NASA Office of Technology Utilization, National Aeronautics and Space Administration, December 1968.

Donnell, M. L. The application of decision-analytic techniques to the test and evaluation phase of the acquisition of a major air system: Phase III. (Decisions and Designs TR 79-6-91) Decisions and Designs, May 1979.

Draper, J. V., & Sundstrom, E. Joint motion clusters in servomanipulator operation. Unpublished manuscript.

Draper, J. V., Sundstrom, E., & Herndon, J. N. Joint motion clusters in servomanipulator operation. Proceedings of the Human Factors Society 30th Annual Meeting, October 1986, 552-555.

Dumbreck, A. A. Unpublished data of the Project Engineering Development Group, Engineering Projects Division, Atomic Energy Research Establishment, Harwell, England.

Finkelman, J. M., Wolf, E. H., & Friend, M. A. Modified discriminant analysis as a multivariate post-comparison extension of MANOVA for interpretation of simultaneous multimodality measures. Human Factors, 1977, 19(3), 253-261.

Funk, K., & McDowell, E. Voice input/output in perspective. Proceedings of the Human Factors Society 26th Annual Meeting, October 1982, 218-222.

Gopher, D. On the training of time sharing skills: An attention viewpoint. Proceedings of the Human Factors Society 24th Annual Meeting, October 1980, 259-263.

Granaas, M. M., McKay, T. D., Laham, R. D., Hurt, L. D., & Juola, J. F. Reading moving text on a CRT screen. Human Factors, 1984, 26(1), 97-104.

Greenberg, L., & Chafin, D. B. Workers and Their Tools: a Guide to the Ergonomic Design of Hand Tools and Small Presses. Midland, Michigan: Pendell Publishing Company, 1974.

Greenhouse, S. W., & Geisser, S. On methods in the analysis of profile data. Psychometrika, 1959, 24(2), 95-112.

Greenstein, J. S., & Revesman, M. E. A Monte-Carlo simulation investigating means of human-computer communication for dynamic task allocation. Proceedings of the IEEE 1981 International Conference on Cybernetics and Society, 1981, 488-494.

Hamel, W. R., & Feldman, M. J. The advancement of remote systems technology: past perspectives and future plans. Proceedings of the 1984 National Topical Meeting on Robotics and Remote Handling in Hostile Environments, 1984, 11-23.

Handel, S. J., & Draper, J. V. Coming to grips with force reflecting servomanipulators. Poster session at the Second International Conference on Event Perception, Vanderbilt University, Nashville, Tennessee, June 9-13, 1983.

Herndon, J. N., & Hamel, W. R. Analysis of the options-rationale for servomanipulator maintenance in future reprocessing plants. Proceedings of the 1984 National Topical Meeting on Robotics and Remote Handling in Hostile Environments, 1984, 439-452.

Hicks, T. G., & Wierwille, W. W. Comparison of five mental workload assessment procedures in a moving-base driving simulator. Human Factors, 1979, 21(2), 129-143.

Huynh, H. Some approximate tests for repeated measures designs. Psychometrika, 1978, 43(2), 161-175.

Huynh, H., & Mandeville, G. K. Validity conditions in repeated measures designs. Psychological Bulletin, 1979, 86(5), 964-973.

Imhof, J. P. Testing the hypothesis of fixed main-effects in Scheffe's mixed model. Annals of Mathematical Statistics, 1962, 33, 1086-1095.

Kirk, R. E. Experimental Design: Procedures for the Behavioral Sciences (2nd ed.). Belmont, CA: Wadsworth, 1982.

Kohler, G. W. Manipulator Type Book. Munich, Federal Republic of Germany: Verlag Karl Thiemig, 1981, 234-238.

Magnusson, D. Test Theory. Reading, MA: Addison-Wesley Publishing Company, 1967, 73-74.

McCauley, M. E. Human factors in voice technology. In F. A. Muckler (Ed.), Human Factors Review: 1984. Santa Monica, California: Human Factors Society, 1984, 131-166.

McCormick, E. J., & Sanders, M. S. Human Factors in Engineering and Design (5th Edition). New York: McGraw-Hill, 1982.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 1956, 63(2), 81-97.

MIL-HDBK-759a. Military handbook: human factors engineering design for Army materiel. Redstone Arsenal, Alabama: Standardization Division, U. S. Army Missile Command, 1981.

MIL-STD-1472c. Military standard: human engineering design criteria for military systems, equipment, and facilities. Washington, D. C.: Department of Defense, 1981.

Mountford, S. J., & North, R. A. Voice entry for reducing pilot workload. Proceedings of the Human Factors Society 24th Annual Meeting, October 1980, 185-189.

Murray, J. T., Gilfoil, D., & Van Praag, J. Voice versus keyboard control of cursor motion. Proceedings of the Human Factors Society 27th Annual Meeting, October 1983, 103.

Navon, D., & Gopher, D. On the economy of the human processing system. Psychological Review, 1979, 86(3), 214-255.

Navon, D., & Gopher, D. Task difficulty, resources, and dual-task performance. In R. S. Nickerson (Ed.) Attention and Performance VIII. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1980, 297-315.

Norman, D. A., & Bobrow, D. G. On data-limited and resource-limited processes. Cognitive Psychology, 1975, 7, 44-64.

O'Brien, R. G., & Kaiser, M. K. The MANOVA method for analyzing repeated measures designs: An extensive primer. Psychological Bulletin, 1985.

Poock, G. K. Experiments with voice input for command and control (NPS-55-80-016). Monterrey, California: Naval Postgraduate School, 1980.

Price, H. E., Maisano, R. E., & Van Cott, H. P. The allocation of functions in man-machine systems: a perspective and literature review (NUREG/CR-2623). Oak Ridge, Tennessee: Instrumentation and Controls Division, Oak Ridge National Laboratory, November 1982.

Rieger, C. A., & Greenstein, J. S. The allocation of tasks between the human and the computer in automated systems. Proceedings of the IEEE 1982 International Conference on Cybernetics and Society, 1982, 204-208.

Riley, M. W., & Cochran, D. J. Dexterity performance and reduced ambient temperature. Human Factors, 1984, 26(2), 207-214.

SAS Institute. SAS Users's Guide: Statistics. Cary, North Carolina: SAS Institute, 1982.

Skriver, C. P. Vocal and manual response modes: comparison using a time-sharing paradigm (NADC-79127-60). Warminster, Pennsylvannia: Naval Air Development Center, 1979.

Stave, A. M. The effects of cockpit environment on long- term pilot performance. Human Factors, 1977, 19(5), 503-514.

Stoughton, R. S. Kinematics and Duty Cycles of the SM-229 Force Reflecting Servomanipulator. ORNL/TM-9655, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 1986.

Stoughton, R. S., Martin, H. L., & Bentz, R. R. Automatic camera tracking for remote manipulators. Proceedings of the National Topical Meeting on Robotics and Remote Handling in Hostile Environments, April 1984, 383-389.

Taplin, P. S., & Reid, J. B. Effects of instructional set and experimenter influence on observer reliability. Child Development, 1973, 44, 547-554.

Tsang, P. Effects of task structures on attention allocation optimality. Proceedings of the Human Factors Society 26th Annual Meeting, October 1982, 465-469.

Vertut, J., Hugon, R., Fournier, R., Espiau, B., & Coiffet, Ph. Human factor aspects of computer enhanced teleoperation. Proceedings of the 31st Conference on Remote Systems Technology, June 1983, 40-43.

Vidulich, M., & Wickens, C. D. The influence of S-C-R compatibility and resource competition on performance of threat evaluation and fault diagnosis. Proceedings of the Human Factors Society 26th Annual Meeting, October 1982, 223-226.

Welch, J. R. Automatic data entry analysis (RADC-TR-77- 306). Griffiss Air Force Base, New York: Rome Air Development Center, 1977.

Wickens, C. D. The structure of attentional resources. In R. S. Nickerson (Ed.) Attention and Performance VIII. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1980, 239-257.

Wickens, C. D. Engineering Psychology and Human Performance. Columbus, Ohio: Charles E. Merrill Publishing Company, 1984, 253.

Wickens, C., Derrick, W., Berringer, D., & Micalizzi, J. The structure of processing resources: Implications for task configuration and workload. Proceedings of the Human Factors Society 24th Annual Meeting, October 1980, 253-256.

Wickens, C. D., Sandry, D. L., & Vidulich, M. Compatibility and resource competition between modalities of input, central processing, and output. Human Factors, 1983, 25(2), 227-248.

Wickens, C. D., Sandry, D. L., Vidulich, M., & Schiflett, S. Factors affecting the performance advantage of speech technology. Proceedings of the Human Factors Society 25th Annual Meeting, October 1981, 705-709.

Wierwille, W. W. Physiological measures of aircrew mental workload. Human Factors, 1979, 21(5), 575-594.

Wierwille, W. W., & Casali, J. G. A validated rating scale for global mental workload measurement and applications. Proceedings of the Human Factors Society 27th Annual Meeting, October 1983, 129-133.

Wierwille, W. W., & Connor, S. A. Evaluation of 20 workload measures using a psychomotor task in a moving-base aircraft simulator. Human Factors, 1983, 25(1), 1-16.

Wierwille, W. W., & Williges, R. C. Survey and analysis of operator workload assessment techniques. (S-78-101) Patuxent, Maryland: Systems Engineering Division, Naval Air Test Center, September 1978.

Willeges, R. C., & Wierwille, W. W. Behavioral measures of aircrew mental workload. Human Factors, 1979, 21(5), 549-574.

Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1962.

# APPENDIX

# APPENDIX

Sums of squares for the operator by contrast interactions, used in the analysis of the camera controls main effect for the logarithm to base 10 of task completion time and for the percent of time engaged in fine-adjusting motions, were calculated using a custom written, BASIC-language program. The program was run on a COMPAQ portable computer using the MS-DOS operating system. The program used a formula published in Kirk (1982) to calculate sums of squares. According to Kirk (1982), the sum of squares for a specific (effect by treatment) error term is calculated by

$$SS\alpha\beta_i = n[\Sigma(c_i x_i^2 \text{ at } \alpha_j) - (\Sigma(c_i x_i \text{ at } a_j)^2/p)]/\Sigma(c_i^2);$$

where $c_i$ = the contrast coefficient for level i of factor B at level j of factor A in the interaction,

$x_i$ = the mean at level i of factor B at level j of factor A in the interaction.

The degrees of freedom of the sums of squares are equal to the omnibus interaction degrees of freedom divided by the total number of orthogonal contrasts required to describe the interaction. In other words, the interaction degrees of freedom are equally partitioned among the orthogonal contrasts for the omnibus interaction.

A BASIC-language computer program which calculates sums of squares for treatment by factor interactions in the RSDF data is listed below. The essential calculations are made in lines 320 through 430. In lines 340 and 350, the means for the interactions are multiplied by the

contrast coefficients; this corresponds to the term $c_i x_i$ in the formula. In lines 370 and 380 the products are combined to furnish the result of the term $\Sigma c_i x_i$. Lines 390 and 400 calculate $\Sigma c_i x_i^2$. Lines 420 and 430 square $\Sigma c_i x_i$; divide the result by p (in this case 4), which yields $(\Sigma c_i x_i)^2/p$; then subtract from $\Sigma c_i x_i^2$; and, finally, multiply by n (which is the 12 in this case) to give the sum of squares. Mean squares are calculated during the printing routine, in lines 550 and 600.

```
100 ' THIS PROGRAM CALCULATES CONTRAST BY OPERATOR SUMS OF SQUARES FOR
THE
110 ' CAMERA CONTROLS MAIN EFFECT ORTHONORMAL CONTRASTS IN THE RSDF DATA
120 KEY OFF
130 CLS
140 PRINT "Do you wish an output based on TASK TIME or FINE MOTIONS"
150 INPUT "(enter T for TASK TIME or F for FINE MOTIONS)";E$
160 IF E$="T" OR E$="t" THEN RESTORE 660 ELSE RESTORE 700
170 CLS
180 READ SOURCE$
190 DIM OPERATOR.BY.CONTROL(4,3)
200 FOR I=1 TO 4
210 FOR J=1 TO 3
220 READ OPERATOR.BY.CONTROL(I,J)
230 NEXT J
240 NEXT I
250 RESTORE 740
```

```
260 FOR I=1 TO 3
270 READ CONTRAST.1(I)
280 NEXT I
290 FOR I=1 TO 3
300 READ CONTRAST.2(I)
310 NEXT I
320 FOR I=1 TO 4
330 FOR J=1 TO 3
340 SUM.1(I)=SUM.1(I)+OPERATOR.BY.CONTROL(I,J)*CONTRAST.1(J)
350 SUM.2(I)=SUM.2(I)+OPERATOR.BY.CONTROL(I,J)*CONTRAST.2(J)
360 NEXT J
370 X.1=X.1+SUM.1(I)
380 X.2=X.2+SUM.2(I)
390 XX.1=XX.1+SUM.1(I)2
400 XX.2=XX.2+SUM.2(I)2
410 NEXT I
420 SUMSQ.1=12*(XX.1-((X.12)/4))
430 SUMSQ.2=12*(XX.2-((X.22)/4))
440 CLS
450 PRINT
460 PRINT
470 PRINT
480 PRINT "              SPECIFIC ERROR TERMS FOR THE ";SOURCE$;" VARI-
ABLE"
490 PRINT
500 PRINT "              Source          Sum of Squares
    Mean Square"
```

```
510     PRINT     "                    ----------------------------------------
------------------"
520 PRINT "              Con 1 by Oper       ";
530 PRINT USING "####.###";SUMSQ.1;
540 PRINT "                ";
550 PRINT USING "####.###";SUMSQ.1/3
560 PRINT
570 PRINT "              Con 2 by Oper       ";
580 PRINT USING "####.###";SUMSQ.2;
590 PRINT "                ";
600 PRINT USING "####.###";SUMSQ.2/3
610 PRINT
620 FOR I=1 TO 7
630 PRINT
640 NEXT I
650 END
660 DATA "TASK TIME",2.5302,2.4594,2.4355
670 DATA 2.4546,2.3202,2.3331
680 DATA 2.4311,2.4165,2.3477
690 DATA 2.5255,2.5423,2.5059
700 DATA "FINE MOTIONS",27.1667,33.9167,31.0833
710 DATA 44.6667,48.3333,46.7273
720 DATA 42.0833,42.2727,44.9167
730 DATA 28.6667,30.4167,29.5833
740 DATA 0,0.70710678,-0.70710678
750 DATA 0.816496,-0.408248,-0.408248
```

# VITA

John Venoy Draper was born in Thief River Falls, Minnesota on October 14, 1957. He attended elementary schools in that city and in Bluff City, Tennessee. He graduated from Sullivan East High School in Bluff City in June 1976. The following September he entered King College, in Bristol, Tennessee. He graduated with a Bachelor of Arts degree in Psychology in August, 1980, with Honors in Independant Study. He began studies leading towards a doctorate at The Graduate School of The University of Tennessee, Knoxville, in September of that year. He received this degree in March, 1987.

Mr. Draper was previously employed (1981-82) as research assistant for Oak Ridge Associated Universities' Human Factors Projects. In October 1983 he accepted a position as Senior Human Factors Specialist with Clarke Ambrose, Incorporated of Knoxville, Tennessee, where his duties involved research and development for remote handling systems at the Oak Ridge National Laboratory. In October 1984 he was promoted to Scientist, and he is currently responsible for human factors research and development at the Oak Ridge National Laboratory's Remote Control Engineering task, where Clarke Ambrose, Inc. works under subcontract to Martin Marietta Energy Systems, Incorporated. He has experience in design and execution of research programs dealing with all facets of manipulator systems, including lighting for remote viewing, television control systems for remote viewing, and evaluation of manipulator system performance. He has also assisted in the design and construction of equipment and mock-ups for advanced teleoperator systems.

The author is a licensed Psychological Examiner in the state of Tennessee, a member in the Human Factors Society, and a member of the American Nuclear Society, in the Human Factors and the Remote Systems Technology Divisions. He has written several technical publications on human factors for remote handling applications, including:


Draper, J. V., Sundstrom, E., & Herndon, J. N. Joint motion clusters in servomanipulator operation. Proceedings of the Human Factors Society 30th Annual Meeting, October 1986, 552-555.

Draper, J. V., & Herndon, J. N. Human factors activities in teleoperator development at the Oak Ridge National Laboratory. Proceedings of the 1986 National Topical Meeting on Advances in Human Factors in Nuclear Power Systems, American Nuclear Society, Knoxville, Tennessee, April 1986.

Draper, J. V. Human Factors Guidelines for Force-reflecting Master/Slave Manipulator Handles. ORNL/TM-9973, Oak Ridge National Laboratory, Oak Ridge, Tennessee, in press.

Draper, J. V. The human-machine interface in mobile teleoperators. Proceedings of the Department of Energy Workshop on Requirements for Mobile Teleoperators for Radiological Emergency Response and Recovery, Dallas, Texas, June 1985. (Invited paper).

Draper, J. V. Lighting for Remote Viewing Systems. Proceedings of the 1984 National Topical Meeting on Robotics and Remote Handling in Hostile Environments, American Nuclear Society, Gatlinburg, Tennessee, April 1984.

Clarke, M. M., Hamel, W. R., & Draper, J. V. Human factors activities in remote control engineering development activities. Proceedings of the 31st Conference on Remote Systems Technology, American Nuclear Society, Detroit, Michigan, June 1983.

Clarke, M. M., Kreifeldt, J. G., & Draper, J. V. Man/Machine Interface for a Nuclear Cask Remote Handling Control Station: System Design Requirements. Technical Report HEDL-7465, Hanford Engineering Development Laboratory, Hanford, Washington, 1984.

Kreifeldt, J. G., Clarke, M. M., & Draper, J. V. The Borg Scale: Validation for a consumer product. Proceedings of the Human Factors Society 28th Annual Meeting, San Antonio, Texas, October 1984.